Pearson New International Edition

Introduction to Environmental Engineering and Science

Gilbert M. Masters   Wendell P. Ela

Third Edition

ALWAYS LEARNING

PEARSON

Pearson New International Edition

Introduction to Environmental
Engineering and Science
Gilbert M. Masters   Wendell P. Ela
Third Edition

PEARSON

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsoned.co.uk

ISBN 10: 1-292-02575-1
ISBN 13: 978-1-292-02575-9

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Printed in the United States of America

Table of Contents

This page intentionally left blank

Mass and Energy Transfer

> When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.
>
> —*William Thomson, Lord Kelvin (1891)*

1 Introduction

This chapter begins with a section on units of measurement. Engineers need to be familiar with both the American units of feet, pounds, hours, and degrees Fahrenheit as well as the more recommended International System of units. Both are used in the practice of environmental engineering.

Next, two fundamental topics, which should be familiar from the study of elementary physics, are presented: the *law of conservation of mass* and the *law of conservation of energy*. These laws tell us that within any environmental system, we theoretically should be able to account for the flow of energy and materials into, and out of, that system. The law of conservation of mass, besides providing an important tool for quantitatively tracking pollutants as they disperse in the environment, reminds us that pollutants have to go somewhere, and that we should be wary of approaches that merely transport them from one medium to another.

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

In a similar way, the law of conservation of energy is also an essential accounting tool with special environmental implications. When coupled with other thermodynamic principles, it will be useful in a number of applications, including the study of global climate change, thermal pollution, and the dispersion of air pollutants.

2 Units of Measurement

In the United States, environmental quantities are measured and reported in both the *U.S. Customary System* (USCS) and the *International System of Units* (SI), so it is important to be familiar with both. Here, preference is given to SI units, although the U.S. system will be used in some circumstances. Table 1 lists conversion factors between the SI and USCS systems for some of the most basic units that will be encountered.

In the study of environmental engineering, it is common to encounter both extremely large quantities and extremely small ones. The concentration of some toxic substance may be measured in parts per billion (ppb), for example, whereas a country's rate of energy use may be measured in thousands of billions of watts (terawatts). To describe quantities that may take on such extreme values, it is useful to have a system of prefixes that accompany the units. Some of the most important prefixes are presented in Table 2.

Often, it is the concentration of some substance in air or water that is of interest. Using the metric system in either medium, concentrations may be based on mass (usually mg or g), volume (usually L or m^3), or number (usually mol), which can lead to some confusion. It may be helpful to recall from chemistry that one mole of any substance has Avogadro's number of molecules in it (6.02×10^{23} molecules/mol) and has a mass equal to its molecular weight.

Liquids

Concentrations of substances dissolved in water are usually expressed in terms of mass or number per unit volume of mixture. Most often the units are milligrams (mg),

TABLE 1

Some Basic Units and Conversion Factors

Quantity	SI units	SI symbol ×	Conversion factor =	USCS units
Length	meter	m	3.2808	ft
Mass	kilogram	kg	2.2046	lb
Temperature	Celsius	°C	1.8 (°C) + 32	°F
Area	square meter	m^2	10.7639	ft^2
Volume	cubic meter	m^3	35.3147	ft^3
Energy	kilojoule	kJ	0.9478	Btu
Power	watt	W	3.4121	Btu/hr
Velocity	meter/sec	m/s	2.2369	mi/hr
Flow rate	$meter^3$/sec	m^3/s	35.3147	ft^3/s
Density	$kilogram/meter^3$	kg/m^3	0.06243	lb/ft^3

TABLE 2

Common Prefixes

Quantity	Prefix	Symbol
10^{-15}	femto	f
10^{-12}	pico	p
10^{-9}	nano	n
10^{-6}	micro	μ
10^{-3}	milli	m
10^{-2}	centi	c
10^{-1}	deci	d
10	deka	da
10^{2}	hecto	h
10^{3}	kilo	k
10^{6}	mega	M
10^{9}	giga	G
10^{12}	tera	T
10^{15}	peta	P
10^{18}	exa	E
10^{21}	zetta	Z
10^{24}	yotta	Y

micrograms (μg), or moles (mol) of substance per liter (L) of mixture. At times, they may be expressed in grams per cubic meter (g/m^3).

Alternatively, concentrations in liquids are expressed as mass of substance per mass of mixture, with the most common units being parts per million (ppm) or parts per billion (ppb). To help put these units in perspective, 1 ppm is about the same as 1 drop of vermouth added to 15 gallons of gin, whereas 1 ppb is about the same as one drop of pollutant in a fairly large (70 m^3) back-yard swimming pool. Since most concentrations of pollutants are very small, 1 liter of mixture has a mass that is essentially 1,000 g, so for all practical purposes, we can write

$$1\,\text{mg/L} = 1\,\text{g/m}^3 = 1\,\text{ppm (by weight)} \tag{1}$$

$$1\,\mu\text{g/L} = 1\,\text{mg/m}^3 = 1\,\text{ppb (by weight)} \tag{2}$$

In unusual circumstances, the concentration of liquid wastes may be so high that the specific gravity of the mixture is affected, in which case a correction to (1) and (2) may be required:

$$\text{mg/L} = \text{ppm (by weight)} \times \text{specific gravity of mixture} \tag{3}$$

EXAMPLE 1 Fluoridation of Water

The fluoride concentration in drinking water may be increased to help prevent tooth decay by adding sodium fluoride; however, if too much fluoride is added, it can cause discoloring (mottling) of the teeth. The optimum dose of fluoride in drinking water is about 0.053 mM (millimole/liter). If sodium fluoride (NaF) is purchased in 25 kg bags, how many gallons of drinking water would a bag treat? (Assume there is no fluoride already in the water.)

Solution Note that the mass in the 25 kg bag is the sum of the mass of the sodium and the mass of the fluoride in the compound. The atomic weight of sodium is 23.0, and fluoride is 19.0, so the molecular weight of NaF is 42.0. The ratio of sodium to fluoride atoms in NaF is 1:1. Therefore, the mass of fluoride in the bag is

$$\text{mass F} = 25\text{ kg} \times \frac{19.0\text{ g/mol}}{42.0\text{ g/mol}} = 11.31\text{ kg}$$

Converting the molar concentration to a mass concentration, the optimum concentration of fluoride in water is

$$\text{F} = \frac{0.053\text{ mmol/L} \times 19.0\text{ g/mol} \times 1{,}000\text{ mg/g}}{1{,}000\text{ mmol/mol}} = 1.01\text{ mg/L}$$

The mass concentration of a substance in a fluid is generically

$$C = \frac{m}{V} \tag{4}$$

where m is the mass of the substance and V is the volume of the fluid. Using (4) and the results of the two calculations above, the volume of water that can be treated is

$$V = \frac{11.31\text{ kg} \times 10^6\text{ mg/kg}}{1.01\text{ mg/L} \times 3.785\text{ L/gal}} = 2.97 \times 10^6\text{ gal}$$

The bag would treat a day's supply of drinking water for about 20,000 people in the United States!

Gases

For most air pollution work, it is customary to express pollutant concentrations in volumetric terms. For example, the concentration of a gaseous pollutant in parts per million (ppm) is the volume of pollutant per million volumes of the air mixture:

$$\frac{1\text{ volume of gaseous pollutant}}{10^6\text{ volumes of air}} = 1\text{ ppm (by volume)} = 1\text{ ppmv} \tag{5}$$

To help remind us that this fraction is based on volume, it is common to add a "v" to the ppm, giving ppmv, as suggested in (5).

At times, concentrations are expressed as mass per unit volume, such as μg/m^3 or mg/m^3. The relationship between ppmv and mg/m^3 depends on the pressure, temperature, and molecular weight of the pollutant. The ideal gas law helps us establish that relationship:

$$PV = nRT \tag{6}$$

where

P = absolute pressure (atm)
V = volume (m^3)
n = mass (mol)

R = ideal gas constant = 0.082056 $\text{L}\cdot\text{atm}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$
T = absolute temperature (K)

The mass in (6) is expressed as moles of gas. Also note the temperature is expressed in kelvins (K), where

$$\text{K} = °\text{C} + 273.15 \tag{7}$$

There are a number of ways to express pressure; in (6), we have used atmospheres. One atmosphere of pressure equals 101.325 kPa (Pa is the abbreviation for Pascals). One atmosphere is also equal to 14.7 pounds per square inch (psi), so 1 psi = 6.89 kPa. Finally, 100 kPa is called a bar, and 100 Pa is a millibar, which is the unit of pressure often used in meteorology.

EXAMPLE 2 Volume of an Ideal Gas

Find the volume that 1 mole of an ideal gas would occupy at standard temperature and pressure (STP) conditions of 1 atmosphere of pressure and 0°C temperature. Repeat the calculation for 1 atm and 25°C.

Solution Using (6) at a temperature of 0°C (273.15 K) gives

$$V = \frac{1\,\text{mol} \times 0.082056\ \text{L}\cdot\text{atm}\cdot\text{K}^{-1}\cdot\text{mol}^{-1} \times 273.15\ \text{K}}{1\ \text{atm}} = 22.414\ \text{L}$$

and at 25°C (298.15 K)

$$V = \frac{1\ \text{mol} \times 0.082056\ \text{L}\cdot\text{atm}\cdot\text{K}^{-1}\cdot\text{mol}^{-1} \times 298.15\ \text{K}}{1\ \text{atm}} = 22.465\ \text{L}$$

From Example 2, 1 mole of an ideal gas at 0°C and 1 atm occupies a volume of 22.414 L ($22.414 \times 10^{-3}\ \text{m}^3$). Thus we can write

$$\text{mg/m}^3 = \text{ppmv} \times \frac{1\ \text{m}^3\ \text{pollutant}/10^6\ \text{m}^3\ \text{air}}{\text{ppmv}} \times \frac{\text{mol wt (g/mol)}}{22.414 \times 10^{-3}\ \text{m}^3/\text{mol}} \times 10^3\ \text{mg/g}$$

or, more simply,

$$\text{mg/m}^3 = \frac{\text{ppmv} \times \text{mol wt}}{22.414} \qquad \text{(at 0°C and 1 atm)} \tag{8}$$

Similarly, at 25°C and 1 atm, which are the conditions that are assumed when air quality standards are specified in the United States,

$$\text{mg/m}^3 = \frac{\text{ppmv} \times \text{mol wt}}{24.465} \qquad \text{(at 25°C and 1 atm)} \tag{9}$$

In general, the conversion from ppm to mg/m^3 is given by

$$\text{mg/m}^3 = \frac{\text{ppmv} \times \text{mol wt}}{22.414} \times \frac{273.15\ \text{K}}{T\ \text{(K)}} \times \frac{P\text{(atm)}}{1\ \text{atm}} \tag{10}$$

EXAMPLE 3 Converting ppmv to mg/m^3

The U.S. Air Quality Standard for carbon monoxide (based on an 8-hour measurement) is 9.0 ppmv. Express this standard as a percent by volume as well as in mg/m^3 at 1 atm and 25°C.

Solution Within a million volumes of this air there are 9.0 volumes of CO, no matter what the temperature or pressure (this is the advantage of the ppmv units). Hence, the percentage by volume is simply

$$\text{percent CO} = \frac{9.0}{1 \times 10^6} \times 100 = 0.0009\%$$

To find the concentration in mg/m^3, we need the molecular weight of CO, which is 28 (the atomic weights of C and O are 12 and 16, respectively). Using (9) gives

$$\text{CO} = \frac{9.0 \times 28}{24.465} = 10.3 \text{ mg/m}^3$$

Actually, the standard for CO is usually rounded and listed as 10 mg/m^3.

The fact that 1 mole of every ideal gas occupies the same volume (under the same temperature and pressure condition) provides several other interpretations of volumetric concentrations expressed as ppmv. For example, 1 ppmv is 1 volume of pollutant per million volumes of air, which is equivalent to saying 1 mole of pollutant per million moles of air. Similarly, since each mole contains the same number of molecules, 1 ppmv also corresponds to 1 molecule of pollutant per million molecules of air.

$$1 \text{ ppmv} = \frac{1 \text{ mol of pollutant}}{10^6 \text{ mol of air}} = \frac{1 \text{ molecule of pollutant}}{10^6 \text{ molecules of air}} \tag{11}$$

3 Materials Balance

Everything has to go somewhere is a simple way to express one of the most fundamental engineering principles. More precisely, the *law of conservation of mass* says that when chemical reactions take place, matter is neither created nor destroyed (though in nuclear reactions, mass can be converted to energy). What this concept allows us to do is track materials, for example pollutants, from one place to another with *mass balance* equations. This is one of the most widely used tools in analyzing pollutants in the environment.

The first step in a mass balance analysis is to define the particular region in space that is to be analyzed. This is often called the control volume. As examples, the control volume might include anything from a glass of water or simple chemical mixing tank, to an entire coal-fired power plant, a lake, a stretch of stream, an air basin above a city, or the globe itself. By picturing an imaginary boundary around

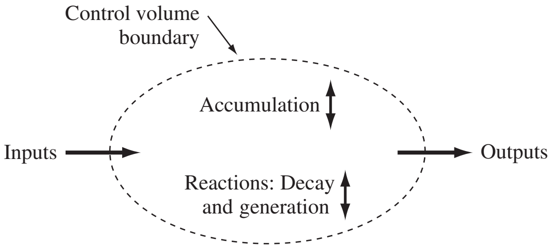


FIGURE 1 A materials balance diagram.

the region, as is suggested in Figure 1, we can then begin to quantify the flow of materials across the boundary as well as the accumulation and reaction of materials within the region.

A substance that enters the control volume has four possible fates. Some of it may leave the region unchanged, some of it may accumulate within the boundary, and some of it may be converted to some other substance (*e.g.,* entering CO may be oxidized to CO_2 within the region). There is also the possibility that more substance may be produced (*e.g.,* CO may be produced by cigarette smoking within the control volume of a room). Often, the conversion and production processes that may occur are lumped into a single category termed *reactions*. Thus, using Figure 1 as a guide, the following materials balance equation can be written for each substance of interest:

$$\begin{pmatrix}\text{Accumulation}\\ \text{rate}\end{pmatrix} = \begin{pmatrix}\text{Input}\\ \text{rate}\end{pmatrix} - \begin{pmatrix}\text{Output}\\ \text{rate}\end{pmatrix} + \begin{pmatrix}\text{Reaction}\\ \text{rate}\end{pmatrix} \tag{12}$$

The reaction rate may be positive if generation of the substance is faster than its decay, or negative if it is decaying faster than it is being produced. Likewise, the accumulation rate may be positive or negative. The *reaction* term in (12) does not imply a violation of the law of conservation of mass. Atoms are conserved, but there is no similar constraint on the chemical compounds, which may chemically change from one substance into another. It is also important to notice that each term in (12) quantifies a mass rate of change (*e.g.,* mg/s, lb/hr) and not a mass. Strictly, then, it is a mass rate balance rather than a mass balance, and (12) denotes that the rate of mass accumulation is equal to the difference between the rate the mass enters and leaves plus the net rate that the mass reacts within the defined control volume.

Frequently, (12) can be simplified. The most common simplification results when *steady state* or *equilibrium* conditions can be assumed. Equilibrium simply means that there is no accumulation of mass with time; the system has had its inputs held constant for a long enough time that any transients have had a chance to die out. Pollutant concentrations are constant. Hence the *accumulation rate* term in (12) is set equal to zero, and problems can usually be solved using just simple algebra.

A second simplification to (12) results when a substance is *conserved* within the region in question, meaning there is no reaction occurring—no radioactive decay, bacterial decomposition, or chemical decay or generation. For such conservative substances, the reaction rate in (12) is 0. Examples of substances that are typically modeled as conservative include total dissolved solids in a body of water, heavy metals in soils, and carbon dioxide in air. Radioactive radon gas in a home or

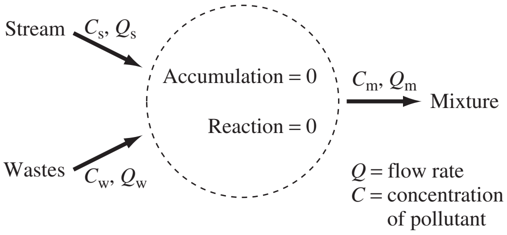


FIGURE 2 A steady-state conservative system. Pollutants enter and leave the region at the same rate.

decomposing organic wastes in a lake are examples of nonconservative substances. Often problems involving nonconservative substances can be simplified when the reaction rate is small enough to be ignored.

Steady-State Conservative Systems

The simplest systems to analyze are those in which steady state can be assumed (so the accumulation rate equals 0), and the substance in question is conservative (so the reaction rate equals 0). In these cases, (12) simplifies to the following:

$$\text{Input rate} = \text{Output rate} \tag{13}$$

Consider the steady-state conservative system shown in Figure 2. The system contained within the boundaries might be a lake, a section of a free flowing stream, or the mass of air above a city. One input to the system is a stream (of water or air, for instance) with a flow rate Q_s (volume/time) and pollutant concentration C_s (mass/volume). The other input is assumed to be a waste stream with flow rate Q_w and pollutant concentration C_w. The output is a mixture with flow rate Q_m and pollutant concentration C_m. If the pollutant is conservative, and if we assume steady state conditions, then a mass balance based on (13) allows us to write the following:

$$C_s Q_s + C_w Q_w = C_m Q_m \tag{14}$$

The following example illustrates the use of this equation. More importantly, it also provides a general algorithm for doing mass balance problems.

EXAMPLE 4 Two Polluted Streams

A stream flowing at 10.0 m^3/s has a tributary feeding into it with a flow of 5.0 m^3/s. The stream's concentration of chloride upstream of the junction is 20.0 mg/L, and the tributary chloride concentration is 40.0 mg/L. Treating chloride as a conservative substance and assuming complete mixing of the two streams, find the downstream chloride concentration.

Solution The first step in solving a mass balance problem is to sketch the problem, identify the "region" or control volume that we want to analyze, and label the variables as has been done in Figure 3 for this problem.

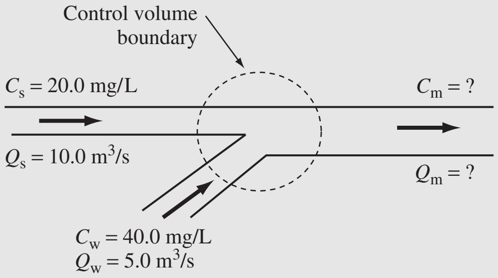


FIGURE 3 Sketch of system, variables, and quantities for a stream and tributary mixing example.

Next the mass balance equation (12) is written and simplified to match the problem's conditions

$$\cancel{\begin{pmatrix}\text{Accumulation}\\ \text{rate}\end{pmatrix}} = \begin{pmatrix}\text{Input}\\ \text{rate}\end{pmatrix} - \begin{pmatrix}\text{Output}\\ \text{rate}\end{pmatrix} + \cancel{\begin{pmatrix}\text{Reaction}\\ \text{rate}\end{pmatrix}}$$

The simplified (12) is then written in terms of the variables in the sketch

$$0 = C_s Q_s + C_w Q_w - C_m Q_m$$

The next step is to rearrange the expression to solve for the variable of interest—in this case, the chloride concentration downstream of the junction, C_m. Note that since the mixture's flow is the sum of the two stream flows, $Q_s + Q_w$ can be substituted for Q_m in this expression.

$$C_m = \frac{C_s Q_s + C_w Q_w}{Q_m} = \frac{C_s Q_s + C_w Q_w}{Q_s + Q_w}$$

The final step is to substitute the appropriate values for the known quantities into the expression, which brings us to a question of units. The units given for C are mg/L, and the units for Q are m^3/s. Taking the product of concentrations and flow rates yields mixed units of mg/L · m^3/s, which we could simplify by applying the conversion factor of $10^3 \text{ L} = 1 \text{ m}^3$. However, if we did so, we should have to reapply that same conversion factor to get the mixture concentration back into the desired units of mg/L. In problems of this sort, it is much easier to simply leave the mixed units in the expression, even though they may look awkward at first, and let them work themselves out in the calculation. The downstream concentration of chloride is thus

$$C_m = \frac{(20.0 \times 10.0 + 40.0 \times 5.0)\text{ mg/L} \cdot \text{m}^3/\text{s}}{(10.0 + 5.0)\text{ m}^3/\text{s}} = 26.7 \text{ mg/L}$$

This stream mixing problem is relatively simple, whatever the approach used. Drawing the system, labeling the variables and parameters, writing and simplifying the mass balance equation, and then solving it for the variable of interest is the same approach that will be used to solve much more complex mass balance problems later in this chapter.

Batch Systems with Nonconservative Pollutants

The simplest system with a nonconservative pollutant is a *batch system*. By definition, there is no contaminant flow into or out of a batch system, yet the contaminants in the system undergo chemical, biological, or nuclear reactions fast enough that they must be treated as nonconservative substances. A batch system (reactor) assumes that its contents are homogeneously distributed and is often referred to as a *completely mixed batch reactor* (CMBR). The bacterial concentration in a closed water storage tank may be considered a nonconservative pollutant in a batch reactor because it will change with time even though no water is fed into or withdrawn from the tank. Similarly, the concentration of carbon dioxide in a poorly ventilated room can be modeled as a nonconservative batch system because the concentration of carbon dioxide increases as people in the room breathe. For a batch reactor, (12) simplifies to

$$\text{Accumulation rate} = \text{Reaction rate} \tag{15}$$

As discussed before, the *reaction rate* is the sum of the *rates of decay,* which are negative, and the *rates of generation,* which are positive. Although the rates of reaction can exhibit many dependencies and complex relationships, most nuclear, chemical, and biochemical reaction rates can be approximated as either *zero-, first-, or second-order reaction rates*. In a zero-order reaction, the rate of reaction, $r(C)$, of the substance is not dependent on the amount of the substance present and can be expressed as

$$r(C) = k \text{ (generation)} \quad \text{or} \quad r(C) = -k \text{ (decay)} \tag{16}$$

where k is a reaction rate coefficient, which has the units of $\text{mass} \cdot \text{volume}^{-1} \cdot \text{time}^{-1}$ (*e.g.,* $\text{mg} \cdot \text{L}^{-1} \cdot \text{s}^{-1}$). The rate of evaporation of water from a bucket is a zero-order reaction because the rate of loss of the water is not dependent on the amount of water in the bucket but is only dependent on the nearly constant surface area of the water exposed to the air.

Using (15) and (16), the mass balance for the zero-order reaction of a substance in a batch reactor is

$$V\frac{dC}{dt} = -Vk$$

The equation is written as a zero-order decay, denoted by the negative sign on the right-hand side of the equation. So that each term in the mass balance has the correct units of mass/time, both the accumulation and reaction terms are multiplied by the volume of the batch reactor. Although in a batch system, the volume coefficient disappears by dividing both sides by V, it is worth remembering its presence in the initial balance equation because in other systems it may not cancel out. To solve the differential equation, the variables are separated and integrated as

$$\int_{C_0}^{C} dC = -k\int_{0}^{t} dt \tag{17}$$

which yields

$$C - C_0 = -kt$$

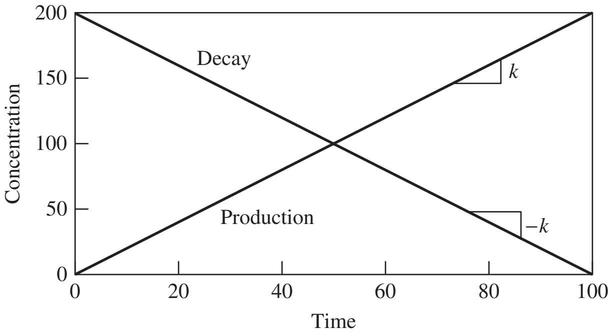


FIGURE 4 Concentration of a substance reacting in a batch system with zero-order kinetics.

Solving for concentration gives us

$$C = C_0 - kt \tag{18}$$

where C_0 is the initial concentration. Using (18) and its analog for a zero-order generation reaction, Figure 4 shows how the concentration of a substance will change with time, if it is reacting (either being generated or destroyed) with zero-order kinetics.

For all nonconservative pollutants undergoing a reaction other than zero-order, the rate of the reaction is dependent on the concentration of the pollutant present. Although decay and generation rates may be any order, the most commonly encountered reaction rate for generation is zero-order, whereas for decay it is first-order. The first-order reaction rate is

$$r(C) = kC \text{ (generation)} \quad \text{or} \quad r(C) = -kC \text{ (decay)} \tag{19}$$

where k is still a reaction rate constant, but now has the units of reciprocal time (time^{-1}). Radioactive decay of radon gas follows first-order decay—the mass that decays per given time is directly proportional to the mass that is originally present. Using (15) and (19), the mass balance for a pollutant undergoing first-order decay in a batch reactor is

$$V\frac{dC}{dt} = -VkC$$

This equation can be integrated by separation of variables and solved similarly to (17). When solved for concentration, it yields

$$C = C_0 e^{-kt} \tag{20}$$

That is, assuming a first-order reaction, the concentration of the substance in question decays exponentially. The first-order time dependence of a nonconservative pollutant's concentration in a batch system can be seen in Figure 5.

Although not nearly as common as first-order processes, sometimes a substance will decay or be generated by a second-order process. For instance, hydroxyl

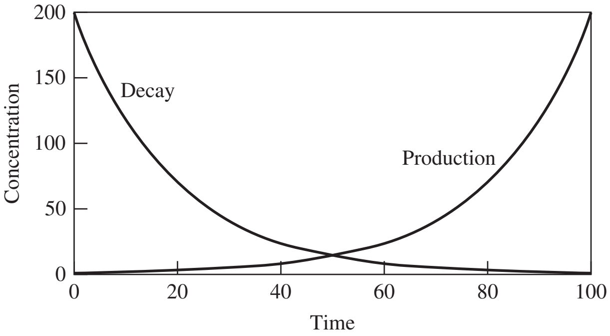


FIGURE 5 Concentration of a substance reacting in a batch system with first-order kinetics.

radical reactions with volatile organic pollutants is a key step in smog generation. However, if two hydroxyl radicals collide and react, they form a much less potent hydrogen peroxide molecule. This is a second-order reaction, since two hydroxyl radicals are consumed for each hydrogen peroxide produced. The second-order reaction rate is

$$r(C) = kC^2 \text{ (generation)} \quad \text{or} \quad r(C) = -kC^2 \text{ (decay)} \tag{21}$$

where k is now a reaction rate constant with units of (volume $\cdot$ mass$^{-1} \cdot$ time^{-1}). Again substituting (21) into (15), we have the differential equation for the second-order decay of a nonconservative substance in a batch reactor

$$V\frac{dC}{dt} = -VkC^2$$

which can be integrated and then solved for the concentration to yield

$$C = \frac{C_0}{1 + C_0kt} \tag{22}$$

Figure 6 shows how the concentration of a substance changes with time if it decays or is produced by a second-order reaction in a batch reactor.

Steady-State Systems with Nonconservative Pollutants

If we assume that steady-state conditions prevail and treat the pollutants as nonconservative, then (12) becomes

$$0 = \text{Input rate} - \text{Output rate} + \text{Reaction rate} \tag{23}$$

The batch reactor, which has just been discussed, can’t describe a steady-state system with a nonconservative substance because now there is input and output. Although there are an infinite number of other possible types of reactors, simply employing two other types of ideal reactors allows us to model a great number of environmental processes. The type of mixing in the system distinguishes between the

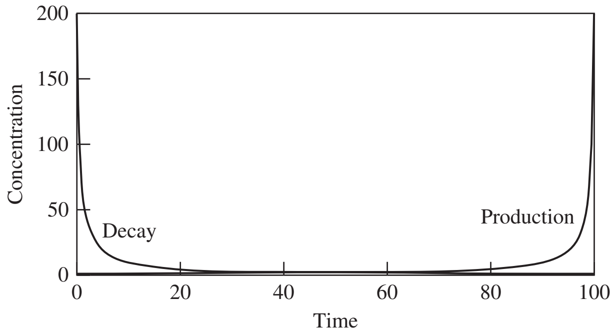


FIGURE 6 Concentration of a substance reacting in a batch system with second-order kinetics.

two other ideal reactors. In the first, the substances within the system are still homogeneously mixed as in a batch reactor. Such a system is variously termed a *continuously stirred tank reactor* (CSTR), a perfectly mixed flow reactor, and a complete mix box model. The water in a shallow pond with an inlet and outlet is typically modeled as a CSTR as is the air within a well-ventilated room. The key concept is that the concentration, C, within the CSTR container is uniform throughout. We'll give examples of CSTR behavior first and later discuss the other ideal reactor model that is often used, the *plug flow reactor* (PFR).

The reaction rate term in the right-hand side of (23) can denote either a substance's decay or generation (by making it positive or negative) and, for most environmental purposes, its rate can be approximated as either zero-, first-, or second-order. Just as for the batch reactor, for a CSTR, we assume the substance is uniformly distributed throughout a volume V, so the total amount of substance is CV. The total rate of reaction of the amount of a nonconservative substance is thus $d(CV)/dt = V\,dC/dt = Vr(C)$. So summarizing (16), (19), and (21), we can write the reaction rate expressions for a nonconservative substance:

$$\text{Zero-order, decay rate} = -Vk \tag{24}$$

$$\text{Zero-order, generation rate} = Vk \tag{25}$$

$$\text{First-order, decay rate} = -VkC \tag{26}$$

$$\text{First-order, generation rate} = VkC \tag{27}$$

$$\text{Second-order, decay rate} = -VkC^2 \tag{28}$$

$$\text{Second-order, generation rate} = VkC^2 \tag{29}$$

Thus, for example, to model a CSTR containing a substance that is decaying with a second-order rate, we combine (23) with (28) to get a final, simple, and useful expression for the mass balance involving a nonconservative pollutant in a steady-state, CSTR system:

$$\text{Input rate} = \text{Output rate} + kC^2V \tag{30}$$

EXAMPLE 5 A Polluted Lake

Consider a $10 \times 10^6\ \text{m}^3$ lake fed by a polluted stream having a flow rate of $5.0\ \text{m}^3/\text{s}$ and pollutant concentration equal to 10.0 mg/L (Figure 7). There is also a sewage outfall that discharges $0.5\ \text{m}^3/\text{s}$ of wastewater having a pollutant concentration of 100 mg/L. The stream and sewage wastes have a decay rate coefficient of 0.20/day. Assuming the pollutant is completely mixed in the lake and assuming no evaporation or other water losses or gains, find the steady-state pollutant concentration in the lake.

Solution We can conveniently use the lake in Figure 7 as our control volume. Assuming that complete and instantaneous mixing occurs in the lake—it acts as a CSTR—implies that the concentration in the lake, C, is the same as the concentration of the mix leaving the lake, C_m. The units (day^{-1}) of the reaction rate constant indicate this is a first-order reaction. Using (23) and (26):

$$\text{Input rate} = \text{Output rate} + kCV \tag{31}$$

We can find each term as follows:

There are two input sources, so the total input rate is

$$\text{Input rate} = Q_s C_s + Q_w C_w$$

The output rate is

$$\text{Output rate} = Q_m C_m = (Q_s + Q_w)C$$

(31) then becomes

$$Q_s C_s + Q_w C_w = (Q_s + Q_w)C + kCV$$

And rearranging to solve for C,

$$C = \frac{Q_s C_s + Q_w C_w}{Q_s + Q_w + kV}$$

$$= \frac{5.0\ \text{m}^3/\text{s} \times 10.0\ \text{mg/L} + 0.5\ \text{m}^3/\text{s} \times 100.0\ \text{mg/L}}{(5.0 + 0.5)\ \text{m}^3/\text{s} + \dfrac{0.20/\text{d} \times 10.0 \times 10^6\ \text{m}^3}{24\ \text{hr/d} \times 3600\ \text{s/hr}}}$$

So,

$$C = \frac{100}{28.65} = 3.5\ \text{mg/L}$$

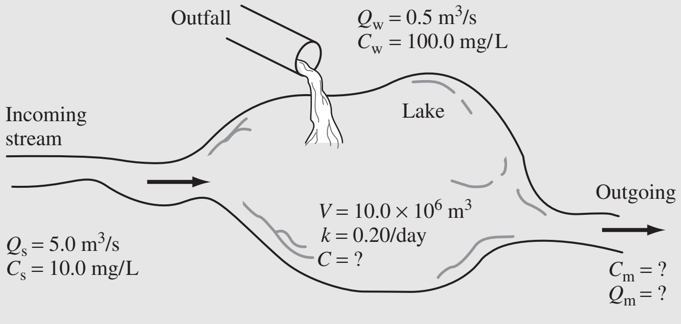


FIGURE 7 A lake with a nonconservative pollutant.

Idealized models involving nonconservative pollutants in completely mixed, steady-state systems are used to analyze a variety of commonly encountered water pollution problems such as the one shown in the previous example. The same simple models can be applied to certain problems involving air quality, as the following example demonstrates.

EXAMPLE 6 A Smoky Bar

A bar with volume 500 m^3 has 50 smokers in it, each smoking 2 cigarettes per hour (see Figure 8). An individual cigarette emits, among other things, about 1.4 mg of formaldehyde (HCHO). Formaldehyde converts to carbon dioxide with a reaction rate coefficient $k = 0.40/\text{hr}$. Fresh air enters the bar at the rate of 1,000 m^3/hr, and stale air leaves at the same rate. Assuming complete mixing, estimate the steady-state concentration of formaldehyde in the air. At 25°C and 1 atm of pressure, how does the result compare with the threshold for eye irritation of 0.05 ppm?

Solution The bar's building acts as a CSTR reactor, and the complete mixing inside means the concentration of formaldehyde C in the bar is the same as the concentration in the air leaving the bar. Since the formaldehyde concentration in fresh air can be considered 0, the input rate in (23) is also 0. Our mass balance equation is then

$$\text{Output rate} = \text{Reaction rate} \tag{32}$$

However, both a generation term (the cigarette smoking) and a decay term (the conversion of formaldehyde to carbon dioxide) are contributing to the reaction rate. If we call the generation rate, G, we can write

$$G = 50 \text{ smokers} \times 2 \text{ cigs/hr} \times 1.4 \text{ mg/cig} = 140 \text{ mg/hr}$$

We can then express (32) in terms of the problem's variables and (26) as

$$QC = G - kCV$$

so

$$C = \frac{G}{Q + kV} = \frac{140 \text{ mg/hr}}{1{,}000 \text{ m}^3\text{/hr} + (0.40/\text{hr}) \times 500 \text{ m}^3}$$
$$= 0.117 \text{ mg/m}^3$$

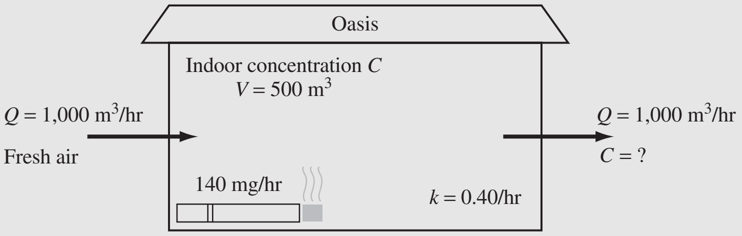


FIGURE 8 Tobacco smoke in a bar.

We will use (9) to convert mg/m^3 to ppm. The molecular weight of formaldehyde is 30, so

$$\text{HCHO} = \frac{C\ (\text{mg/m}^3) \times 24.465}{\text{mol wt}} = \frac{0.117 \times 24.465}{30} = 0.095\ \text{ppm}$$

This is nearly double the 0.05 ppm threshold for eye irritation.

Besides a CSTR, the other type of ideal reactor that is often useful to model pollutants as they flow through a system is a *plug flow reactor* (PFR). A PFR can be visualized as a long pipe or channel in which there is no mixing of the pollutant along its length between the inlet and outlet. A PFR could also be seen as a conveyor belt carrying a single-file line of bottles in which reactions can happen within each bottle, but there is no mixing of the contents of one bottle with another. The behavior of a pollutant being carried in a river or in the jet stream in the Earth's upper atmosphere could be usefully represented as a PFR system. The key difference between a PFR and CSTR is that in a PFR, there is no mixing of one parcel of fluid with other parcels in front or in back of it in the control volume, whereas in a CSTR, all of the fluid in the container is continuously and completely mixed. (23) applies to both a CSTR and PFR at steady state, but for a PFR, we cannot make the simplification that the concentration everywhere in the control volume and in the fluid leaving the region is the same, as we did in the CSTR examples. The pollutant concentration in a parcel of fluid changes as the parcel progresses through the PFR. Intuitively, it can then be seen that a PFR acts like a conveyor belt of differentially thin batch reactors translating through the control volume. When the pollutant enters the PFR at concentration, C_0, then it will take a given time, t, to move the length of the control volume and will leave the PFR with a concentration just as it would if it had been in a batch reactor for that period of time. Thus, a substance decaying with a zero-, first-, or second-order rate will leave a PFR with a concentration given by (18), (20), and (22), respectively, with the understanding that t in each equation is the residence time of the fluid in the control volume and is given by

$$t = l/v = V/Q \tag{33}$$

where l is the length of the PFR, v is the fluid velocity, V is the PFR control volume, and Q is the fluid flowrate.

EXAMPLE 7 Young Salmon Migration

Every year, herons, seagulls, eagles, and other birds mass along a 4.75 km stretch of stream connecting a lake to the ocean to catch the fingerling salmon as they migrate downstream to the sea. The birds are efficient fishermen and will consume 10,000 fingerlings per kilometer of stream each hour regardless of the number of the salmon in the stream. In other words, there are enough salmon; the birds are only limited by how fast they can catch and eat the fish. The stream's average cross-sectional area is 20 m^2, and the salmon move downstream with the stream's flow rate of 700 m^3/min. If there are 7 fingerlings per m^3 in the water entering the stream, what is the concentration of salmon that reach the ocean when the birds are feeding?

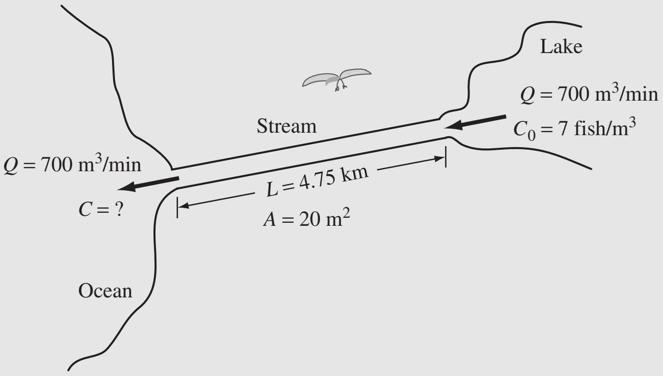


FIGURE 9 Birds fishing along a salmon stream.

Solution First we draw a figure depicting the stream as the control volume and the variables (Figure 9).

Since the birds eat the salmon at a steady rate that is not dependent on the concentration of salmon in the stream, the rate of salmon consumption is zero-order. So,

$$k = \frac{10{,}000\ \text{fish} \cdot \text{km}^{-1} \cdot \text{hr}^{-1}}{20\ \text{m}^2 \times 1{,}000\ \text{m/km}} = 0.50\ \text{fish} \cdot \text{m}^{-3} \cdot \text{hr}^{-1}$$

For a steady-state PFR, (23) becomes (18),

$$C = C_0 - kt$$

The residence time, t, of the stream can be calculated using (33) as

$$t = \frac{V}{Q} = \frac{4.75\ \text{km} \times 20\ \text{m}^2 \times 1{,}000\ \text{m/km}}{700\ \text{m}^3/\text{min} \times 60\ \text{min/hr}} = 2.26\ \text{hr}$$

and the concentration of fish reaching the ocean is

$$C = 7\ \text{fish/m}^3 - 0.50\ \text{fish} \cdot \text{m}^{-3} \cdot \text{hr}^{-1} \times 2.26\ \text{hr} = 5.9\ \text{fish/m}^3$$

Step Function Response

So far, we have computed steady-state concentrations in environmental systems that are contaminated with either conservative or nonconservative pollutants. Let us now extend the analysis to include conditions that are not steady state. Quite often, we will be interested in how the concentration will change with time when there is a sudden change in the amount of pollution entering the system. This is known as the *step function response* of the system.

In Figure 10, the environmental system to be modeled has been drawn as if it were a box of volume V that has flow rate Q in and out of the box.

Let us assume the contents of the box are at all times completely mixed (a CSTR model) so that the pollutant concentration C in the box is the same as the concentration leaving the box. The total mass of pollutant in the box is therefore

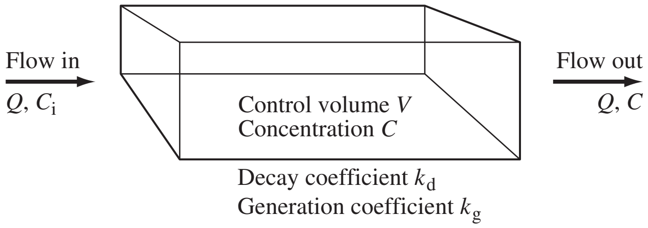


FIGURE 10 A box model for a transient analysis.

VC, and the rate of accumulation of pollutant in the box is VdC/dt. Let us designate the concentration of pollutant entering the box as C_i. We'll
also assume there are both production and decay components of the reaction rate and designate the decay coefficient k_d and the generation coefficient k_g. However, as is most common, the decay will be first-order, so k_d's units are time^{-1}, whereas the generation is zero-order, so k_g's units are mass · volume^{-1} · time^{-1}. From (12), we can write

$$\begin{pmatrix}\text{Accumulation}\\\text{rate}\end{pmatrix} = \begin{pmatrix}\text{Input}\\\text{rate}\end{pmatrix} - \begin{pmatrix}\text{Output}\\\text{rate}\end{pmatrix} + \begin{pmatrix}\text{Reaction}\\\text{rate}\end{pmatrix}$$

$$V\frac{dC}{dt} = QC_i - QC - Vk_dC + k_gV \tag{34}$$

where

V = box volume (m^3)
C = concentration in the box and exiting waste stream (g/m^3)
C_i = concentration of pollutants entering the box (g/m^3)
Q = the total flow rate in and out of the box (m^3/hr)
k_d = decay rate coefficient (hr^{-1})
k_g = production rate coefficient (g · m · hr^{-1})

The units given in the preceding list are representative of those that might be encountered; any consistent set will do.

An easy way to find the steady-state solution to (34) is simply to set $dC/dt = 0$, which yields

$$C_\infty = \frac{QC_i + k_gV}{Q + k_dV} \tag{35}$$

where C_∞ is the concentration in the box at time $t = \infty$. Our concern now, though, is with the concentration before it reaches steady state, so we must solve (34). Rearranging (34) gives

$$\frac{dC}{dt} = -\left(\frac{Q}{V} + k_d\right)\cdot\left(C - \frac{QC_i + k_gV}{Q + k_dV}\right) \tag{36}$$

which, using (35), can be rewritten as

$$\frac{dC}{dt} = -\left(\frac{Q}{V} + k_d\right)\cdot(C - C_\infty) \tag{37}$$

One way to solve this differential equation is to make a change of variable. If we let

$$y = C - C_\infty \tag{38}$$

then

$$\frac{dy}{dt} = \frac{dC}{dt} \tag{39}$$

so (37) becomes

$$\frac{dy}{dt} = -\left(\frac{Q}{V} + k_{\mathrm{d}}\right)y \tag{40}$$

This is a differential equation, which we can solve by separating the variables to give

$$\int_{y_0}^{y} dy = -\left(\frac{Q}{V} + k_{\mathrm{d}}\right)\int_{0}^{t} dt \tag{41}$$

where y_0 is the value of y at $t = 0$. Integrating gives

$$y = y_0 e^{-\left(k_{\mathrm{d}} + \frac{Q}{V}\right)t} \tag{42}$$

If C_0 is the concentration in the box at time $t = 0$, then from (38) we get

$$y_0 = C_0 - C_\infty \tag{43}$$

Substituting (38) and (43) into (42) yields

$$C - C_\infty = (C_0 - C_\infty)e^{-\left(k_{\mathrm{d}} + \frac{Q}{V}\right)t} \tag{44}$$

Solving for the concentration in the box, writing it as a function of time $C(t)$, and expressing the exponential as exp( ) gives

$$C(t) = C_\infty + (C_0 - C_\infty)\exp\left[-\left(k_{\mathrm{d}} + \frac{Q}{V}\right)t\right] \tag{45}$$

Equation (45) should make some sense. At time $t = 0$, the exponential function equals 1 and $C = C_0$. At $t = \infty$, the exponential term equals 0, and $C = C_\infty$. Equation (45) is plotted in Figure 11.

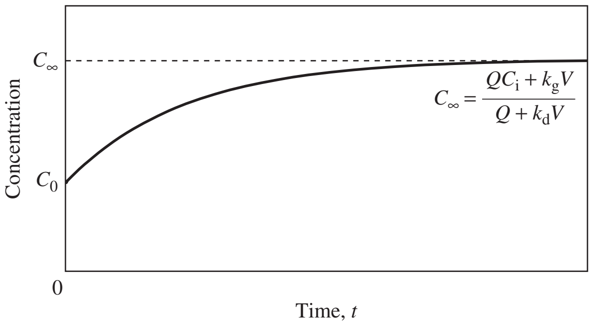


FIGURE 11 Step function response for the CSTR box model.

EXAMPLE 8 **The Smoky Bar Revisited**

The bar in Example 6 had volume 500 m^3 with fresh air entering at the rate of 1,000 m^3/hr. Suppose when the bar opens at 5 P.M., the air is clean. If formaldehyde, with decay rate $k_d = 0.40$/hr, is emitted from cigarette smoke at the constant rate of 140 mg/hr starting at 5 P.M., what would the concentration be at 6 P.M.?

Solution In this case, $Q = 1{,}000\ m^3/hr$, $V = 500\ m^3$, $G = k_g, V = 140$ mg/hr, $C_i = 0$, and $k_d = 0.40$/hr. The steady-state concentration is found using (35):

$$C_\infty = \frac{QC_i + k_gV}{Q + k_dV} = \frac{G}{Q + k_dV} = \frac{140\ \text{mg/hr}}{1{,}000\ \text{m}^3/\text{hr} + 0.40/\text{hr} \times 500\ \text{m}^3}$$

$$= 0.117\ \text{mg/m}^3$$

This agrees with the result obtained in Example 6. To find the concentration at any time after 5 P.M., we can apply (45) with $C_0 = 0$.

$$C(t) = C_\infty\left\{1 - \exp\left[-\left(k_d + \frac{Q}{V}\right)t\right]\right\}$$

$$= 0.117\{1 - \exp[-(0.40 + 1{,}000/500)t]\}$$

at 6 P.M., $t = 1$ hr, so

$$C(1\,\text{hr}) = 0.117[1 - \exp(-2.4 \times 1)] = 0.106\ \text{mg/m}^3$$

To further demonstrate the use of (45), let us reconsider the lake analyzed in Example 5. This time we will assume that the outfall suddenly stops draining into the lake, so its contribution to the lake's pollution stops.

EXAMPLE 9 **A Sudden Decrease in Pollutants Discharged into a Lake**

Consider the $10 \times 10^6\ m^3$ lake analyzed in Example 5 which, under a conditions given, was found to have a steady-state pollution concentration of 3.5 mg/L. The pollution is nonconservative with reaction-rate constant $k_d = 0.20$/day. Suppose the condition of the lake is deemed unacceptable. To solve the problem, it is decided to completely divert the sewage outfall from the lake, eliminating it as a source of pollution. The incoming stream still has flow $Q_s = 5.0\ m^3/s$ and concentration $C_s = 10.0$ mg/L. With the sewage outfall removed, the outgoing flow Q is also 5.0 m^3/s. Assuming complete-mix conditions, find the concentration of pollutant in the lake one week after the diversion, and find the new final steady-state concentration.

Solution For this situation,

$$C_0 = 3.5\ \text{mg/L}$$
$$V = 10 \times 10^6\ \text{m}^3$$
$$Q = Q_s = 5.0\ \text{m}^3/\text{s} \times 3{,}600\ \text{s/hr} \times 24\,\text{hr/day} = 43.2 \times 10^4\ \text{m}^3/\text{day}$$

$$C_s = 10.0 \text{ mg/L} = 10.0 \times 10^3 \text{ mg/m}^3$$

$$k_d = 0.20/\text{day}$$

The total rate at which pollution is entering the lake from the incoming stream is

$$Q_s C_s = 43.2 \times 10^4 \text{ m}^3/\text{day} \times 10.0 \times 10^3 \text{ mg/m}^3 = 43.2 \times 10^8 \text{ mg/day}$$

The steady-state concentration can be obtained from (35)

$$C_\infty = \frac{QC_s}{Q + k_d V} = \frac{43.2 \times 10^8 \text{ mg/day}}{43.2 \times 10^4 \text{ m}^3/\text{day} + 0.20/\text{day} \times 10^7 \text{ m}^3}$$

$$= 1.8 \times 10^3 \text{ mg/m}^3 = 1.8 \text{ mg/L}$$

Using (45), we can find the concentration in the lake one week after the drop in pollution from the outfall:

$$C(t) = C_\infty + (C_0 - C_\infty)\exp\left[-\left(k_d + \frac{Q}{V}\right)t\right]$$

$$C(7 \text{ days}) = 1.8 + (3.5 - 1.8)\exp\left[-\left(0.2/\text{day} + \frac{43.2 \times 10^4 \text{ m}^3/\text{day}}{10 \times 10^6 \text{ m}^3}\right) \times 7 \text{ days}\right]$$

$$C(7 \text{ days}) = 2.1 \text{ mg/L}$$

Figure 12 shows the decrease in contaminant concentration for this example.

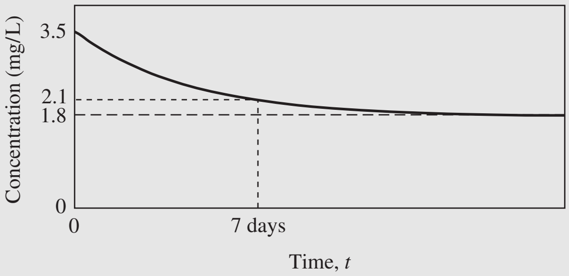


FIGURE 12 The contaminant concentration profile for Example 9.

4 Energy Fundamentals

Just as we are able to use the law of conservation of mass to write mass balance equations that are fundamental to understanding and analyzing the flow of materials, we can use the *first law of thermodynamics* to write energy balance equations that will help us analyze energy flows.

One definition of *energy* is that it is the capacity for doing work, where *work* can be described by the product of force and the displacement of an object caused by that force. A simple interpretation of the *second law of thermodynamics* suggests that when work is done, there will always be some inefficiency, that is, some portion of the energy put into the process will end up as waste heat. How that waste heat affects the environment is an important consideration in the study of environmental engineering and science.

Another important term to be familiar with is *power*. Power is the *rate* of doing work. It has units of energy per unit of time. In SI units, power is given in joules per second (J/s) or kilojoules per second (kJ/s). To honor the Scottish engineer James Watt, who developed the reciprocating steam engine, the joule per second has been named the watt (1 J/s = 1 W = 3.412 Btu/hr).

The First Law of Thermodynamics

The first law of thermodynamics says, simply, that energy can neither be created nor destroyed. Energy may change forms in any given process, as when chemical energy in a fuel is converted to heat and electricity in a power plant or when the potential energy of water behind a dam is converted to mechanical energy as it spins a turbine in a hydroelectric plant. No matter what is happening, the first law says we should be able to account for every bit of energy as it takes part in the process under study, so that in the end, we have just as much as we had in the beginning. With proper accounting, even nuclear reactions involving conversion of mass to energy can be treated.

To apply the first law, it is necessary to define the system being studied, much as was done in the analysis of mass flows. The system (control volume) can be anything that we want to draw an imaginary boundary around; it can be an automobile engine, or a nuclear power plant, or a volume of gas emitted from a smokestack. Later when we explore the topic of global temperature equilibrium, the system will be the Earth itself. After a boundary has been defined, the rest of the universe becomes the *surroundings*. Just because a boundary has been defined, however, does not mean that energy and/or materials cannot flow across that boundary. Systems in which both energy and matter can flow across the boundary are referred to as *open systems,* whereas those in which energy is allowed to flow across the boundary, but matter is not, are called *closed systems*.

Since energy is conserved, we can write the following for whatever system we have defined:

$$\begin{pmatrix}\textit{Total}\text{ energy}\\ \text{crossing boundary}\\ \text{as heat and work}\end{pmatrix} + \begin{pmatrix}\textit{Total}\text{ energy}\\ \text{of mass}\\ \text{entering system}\end{pmatrix} - \begin{pmatrix}\textit{Total}\text{ energy}\\ \text{of mass}\\ \text{leaving system}\end{pmatrix} = \begin{pmatrix}\textit{Net}\text{ change}\\ \text{of energy in}\\ \text{the system}\end{pmatrix} \tag{46}$$

For closed systems, there is no movement of mass across the boundary, so the second and third term drop out of the equation. The accumulation of energy represented by the right side of (46) may cause changes in the observable, macroscopic forms of energy, such as kinetic and potential energies, or microscopic forms related to the atomic and molecular structure of the system. Those microscopic forms of energy include the kinetic energies of molecules and the energies associated with the forces acting between molecules, between atoms within molecules, and within atoms. The sum of those microscopic forms of energy is called the system's *internal energy* and is represented by the symbol U. The *total energy* E that a substance possesses can be described then as the sum of its internal energy U, its kinetic energy KE, and its potential energy PE:

$$E = U + KE + PE \tag{47}$$

In many applications of (46), the net energy added to a system will cause an increase in temperature. Waste heat from a power plant, for example, will raise the temperature of cooling water drawn into its condenser. The amount of energy needed to raise the temperature of a unit mass of a substance by 1 degree is called the *specific heat*. The specific heat of water is the basis for two important units of energy, namely the *British thermal unit* (Btu), which is defined to be the energy required to raise 1 lb of water by 1°F, and the *kilocalorie,* which is the energy required to raise 1 kg of water by 1°C. In the definitions just given, the assumed temperature of the water is 15°C (59°F). Since kilocalories are no longer a preferred energy unit, values of specific heat in the SI system are given in kJ/kg°C, where 1 kcal/kg°C = 1 Btu/lb°F = 4.184 kJ/kg°C.

For most applications, the specific heat of a liquid or solid can be treated as a simple quantity that varies slightly with temperature. For gases, on the other hand, the concept of specific heat is complicated by the fact that some of the heat energy absorbed by a gas may cause an increase in temperature, and some may cause the gas to expand, doing work on its environment. That means it takes more energy to raise the temperature of a gas that is allowed to expand than the amount needed if the gas is kept at constant volume. The *specific heat at constant volume* c_v is used when a gas does not change volume as it is heated or cooled, or if the volume is allowed to vary but is brought back to its starting value at the end of the process. Similarly, the *specific heat at constant pressure* c_p applies for systems that do not change pressure. For incompressible substances, that is, liquids and solids under the usual circumstances, c_v and c_p are identical. For gases, c_p is greater than c_v.

The added complications associated with systems that change pressure and volume are most easily handled by introducing another thermodynamic property of a substance called *enthalpy*. The enthalpy H of a substance is defined as

$$H = U + PV \tag{48}$$

where U is its internal energy, P is its pressure, and V is its volume. The enthalpy of a unit mass of a substance depends only on its temperature. It has energy units (kJ or Btu) and historically was referred to as a system's "heat content." Since heat is correctly defined only in terms of energy transfer across a system's boundaries, heat content is a somewhat misleading descriptor and is not used much anymore.

When a process occurs without a change of volume, the relationship between internal energy and temperature change is given by

$$\Delta U = m c_v \Delta T \tag{49}$$

The analogous equation for changes that occur under constant pressure involves enthalpy

$$\Delta H = m c_p \Delta T \tag{50}$$

For many environmental systems, the substances being heated are solids or liquids for which $c_v = c_p = c$ and $\Delta U = \Delta H$. We can write the following equation for the energy needed to raise the temperature of mass m by an amount ΔT:

$$\text{Change in stored energy} = m c \Delta T \tag{51}$$

Table 3 provides some examples of specific heat for several selected substances. It is worth noting that water has by far the highest specific heat of the substances listed; in fact, it is higher than almost all common substances. This is one

TABLE 3

Specific Heat Capacity *c* of Selected Substances

	(kcal/kg°C, Btu/lb°F)	(kJ/kg°C)
Water (15°C)	1.00	4.18
Air	0.24	1.01
Aluminum	0.22	0.92
Copper	0.09	0.39
Dry soil	0.20	0.84
Ice	0.50	2.09
Steam (100°C)[a]	0.48	2.01
Water vapor (20°C)[a]	0.45	1.88

[a] Constant pressure values.

of water's very unusual properties and is in large part responsible for the major effect the oceans have on moderating temperature variations of coastal areas.

EXAMPLE 10 A Water Heater

How long would it take to heat the water in a 40-gallon electric water heater from 50°F to 140°F if the heating element delivers 5 kW? Assume all of the electrical energy is converted to heat in the water, neglect the energy required to raise the temperature of the tank itself, and neglect any heat losses from the tank to the environment.

Solution The first thing to note is that the electric input is expressed in kilowatts, which is a measure of the *rate* of energy input (*i.e.*, power). To get total energy delivered to the water, we must multiply rate by time. Letting Δt be the number of hours that the heating element is on gives

$$\text{Energy input} = 5\text{ kW} \times \Delta t\text{ hrs} = 5\Delta t\text{ kWhr}$$

Assuming no losses from the tank and no water withdrawn from the tank during the heating period, there is no energy output:

$$\text{Energy output} = 0$$

The change in energy stored corresponds to the water warming from 50°F to 140°F. Using (51) along with the fact that water weighs 8.34 lb/gal gives

$$\begin{aligned}\text{Change in stored energy} &= mc\Delta T\\ &= 40\text{ gal} \times 8.34\text{ lb/gal} \times 1\text{ Btu/lb°F} \times (140 - 50)\text{°F}\\ &= 30 \times 10^3\text{ Btu}\end{aligned}$$

Setting the energy input equal to the change in internal energy and converting units using Table 1 yields

$$\begin{aligned}5\Delta t\text{ kWhr} \times 3{,}412\text{ Btu/kWhr} &= 30 \times 10^3\text{ Btu}\\ \Delta t &= 1.76\text{ hr}\end{aligned}$$

There are two key assumptions implicit in (51). First, the specific heat is assumed to be constant over the temperature range in question, although in actuality it does vary slightly. Second, (51) assumes that there is no change of *phase* as would occur if the substance were to freeze or melt (liquid-solid phase change) or evaporate or condense (liquid-gas phase change).

When a substance changes phase, energy is absorbed or released without a change in temperature. The energy required to cause a phase change of a unit mass from solid to liquid (melting) at the same pressure is called the *latent heat of fusion* or, more correctly, the *enthalpy of fusion*. Similarly, the energy required to change phase from liquid to vapor at constant pressure is called the *latent heat of vaporization* or the *enthalpy of vaporization*. For example, 333 kJ will melt 1 kg of ice (144 Btu/lb), whereas 2,257 kJ are required to convert 1 kg of water at 100°C to steam (970 Btu/lb). When steam condenses or when water freezes, those same amounts of energy are released. When a substance changes temperature as heat is added, the process is referred to as *sensible heating*. When the addition of heat causes a phase change, as is the case when ice is melting or water is boiling, the addition is called *latent heat*. To account for the latent heat stored in a substance, we can include the following in our energy balance:

$$\text{Energy released or absorbed in phase change} = mL \qquad (52)$$

where m is the mass and L is the latent heat of fusion or vaporization.

Figure 13 illustrates the concepts of latent heat and specific heat for water as it passes through its three phases from ice, to water, to steam.

Values of specific heat, heats of vaporization and fusion, and density for water are given in Table 4 for both SI and USCS units. An additional entry has been included in the table that shows the heat of vaporization for water at 15°C. This is a useful number that can be used to estimate the amount of energy required to cause

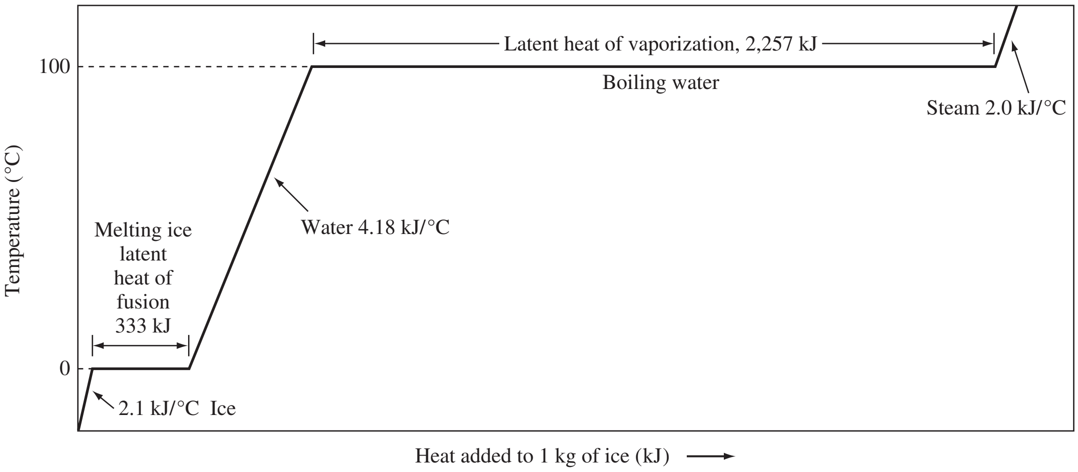


FIGURE 13 Heat needed to convert 1 kg of ice to steam. To change the temperature of 1 kg of ice, 2.1 kJ/°C are needed. To completely melt that ice requires another 333 kJ (heat of fusion). Raising the temperature of that water requires 4.184 kJ/°C, and converting it to steam requires another 2,257 kJ (latent heat of vaporization). Raising the temperature of 1 kg of steam (at atmospheric pressure) requires another 2.0 kJ/°C.

TABLE 4

Important Physical Properties of Water

Property	SI Units	USCS Units
Specific heat (15°C)	4.184 kJ/kg°C	1.00 Btu/lb°F
Heat of vaporization (100°C)	2,257 kJ/kg	972 Btu/lb
Heat of vaporization (15°C)	2,465 kJ/kg	1,060 Btu/lb
Heat of fusion	333 kJ/kg	144 Btu/lb
Density (at 4°C)	1,000 kg/m^3	62.4 lb/ft^3 (8.34 lb/gal)

surface water on the Earth to evaporate. The value of 15°C has been picked as the starting temperature since that is approximately the current average surface temperature of the globe.

One way to demonstrate the concept of the heat of vaporization, while at the same time introducing an important component of the global energy balance, is to estimate the energy required to power the global hydrologic cycle.

EXAMPLE 11 Power for the Hydrologic Cycle

Global rainfall has been estimated to average about 1 m of water per year across the entire 5.10×10^{14} m^2 of the Earth's surface. Find the energy required to cause that much water to evaporate each year. Compare this to the estimated 2007 world energy consumption of 4.7×10^{17} kJ and compare it to the average rate at which sunlight strikes the surface of the Earth, which is about 168 W/m^2.

Solution In Table 4, the energy required to vaporize 1 kg of 15°C water (roughly the average global temperature) is given as 2,465 kJ. The total energy required to vaporize all of that water is

$$\begin{aligned}\text{Increase in stored energy} &= 1\,\text{m/yr} \times 5.10 \times 10^{14}\,\text{m}^2 \times 10^3\,\text{kg/m}^3 \times 2{,}465\,\text{kJ/kg} \\ &= 1.25 \times 10^{21}\,\text{kJ/yr}\end{aligned}$$

This is roughly 2,700 times the 4.7×10^{17} kJ/yr of energy we use to power our society.

Averaged over the globe, the energy required to power the hydrologic cycle is

$$\frac{1.25 \times 10^{24}\,\text{J/yr} \times 1\dfrac{\text{W}}{\text{J/s}}}{365\,\text{day/yr} \times 24\,\text{hr/day} \times 3{,}600\,\text{s/hr} \times 5.10 \times 10^{14}\,\text{m}^2} = 78.0\ \text{W/m}^2$$

which is equivalent to almost half of the 168 W/m^2 of incoming sunlight striking the Earth's surface. It might also be noted that the energy required to raise the water vapor high into the atmosphere after it has evaporated is negligible compared to the heat of vaporization (see Problem 27 at the end of this chapter).

Many practical environmental engineering problems involve the flow of both matter and energy across system boundaries (open systems). For example, it is common for a hot liquid, usually water, to be used to deliver heat to a pollution control process or, the opposite, for water to be used as a coolant to remove heat from a process. In such cases, there are energy flow rates and fluid flow rates, and (51) needs to be modified as follows:

$$\text{Rate of change of stored energy} = \dot{m}\, c\, \Delta T \tag{53}$$

where $\dot{m}$ is the mass flow rate across the system boundary, given by the product of fluid flow rate and density, and ΔT is the change in temperature of the fluid that is carrying the heat to, or away from, the process. For example, if water is being used to cool a steam power plant, then $\dot{m}$ would be the mass flow rate of coolant, and ΔT would be the increase in temperature of the cooling water as it passes through the steam plant's condenser. Typical units for energy rates include watts, Btu/hr, or kJ/s, whereas mass flow rates might typically be in kg/s or lb/hr.

The use of a local river for power plant cooling is common, and the following example illustrates the approach that can be taken to compute the increase in river temperature that results.

EXAMPLE 12 Thermal Pollution of a River

A coal-fired power plant converts one-third of the coal's energy into electrical energy. The electrical power output of the plant is 1,000 MW. The other two-thirds of the energy content of the fuel is rejected to the environment as waste heat. About 15 percent of the waste heat goes up the smokestack, and the other 85 percent is taken away by cooling water that is drawn from a nearby river. The river has an upstream flow of 100.0 m^3/s and a temperature of 20.0°C.

a. If the cooling water is only allowed to rise in temperature by 10.0°C, what flow rate from the stream would be required?
b. What would be the river temperature just after it receives the heated cooling water?

Solution Since 1,000 MW represents one-third of the power delivered to the plant by fuel, the total rate at which energy enters the power plant is

$$\text{Input power} = \frac{\text{Output power}}{\text{Efficiency}} = \frac{1{,}000\,\text{MW}_e}{1/3} = 3{,}000\,\text{MW}_t$$

Notice the subscript on the input and output power in the preceding equation. To help keep track of the various forms of energy, it is common to use MW_t for thermal power and MW_e for electrical power.

Total losses to the cooling water and stack are therefore 3,000 MW − 1,000 MW = 2,000 MW. Of that 2,000 MW,

$$\text{Stack losses} = 0.15 \times 2{,}000\,\text{MW}_t = 300\,\text{MW}_t$$

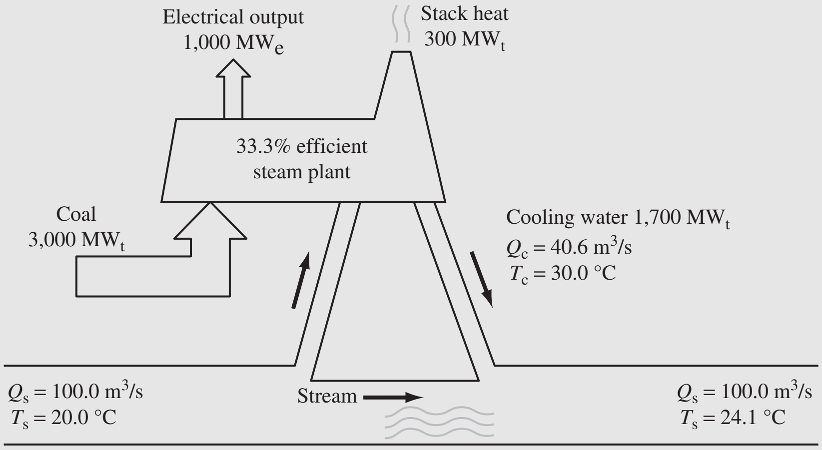


FIGURE 14 Cooling water energy balance for the 33.3 percent efficient, 1,000 MW$_e$ power plant in Example 12.

and

$$\text{Coolant losses} = 0.85 \times 2{,}000\ \text{MW}_t = 1{,}700\ \text{MW}_t$$

a. Finding the cooling water needed to remove 1,700 MW$_t$ with a temperature increase ΔT of 10.0°C will require the use of (1.53) along with the specific heat of water, 4,184 J/kg°C,given in Table 4:

$$\text{Rate of change in internal energy} = \dot{m}\,c\,\Delta T$$

$$1{,}700\ \text{MW}_t = \dot{m}\,\text{kg/s} \times 4{,}184\,\text{J/kg°C} \times 10.0\text{°C} \times 1\ \text{MW}/(10^6\,\text{J/s})$$

$$\dot{m} = \frac{1{,}700}{4{,}184 \times 10.0 \times 10^{-6}} = 40.6 \times 10^3\,\text{kg/s}$$

or, since 1,000 kg equals 1 m^3 of water, the flow rate is 40.6 m^3/s.

b. To find the new temperature of the river, we can use (53) with 1,700 MW$_t$ being released into the river, which again has a flow rate of 100.0 m^3/s.

$$\text{Rate of change in internal energy} = \dot{m}\,c\,\Delta T$$

$$\Delta T = \frac{1{,}700\,\text{MW} \times \left(\dfrac{1 \times 10^6\,\text{J/s}}{\text{MW}}\right)}{100.00\,\text{m}^3/\text{s} \times 10^3\,\text{kg/m}^3 \times 4{,}184\,\text{J/kg°C}} = 4.1\text{°C}$$

so the temperature of the river will be elevated by 4.1°C making it 24.1°C. The results of the calculations just performed are shown in Figure 14.

The Second Law of Thermodynamics

In Example 12, you will notice that a relatively modest fraction of the fuel energy contained in the coal actually was converted to the desired output, electrical power, and a rather large amount of the fuel energy ended up as waste heat rejected to the environment. The second law of thermodynamics says that there will always be some waste heat; that is, it is impossible to devise a machine that can convert heat

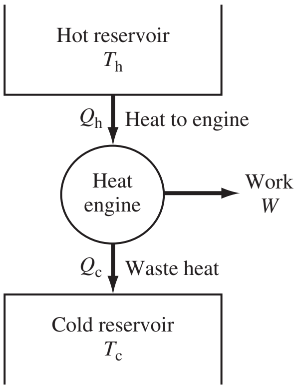


FIGURE 15 Definition of terms for a Carnot engine.

to work with 100 percent efficiency. There will always be "losses" (although, by the first law, the energy is not lost, it is merely converted into the lower quality, less useful form of low-temperature heat).

The steam-electric plant just described is an example of a *heat engine,* a device studied at some length in thermodynamics. One way to view the steam plant is that it is a machine that takes heat from a high-temperature source (the burning fuel), converts some of it into work (the electrical output), and rejects the remainder into a low-temperature reservoir (the river and the atmosphere). It turns out that the maximum efficiency that our steam plant can possibly have depends on how high the source temperature is and how low the temperature is of the reservoir accepting the rejected heat. It is analogous to trying to run a turbine using water that flows from a higher elevation to a lower one. The greater the difference in elevation, the more power can be extracted.

Figure 15 shows a theoretical heat engine operating between two heat reservoirs, one at temperature T_h and one at T_c. An amount of heat energy Q_h is transferred from the hot reservoir to the heat engine. The engine does work W and rejects an amount of waste heat Q_c to the cold reservoir.

The efficiency of this engine is the ratio of the work delivered by the engine to the amount of heat energy taken from the hot reservoir:

$$\text{Efficiency}\,\eta = \frac{W}{Q_h} \tag{54}$$

The most efficient heat engine that could possibly operate between the two heat reservoirs is called a *Carnot* engine after the French engineer Sadi Carnot, who first developed the explanation in the 1820s. Analysis of Carnot engines shows that the most efficient engine possible, operating between two temperatures, T_h and T_c, has an efficiency of

$$\eta_{\max} = 1 - \frac{T_c}{T_h} \tag{55}$$

where these are absolute temperatures measured using either the Kelvin scale or Rankine scale. Conversions from Celsius to Kelvin, and Fahrenheit to Rankine are

$$\text{K} = {}^\circ\text{C} + 273.15 \tag{56}$$
$$\text{R} = {}^\circ\text{F} + 459.67 \tag{57}$$

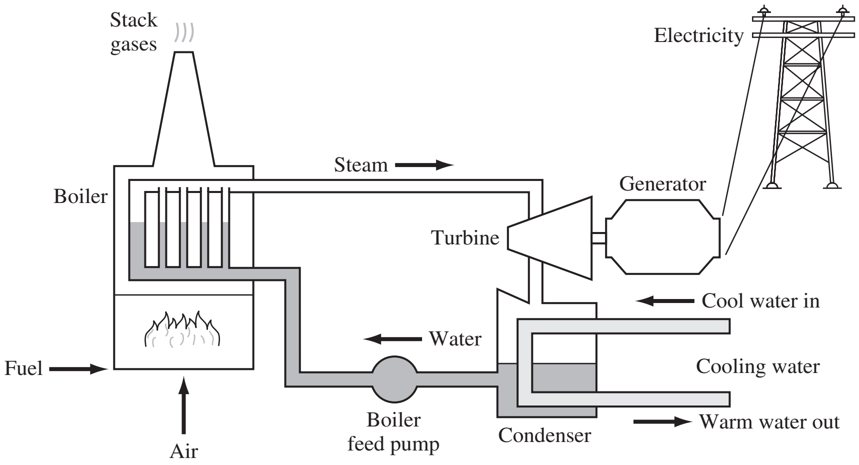


FIGURE 16 A fuel-fired, steam-electric power plant.

One immediate observation that can be made from (55) is that the maximum possible heat engine efficiency increases as the temperature of the hot reservoir increases or the temperature of the cold reservoir decreases. In fact, since neither infinitely hot temperatures nor absolute zero temperatures are possible, we must conclude that no real engine has 100 percent efficiency, which is just a restatement of the second law.

Equation (55) can help us understand the seemingly low efficiency of thermal power plants such as the one diagrammed in Figure 16. In this plant, fuel is burned in a firing chamber surrounded by metal tubing. Water circulating through this boiler tubing is converted to high-pressure, high-temperature steam. During this conversion of chemical to thermal energy, losses on the order of 10 percent occur due to incomplete combustion and loss of heat up the smokestack. Later, we shall consider local and regional air pollution effects caused by these emissions as well as their possible role in global warming.

The steam produced in the boiler then enters a steam turbine, which is in some ways similar to a child's pinwheel. The high-pressure steam expands as it passes through the turbine blades, causing a shaft that is connected to the generator to spin. Although the turbine in Figure 16 is shown as a single unit, in actuality, turbines have many stages with steam exiting one stage and entering another, gradually expanding and cooling as it goes. The generator converts the rotational energy of a spinning shaft into electrical power that goes out onto transmission lines for distribution. A well-designed turbine may have an efficiency that approaches 90 percent, whereas the generator may have a conversion efficiency even higher than that.

The spent steam from the turbine undergoes a phase change back to the liquid state as it is cooled in the condenser. This phase change creates a partial vacuum that helps pull steam through the turbine, thereby increasing the turbine efficiency. The condensed steam is then pumped back to the boiler to be reheated.

The heat released when the steam condenses is transferred to cooling water that circulates through the condenser. Usually, cooling water is drawn from a lake or river, heated in the condenser, and returned to that body of water, which is called *once-through cooling*. A more expensive approach, which has the advantage of requiring less water, involves the use of cooling towers that transfer the heat directly into the atmosphere rather than into a receiving body of water. In either case, the rejected heat is released into the environment. In terms of the heat engine concept shown in Figure 15, the cold reservoir temperature is thus determined by the temperature of the environment.

Let us estimate the maximum possible efficiency that a thermal power plant such as that diagrammed in Figure 16 can have. A reasonable estimate of T_h might be the temperature of the steam from the boiler, which is typically around 600°C. For T_c, we might use an ambient temperature of about 20°C. Using these values in (55) and remembering to convert temperatures to the absolute scale, gives

$$\eta_{\max} = 1 - \frac{(20 + 273)}{(600 + 273)} = 0.66 = 66 \text{ percent}$$

New fossil fuel-fired power plants have efficiencies around 40 percent. Nuclear plants have materials constraints that force them to operate at somewhat lower temperatures than fossil plants, which results in efficiencies of around 33 percent. The average efficiency of all thermal plants actually in use in the United States, including new and old (less efficient) plants, fossil and nuclear, is close to 33 percent. That suggests the following convenient rule of thumb:

> *For every 3 units of energy entering the average thermal power plant, approximately 1 unit is converted to electricity and 2 units are rejected to the environment as waste heat.*

The following example uses this rule of thumb for power plant efficiency combined with other emission factors to develop a mass and energy balance for a typical coal-fired power plant.

EXAMPLE 13 Mass and Energy Balance for a Coal-Fired Power Plant

Typical coal burned in power plants in the United States has an energy content of approximately 24 kJ/g and an average carbon content of about 62 percent. For almost all new coal plants, Clean Air Act emission standards limit sulfur emissions to 260 g of sulfur dioxide (SO_2) per million kJ of heat input to the plant (130 g of elemental sulfur per 10^6 kJ). They also restrict particulate emissions to 13 g/10^6 kJ. Suppose the average plant burns fuel with 2 percent sulfur content and 10 percent unburnable minerals called *ash*. About 70 percent of the ash is released as *fly ash,* and about 30 percent settles out of the firing chamber and is collected as *bottom ash*. Assume this is a typical coal plant with 3 units of heat energy required to deliver 1 unit of electrical energy.

a. Per kilowatt-hour of electrical energy produced, find the emissions of SO_2, particulates, and carbon (assume all of the carbon in the coal is released to the atmosphere).
b. How efficient must the sulfur emission control system be to meet the sulfur emission limitations?
c. How efficient must the particulate control system be to meet the particulate emission limits?

Solution

a. We first need the heat input to the plant. Because 3 kWhr of heat are required for each 1 kWhr of electricity delivered,

$$\frac{\text{Heat input}}{\text{kWhr electricity}} = 3\ \text{kWhr heat} \times \frac{1\ \text{kJ/s}}{\text{kW}} \times 3{,}600\ \text{s/hr} = 10{,}800\ \text{kJ}$$

The sulfur emissions are thus restricted to

$$\text{S emissions} = \frac{130\ \text{g S}}{10^6\ \text{kJ}} \times 10{,}800\ \text{kJ/kWhr} = 1.40\ \text{g S/kWhr}$$

The molecular weight of SO_2 is $32 + 2 \times 16 = 64$, half of which is sulfur. Thus, 1.4 g of S corresponds to 2.8 g of SO_2, so 2.8 g SO_2/kWhr would be emitted. Particulate emissions need to be limited to:

$$\text{Particulate emissions} = \frac{13\ \text{g}}{10^6\ \text{kJ}} \times 10{,}800\ \text{kJ/kWhr} = 0.14\ \text{g/kWhr}$$

To find carbon emissions, first find the amount of coal burned per kWhr

$$\text{Coal input} = \frac{10{,}800\ \text{kJ/kWhr}}{24\ \text{kJ/g coal}} = 450\ \text{g coal/kWhr}$$

Therefore, since the coal is 62 percent carbon

$$\text{Carbon emissions} = \frac{0.62\ \text{g C}}{\text{g coal}} \times \frac{450\ \text{g coal}}{\text{kWhr}} = 280\ \text{g C/kWhr}$$

b. Burning 450 g coal containing 2 percent sulfur will release $0.02 \times 450 = 9.0$ g of S. Since the allowable emissions are 1.4 g, the removal efficiency must be

$$\text{S removal efficiency} = 1 - \frac{1.4}{9.0} = 0.85 = 85\ \text{percent}$$

c. Since 10 percent of the coal is ash, and 70 percent of that is fly ash, the total fly ash generated will be

$$\begin{aligned}\text{Fly ash generated} &= 0.70 \times 0.10 \times 450\ \text{g coal/kWhr}\\ &= 31.5\ \text{g fly ash/kWhr}\end{aligned}$$

The allowable particulate matter is restricted to 0.14 g/kWhr, so controls must be installed that have the following removal efficiency:

$$\text{Particulate removal efficiency} = 1 - \frac{0.14}{31.5} = 0.995 = 99.5\ \text{percent}$$

The complete mass and energy balance for this coal plant is diagrammed in Figure 17. In this diagram, it has been assumed that 85 percent of the waste heat is removed by cooling water, and the remaining 15 percent is lost in stack gases (corresponding to the conditions given in Example 12).

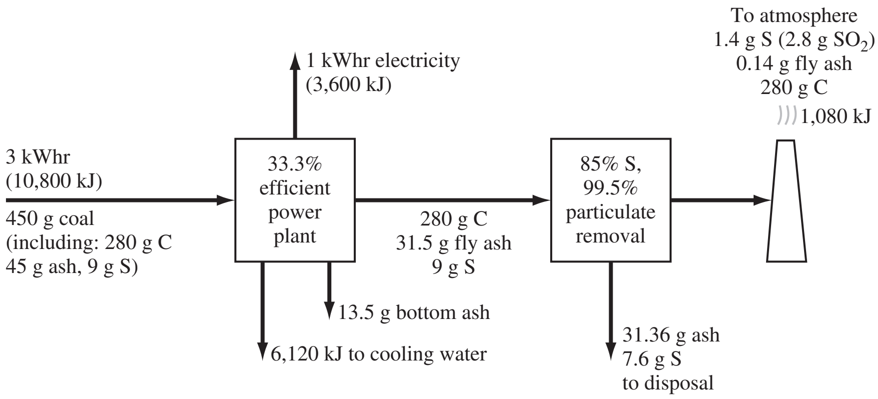


FIGURE 17 Energy and mass balance for a coal-fired power plant generating 1 kWhr of electricity (see Example 13).

The Carnot efficiency limitation provides insight into the likely performance of other types of thermal power plants in addition to the steam plants just described. For example, there have been many proposals to build power plants that would take advantage of the temperature difference between the relatively warm surface waters of the ocean and the rather frigid waters found below. In some locations, the sun heats the ocean's top layer to as much as 30°C, whereas several hundred meters down, the temperature is a constant 4 or 5°C. Power plants, called *ocean thermal energy conversion* (OTEC) systems, could be designed to operate on these small temperature differences in the ocean; however, as the following example shows, they would be inefficient.

EXAMPLE 14 **OTEC System Efficiency**

Consider an OTEC system operating between 30°C and 5°C. What would be the maximum possible efficiency for an electric generating station operating with these temperatures?

Solution Using (55), we find

$$\eta_{\text{max}} = 1 - \frac{(5 + 273)}{(30 + 273)} = 0.08 = 8 \text{ percent}$$

An even lower efficiency, estimated at 2 to 3 percent for a real plant, would be expected.

Conductive and Convective Heat Transfer

When two objects are at different temperatures, heat will be transferred from the hotter object to the colder one. That heat transfer can be by *conduction,* by *convection,* or by *radiation,* which can take place even in the absence of any physical medium between the objects.

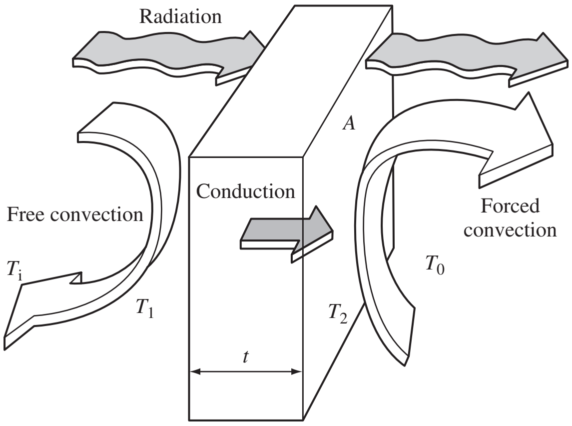


FIGURE 18 Heat transfer through a simple wall.

Conductive heat transfer is usually associated with solids, as one molecule vibrates the next in the lattice. The rate of heat transfer in a solid is proportional to the thermal conductivity of the material. Metals tend to be good thermal conductors, which makes them very useful when high heat-transfer rates are desired. Other materials are much less so, with some being particularly poor thermal conductors, which makes them potentially useful as thermal insulation.

Convective heat transfer occurs when a fluid at one temperature comes in contact with a substance at another temperature. For example, warm air in a house in the winter that comes in contact with a cool wall surface will transfer heat to the wall. As that warm air loses some of its heat, it becomes cooler and denser, and it will sink and be replaced by more warm air from the interior of the room. Thus there is a continuous movement of air around the room and with it a transference of heat from the warm room air to the cool wall. The cool wall, in turn, conducts heat to the cold exterior surface of the house where outside air removes the heat by convection.

Figure 18 illustrates the two processes of convection and conduction through a hypothetical wall. In addition, there is radiative heat transfer from objects in the room to the wall, and from the wall to the ambient outside. It is conventional practice to combine all three processes into a single, overall heat-transfer process that is characterized by the following simple equation:

$$q = \frac{A(T_i - T_o)}{R} \tag{58}$$

where

q = heat transfer rate through the wall (W) or (Btu/hr)
A = wall area (m^2) or (ft^2)
T_i = air temperature on one side of the wall (°C) or (°F)
T_o = ambient air temperature (°C) or (°F)
R = overall thermal resistance (m^2-°C/W) or (hr-ft^2-°F/Btu)

The overall thermal resistance R is called the *R-value*. If you buy insulation at the hardware store, it will be designated as having an R-value in the American unit system (hr-ft^2-°F/Btu). For example, 3½-inch-thick fiberglass insulation is usually marked *R-11,* whereas 6 inches of the same material is *R-19*.

As the following example illustrates, improving the efficiency with which we use energy can save money as well as reduce emissions of pollutants associated with energy consumption. This important connection between energy efficiency and pollution control has in the past been overlooked and underappreciated. However, that situation has been changing. The 1990 Amendments to the Clean Air Act, for example, provide SO_2 emission credits for energy efficiency projects.

EXAMPLE 15 Reducing Pollution by Adding Ceiling Insulation

A home with 1,500 ft^2 of poorly insulated ceiling is located in an area with an 8-month heating season during which time the outdoor temperature averages 40°F while the inside temperature is kept at 70°F (this could be Chicago, for example). It has been proposed to the owner that \$1,000 be spent to add more insulation to the ceiling, raising its total *R-value* from 11 to 40 (ft^2-°F-hr/Btu). The house is heated with electricity that costs 8 cents/kWhr.

a. How much money would the owner expect to save each year, and how long would it take for the energy savings to pay for the cost of insulation?

b. Suppose 1 million homes served by coal plants like the one analyzed in Example 13 could achieve similar energy savings. Estimate the annual reduction in SO_2, particulate, and carbon emissions that would be realized.

Solution

a. Using (58) to find the heat loss rate with the existing insulation gives

$$q = \frac{A(T_i - T_o)}{R} = \frac{1{,}500\ \text{ft}^2 \times (70 - 40)°\text{F}}{11\ (\text{ft}^2\text{-°F-hr/Btu})} = 4{,}090\ \text{Btu/hr}$$

After adding the insulation, the new heat loss rate will be

$$q = \frac{A(T_i - T_o)}{R} = \frac{1{,}500\ \text{ft}^2 \times (70 - 40)°\text{F}}{40\ (\text{ft}^2\text{-°F-hr/Btu})} = 1{,}125\ \text{Btu/hr}$$

The annual energy savings can be found by multiplying the rate at which energy is being saved by the number of hours in the heating season. If we assume the electric heating system in the house is 100 percent efficient at converting electricity to heat (reasonable) and that it delivers all of that heat to the spaces that need heat (less reasonable, especially if there are heating ducts, which tend to leak), then we can use the conversion 3,412 Btu = 1 kWhr.

$$\begin{aligned}\text{Energy saved} &= \frac{(4{,}090 - 1{,}125)\ \text{Btu/hr}}{3{,}412\ \text{Btu/kWhr}} \times 24\ \text{hr/day} \times 30\ \text{day/mo} \times 8\ \text{mo/yr}\\ &= 5{,}005\ \text{kWhr/yr}\end{aligned}$$

The annual savings in dollars would be

$$\text{Dollar savings} = 5{,}005\ \text{kWhr/yr} \times \$0.08/\text{kWhr} = \$400/\text{yr}$$

Since the estimated cost of adding extra insulation is \$1,000, the reduction in electricity bills would pay for this investment in about 2½ heating seasons.

b. One million such houses would save a total of 5 billion kWhr/yr (nearly the entire annual output of a typical 1,000 MW_e power plant). Using the emission factors derived in Example 13, the reduction in air emissions would be

$$\begin{aligned}\text{Carbon reduction} &= 280\,\text{g C/kWhr} \times 5 \times 10^9\,\text{kWhr/yr} \times 10^{-3}\,\text{kg/g}\\ &= 1{,}400 \times 10^6\ \text{kg/yr}\\ \text{SO}_2\ \text{reduction} &= 2.8\,\text{g SO}_2\text{/kWhr} \times 5 \times 10^9\,\text{kWhr/yr} \times 10^{-3}\,\text{kg/g}\\ &= 14 \times 10^6\,\text{kg/yr}\\ \text{Particulate reduction} &= 0.14\,\text{g/kWhr} \times 5 \times 10^9\,\text{kWhr/yr} \times 10^{-3}\,\text{kg/g}\\ &= 0.7 \times 10^6\,\text{kg/yr}\end{aligned}$$

Radiant Heat Transfer

Heat transfer by thermal radiation is the third way that one object can warm another. Unlike conduction and convection, radiant energy is transported by electromagnetic waves and does not require a medium to carry the energy. As is the case for other forms of electromagnetic phenomena, such as radio waves, x-rays, and gamma rays, thermal radiation can be described either in terms of wavelengths or, using the particle nature of electromagnetic radiation, in terms of discrete photons of energy. All electromagnetic waves travel at the speed of light. They can be described by their wavelength or their frequency, and the two are related as follows:

$$c = \lambda \nu \tag{59}$$

where

c = speed of light (3×10^8 m/s)
λ = wavelength (m)
ν = frequency (hertz, *i.e.,* cycles per second)

When radiant energy is described in terms of photons, the relationship between frequency and energy is given by

$$E = h\nu \tag{60}$$

where

E = energy of a photon (J)
h = Planck's constant (6.6×10^{-34} J-s)

Equation (60) points out that higher-frequency, shorter-wavelength photons have higher energy content, which makes them potentially more dangerous when living things are exposed to them.

Consider the effects of various gases on global climate and depletion of the stratospheric ozone layer. The wavelengths of importance in that context are roughly in the range of about 0.1 μm up to about 100 μm (1 μm is 10^{-6} m, also called 1 micron). For perspective, Figure 19 shows a portion of the electromagnetic spectrum.

Every object emits thermal radiation. The usual way to describe how much radiation a real object emits, as well as other characteristics of the wavelengths

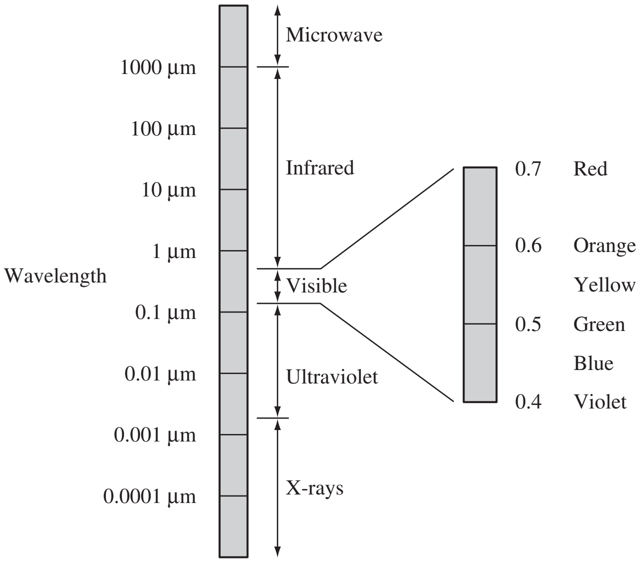


FIGURE 19 A portion of the electromagnetic spectrum. The wavelengths of greatest interest for this text are in the range of about 0.1 μm to 100 μm.

emitted, is to compare it to a theoretical abstraction called a *blackbody*. A blackbody is defined as a perfect emitter as well as a perfect absorber. As a perfect emitter, it radiates more energy per unit of surface area than any real object at the same temperature. As a perfect absorber, it absorbs all radiation that impinges upon it; that is, none is reflected, and none is transmitted through it. Actual objects do not emit as much radiation as this hypothetical blackbody, but most are close to this theoretical limit. The ratio of the amount of radiation an actual object would emit to the amount that a blackbody would emit at the same temperature is known as the *emissivity*, ε. The emissivity of desert sand, dry ground, and most woodlands is estimated to be approximately 0.90, whereas water, wet sand, and ice all have estimated emissivities of roughly 0.95. A human body, no matter what pigmentation, has an emissivity of around 0.96.

The wavelengths radiated by a blackbody depend on its temperature, as described by *Planck's law:*

$$E_\lambda = \frac{C_1}{\lambda^5\left(e^{C_2/\lambda T} - 1\right)} \tag{61}$$

where

E_λ = emissive power of a blackbody (W/m^2-μm)
T = absolute temperature of the body (K)
λ = wavelength (μm)
$C_1 = 3.74 \times 10^8$ W-μm^4/m^2
$C_2 = 1.44 \times 10^4$ μm-K

Figure 20 is a plot of the emissive power of radiation emitted from blackbodies at various temperatures. Curves such as these, which show the spectrum of wave-

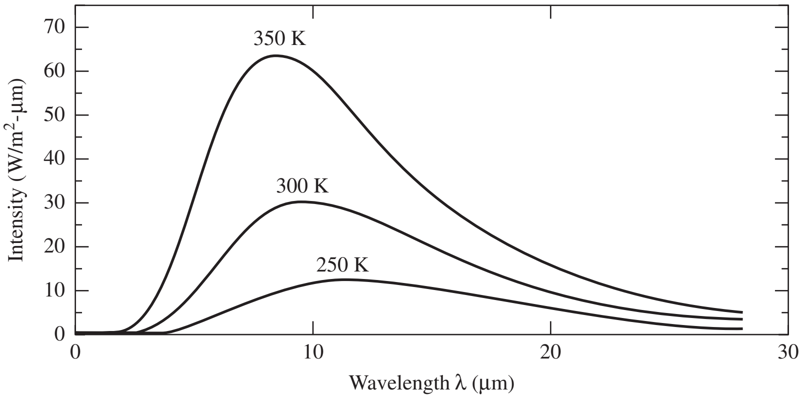


FIGURE 20 The spectral emissive power of a blackbody with various temperatures.

lengths emitted, have as their vertical axis an amount of power per unit-wavelength. The way to interpret such spectral diagrams is to realize that the area under the curve between any two wavelengths is equal to the power radiated by the object within that band of wavelengths. Hence the total area under the curve is equal to the total power radiated. Objects at higher temperatures emit more power and, as the figure suggests, their peak intensity occurs at shorter wavelengths.

Extraterrestrial solar radiation (just outside the Earth's atmosphere) shows spectral characteristics that are well approximated by blackbody radiation. Although the temperature deep within the sun is many millions of degrees, its effective surface temperature is about 5,800 K. Figure 21 shows the close match between the actual solar spectrum and that of a 5,800-K blackbody.

As has already been mentioned, the area under a spectral curve between any two wavelengths is the total power radiated by those wavelengths. For the solar spectrum of Figure 21, the area under the curve between 0.38 and 0.78 μm (the wavelengths visible to the human eye) is 47 percent of the total area. That is, 47 percent of the solar energy striking the outside of our atmosphere is in the visible portion of the spectrum. The ultraviolet range contains 7 percent of the energy, and the infrared wavelengths deliver the remaining 46 percent. The total area under the solar spectrum curve is called the *solar constant,* and it is estimated to be 1,372 watts per m^2. The solar constant plays an important role in determining the surface temperature of the Earth.

The equation describing Planck's law (61) is somewhat tricky to manipulate, especially for calculations involving the area under a spectral curve. Two other radiation equations, however, are straightforward and are often all that is needed. The first, known as the *Stefan-Boltzmann law of radiation,* gives the total radiant energy emitted by a blackbody with surface area A and absolute temperature T:

$$E = \sigma A T^4 \tag{62}$$

where

E = total blackbody emission rate (W)
σ = the Stefan-Boltzmann constant = 5.67×10^{-8} W/m^2-K^4
T = absolute temperature (K)
A = surface area of the object (m^2)

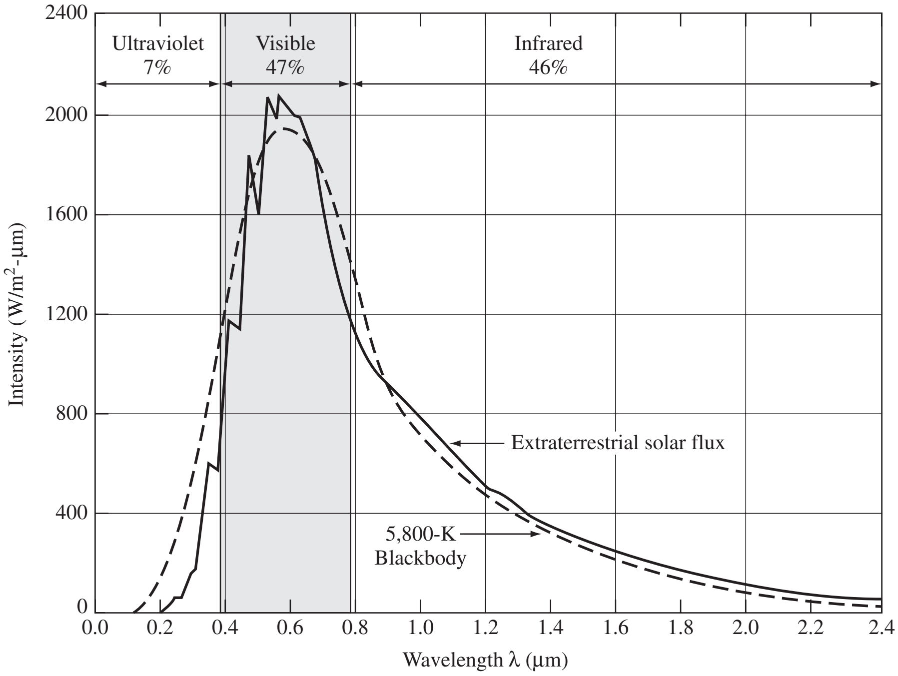


FIGURE 21 The extraterrestrial solar spectrum (solid line) compared with the spectrum of a 5,800-K blackbody (dashed).

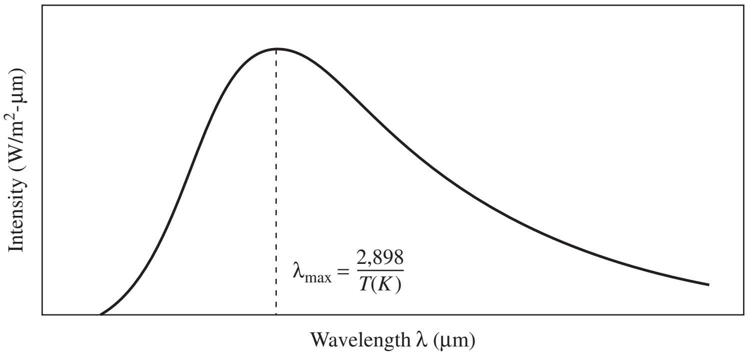


FIGURE 22 Wien's rule for finding the wavelength at which the spectral emissive power of a blackbody reaches its maximum value.

The second is *Wien's displacement rule,* which tells us the wavelength at which the spectrum reaches its maximum point:

$$\lambda_{\max}(\mu\text{m}) = \frac{2,898}{T(\text{K})} \tag{63}$$

where wavelength is specified in micrometers and temperature is in kelvins. Figure 22 illustrates this concept, and Example 16 shows how it can be used.

EXAMPLE 16 The Earth's Spectrum

Consider the Earth to be a blackbody with average temperature 15°C and surface area equal to $5.1 \times 10^{14}\ \text{m}^2$. Find the rate at which energy is radiated by the Earth and the wavelength at which maximum power is radiated. Compare this peak wavelength with that for a 5,800-K blackbody (the sun).

Solution Using (62), the Earth radiates

$$\begin{aligned} E &= \sigma A T^4 \\ &= 5.67 \times 10^{-8}\ \text{W/m}^2\text{-K}^4 \times 5.1 \times 10^{14}\ \text{m}^2 \times (15.0 + 273.15\ \text{K})^4 \\ &= 2.0 \times 10^{17}\ \text{Watts} \end{aligned}$$

The wavelength at which the maximum point is reached in the Earth's spectrum is

$$\lambda_{\max}(\mu\text{m}) = \frac{2{,}898}{T(\text{K})} = \frac{2{,}898}{288.15} = 10.1\,\mu\text{m (Earth)}$$

For the 5,800-K sun,

$$\lambda_{\max}(\mu\text{m}) = \frac{2{,}898}{5{,}800} = 0.48\,\mu\text{m (sun)}$$

This tremendous rate of energy emission by the Earth is balanced by the rate at which the Earth absorbs energy from the sun. As shown in Example 16, however, the solar energy striking the Earth has much shorter wavelengths than energy radiated back to space by the Earth. This wavelength shift plays a crucial role in the greenhouse effect. Carbon dioxide and other greenhouse gases are relatively transparent to the incoming short wavelengths from the sun, but they tend to absorb the outgoing, longer wavelengths radiated by the Earth. As those greenhouse gases accumulate in our atmosphere, they act like a blanket that envelops the planet, upsets the radiation balance, and raises the Earth's temperature.

PROBLEMS

1 The proposed air quality standard for ozone (O_3) is 0.08 ppm.
   (a) Express that standard in μg/m^3 at 1 atm of pressure and 25°C.
   (b) At the elevation of Denver, the pressure is about 0.82 atm. Express the ozone standard at that pressure and at a temperature of 15°C.

2 Suppose the exhaust gas from an automobile contains 1.0 percent by volume of carbon monoxide. Express this concentration in mg/m^3 at 25°C and 1 atm.

3 Suppose the average concentration of SO_2 is measured to be 400 μg/m^3 at 25°C and 1 atm. Does this exceed the (24-hr) air quality standard of 0.14 ppm?

4 A typical motorcycle emits about 20 g of CO per mile.
   (a) What volume of CO would a 5-mile trip produce after the gas cools to 25°C (at 1 atm)?
   (b) Per meter of distance traveled, what volume of air could be polluted to the air quality standard of 9 ppm?

5 If we approximate the atmosphere to be 79 percent nitrogen (N_2) by volume and 21 percent oxygen (O_2), estimate the density of air (kg/m^3) at STP conditions (0°C, 1 atm).

6 Five million gallons per day (MGD) of wastewater, with a concentration of 10.0 mg/L of a conservative pollutant, is released into a stream having an upstream flow of 10 MGD and pollutant concentration of 3.0 mg/L.

(a) What is the concentration in ppm just downstream?

(b) How many pounds of substance per day pass a given spot downstream? (The conversions 3.785 L/gal and 2.2 kg/lbm may be helpful.)

7 A river with 400 ppm of salts (a conservative substance) and an upstream flow of 25.0 m^3/s receives an agricultural discharge of 5.0 m^3/s carrying 2,000 mg/L of salts (see Figure P7). The salts quickly become uniformly distributed in the river. A municipality just downstream withdraws water and mixes it with enough pure water (no salt) from another source to deliver water having no more than 500 ppm salts to its customers.

What should be the mixture ratio *F* of pure water to river water?

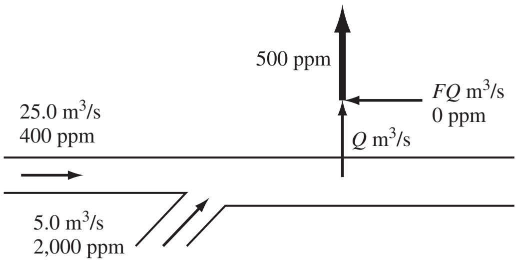


FIGURE P7

8 A home washing machine removes grease and dirt from clothes in a nearly first-order process in which 12 percent of the grease on the clothes is removed per minute. The washing machine holds 50.0 L of water and has a wash cycle of 5.00 minutes before discharging the wash water. What will be the grease concentration (in mg/L) in the discharge water if the clothes initially contain 0.500 g of grease?

9 Plateau Creek carries 5.0 m^3/s of water with a selenium (Se) concentration of 0.0015 mg/L. A farmer starts withdrawing 1.0 m^3/s of the creek water to irrigate the land. During irrigation, the water picks up selenium from the salts in the soil. One-half of the irrigation water is lost to the ground and plants, and the other half is returned to Plateau Creek. The irrigation run-off to the creek contains 1.00 mg/L of selenium. Selenium is a conservative, nonreactive substance (it does not degrade in the stream), and the stream does not pick up more selenium from any other source.

(a) If the farmer irrigates continuously, what will be the steady-state concentration of selenium in the stream downstream from the farm (after the irrigation run-off returns to the stream)?

(b) Fish are sensitive to selenium levels over 0.04 mg/L. The farmer agrees not to use more water than will keep the stream selenium level below this critical concentration. How much water can the farmer withdraw from the stream to use for irrigation?

10 When methanol is used to generate hydrogen, it reacts with the following reaction:

$$2CH_2OH \rightarrow 2CO + 3H_2$$

The reaction is second order in methanol (CH_2OH), and it is observed that 100 g of carbon monoxide (CO) can be produced in one day in a batch reactor, if you start with 200 g of methanol. What is the rate constant for this reaction?

11 A lake with constant volume 10×10^6 m^3 is fed by a pollution-free stream with flow rate 50 m^3/s. A factory dumps 5 m^3/s of a nonconservative waste with concentration 100 mg/L into the lake. The pollutant has a reaction rate coefficient K of 0.25/day. Assuming the pollutant is well mixed in the lake, find the steady-state concentration of pollutant in the lake.

12 The two-pond system shown in Figure P12 is fed by a stream with flow rate 1.0 MGD (millions gallons per day) and BOD (a nonconservative pollutant) concentration 20.0 mg/L. The rate of decay of BOD is 0.30/day. The volume of the first pond is 5.0 million gallons, and the second is 3.0 million.

Assuming complete mixing within each pond, find the BOD concentration leaving each pond.

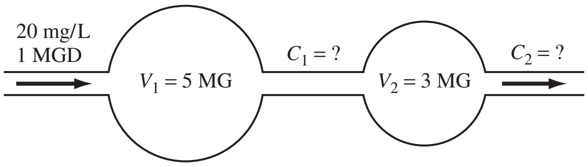


FIGURE P12

13 A lagoon is to be designed to accomodate an input flow of 0.10 m^3/s of nonconservative pollutant with concentration 30.0 mg/L and reaction rate 0.20/day. The effluent from the lagoon must have pollutant concentration of less than 10.0 mg/L. Assuming complete mixing, how large must the lagoon be?

14 Ozone is sometimes used as a disinfectant for drinking water. Ozone is highly reactive. It will react with many benign species found in water, as well as the pathogens it is intended to kill. It is found that ozone reaction in water nearly follows first-order kinetics in ozone concentration, such that its concentration decreases by 50 percent in 12 minutes ($t_{1/2} = 12$ min). A water supplier wants to inject ozone into a pipe bringing water to the water treatment plant to predisinfect the influent. The 3.0-ft. diameter pipe is 3,400 feet long with a steady flow rate of 10,000 gal/min. What concentration of ozone (in mg/L) should be injected at the head of the pipe so that there will be an ozone concentration of 1.0 mg/L at the pipe's exit into the plant? The pipe may be realistically assumed to act as an ideal PFR.

15 Your stomach is a chemical reactor. When you devour a fast food 99¢ special hamburger in about a minute, it acts like an instantaneous input of 325 g of food entering the stomach. In response, the stomach starts producing gastric liquids (acids), which are continuously excreted into the stomach at a rate of 12.0 mL/min as the hamburger is digested. The fluid also leaves the stomach to the small intestine at a flow rate of 12.0 mL/min, so the volume of liquid in the stomach stays constant at 1.15 L. The hamburger digestion rate constant is 1.33 hr^{-1}.

(a) What kind of ideal reactor would you model your stomach as?

(b) What fraction of the hamburger's mass will remain undigested in your stomach one hour after you eat the hamburger?

16 A simple way to model air pollution over a city is with a box model that assumes complete mixing and limited capability for the pollution to disperse horizontally or vertically except in the direction of the prevailing winds (for example, a town located in a valley

with an inversion layer above it). Consider a town having an inversion at 250 m, a 20-km horizontal distance perpendicular to the wind, a windspeed of 2 m/s, and a carbon monoxide (CO) emission rate of 60 kg/s (see Figure P16). Assume the CO is conservative and completely mixed in the box.

What would be the CO concentration in the box?

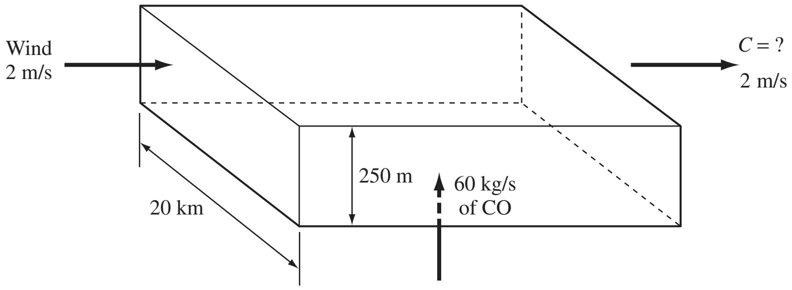


FIGURE P16

17 Consider the air over a city to be a box 100 km on a side that reaches up to an altitude of 1.0 km. Clean air is blowing into the box along one of its sides with a speed of 4 m/s. Suppose an air pollutant with reaction rate $k = 0.20/\text{hr}$ is emitted into the box at a total rate of 10.0 kg/s. Find the steady-state concentration if the air is assumed to be completely mixed.

18 If the windspeed in Problem 17 suddenly drops to 1 m/s, estimate the concentration of pollutants two hours later.

19 A lagoon with volume 1,200 m^3 has been receiving a steady flow of a conservative waste at a rate of 100 m^3/day for a long enough time to assume that steady-state conditions apply. The waste entering the lagoon has a concentration of 10 mg/L. Assuming completely mixed conditions,

(a) What would be the concentration of pollutant in the effluent leaving the lagoon?

(b) If the input waste concentration suddenly increased to 100 mg/L, what would the concentration in the effluent be 7 days later?

20 Repeat Problem 19 for a nonconservative pollutant with reaction rate $k = 0.20/\text{d}$.

21 A nuclear power station is situated in Coal Valley, which is a roughly rectangular valley that is 5 km long, 2 km wide, and 200 m deep. You have been asked to evaluate the effects of a worst-case scenario where the reactor housing fails, and radiation is released to the atmosphere. In your evaluation, you determine that 120 kg of Iodine-131 (a radioisotope that causes thyroid gland and liver damage) could be released into the atmosphere.

(a) Assuming the release of Iodine-131 was very rapid and all of it was uniformly distributed through the valley's atmosphere with none escaping the valley, what would the concentration of Iodine-131 be in the valley's air? Your answer should be expressed in units of ppmv, and you may assume an atmopheric pressure of 1.0 atm and a temperature of 20°C.

(b) Assuming the Iodine-131 concentration you calculated in part (a) is the initial concentration in the valley, you now want to determine the time it will take for the concentration to decrease to the safe limit of 1.0×10^{-5} ppmv. The average windspeed through the valley (entering at one end and exiting at the other) is only 1.5 m/min. However, Iodine-131 also is removed by two other processes: 1) radioactive decay with a half life of 8.1 days, and 2) sedimentation to the ground with a rate constant of $0.02\ d^{-1}$. Draw a picture of the situation, and label the appropriate variables.

(c) Now derive an equation that expresses the concentration of Iodine-131 in the valley's air as a function of the time since the accident. Use the equation to determine the time needed for the concentration to decrease to a safe level.

22 When a strong acid is added to water, it completely dissociates its protons (H^+). For example, the dissociation of sulfuric acid in water is

$$H_2SO_4 + H_2O \rightarrow 2H^+ + SO_4^{2-} + H_2O$$

Or more simply:

$$H_2SO_4 \rightarrow 2H^+ + SO_4^{2-} \qquad (P1)$$

A similar type of reaction occurs if a strong base is added to water, except hydroxide ions (OH^-) are dissociated. For example, the dissociation of the strong base, sodium hydroxide (NaOH), in water is

$$NaOH + H_2O \rightarrow OH^- + H_2O + Na^+$$

Or more simply:

$$NaOH \rightarrow OH^- + Na^+ \qquad (P2)$$

When both a strong acid and a strong base are added to water together, the protons from the acid react with the hydroxides from the base to produce more water molecules until the supply of either excess protons or hydroxides is used up. The reaction is

$$H^+ + OH^- \rightarrow H_2O \qquad (P3)$$

The rate constant for reaction (P3) is $1.4 \times 10^{11}\ L \cdot mol^{-1} \cdot s^{-1}$*.

(a) 8.00 mg of H_2SO_4 is added to 1.00 L of water in a flask. How much NaOH (in mg) must be added to neutralize the acid? Neutralization means conversion of all of the protons from the acid to water as shown in reaction (P3).

(b) If the 8.00 mg of H_2SO_4 and the amount of base you calculated in part (a) are added to the flask's water at the same instant, what will be the half-life of the added protons in the water?

23 A 4 × 8-ft solar collector has water circulating through it at the rate of 1.0 gallon per minute (gpm) while exposed to sunlight with intensity 300 Btu/ft²-hr (see Figure P23).

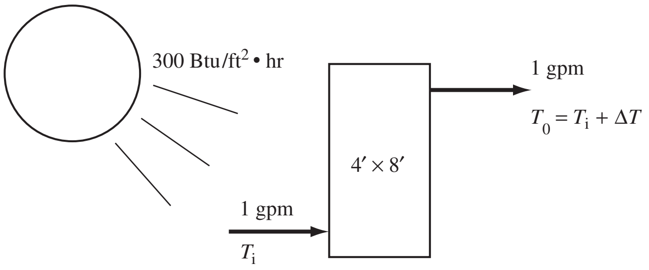


FIGURE P23

---

*From Moore and Pearson, *Kinetics and Mechanism,* 3rd ed., 1981, Wiley-Interscience, New York, NY.

Fifty percent of that sunlight is captured by the collector and heats the water flowing through it. What would be the temperature rise of the water as it leaves the collector?

24 An uncovered swimming pool loses 1.0 in of water off of its 1,000 ft^2 surface each week due to evaporation. The heat of vaporization for water at the pool temperature is 1,050 Btu/lb. The cost of energy to heat the pool is \$10.00 per million Btu. A salesman claims that a \$500 pool cover that reduces evaporizative losses by two-thirds will pay for itself in one 15-week swimming season. Can it be true?

25 Two-thirds of the energy content of fuel entering a 1,000-MW_e nuclear power plant is removed by condenser cooling water that is withdrawn from a local river (there are no stack losses, as is the case for a fossil fuel-fired plant). The river has an upstream flow of 100 m^3/s and a temperature of 20°C.

(a) If the cooling water is only allowed to rise in temperature by 10°C, what flow rate from the river would be required? Compare it to the coal plant in Example 12.

(b) How much would the river temperature rise as it receives the heated cooling water? Again, compare it to Example 12.

26 Consider a 354-mL can of soda cold enough to cause moisture from the air to condense on the outside of the can. If all of the heat released when 5 ml of vapor condenses on the can is transferred into the soda, how much would the soda temperature increase? Assume the density and specific heat of the soda are the same as water, neglect the thermal capacitance of the can itself, and use 2,500 kJ/kg as the latent heat of vaporization (condensation).

27 Compare the energy required to evaporate a kilogram of water at 15°C to that required to raise it 3 km into the air. (Recall that 1 kg on Earth weighs 9.8 N, and 1 J = 1 N-m.)

28 Compare the potential energy represented by 1 lb of water vapor at 59°F (15°C) and an elevation of 5,000 ft to the energy that would be released when it condenses into rain (1 Btu = 778 ft-lb).

29 A 600-MW_e power plant has an efficiency of 36 percent with 15 percent of the waste heat being released to the atmosphere as stack heat and the other 85 percent taken away in the cooling water (see Figure P29). Instead of drawing water from a river, heating it, and returning it to the river, this plant uses an evaporative cooling tower wherein heat is released to the atmosphere as cooling water is vaporized.

At what rate must 15°C makeup water be provided from the river to offset the water lost in the cooling tower?

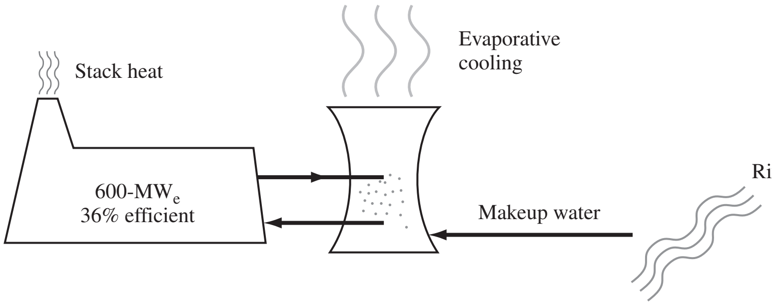


FIGURE P29

30 An electric water heater held at 140°F is kept in a 70°F room. When purchased, its insulation is equivalent to *R*-5. An owner puts a 25-ft^2 blanket on the water heater, raising its

total R value to 15. Assuming 100 percent conversion of electricity into heated water, how much energy (kWhr) will be saved each year? If electricity costs 8.0 cents/kWhr, how much money will be saved in energy each year?

31 A 15-W compact fluorescent lightbulb (CFL) produces the same amount of light as a 60-W incandescent while using only one-fourth the power. Over the 9,000-hr lifetime of one CFL, compute carbon, SO_2, and particulate emissions that would be saved if one CFL replaces incandescents and the electricity comes from the coal-fired power plant described in Example 13.

32 Suppose a utility generates electricity with a 36 percent efficient coal-fired power plant emitting the legal limit of 0.6 lb of SO_2 per million Btus of heat into the plant. Suppose the utility encourages its customers to replace their 75-W incandescents with 18-W compact fluorescent lamps (CFLs) that produce the same amount of light. Over the 10,000-hr lifetime of a single CFL,

(a) How many kilowatt-hours of electricity would be saved?

(b) How many 2,000-lb tons of SO_2 would not be emitted?

(c) If the utility can sell its rights to emit SO_2 at \$800 per ton (these are called "allowances"; see Chapter 7), how much money could the utility earn by selling the SO_2 saved by a single CFL?

33 No. 6 fuel oil has a carbon content of 20 kg carbon per 10^9 J. If it is burned in a 40 percent efficient power plant, find the carbon emissions per kilowatt-hour of electricity produced, assuming all of the carbon in the fuel is released into the atmosphere. By law, new oil-fired power plant emissions are limited to 86 mg of SO_2 per million joules (MJ) of thermal input and 130 mg NO_x/MJ. Estimate the maximum allowable SO_2 and NO_x emissions per kilowatt-hour.

34 Mars radiates energy with a peak wavelength of 13.2 μm.

(a) Treating it as a blackbody, what would its temperature be?

(b) What would be the frequency and energy content of a photon at that wavelength?

35 The rate at which sunlight reaches the outer edge of the atmosphere of Earth is 1,370 W/m^2 (the solar constant for Earth). The Earth's orbit has an average radius of 150×10^6 km. Solar radiation decreases as the square of the distance from the sun. Estimate the solar constants for

(a) Mars, whose orbit has a radius of 228×10^6 km.

(b) Venus, whose orbit has a radius of 108×10^6 km.

36 Objects not only radiate energy, but they absorb radiant energy as well. The net blackbody radiation for an object at temperature T_1 in an environment with temperature T_2 is given by

$$E_{\text{net}} = \sigma A[(T_1)^4 - (T_2)^4]$$

Suppose an unclothed human body has a surface area of 1.35 m^2, an average skin temperature of 32°C, and is in a room with surfaces at 15°C. Treating this person as a blackbody (a very good approximation), find the net heat loss by radiation (watts).

37 A hot-water radiator has a surface temperature of 80°C and a surface area of 2 m^2. Treating it as a blackbody, find the net rate at which it will radiate energy to a 20°C room (see Problem 36).

Environmental Chemistry

> It often matters much how given atoms combine, in what arrangement, with what others, what impulse they receive, and what impart. The same ones make up earth, sky, sea, and stream; the same the sun, the animals, grain and trees, but mingling and moving in ever different ways.
>
> —*Lucretius (95–52 B.C.) in* The Nature of Things

1 Introduction

Almost every pollution problem that we face has a chemical basis. Even the most qualitative descriptions of such problems as the greenhouse effect, ozone depletion, toxic wastes, groundwater contamination, air pollution, and acid rain, to mention a few, require at least a rudimentary understanding of some basic chemical concepts. And, of course, an environmental engineer who must design an emission control system or a waste treatment plant must be well grounded in chemical principles and the techniques of chemical engineering. In this brief chapter, the topics have been selected with the goal of providing only the essential chemical principles required to understand the nature of the pollution problems that we face and the engineering approaches to their solutions.

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

2 Stoichiometry

When a chemical reaction is written down, it provides both qualitative and quantitative information. Qualitatively, we can see which chemicals are interacting to produce which end products. Quantitatively, the principle of conservation of mass can be applied to give information about how much of each compound is involved to produce the results shown. The balancing of equations so that the same number of each kind of atom appears on each side of the equation and the subsequent calculations, which can be used to determine amounts of each compound involved, is known as *stoichiometry*.

The first step is to balance the equation. For example, suppose we want to investigate the combustion of methane (CH_4), the principal component of natural gas and a major greenhouse gas. Methane combines with oxygen to produce carbon dioxide and water as the following reaction suggests:

$$CH_4 + O_2 \rightarrow CO_2 + H_2O$$

The equation is not balanced. One atom of carbon appears on each side, which is fine, but there are four atoms of hydrogen on the left and only two on the right, and there are only two atoms of oxygen on the left while there are three on the right. We might try to double the water molecules on the right to balance the hydrogen on each side, but then there would be an imbalance of oxygen with two on the left and four on the right. So try doubling the oxygen on the left. This sort of trial-and-error approach to balancing simple reactions usually converges pretty quickly. In this instance, the following is a balanced equation with the same number of C, H, and O atoms on each side of the arrow:

$$CH_4 + 2\,O_2 \rightarrow CO_2 + 2\,H_2O \tag{1}$$

This balanced chemical equation can be read as: One molecule of methane reacts with two molecules of oxygen to produce one molecule of carbon dioxide and two molecules of water. It is of more use, however, to be able to describe this reaction in terms of the mass of each substance, that is, how many grams of oxygen are required to react with how many grams of methane, and so on. To do so requires that we know something about the mass of individual atoms and molecules.

The *atomic weight* of an atom is the mass of the atom measured in *atomic mass units* (amu) where one amu is defined to be exactly one-twelfth the mass of a carbon atom having six protons and six neutrons in its nucleus. Although this might suggest that if we look up the atomic weight of carbon, we would expect to find it to be exactly 12 amu, that is not the case. All carbon atoms do have six protons, but they do not all have six neutrons, so they do not all have the same atomic weight. Atoms having the same number of protons but differing numbers of neutrons are called *isotopes*. What is reported in tables of atomic weights, is the average based on the relative abundance of different isotopes found in nature. The *atomic number* is the number of protons in the nucleus. All isotopes of a given element have the same atomic number.

The *molecular weight* of a molecule is simply the sum of the atomic weights of all of the constituent atoms. If we divide the mass of a substance by its molecular weight, the result is the mass expressed in *moles* (mol). Usually the mass is expressed in grams, in which case 1 mole contains 6.02×10^{23} molecules (Avogadro's number).

$$\text{Moles} = \frac{\text{Mass (g)}}{\text{Molecular weight}} \tag{2}$$

The special advantage of expressing amounts in moles is that 1 mole of any substance contains exactly the same number of molecules, which gives us another way to interpret a chemical equation. Consider (1), repeated here:

$$CH_4 + 2\,O_2 \rightarrow CO_2 + 2\,H_2O$$

On a molecular level, we can say that 1 molecule of methane reacts with 2 molecules of oxygen to produce 1 molecule of carbon dioxide and 2 molecules of water. On a larger (and more environmentally relevant) scale, we can say that 1 mole of methane reacts with 2 moles of oxygen to produce 1 mole of carbon dioxide and 2 moles of water. Because we know how many grams are contained in each mole, we can express our mass balance in those terms as well.

To express the preceding methane reaction in grams, we need first to find the number of grams per mole for each substance. The atomic weight of C is 12, H is 1, and O is 16. Notice that these values have been rounded slightly, which is common engineering practice. Thus, the molecular weights and hence the number of grams per mole, are

$$\begin{aligned} CH_4 &= 12 + 4 \times 1 &&= 16 \text{ g/mol} \\ O_2 &= 2 \times 16 &&= 32 \text{ g/mol} \\ CO_2 &= 12 + 2 \times 16 &&= 44 \text{ g/mol} \\ H_2O &= 2 \times 1 + 16 &&= 18 \text{ g/mol} \end{aligned}$$

Summarizing these various ways to express the oxidation of methane, we can say

	CH_4	+	$2\,O_2$	→	CO_2	+	$2\,H_2O$
	1 molecule of methane	+	2 molecules of oxygen	→	1 molecule of carbon dioxide	+	2 molecules of water
or,							
	1 mol of methane	+	2 mol of oxygen	→	1 mole of carbon dioxide	+	2 mol of water
or,							
	16 g of methane	+	64 g of oxygen	→	44 g of carbon dioxide	+	36 g of water

Notice that mass is conserved in the last expression; that is, there are 80 grams on the left and 80 grams on the right.

EXAMPLE 1 **Combustion of Butane**

What mass of carbon dioxide would be produced if 100 g of butane (C_4H_{10}) is completely oxidized to carbon dioxide and water?

Solution First write down the reaction:

$$C_4H_{10} + O_2 \rightarrow CO_2 + H_2O$$

then balance it

$$2\,C_4H_{10} + 13\,O_2 \rightarrow 8\,CO_2 + 10\,H_2O$$

Find the grams per mole for butane:

$$C_4H_{10} = 4 \times 12 + 10 \times 1 = 58\,\text{g/mol}$$

We already know that there are 44 grams per mole of CO_2, so we don't need to recalculate that. Two moles of butane (2 mol × 58 g/mol = 116 g) yields 8 moles of carbon dioxide (8 mol × 44 g/mol = 352 g CO_2). So, we can set up the following proportion:

$$\frac{116\,\text{g } C_4H_{10}}{352\,\text{g } CO_2} = \frac{100\,\text{g } C_4H_{10}}{X\text{ g } CO_2}$$

Thus,

$$X = 100 \times 352/116 = 303 \text{ g of } CO_2 \text{ produced}$$

Many environmental problems involve concentrations of substances dissolved in water. We introduce two common sets of units: mg/L and ppm. However, it is also useful to express concentrations in terms of *molarity,* which is simply the number of moles of substance per liter of solution. A 1 molar (1 M) solution has 1 mole of substance dissolved into enough water to make the mixture have a volume of 1 L. Molarity is related to mg/L concentrations by the following:

$$\text{mg/L} = \text{Molarity (mol/L)} \times \text{Molecular Weight (g/mol)} \times 10^3\ \text{(mg/g)} \qquad (3)$$

The following example illustrates the use of molarity, and at the same time, introduces another important concept having to do with the amount of oxygen required to oxidize a given substance.

EXAMPLE 2 **Theoretical Oxygen Demand**

Consider a 1.67×10^{-3} M glucose solution ($C_6H_{12}O_6$) that is completely oxidized to CO_2 and H_2O. Find the amount of oxygen required to complete the reaction.

Solution To find the oxygen required to completely oxidize this glucose, we first write a balanced equation, determine molecular weights, and find the mass of each constituent in the reaction:

$$\begin{array}{ccccccc} C_6H_{12}O_6 & + & 6\,O_2 & \rightarrow & 6\,CO_2 & + & 6\,H_2O \\ 6 \times 12 + 12 \times 1 + 6 \times 16 = 180 & & 6 \times 32 = 192 & & 6 \times 44 = 264 & & 6 \times 18 = 108 \end{array}$$

Thus it takes 192 g of oxygen to oxidize 180 g of glucose. From (3), the concentration of glucose is

$$\text{mg/L} = 1.67 \times 10^{-3}\ \text{mol/L} \times 180\ \text{g/mol} \times 10^{3}\ \text{mg/g} = 301\ \text{mg/L}$$

so the oxygen requirement would be

$$300\ \text{mg/L glucose} \times \frac{192\ \text{g}\,O_2}{180\,\text{g glucose}} = 321\ \text{mg/L}\ \ O_2$$

If the chemical composition of a substance is known, then the amount of oxygen required to oxidize it to carbon dioxide and water can be calculated using stoichiometry, as was done in the preceding example. That oxygen requirement is known as the *theoretical oxygen demand*. If that oxidation is carried out by bacteria using the substance for food, then the amount of oxygen required is known as the *biochemical oxygen demand* or *BOD*. The BOD will be somewhat less than the theoretical oxygen demand because some of the original carbon is incorporated into bacterial cell tissue rather than being oxidized to carbon dioxide. Oxygen demand is an important measure of the likely impact that wastes will have on a receiving body of water.

The convenience of using moles to describe amounts of substances also helps when calculating atmospheric concentrations of pollutants. It was Avogadro's hypothesis, made in 1811, that equal volumes of all gases, at a specified temperature and pressure, contain equal numbers of molecules. In fact, since 1 mole of any substance has Avogadro's number of molecules, it follows that 1 mole of gas, at a specified temperature and volume, will occupy a predictable volume. At standard temperature and pressure (STP), corresponding to 0°C and 1 atm. (760 mm of mercury, 101.3 kPa), 1 mole of an ideal gas occupies 22.4 L, or $0.0224\,\text{m}^3$, and contains 6.02×10^{23} molecules.

Let us demonstrate the usefulness of Avogadro's hypothesis for gases by applying it to a very modern concern; that is, the rate at which we are pouring carbon dioxide into the atmosphere as we burn up our fossil fuels.

EXAMPLE 3 Carbon Emissions from Natural Gas

Worldwide combustion of methane, CH_4 (natural gas), provides about $10.9 \times 10^{16}\,\text{kJ}$ of energy per year. If methane has an energy content of $39 \times 10^{3}\,\text{kJ/m}^3$ (at STP), what mass of CO_2 is emitted into the atmosphere each year? Also, express that emission rate as metric tons of carbon (not CO_2) per year. A metric ton, which is 1,000 kg, is usually written as *tonne* to distinguish it from the 2,000-lb American, or short, *ton*.

Solution We first need to express that consumption rate in moles. Converting kJ of energy into moles of methane is straightforward:

$$\text{moles } CH_4 = \frac{10.9 \times 10^{16}\ \text{kJ/yr}}{39 \times 10^{3}\ \text{kJ/m}^3} \times \frac{1}{22.4 \times 10^{-3}\ \text{m}^3\text{/mol}} = 12.5 \times 10^{13}\ \text{mol/yr}$$

We know from the balanced chemical reaction given in (1) that each mole of CH_4 yields 1 mole of CO_2, so there will be 12.5×10^{13} moles of CO_2 emitted. Since the molecular weight of CO_2 is 44, the mass of CO_2 emitted is

$$\text{mass } CO_2 = 12.5 \times 10^{13} \text{ mol/yr} \times 44 \text{ g/mol} = 5.5 \times 10^{15} \text{ g/yr}$$

To express these emissions as tonnes of C per year, we must convert grams to tonnes and then sort out the fraction of CO_2 that is carbon. The fraction of CO_2 that is C is simply the ratio of the atomic weight of carbon (12) to the molecular weight of carbon dioxide (44):

$$\begin{aligned}\text{C emissions} &= 5.5 \times 10^{15} \text{ g } CO_2\text{/yr} \times \frac{1 \text{ kg}}{1{,}000 \text{ g}} \times \frac{1 \text{ tonne}}{1{,}000 \text{ kg}} \times \frac{12 \text{ g C}}{44 \text{ g } CO_2} \\ &= 1.5 \times 10^{9} \text{ tonnes/yr} = 1.5 \text{ gigatonne/yr} = 1.5 \text{ GtC/yr}\end{aligned}$$

The 1.5×10^9 tonnes of carbon found in the preceding example is about 20 percent of the total, worldwide carbon emissions entering the atmosphere each year when fossil fuels (coal, oil, natural gas) are burned. "Global Atmospheric Change" the main worry about these emissions is their potential to enhance the Earth's greenhouse effect.

3 Enthalpy in Chemical Systems

Just as we used conservation of mass to balance chemical equations, we can use the conservation of energy to learn something about heat absorbed or released during chemical reactions. Since energy must be conserved, we should be able to track it from beginning to end. The first law of thermodynamics lets us say that the energy in the reactants on the left side of the equation, plus any heat added to the system, should equal the energy contained in the reaction products on the right side, plus any work done during the reaction.

$$U_1 + Q = U_2 + W \tag{4}$$

where

U_1 = internal energy of the chemical system at the beginning
U_2 = internal energy at the end
Q = heat absorbed during the reaction
W = work done by the system during the reaction

Although many forms of work could be included in (4), our concern will be only with work that occurs when a system changes volume under constant pressure, which is typical of chemical reactions. Any possibility of electrical, magnetic, gravitational, or other forms of work will be ignored. To analyze this work done by expansion, consider the cylinder in Figure 1 containing a volume of gas V exerting a pressure P against a piston with area A. The force exerted by the gas on the piston is

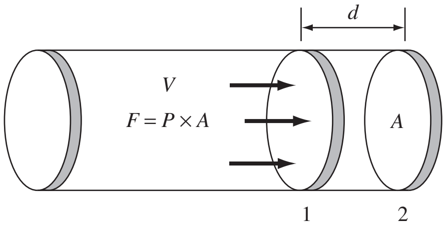


FIGURE 1 Work done when a substance expands at constant pressure is $P\Delta V$.

$P \times A$. If the piston moves a distance d, then, since work is force × distance, we can write

$$W = Fd = PAd = P(V_2 - V_1) \tag{5}$$

Substituting (5) into (4) and rearranging terms gives

$$(U_2 + PV_2) - (U_1 + PV_1) = Q \tag{6}$$

If we define the *enthalpy, H*, of a system as

$$H = U + PV \tag{7}$$

then

$$H_2 - H_1 = \Delta H = Q \tag{8}$$

That is, the change in enthalpy during a constant pressure reaction is equal to the heat absorbed by the system. When ΔH is positive, heat is absorbed and the reaction is said to be *endothermic*. When ΔH is negative, heat is liberated and the reaction is called *exothermic*. The change in enthalpy, $H_2 - H_1$, is called the *heat of reaction*.

As is usually the case for discussions about energy, it is changes in energy, or in this case enthalpy, that are of interest. For example, we can talk about the potential energy of an object as being its weight times its height above some reference elevation. Our choice of reference elevation doesn't matter as long as we are only interested in the change in potential energy as an object is raised against gravity from one height to another. Similarly, since changes in enthalpy during a chemical reaction are of interest, it doesn't really matter what reference conditions are chosen. Tabulated values of enthalpy are usually based on 1 atm of pressure and 25°C (298 K); in which case, they are designated with the notation

$$H^0_{298} = \text{standard enthalpy} = \text{enthalpy at 1 atm and 298 K (kJ/mol)} \tag{9}$$

It is also assumed that the reference condition for pure elements is the stable state of the substance at 1 atm and 25°C. For example, the stable state of oxygen at 1 atm and 25°C, is gaseous O_2, so the standard enthalpy for $O_2(g)$, where (g) just means it is in the gaseous state, is defined to be 0. Similarly, mercury under those conditions is a liquid, so the standard enthalpy for $Hg(l)$ is 0, where (l) means the liquid state. A table of standard enthalpies for a number of substances is presented in Table 1. More extensive lists can be found in chemical handbooks or more advanced environmental engineering texts.

The sum of the enthalpies of the reaction products minus the sum of the enthalpies of the reactants is called the *heat of reaction*. When it is negative, heat is

TABLE 1

Standard Enthalpies for Selected Species (kJ/mol)

Substance	State[a]	H^0_{298}	Substance	State[a]	H^0_{298}
Ca^{2+}	*aq*	−543.0	HCO_3^-	*aq*	−691.1
$CaCO_3$	*s*	−1207	H_2O	*l*	−285.8
$Ca(OH)_2$	*s*	−986.6	H_2O	*g*	−241.8
C	*s*	0	NO	*g*	90.4
CO	*g*	−110.5	NO_2	*g*	33.9
CO_2	*g*	−393.5	NO_2^-	*aq*	−106.3
CO_2	*aq*	−412.9	N	*g*	472.6
CH_4	*g*	−74.9	N_2	*g*	0
C_2H_4	*g*	52.3	N_2O	*g*	81.55
CH_3COOH	*aq*	−488.5	NH_3	*aq*	−80.8
n-C_4H_{10}	*g*	−124.7	O	*g*	247.5
i-C_4H_{10}	*g*	−131.6	O_2	*g*	0
H	*g*	217.9	O_3	*g*	142.9
H^+	*aq*	0	OH^-	*aq*	−229.9
H_2	*g*	0	SO_2	*g*	−296.9

[a] *g* = gas, *aq* = aqueous, *s* = solid, *l* = liquid.

liberated during the reaction; when it is positive, heat is absorbed. Since standard enthalpies are expressed as energy per mole, we must first balance the chemical reaction, and then for each species, multiply the number of moles by the standard enthalpy to get its total enthalpy.

EXAMPLE 4 Gross Heat of Combustion for Methane

Find the heat of reaction when methane CH_4 is oxidized to CO_2 and liquid H_2O.

Solution The reaction is written here, and enthalpies taken from Table 1 appear below it.

$$\begin{array}{ccccccc} CH_4\,(g) & + & 2\,O_2\,(g) & \rightarrow & CO_2\,(g) & + & 2\,H_2O\,(l) \\ (-74.9) & & 2 \times (0) & & (-393.5) & & 2 \times (-285.8) \end{array}$$

Notice we have used the enthalpy of liquid water for this calculation. The heat of reaction is the difference between the total enthalpy of the reaction products and reactants:

$$[(-393.5) + 2 \times (-285.8)] - [(-74.9) + 2 \times (0)] = -890.2 \text{ kJ/mol of } CH_4$$

Since the heat of reaction is negative, heat is released during combustion (*i.e.*, it is exothermic).

When a fuel is burned, some of the energy released ends up as latent heat in the water vapor produced. Usually that water vapor, along with the latent heat it contains, exits the stack along with all the other combustion gases, and its heating value is, in essence, lost. That leads to two different values of what is called the *heat of combustion*. The *higher heating value* (HHV) includes the heat released when water condenses to liquid form, as was the case in Example 4. The HHV is also known as the *gross heat of combustion*. In the United States, the heat content of fossil fuels is usually expressed as this gross amount. The *lower heating value* (LHV), or *net heat of combustion,* is based on the heat of reaction when water is assumed to remain in the vapor state. The most fuel-efficient, modern furnaces used for space-heating buildings achieve their high efficiencies (above 90 percent) by causing the combustion gases to cool enough to condense the water vapor before it leaves the stack. Not unexpectedly, these are called *condensing furnaces*.

EXAMPLE 5 The Net Heat of Combustion

Find the net heat of combustion when methane is burned.

Solution We just repeat the procedure of Example 4, but this time, the water will remain in the gaseous state. Again, using values from Table 1,

$$\begin{array}{ccccccccc} CH_4\,(g) & + & 2\,O_2\,(g) & \rightarrow & CO_2\,(g) & + & 2\,H_2O\,(g) \\ (-74.9) & & 2 \times (0) & & (-393.5) & & 2 \times (-241.8) \end{array}$$

The net heat of combustion is

$$[(-393.5) + 2 \times (-241.8)] - [(-74.9) + 2 \times (0)] = -802.2 \text{ kJ/mol of } CH_4$$

Again, the sign tells us this is an exothermic reaction. Notice about 10 percent of the gross heating value is lost when water vapor is not condensed.

There is another application of the concept of enthalpy that is useful in describing photochemical smog reactions in the atmosphere and for analyzing the chemical reactions that affect the stratospheric ozone layer. When a molecule is dissociated by absorbing a photon of energy, the process is called a *photochemical reaction, photochemical dissociation,* or more simply *photolysis*. We can use enthalpies to determine the amount of energy a photon must have to cause photolysis, and from the photon's energy, we can determine the maximum wavelength that the photon can have.

For a photon to be able to cause photolysis, it must have at least as much energy as the change in enthalpy for the reaction. It is important to realize that when a molecule absorbs a photon, the energy must be used almost immediately for photolysis or else the energy will be dissipated as waste heat as the molecule collides with neighboring molecules. Molecules cannot store up energy from a series of photon encounters, waiting until enough energy is accumulated to cause photolysis. That means a *single* photon has to have sufficient energy to cause photolysis all by itself.

The following relationship between energy contained in 1 photon and the wavelength and frequency associated with that photon is introduced:

$$E = h\nu = \frac{hc}{\lambda} \tag{10}$$

where

E = Energy of a photon (J)
h = Planck's constant (6.6×10^{-34} J-s)
ν = frequency (hertz, *i.e.*, cycles per second)
c = speed of light (3×10^8 m/s)
λ = wavelength (m)

We are interested in the maximum wavelength that a photon can have and still have enough energy for photolysis, so rearranging (10) gives

$$\lambda \leq \frac{hc}{E} \tag{11}$$

Before we can equate the energy in a photon with the enthalpy change, we have to be sure the units are consistent. The reaction's enthalpy change ΔH^0 has units of kJ/mol, and E is joules per photon. Using Avogadro's number, along with the fact that 1 photon dissociates 1 molecule, lets us write:

$$\lambda \leq \frac{6.6 \times 10^{-34}\ \text{J} \cdot \text{s} \times 3 \times 10^{8}\ \text{m/s} \times 6.02 \times 10^{23}\ \text{molecules/mol} \times 1\ \text{photon/molecule}}{\Delta H^0\ (\text{kJ/mol}) \times 10^{3}\ \text{J/kJ}}$$

$$\lambda(\text{m}) \leq \frac{1.19 \times 10^{-4}\ (\text{kJ} \cdot \text{m/mol})}{\Delta H^0\ (\text{kJ/mol})} \tag{12}$$

EXAMPLE 6 Photolysis of Ozone

What maximum wavelength of light would be capable of causing photolysis of ozone O_3 into O_2 and O?

Solution First write the reaction, including enthalpies from Table 1. Even though those enthalpies are for standard conditions (1 atm, 298 K), they can be used under stratospheric pressure and temperature conditions with only modest error.

$$\begin{array}{ccccccc} O_3 & + & h\nu & \rightarrow & O_2 & + & O \\ (142.9) & & & & (0) & & (247.5) \end{array}$$

The enthalpy change is

$$\Delta H^0 = 247.5 - 142.9 = 104.6\ \text{kJ/mol}$$

Since the sign of ΔH^0 is positive, this is an endothermic reaction that needs to absorb energy for it to take place. That energy comes from the photon.

From Equation (12),

$$\lambda \leq \frac{1.19 \times 10^{-4}\ \text{kJ} \cdot \text{m/mol}}{104.6\ \text{kJ/mol}} = 1.13 \times 10^{-6}\ \text{m} = 1.13\ \mu\text{m}$$

Absorption of incoming solar radiation by ozone, as described in Example 6, is part of the shielding that the stratospheric ozone layer provides for Earth.

4 Chemical Equilibria

In the reactions considered so far, the assumption has been that they proceed in one direction only. Most chemical reactions are, to some extent, reversible, proceeding in both directions at once. When the rates of reaction are the same—that is products are being formed on the right at the same rate as they are being formed on the left—the reaction is said to have reached *equilibrium*.

In this generalized reversible reaction

$$aA + bB \rightleftharpoons cC + dD \tag{13}$$

the small letters a, b, c, and d are coefficients corresponding to the number of molecules or ions of the respective substances that result in a balanced equation. The capital letters A, B, C, and D are the chemical species. The double arrow designation indicates that the reaction proceeds in both directions at the same time.

At equilibrium, we can write that

$$\frac{[C]^c[D]^d}{[A]^d[B]^b} = K \tag{14}$$

where the [ ] designation represents concentrations of the substances at equilibrium, expressed in moles per liter (molarity). Do not use concentrations in mg/L! K is called the *equilibrium constant*. Informally, the equilibrium constant is spoken of as the products over reactants or the ratio of products to reactants. It should also be emphasized that (14) is valid only when chemical equilibrium is established, if ever. Natural systems are often subject to constantly changing inputs, and since some reactions occur very slowly, equilibrium may never be established. A practicing environmental engineer must therefore use this important equation with a certain degree of caution.

Equilibrium constants are used in so many ways in water chemistry that they often are given specific names to describe a particular equilibrium process. Many molecules, when dissolved in water, separate into positively charged ions, called *cations,* and negatively charged ions, called *anions*. Equation (14) can be applied to the dissociation of such molecules, in which case K is referred to as a *dissociation constant* or an *ionization constant*. The quantity on the left in (13) may be a solid, and the problem is to determine the degree to which that solid enters solution. In such cases, the equilibrium constant is called the *solubility product* and is often denoted K_{SP}. In another process of environmental interest where a gas dissolves in water, the equilibrium constant describing the concentration of gas in the water in equilibrium with the concentration of the gas in the air is termed *Henry's constant* and denoted K_H. These and other equilibrium constants will be discussed in the next few sections. To lessen confusion, the kinetic rate constants are typically signified using a lowercase k, whereas equilibrium constants are written with an uppercase K.

Finally, often when dealing with very large and very small numbers, it is helpful to introduce the following logarithmic measure:

$$X = 10^{-pX} \tag{15}$$

or

$$pX = -\log X \tag{16}$$

where X is any number or a variable denoting any number.

Acid-Base Reactions

Water dissociates slightly into hydrogen ions (H^+) and hydroxide ions (OH^-) as the following reaction suggests

$$H_2O \rightleftharpoons H^+ + OH^- \tag{17}$$

The hydrogen ions, H^+, are also known as protons because when a hydrogen atom loses an electron—is ionized—all that remains is a proton. The corresponding equilibrium expression for this reaction is

$$\frac{[H^+][OH^-]}{[H_2O]} = K \tag{18}$$

The molar concentration of water $[H_2O]$ is 1,000 g/L divided by 18 g/mol, or 55.56 mol/L. Since water dissociates only slightly, the molar concentration after ionization is not changed enough to be of significance, so $[H_2O]$ is essentially a constant that can be included in the equilibrium constant. The result is

$$[H^+][OH^-] = K_w = 1 \times 10^{-14} \quad \text{at } 25°\text{C} \tag{19}$$

where K_w is the dissociation constant for water. For dilute aqueous solutions in general, $[H_2O]$ is considered constant and is included in the equilibrium constant. K_w is slightly temperature dependent, but unless otherwise stated, the value given in (19) at 25°C is the assumed value.

Note that (19) holds, no matter what the source of hydrogen ions or hydroxide ions. That is, the equation is valid even if other substances dissolved in the water make their own contributions to the hydrogen and hydroxide supplies. It is always one of the equations that must be satisfied when chemical equilibria problems in water are analyzed.

$[H^+]$ and $[OH^-]$ concentrations are customarily expressed using the logarithmic measure introduced in (15) and (16). To express hydrogen ion concentrations (also known as proton concentrations), the pH scale is used, where

$$\text{pH} = -\log[H^+] \tag{20}$$

or,

$$[H^+] = 10^{-\text{pH}} \tag{21}$$

With the pH scale, it is easy to specify whether a solution is acidic, basic, or neutral. In a *neutral* solution, the concentration of hydrogen ions $[H^+]$ equals the concentration of hydroxide ions $[OH^-]$. From (19), for a neutral solution

$$[H^+][OH^-] = [H^+][H^+] = [H^+]^2 = 10^{-14}$$

so

$$[H^+] = 10^{-7}$$

and a neutral solution has a pH of 7 (written pH 7).

An *acidic* solution is one in which $[H^+]$ is greater than $[OH^-]$; that is, the hydrogen ion concentration is greater than 10^{-7} mol/L, and its pH is less than 7. A *basic* solution is the other way around with more hydroxide ions than hydrogen ions and a pH greater than 7. Notice for every unit change in pH, the concentration of hydrogen ions changes by a factor of 10.

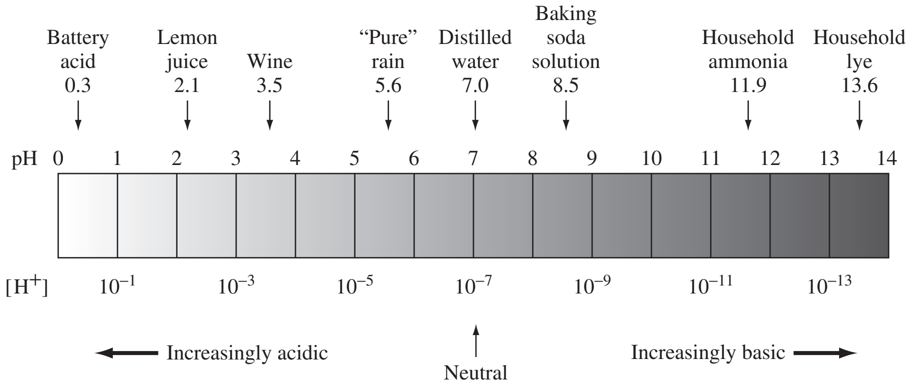


FIGURE 2 The pH scale.

Figure 2 illustrates the pH scale, showing example values of pH for several common solutions. Notice in the figure that a distinction is made between distilled water and "pure" rainfall. As will be seen in a later section on carbonates, as rainwater falls, it absorbs carbon dioxide from the air and carbonic acid is formed. Unpolluted rainfall then has a natural pH of around 5.6. By the usual definition, acid rain, caused by industrial pollutants, has pH lower than 5.6. Actual acid rain and acid fog have been recorded with pH below 2.0.

EXAMPLE 7 pH of Tomato Juice

Find the hydrogen ion concentration and the hydroxide ion concentration in tomato juice having a pH of 4.1.

Solution From (21), the hydrogen ion concentration is

$$[H^+] = 10^{-pH} = 10^{-4.1} = 7.94 \times 10^{-5} \text{ mol/L}$$

From (19), the hydroxide ion concentration is

$$[OH^-] = \frac{10^{-14}}{[H^+]} = \frac{10^{-14}}{7.94 \times 10^{-5}} = 1.26 \times 10^{-10} \text{ mol/L}$$

Water is not the only source of hydrogen ions or hydroxide ions. When a chemical dissociates in water and produces a hydrogen ion, it is an acid. As you would expect, any chemical that dissociates and gives off a hydroxide ion is a base. In other words, acids add protons (or subtract hydroxide ions) in water, whereas bases add hydroxide ions (or subtract protons). For instance, when acetic acid ($C_2H_4O_2$), the main component in vinegar, is dissolved in water, some of it dissociates to protons (H^+) and acetate ($C_2H_3O_2^-$). At equilibrium,

$$CH_3COOH + H_2O \rightleftharpoons CH_3COO^- + H^+ + H_2O$$

The chemical composition of acetic acid is usually written as CH_3COOH rather than $C_2H_4O_2$ to emphasize the presence of the carboxylic acid group (—COOH), an organic functional group that will be explained later in this chapter. Since H_2O

appears as both a reactant and product, it is usually dropped from the equation, so the equilibrium expression is written as

$$CH_3COOH \rightleftharpoons CH_3COO^- + H^+ \tag{22}$$

and the dissociation constant (with a subscript "a" for acidity constant) is

$$K_a = \frac{[CH_3COO^-][H^+]}{[CH_3COOH]} = 10^{-4.76} \tag{23}$$

The degree of dissociation of an acid is a function of pH. This can be seen by rearranging (23) and using (21) to get

$$\frac{K_a}{[H^+]} = \frac{K_a}{10^{-pH}} = \frac{[CH_3COO^-]}{[CH_3COOH]} \tag{24}$$

As pH goes up, that is, the H^+ concentration goes down, then the ratio of concentration of acetate to acetic acid goes up. This is true for all acids; as the pH increases the degree of dissociation increases. In addition, when $[H^+] = K_a$ ($pH = pK_a$), then by (24), $[CH_3COOH] = [CH_3COO^-]$. Figure 3 shows this graphically in what is called an *α (alpha) diagram.* In an α diagram, α_0 denotes the fraction of total acid-derived species present (for this example, $[CH_3COOH] + [CH_3COO^-]$) in the undissociated form (in this example, CH_3COOH), whereas α_1 is the fraction that is in the dissociated or singly deprotonated form (in this example, CH_3COO^-). Remember if we had added acetate rather than acetic acid to the water, the acetate would have been a base because some of it would have withdrawn hydrogen ions from the water to form acetic acid to satisfy the (24) equilibrium between CH_3COO^- and CH_3COOH. In fact, we call acetate the conjugate base of acetic acid.

Acetic acid is an example of a *monoprotic acid,* meaning it can lose 1 proton (H^+) when dissolved in water. There are also *diprotic, triprotic,* and even *quadriprotic acids.* Table 2 shows some of the acids of environmental importance. As you would expect, the number of dissociation constants corresponds to the number of hydrogen ions an acid can lose, so the dissociation constant for dissociating the first proton is K_{a1}, for the second proton is K_{a2}, and so on. The behavior of a very important diprotic acid system, the carbonate system, will be discussed later in this chapter.

Acid-base reactions are among the most important in environmental engineering. Most aquatic forms of life, for example, are very sensitive to the pH of their

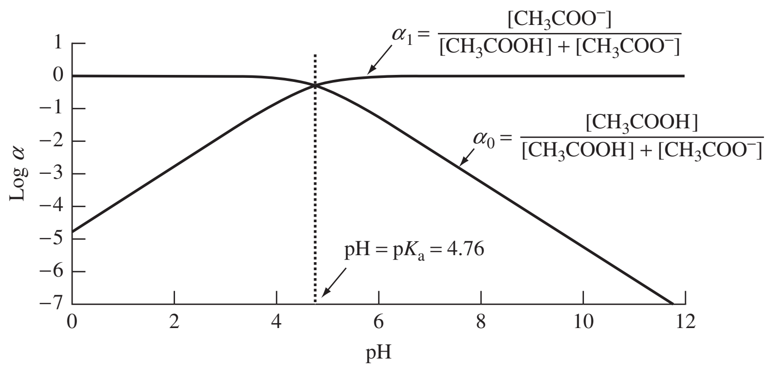


FIGURE 3 α diagram for acetic acid in water.

TABLE 2

Dissociation Constants and Chemical Formulas for Some Acids of Environmental Importance

Acid	Formula	pK_{a1}	pK_{a2}	pK_{a3}
Nitric acid	HNO_3	−1.30	—	—
Hydrochloric acid	HCl	<0	—	—
Sulfuric acid	H_2SO_4	<0	1.99	—
Phosphoric acid	H_3PO_4	2.16	7.20	12.35
Arsenic acid	H_3AsO_4	2.24	6.76	—
Citric acid	$C_3H_4OH(COOH)_3$	3.13	4.72	6.33
Acetic acid	CH_3COOH	4.76	—	—
Carbonic acid	H_2CO_3	6.35	10.33	—
Hydrogen sulfide	H_2S	6.99	12.92	—
Hypochlorous acid	$HOCl$	7.60	—	—
Ammonium ion	NH_4^+	9.26	—	—
Silicic acid	H_4SiO_4	9.84	13.20	—

habitat. To protect local ecosystems, waste neutralization before release is common. In other circumstances, by manipulating pH, unwanted substances can be driven out of a solution as precipitates or gases before the effluent is released.

EXAMPLE 8 Drinking Water Disinfection Using Chlorine

Chlorine is the active ingredient in most household bleach and is one of the most commonly used and inexpensive chemical disinfectants for water. The chlorine is in the hypochlorous acid form, HOCl, and hypochlorous acid is a much better disinfectant than hypochlorite, OCl^-, its conjugate base. If bleach is used to disinfect water, below what pH should the water be maintained so that at least 95 percent of the chlorine added is in the hypochlorous acid form?

Solution From Table 2 and inspection, we know hypochlorous acid is a monoprotic acid. Its equilibrium reaction is

$$HOCl \rightleftharpoons OCl^- + H^+ \tag{25}$$

The equilibrium constant equation is

$$K_{a1} = \frac{[OCl^-][H^+]}{[HOCl]} = 10^{-7.60} \tag{26}$$

which rearranges to

$$[H^+] = \frac{[HOCl]K_{a1}}{[OCl^-]}$$

We can then solve for $[H^+]$ when $[HOCl]/[OCl] = 95/5 = 19$, so

$$[H^+] = 19 \times 10^{-7.60} = 4.77 \times 10^{-7} = 10^{-6.32}$$

If the pH is kept below 6.3, then more than 95 percent of the chlorine will be in the HOCL form.

TABLE 3

Selected Solubility Product Constants at 25°C

Equilibrium Equation	K_{sp} at 25°C	
$CaCO_3 \rightleftharpoons Ca^{2+} + CO_3^{2-}$	5×10^{-9}	Hardness removal, scaling
$CaSO_4 \rightleftharpoons Ca^{2+} + SO_4^{2-}$	2×10^{-5}	Flue gas desulfurization
$Cu(OH)_2 \rightleftharpoons Cu^{2+} + 2OH^-$	2×10^{-19}	Heavy metal removal
$Al(OH)_3 \rightleftharpoons Al^{3+} + 3OH^-$	1×10^{-32}	Coagulation
$Ca_3(PO_4)_2 \rightleftharpoons 3Ca^{2+} + 2PO_4^{3-}$	1×10^{-27}	Phosphate removal
$CaF_2 \rightleftharpoons Ca^{2+} + 2F^-$	3×10^{-11}	Fluoridation

Source: Sawyer et al., 1994.

Solubility Product

All solids are to some degree soluble, although some are much more so than others. A generalized equation describing the equilibrium condition in which a solid is dissociating into its ionic components (*dissolution*) at the same rate that ionic components are recombining into the solid form (*precipitation*) is

$$\text{Solid} \rightleftharpoons aA + bB \tag{27}$$

where A and B are the ionic components that make up the solid. Applying (14) yields

$$\frac{[A]^a[B]^b}{[\text{solid}]} = K \tag{28}$$

As long as there is still solid present at equilibrium, its effect can be incorporated into the equilibrium constant

$$[A]^a[B]^b = K_{sp} \tag{29}$$

where K_{sp} is called the *solubility product*. Table 3 gives a short list of solubility products of particular importance in environmental engineering.

As an example of the use of (29), consider the fluoridation of water with calcium fluoride, CaF_2.

EXAMPLE 9 Fluoride Solubility

Find the equilibrium concentration of fluoride ions in pure water caused by the dissociation of CaF_2. Express the answer both in units of mol/L and mg/L.

Solution From Table 3, the reaction and solubility product are

$$CaF_2 \rightleftharpoons Ca^{2+} + 2F^- \qquad K_{sp} = 3 \times 10^{-11}$$

Remembering to square the term for the fluoride ion, the mass action equation (29) becomes

$$[Ca^{2+}][F^-]^2 = 3 \times 10^{-11} \tag{30}$$

If we let s (mol/L) represent the concentration of Ca^{2+}, then the concentration of F^- will be $2s$, since 2 moles of F^- are dissociated for every mole of Ca^{2+} dissociated. Thus

$$[Ca^{2+}] = s$$

$$[F^-] = 2s$$

and from (30)

$$K_{sp} = s \times (2s)^2 = 4s^3 = 3 \times 10^{-11}$$

$$s = [Ca^{2+}] = 2 \times 10^{-4} \text{ mol/L}$$

$$2s = [F^-] = 4 \times 10^{-4} \text{ mol/L}$$

To find the concentration of fluoride ions in mg/L, we need the atomic weight of fluorine, which is 19.

$$[F^-] = 4 \times 10^{-4} \text{ mol/L} \times 19 \text{ g/mol} \times 10^3 \text{ mg/g} = 7.6 \text{ mg/L}$$

The fluoride concentration obtained in Example 9 is above recommended drinking water levels of 1.8 mg/L. Fluoride concentrations of approximately 1 mg/L in drinking water help prevent cavities in children, but discoloration of teeth, called *mottling,* is relatively common when concentrations exceed 2.0 mg/L.

Solubility of Gases in Water

When air comes in contact with water, some of it dissolves into the water. Different constituents of air dissolve to different degrees and in amounts that vary with temperature and water purity. The behavior of gases in contact with water was reported by W. Henry in England in 1903, and the resulting relationship is known as *Henry's law*:

$$[\text{gas}] = K_H P_g \tag{31}$$

where

$[\text{gas}]$ = concentration of dissolved gas (mol/L)
K_H = Henry's law constant (mol/L·atm)
P_g = the partial pressure of the gas in air (atm)

The quantity P_g is the partial pressure of the gas in air, which is simply its volumetric concentration times the atmospheric pressure. The units suggested in the preceding list for pressure are atmospheres (atm), where 1 atm corresponds to 101,325 pascals (Pa), and 1 Pa equals 1 newton per square meter. For example, oxygen makes up about 21 percent of the atmosphere, so at 1 atm P_g would be $0.21 \times 1\,\text{atm} = 0.21\,\text{atm}$.

Each gas-liquid system has its own value for Henry's coefficient. The coefficient varies both with temperature (solubility decreases as temperature increases) and with concentration of other dissolved gases and solids (the solubility decreases as other dissolved material in the liquid increases). Note that Henry's law is expressed in various ways with different units for the coefficient depending on the

TABLE 4

Henry's Law Coefficients, K_H (mol/L · atm)

T (°C)	CO_2	O_2
0	0.076425	0.0021812
5	0.063532	0.0019126
10	0.053270	0.0016963
15	0.045463	0.0015236
20	0.039172	0.0013840
25	0.033363	0.0012630

method of expression. The user must be careful, then, to check the units given for Henry's constant before applying the law. Table 4 gives some values of K_H for CO_2 and O_2, two of the gases that we will be most concerned with.

Because atmospheric pressure changes with altitude, P_g will change as well. One estimate for atmospheric pressure as a function of altitude is the following (Thomann and Mueller, 1987):

$$P = P_o - 1.15 \times 10^{-4} H \tag{32}$$

where

P = atmospheric pressure at altitude H (atm)
H = altitude (m)
P_o = atmospheric pressure at sea level (atm)

EXAMPLE 10 **Solubility of Oxygen in Water**

By volume, the concentration of oxygen in air is about 21 percent. Find the equilibrium concentration of O_2 in water (in mol/L and mg/L) at 25°C and 1 atmosphere of pressure. Recalculate it for Denver at an altitude of 1,525 m.

Solution Air is 21 percent oxygen, so its partial pressure at 1 atm is

$$P_g = 0.21 \times 1 \text{ atm} = 0.21 \text{ atm}$$

From Table 4, at 25°C, K_H = 0.0012630 mol/L · atm; so, from (31),

$$\begin{aligned}[O_2] &= K_H\ P_g = 0.0012630 \text{ mol/L} \cdot \text{atm} \times 0.21 \text{ atm} \\ &= 2.65 \times 10^{-4} \text{ mol/L} \\ &= 2.65 \times 10^{-4} \text{ mol/L} \times 32 \text{ g/mol} \times 10^3 \text{ mg/g} = 8.5 \text{ mg/L}\end{aligned}$$

In Denver, at 1,525 m, atmospheric pressure can be estimated using (32):

$$P = P_o - 1.15 \times 10^{-4} H = 1 - 1.15 \times 10^{-4} \times 1{,}525 = 0.825 \text{ atm}$$

so

$$\begin{aligned}[O_2] &= 0.0012630 \text{ mol/L} \cdot \text{atm} \times 0.21 \times 0.825 \text{ atm} = 2.19 \times 10^{-4} \text{ mol/L} \\ &= 2.19 \times 10^{-4} \text{ mol/L} \times 32 \text{ g/mol} \times 10^3 \text{ mg/g} = 7.0 \text{ mg/L}\end{aligned}$$

Calculations based on Henry's law provide equilibrium concentrations of dissolved gases, or, as they are frequently called, *saturation* values. Often, actual values differ considerably from those at equilibrium. There may be more than the saturation value of a dissolved gas, as when photosynthesis by plants pumps oxygen into the water at a faster rate than it can leave through the air/water interface. It is more common for dissolved gases to be less than the saturation value, as occurs when bacteria decompose large quantities of waste, drawing oxygen from the water (possibly leading to an oxygen deficiency that can kill fish). In either case, when an excess or a deficiency of a dissolved gas occurs, pressures act to try to bring the amount dissolved back to the saturation level.

As an example of the value of being able to engineer pH and utilize gas transfer across the air/water interface, consider the problem of removing nitrogen from wastewater. Nitrogen is a nutrient that can degrade water quality by stimulating excessive algal growth in receiving bodies of water. Nitrogen in the form of nitrate NO_3^- also poses another problem, especially in groundwater. When babies drink water with high nitrate content, a potentially lethal condition known as methemoglobinemia can result.

One way to remove nitrogen during wastewater treatment is with a process known as *ammonia stripping*. When organic matter decomposes, nitrogen is first released in the form of ammonia NH_3 or ammonium ion NH_4^+. Ammonium ions are acids, whereas NH_3 is ammonia gas, which dissolves in water with a Henry's constant of 57 mol/L · atm at 25°C. Even though ammonia is much more soluble in water than, for instance, the oxygen in the preceding example, the concentration of ammonia in the air is near zero so ammonia is readily driven from water (stripped) and transferred to the air as the system tries to reach the air/water equilibrium state. By driving the acid/base equilibrium reaction

$$NH_4^+ \rightleftharpoons NH_3 + H^+ \tag{33}$$

toward the right, more of the nitrogen will be in the ammonia gas form, which can then be stripped from solution and enter the air in a gas stripping tower. In such a tower, contaminated water is allowed to trickle downward over slats or corrugated surfaces while clean air is blown in from the bottom to aerate the dripping water. Gas stripping can be used in wastewater treatment facilities to remove such gases as ammonia and hydrogen sulfide, and it can also be used to remove *volatile organic chemicals* (VOCs) from contaminated groundwater.

To strip ammonia, the reaction given in (33) must be driven toward the right, which can be done by increasing the concentration of OH^-. This may be more easily seen by adding a hydroxide ion (OH^-) to both sides of (33), which gives

$$NH_4^+ + OH^- \rightleftharpoons NH_3 + H_2O \tag{34}$$

Thus, by adding OH^-, in other words raising the pH, (34) is driven to the right so that more of the nitrogen is in the ammonia and strippable form.

EXAMPLE 11 Ammonia Stripping

Nitrogen in a wastewater treatment plant is in the form of ammonia and ammonium ion and has a total concentration of 7.1×10^{-4} M. The Henry's constant (at 25°C) is 57 mol/L · atm.

a. Find the fraction of the nitrogen that is in the ammonia form (and hence strippable) as a function of pH, and draw a graph.

b. If the wastewater pH is raised to 10 and the atmospheric ammonia concentration is 5.0×10^{-10} atm, what would be the equilibrium concentration of total nitrogen in the wastewater after air stripping?

Solution From (33) and Table 2, the equilibrium equation for the reaction is

$$K_a = \frac{[NH_3][H^+]}{[NH_4^+]} = 10^{-9.26} \tag{35}$$

a. What we want to find is the fraction of nitrogen in the form of ammonia, or

$$NH_3 \text{ fraction} = \frac{[NH_3]}{[NH_3] + [NH_4^+]} = \frac{1}{1 + [NH_4^+]/[NH_3]} \tag{36}$$

Rearranging (35) gives

$$\frac{[NH_4^+]}{[NH_3]} = \frac{[H^+]}{K_a} \tag{37}$$

and putting this into (36) gives

$$\begin{aligned} NH_3 \text{ fraction} &= \frac{1}{1 + [H^+]/K_a} \\ &= \frac{1}{1 + 10^{-pH}/10^{-9.26}} \\ &= \frac{1}{1 + 10^{(9.26-pH)}} \end{aligned} \tag{38}$$

A table of values for this fraction can easily be generated from (38); the results are plotted in Figure 4.

pH	$\frac{[NH_3]}{[NH_3] + [NH_4^+]}$
4	0.000005
6	0.0005
8	0.052
9	0.35
10	0.85
12	0.99

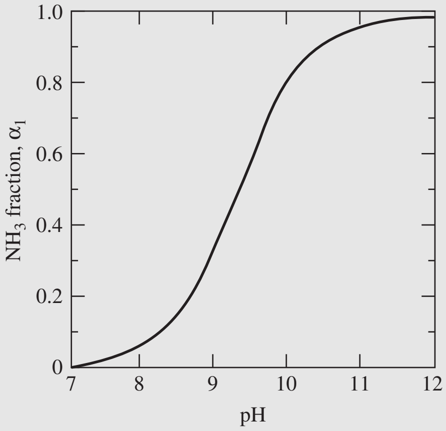


FIGURE 4 Dependence of the ammonia fraction, α_1, on pH.

b. From (31) and the values given

$$[NH_3] = K_H P_g = 57\frac{\text{mol}}{\text{L}\cdot\text{atm}} \times 5.0 \times 10^{-10}\ \text{atm}$$
$$= 2.85 \times 10^{-8}\ \text{mol/L}$$

Using the pH 10 value in Figure 4 and (36)

$$NH_3\,\text{fraction} = 0.85 = \frac{[NH_3]}{[NH_3] + [NH_4^+]}$$

so,

$$[NH_3] + [NH_4^+] = \frac{[NH_3]}{0.85} = \frac{2.85 \times 10^{-8}\ \text{mol/L}}{0.85}$$
$$= 3.35 \times 10^{-8}\ \text{mol/L}$$

and the percentage removed would be

$$\%\,\text{removal} = \left(1 - \frac{3.35 \times 10^{-8}\ \text{M}}{7.1 \times 10^{-4}\ \text{M}}\right) \times 100 = 99.995\%$$

As Example 11 suggests, to drive the reaction significantly toward the ammonia side requires a pH in excess of about 10. Because typical wastewaters seldom have a pH that high, it is necessary to add chemicals, usually lime (CaO), to sufficiently raise the pH. The lime, unfortunately, reacts with CO_2 in the air to form a calcium carbonate scale that can accumulate on stripping surfaces and which must be removed periodically, creating a sludge disposal problem of its own.

The preceding example is just one illustration of the value of being able to control pH in treatment processes. In other circumstances, pollution affects pH, driving reactions in directions that may be undesirable. A notable example of this phenomenon is the mobilization of trace metals, especially aluminum, when pH drops during acidification of a watershed. Aluminum is very toxic to fish, but its normal form is an insoluble solid and relatively safe. However, when acid deposition acidifies a lake, the reduced pH drives equilibria reactions toward liberation of mobile Al^{3+} ions, which are very toxic.

The Carbonate System

The carbonate system is the most important acid-base system in natural waters because it largely controls pH. It is comprised of the following chemical species:

Aqueous carbon dioxide	$CO_2\,(aq)$
Carbonic acid	H_2CO_3
Bicarbonate ion	HCO_3^-
Carbonate ion	CO_3^{2-}

Aqueous CO_2 is formed when atmospheric CO_2 dissolves in water; we can find its concentration in fresh water using Henry's law (31):

$$[CO_2]_{aq} = K_H P_g \tag{39}$$

The concentration is in (mol/L) and P_{CO_2} is the partial pressure of gaseous CO_2 in the atmosphere (about 360 ppmv or 360×10^{-6} atm). Aqueous CO_2 then forms carbonic acid (H_2CO_3). Carbonic acid is a diprotic acid, which can lose two protons (hydrogen ions, H^+). The first proton is released when it ionizes to form bicarbonate (HCO_3^-):

$$CO_2(aq) + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^- \tag{40}$$

The bicarbonate (HCO_3^-) further dissociates to release another hydrogen ion and carbonate (CO_3^{2-}):

$$HCO_3^- \rightleftharpoons H^+ + CO_3^{2-} \tag{41}$$

In addition, if there is a source of solid calcium carbonate ($CaCO_3$), such as occurs when water is in contact with limestone rocks, then the solubility reaction for $CaCO_3(s)$ applies in the system. Such waters are referred to as *calcareous* waters.

$$CaCO_3(s) \rightleftharpoons Ca^{2+} + CO_3^{2-} \tag{42}$$

If sufficient time is allowed for the system to reach equilibrium, then the equilibrium constants given in Table 2 for reactions (40–41) and in Table 3 for reaction (42) can be used to analyze the system. Reaction (40) results in

$$\frac{[H^+][HCO_3^-]}{[CO_2(aq)]} = K_1 = 4.47 \times 10^{-7} \text{ mol/L} \tag{43}$$

and (41) yields

$$\frac{[H^+][CO_3^{2-}]}{[HCO_3^-]} = K_2 = 4.68 \times 10^{-11} \text{ mol/L} \tag{44}$$

while (42) is governed by

$$[Ca^{2+}][CO_3^{2-}] = K_{sp} = 4.57 \times 10^{-9} \text{ mol}^2/\text{L}^2 \tag{45}$$

The values of K_1, K_2, and K_{sp} are temperature dependent. The values given in (43–45) are at 25°C.

Four possible conditions determine the behavior of the carbonate system. There may or may not be a source of solid carbonate as indicated in (42), and the solution may or may not be exposed to the atmosphere. If it is exposed to the atmosphere, the system is said to be *open*; if it is not, the system is *closed*. We will analyze the common situation of an open system not in contact with a carbonate solid, such as occurs when rainwater passes through the atmosphere absorbing CO_2 and forming carbonic acid.

Before proceeding to that calculation, it is useful to compare the relative concentrations of carbonate and bicarbonate as a function of pH. Dividing (44) by $[H^+]$ gives

$$\frac{[CO_3^{2-}]}{[HCO_3^-]} = \frac{K_2}{[H^+]} = \frac{K_2}{10^{-pH}} \tag{46}$$

and then incorporating the value of K_2 given in (44) yields:

$$\frac{[CO_3^{2-}]}{[HCO_3^-]} = \frac{4.68 \times 10^{-11}}{10^{-pH}} = 4.68 \times 10^{(pH-11)} \tag{47}$$

In (47), unless pH is extremely high, the carbonate concentration is usually negligible compared to the bicarbonate concentration, a fact that is useful in the

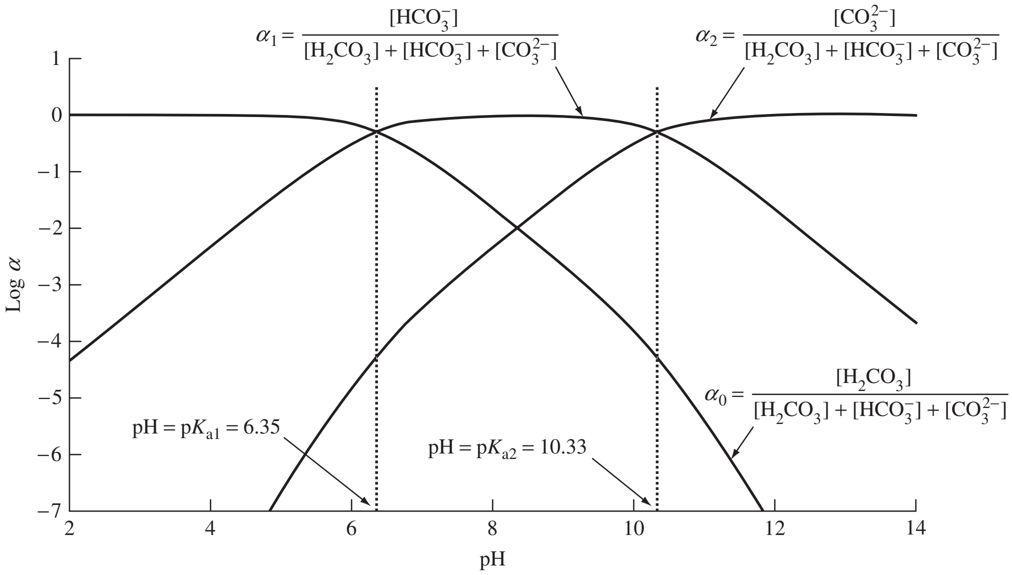


FIGURE 5 α diagram for the carbonate system. α_0 is the fraction of the total carbonate in the system that is in the carbonic acid form (H_2CO_3), α_1 is the fraction in the bicarbonate anion form (HCO_3^-), and α_2 is the fraction in the carbonate anion form (CO_3^{2-}).

following example where we calculate the pH of pristine rainwater. The pH dependency of the relative fractions of carbonate, bicarbonate, and carbonate acid is easily seen in the α diagram (Figure 5) for the carbonate system.

EXAMPLE 12 The pH of Natural Rainwater

Estimate the pH of natural rainwater, assuming that the only substance affecting it is the absorption of CO_2 from the atmosphere. Assume that the concentration of CO_2 is 390 ppmv, and the temperature and pressure are 25°C and 1 atmosphere.

Solution We have a number of equations to work with that are based on equilibrium constants, but another relationship needs to be introduced based on charge neutrality. The rainwater is assumed to have started without any electrical charge, and no net charge is added to it by the absorption of carbon dioxide. To maintain this neutrality, the total positive charge contributed by the hydrogen ions (H^+) must equal the total negative charge of the bicarbonate (HCO_3^-), carbonate (CO_3^{2-}), and hydroxyl (OH^-) ions:

$$[H^+] = [HCO_3^-] + 2[CO_3^{2-}] + [OH^-] \tag{48}$$

Notice that since each carbonate ion (CO_3^{2-}) has two negative charges, its charge contribution is twice as much per mole; hence, its coefficient is 2.

Knowing that rainfall is likely to be slightly acidic will let us simplify (48). Equation (41) indicates that for an acidic solution, $[CO_3^{2-}] \ll [HCO_3^-]$, so (48) becomes

$$[H^+] = [HCO_3^-] + [OH^-] \tag{49}$$

Another equation that must be satisfied is (19):

$$[H^+][OH^-] = K_w = 10^{-14} \tag{19}$$

Substituting (19) into (49) gives

$$[H^+] = [HCO_3^-] + \frac{10^{-14}}{[H^+]} \tag{50}$$

Placing $[HCO_3^-]$ from (43) into (50) yields

$$[H^+] = \frac{K_1[CO_2(aq)] + 10^{-14}}{[H^+]} \tag{51}$$

So,

$$[H^+]^2 = K_1[CO_2(aq)] + 10^{-14} \tag{52}$$

We can find the saturation value of dissolved CO_2 from (39):

$$\begin{aligned}[CO_2] &= K_H P_g \\ &= 0.033363\,\text{mol/L}\cdot\text{atm} \times 390 \times 10^{-6}\,\text{atm} = 1.3 \times 10^{-5}\,\text{mol/L}\end{aligned} \tag{39}$$

where K_H was taken from Table 4, and P_{CO_2} is given as 390 ppm or 390×10^{-6} atm. Equation (52) now gives

$$[H^+]^2 = (4.47 \times 10^{-7} \times 1.3 \times 10^{-5} + 10^{-14})(\text{mol/L})^2$$

so

$$[H^+] = 2.42 \times 10^{-6}\,\text{mol/L}$$
$$pH = -\log[H^+] = 5.62$$

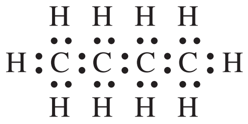
Example 12 indicates that the pH of pristine rainwater is not the same as the pH of pure water due to the presence of carbon dioxide in the atmosphere. Other naturally occurring substances in the atmosphere also affect pH. Sulfur dioxide (SO_2) lowers pH, and ammonia and alkaline dust raise it. When all such natural substances are accounted for, rainfall is likely to have pH somewhere between 5 and 6 before it is influenced by human activities. As a result, some define acid rain caused by human activities to be rainfall with a pH of 5 or below, whereas others prefer to define it in terms of the preceding carbonate calculation, that is, as precipitation with pH less than 5.6.

5 Organic Chemistry

Organic chemistry has come to mean the chemistry of the compounds of carbon. This term is broader than the term *biochemistry,* which can be described as the chemistry of life. The need for a distinction between the two began with a discovery by Fredrich Wöhler in 1828. He accidentally converted an inorganic compound, ammonium cyanate, into urea, which until that time had been thought to be strictly associated with living things. That discovery demolished the *vital-force theory,* which held that organic and inorganic compounds were distinguished by some sort of "life force." With carbon at the heart of the definition of organic chemistry, even DDT is just as organic as yogurt.

The science of organic chemistry is incredibly complex and varied. Literally millions of different organic compounds are known today, and 100,000 or so of these are products of synthetic chemistry, unknown in nature. About all that can be done here in a few pages is to provide the barest introduction to the origins of some of their names and structures, so that they will appear a bit less alien when you encounter them in the rest of the book.

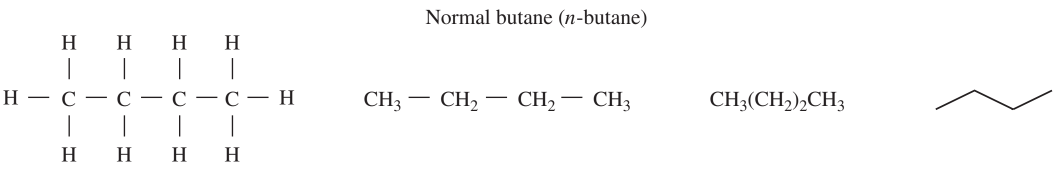
One way to visualize the bonding of atoms is with electron-dot formulas that represent valence electrons in the outermost orbitals. These diagrams were developed by G. N. Lewis in 1916 and are referred to as *Lewis structures*. According to Lewis' theory of covalence, atoms form bonds by losing, gaining, or sharing enough electrons to achieve the outer electronic configuration of a noble gas. For hydrogen, that means bonding with two shared electrons; for many other elements, including carbon, it means achieving a pattern with eight outermost electrons. For example, the following Lewis structure for the compound butane (C_4H_{10}) clearly shows each hydrogen atom sharing a pair of electrons and each carbon atom sharing four pairs of electrons.



n-butane

A common way to simplify Lewis structures is to use a single straight line to represent pairs of shared electrons, as shown next. For convenience, the structural formula is often written linearly by combining hydrogens bonded to a given carbon and then showing those subunits connected by carbon-carbon bonds. Thus, for example, butane becomes $CH_3—CH_2—CH_2—CH_3$. Further simplification, referred to as a *condensed formula*, can be gained by not actually showing the C—C bonds but merely grouping atoms connected to each carbon and writing the result. Butane becomes $CH_3(CH_2)_2CH_3$.

Another way to simplify the structural formula for some organic chemicals is with a symbolic "kinky" diagram. Each carbon in such a diagram is indicated with a kink, and the hydrogens are not shown at all. The number of hydrogens is easy to determine since the number of bonds to the carbon kink must be 4. A kinky diagram for *n*-butane is shown here.



Normal butane (*n*-butane)

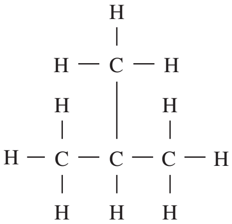
Structural

Linear: $CH_3 — CH_2 — CH_2 — CH_3$

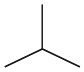
Condensed: $CH_3(CH_2)_2CH_3$

Kinky

When butane has the structure shown, that is with all of its carbons lined up in a straight chain (no side branches), it is called *normal*-butane, or *n*-butane. Butane also occurs in another form known as isobutane. Isobutane (*i*-butane) has the same molecular formula, C_4H_{10}, but has a different structural formula as shown here:



Isobutane (*i*-butane)



Kinky diagram

Compounds having the same molecular formula but different structural formulas are known as structural *isomers*. Isomers may have very different chemical and physical properties. As the number of carbon atoms per molecule increases, the number of possible isomers increases dramatically. Decane ($C_{10}H_{22}$), for example, has 75 possible isomers.

Butane and decane are examples of an important class of compounds called *hydrocarbons,* compounds containing only H and C atoms. Hydrocarbons in which each carbon atom forms four single bonds to other atoms are called *saturated hydrocarbons, paraffins,* or *alkanes*. Alkanes form a series beginning with methane (CH_4), ethane (C_2H_6), propane (C_3H_8), and butane (C_4H_{10}), which are all gases. The series continues with pentane (C_5H_{12}) through ($C_{20}H_{42}$), which are liquids and include gasoline and diesel fuel. Beyond that, the alkanes are waxy solids. Notice that the name for each of these compounds ends in *-ane*. When one of the hydrogens is removed from an alkane, the result is called a *radical*, and the suffix becomes *-yl*. Thus, for example, removing one hydrogen from methane (CH_4) produces the methyl radical ($CH_3—$). Table 5 lists some of the methane-series radicals.

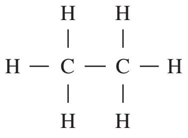
Hydrocarbons appear as building blocks in a great number of chemicals of environmental importance. By substituting other atoms or groups of atoms for some of the hydrogens, new compounds are formed. To get a feel for the naming of these compounds, consider the replacement of some of the hydrogen atoms with chlorine in an ethane molecule. Each carbon atom is given a number, and the corresponding hydrogen replacements are identified by that number. For example, 1,1,1-trichloroethane, better known as TCA, has three chlorines all attached to the same carbon, whereas 1,2-dichloroethane (1,2 DCA) has one chlorine attached to

TABLE 5

Methane-Series Radicals

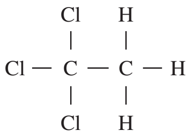
Parent Compound	Radical	Formula
Methane	Methyl	$CH_3—$
Ethane	Ethyl	$C_2H_5—$
Propane	*n*-Propyl	$C_3H_7—$
Propane	Isopropyl	$(CH_3)_2CH—$
n-Butane	*n*-Butyl	$C_4H_9—$

each of the two carbon atoms. Similarly, 1,1,2-trichloro-1,2,2-trifluoroethane has chlorine and fluorine atoms attached to each carbon.



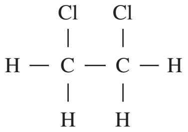
$$\begin{array}{ccccccc} & & \mathrm{H} & & \mathrm{H} & & \\ & & | & & | & & \\ \mathrm{H} & - & \mathrm{C} & - & \mathrm{C} & - & \mathrm{H} \\ & & | & & | & & \\ & & \mathrm{H} & & \mathrm{H} & & \end{array}$$

Ethane



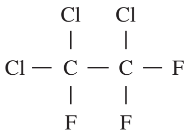
$$\begin{array}{ccccccc} & & \mathrm{Cl} & & \mathrm{H} & & \\ & & | & & | & & \\ \mathrm{Cl} & - & \mathrm{C} & - & \mathrm{C} & - & \mathrm{H} \\ & & | & & | & & \\ & & \mathrm{Cl} & & \mathrm{H} & & \end{array}$$

1,1,1-Trichloroethane (TCA)



$$\begin{array}{ccccccc} & & \mathrm{Cl} & & \mathrm{Cl} & & \\ & & | & & | & & \\ \mathrm{H} & - & \mathrm{C} & - & \mathrm{C} & - & \mathrm{H} \\ & & | & & | & & \\ & & \mathrm{H} & & \mathrm{H} & & \end{array}$$

1,2-Dichloroethane
(1,2-DCA)



$$\begin{array}{ccccccc} & & \mathrm{Cl} & & \mathrm{Cl} & & \\ & & | & & | & & \\ \mathrm{Cl} & - & \mathrm{C} & - & \mathrm{C} & - & \mathrm{F} \\ & & | & & | & & \\ & & \mathrm{F} & & \mathrm{F} & & \end{array}$$

1,1,2-Trichloro-1,2,2-trifluoroethane
(Freon 113)

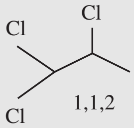
Other names for TCA include methyl chloroform and methyltrichloromethane, which remind us that it contains the methyl radical (CH_3—). TCA is commonly used in the electronics industry for flux cleaning and degreasing operations.

EXAMPLE 13 **Isomers of Trichloropropane**

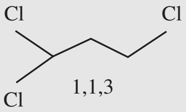
How many different isomers are there of trichloropropane, and how would they be numbered? Draw their kinky diagrams.

Solution Propane has three carbons. The three chlorine atoms can be (1) all attached to an end carbon atom, (2) two attached to an end and one in the middle, (3) two attached to an end and one on the other end, (4) one on the end and two in the middle, or (5) one on each carbon atom. They would be numbered 1,1,1; 1,1,2; 1,1,3; 1,2,2; and 1,2,3, respectively.

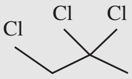
1,1,1



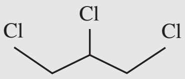
1,1,2



1,1,3

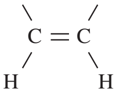


1,2,2

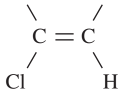


1,2,3

All of the examples so far have been of *saturated* hydrocarbons. *Unsaturated* compounds contain at least two carbon atoms joined by a double bond in which four electrons are shared. When there are double bonds, the chemical name ends in *-ene,* and this category of hydrocarbons is called *alkenes*. The ethylene series of alkenes is analogous to the methane series of alkanes. It begins with ethene (ethylene), propene (propylene), butene (butylene), and so forth.



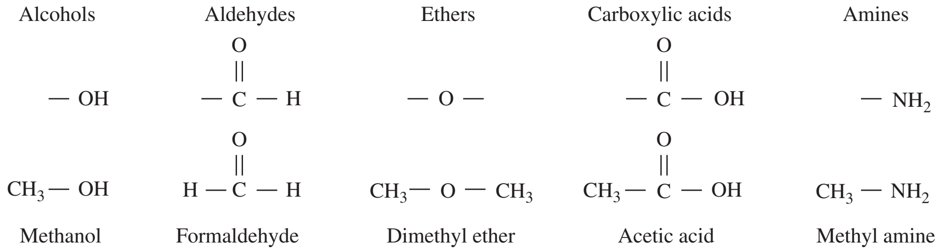
Ethene (ethylene)



Trichloroethylene (TCE)

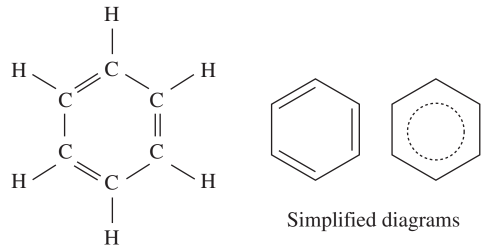
A number of alkanes and alkenes of importance in environmental engineering have some of their hydrogen atoms replaced with chlorine atoms. For example, trichloroethylene, which is more commonly known as TCE, has been used as an industrial solvent for organic substances that are not soluble in water. Principle applications have included cleaning grease from metals, removing solder flux from printed circuit boards in the electronics industry, dry-cleaning clothes, extracting caffeine from coffee, and stripping paint. It was widely used until it was shown to cause cancer in animals. It has the dubious distinction of being the most commonly found contaminant in groundwater at Superfund hazardous waste sites.

Alkane and alkene hydrocarbons often have attachments of various functional groups of atoms. The attachment of the hydroxyl group $-OH$ produces *alcohols;* the $H-C=O$ group produces *aldehydes;* an oxygen $-O-$ between carbons yields *ethers;* connecting an $-OH$ group to a $C=O$ group yields *carboxylic acids;* and attachments of the $-NH_2$ group result in *amines*. Examples of these hydrocarbon derivatives are shown here:



Alcohols	Aldehydes	Ethers	Carboxylic acids	Amines
$-OH$	$-\overset{\overset{O}{\|}}{C}-H$	$-O-$	$-\overset{\overset{O}{\|}}{C}-OH$	$-NH_2$
CH_3-OH	$H-\overset{\overset{O}{\|}}{C}-H$	CH_3-O-CH_3	$CH_3-\overset{\overset{O}{\|}}{C}-OH$	CH_3-NH_2
Methanol	Formaldehyde	Dimethyl ether	Acetic acid	Methyl amine

Hydrocarbons and their derivatives may have linear structures (*e.g.*, *n*-butane), branch structures (*e.g.*, isobutane), or ring structures. The compound benzene (C_6H_6), shown next, is one such ring structure.



Benzene

In the figure for benzene, each carbon atom has been shown forming a single bond with one neighboring carbon and a double bond with the other. That representation is a gross simplification that should not be interpreted too literally; that is, the diagram is really a kind of average value of an actual benzene molecule. As shown, the benzene ring is usually simplified by leaving off the hydrogen atoms.

Molecules in which the hexagonal benzene ring occur are called *aromatic compounds*. This category includes DDT; polychlorinated biphenyls (PCBs); and the defoliants used in the Vietnam War, 2,4,5-T and 2,4-D. When a benzene ring attaches

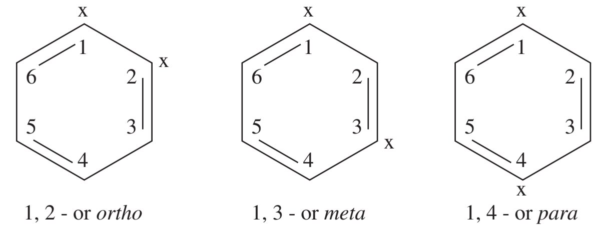


FIGURE 6 Designating the attachment points for phenyl.

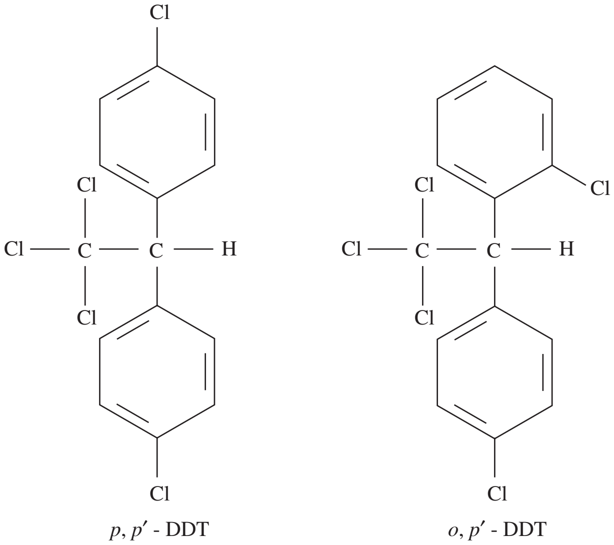


FIGURE 7 Two isomers of dichlorodiphenyltrichloroethane (DDT).

to another molecule, the C_6H_5— that remains of the benzene ring is called a *phenyl* group, so the name *phenyl* frequently appears in the compound name. Another physical characteristic of these compounds that is described by the nomenclature is the location of the attachments to the benzene ring. The terms *ortho, meta,* or *para,* or a numerical designation, specify the relative locations of the attachments as suggested in Figure 6.

Figure 7 shows examples of the *o, m, p* nomenclature for two isomers of dichlorodiphenyltrichloroethane (DDT). Hopefully, this 31-letter name is somewhat less intimidating than it was before you read this section. Reading the name from right to left, the *ethane* on the end tells us it starts as a simple 2-carbon ethane molecule; the *trichloro* refers to the 3 chlorines around one of the carbons; the *diphenyl* tells us there are 2 phenyl groups; and the *dichloro* refers to the chlorines on the phenyls. *p*,*p*′-DDT is an isomer with the chlorine and carbon attachments to each benzene located in the *para* position; the *o*,*p*′-DDT has 1 chlorine in the *ortho* position and 1 in the *para* position. Actually, dichlorodiphenyltrichloroethane as a name is insufficient to unambiguously describe DDT. An even more informative name is 1,1,1-trichloro-2-2-bis(*p*-chlorophenyl)-ethane, where *-bis* means "taken twice."

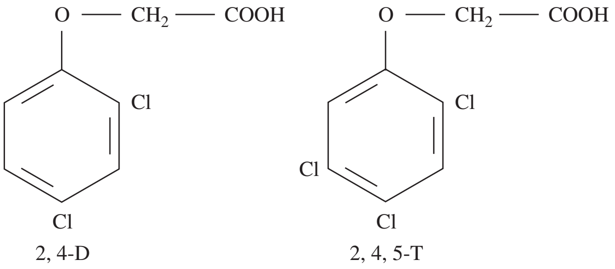


FIGURE 8 2,4-dichlorophenoxyacetic acid (2,4-D) and 2,4,5-trichlorophenoxyacetic acid (2,4,5-T).

In Figure 6, the corners of a phenyl group are numbered, which also provides a way to help describe the structure of a molecule. The herbicides 2,4-dichlorophenoxyacetic acid (2,4-D), and 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) shown in Figure 8 provide examples of this numerical designation. Notice the carboxylic acid functional group $-COOH$ connected to a $-CH_2-$ looks like acetic acid, which is contained in the names of these complex molecules.

Providing unambiguous names for complex organic molecules is extremely difficult and what has been provided here is obviously just a very brief introduction. Even the structural formulas, which convey much more information than a name like DDT, are still highly simplified representations, and there are yet higher level descriptions based on quantum mechanics.

6 Nuclear Chemistry

As a simple but adequate model, we may consider an atom to consist of a dense nucleus containing a number of uncharged neutrons and positively charged protons, surrounded by a cloud of negatively charged electrons. The number of protons is the *atomic number,* which defines a particular element. The sum of the number of protons and neutrons is the *mass number*. Elements with the same atomic number but differing mass numbers are called *isotopes*. The usual way to identify a given isotope is by giving its chemical symbol with the mass number written at the upper left and the atomic number at the lower left. For example, the two most important isotopes of uranium (which has 92 protons) are

$$\underset{\text{Uranium-235}}{{}^{235}_{92}\mathrm{U}} \qquad \text{and} \qquad \underset{\text{Uranium-238}}{{}^{238}_{92}\mathrm{U}}$$

When referring to a particular element, it is common to drop the atomic number subscript since it adds little information to the chemical symbol. Thus, U-238 is a common way to describe that isotope. When describing changes to an element as it decays, the full notation is usually helpful.

Some atomic nuclei are unstable, that is, *radioactive,* and during the spontaneous changes that take place within the nucleus, various forms of *radiation* are emitted. Although all elements having more than 83 protons are naturally radioactive, it is possible to artificially produce unstable isotopes, or *radionuclides,* of virtually every element in the periodic table. Radioactivity is of interest in the study of environmental engineering both because it poses a hazard to exposed organisms, and because it allows radionuclides to act as tracers to help measure the flow of materials in the environment.

The three kinds of radiation are *alpha, beta,* and *gamma* radiation. *Alpha* radiation consists of a heavy 2-proton, 2-neutron particle emitted from the nucleus. When an unstable nucleus emits an alpha particle, its atomic number decreases by 2 units, and its mass number decreases by 4 units. The following example shows the decay of plutonium into uranium:

$$^{239}_{94}\mathrm{Pu} \rightarrow {}^{235}_{92}\mathrm{U} + {}^{4}_{2}\alpha + \gamma$$

In this reaction, not only is an alpha (α) particle ejected from the nucleus, but electromagnetic radiation is emitted as well, indicated by the γ.

As an alpha particle passes through an object, its energy is gradually dissipated as it interacts with other atoms. Its positive charge attracts electrons in its path, raising their energy levels and possibly removing them completely from their nuclei (*ionization*). Alpha particles are relatively massive and easy to stop. Our skin is sufficient protection for sources, which are external to the body, but taken internally, such as by inhalation, α particles can be extremely dangerous.

Beta (β) particles are electrons that are emitted from an unstable nucleus as a result of the spontaneous transformation of a neutron into a proton plus an electron. Emission of a β particle results in an increase in the atomic number by one, while the mass number remains unchanged. A gamma ray may or may not accompany the transformation. The following example shows the decay of strontium-90 into yttrium:

$$^{90}_{38}\mathrm{Sr} \rightarrow {}^{90}_{39}\mathrm{Y} + \beta$$

As negatively charged β particles pass through materials, they encounter swarms of electrons around individual atoms. The resulting Coulomb repulsive force between orbital electrons and β particles can raise the energy level of those electrons, possibly kicking them free of their corresponding atoms. Although such ionizations occur less frequently than with α particles, β penetration is much deeper. Alpha particles travel less than $100\,\mu$m into tissue, whereas β radiation may travel several centimeters. They can be stopped with a modest amount of shielding, however. For example, to stop a β particle, a 1-cm thickness of aluminum is sufficient.

Gamma (γ) rays have no charge or mass, being simply a form of electromagnetic radiation that travels at the speed of light. As such, they can be thought of either as waves with characteristic wavelengths, or as photon particles. Gamma rays have very short wavelengths in the range of 10^{-3} to $10^{-7}\,\mu$m. Having such short wavelengths means individual photons are highly energetic, and they easily cause biologically damaging ionizations. These rays are very difficult to contain and may require several centimeters of lead to provide adequate shielding.

All of these forms of radiation are dangerous to living things. The electron excitations and ionizations that are caused by such radiation cause molecules to become unstable, resulting in the breakage of chemical bonds and other molecular damage. The chain of chemical reactions that follows will result in the formation of new molecules that did not exist before the irradiation. Then, on a much slower time scale, the organism responds to the damage; so slow, in fact, that it may be

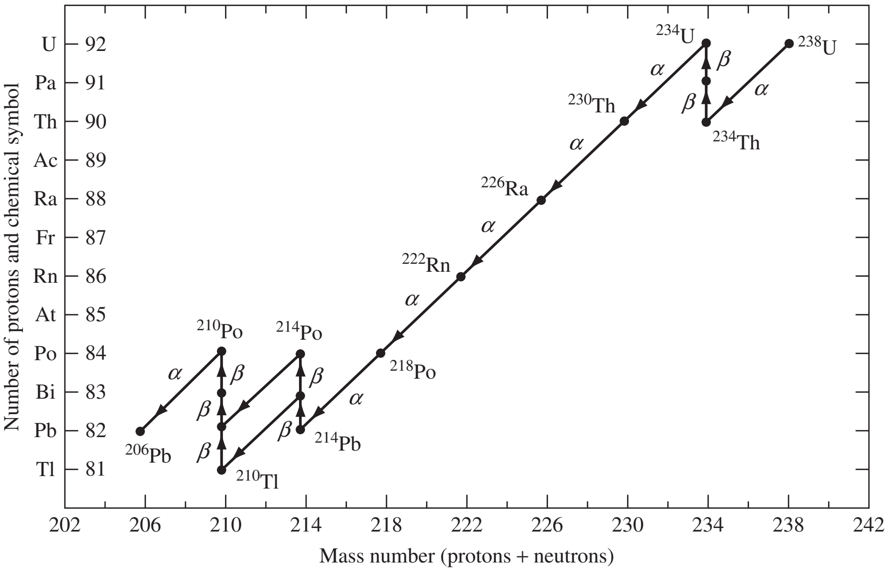


FIGURE 9 The uranium-238 decay chain.

years before the final effects become visible. Low-level exposures can cause *somatic* and/or *genetic* damage. Somatic effects may be observed in the organism that has been exposed and include higher risk of cancer, leukemia, sterility, cataracts, and a reduction in lifespan. Genetic damage, by increasing the mutation rate in chromosomes and genes, affects future generations.

One naturally occurring source of radiation exposure that is thought to be an important cause of lung cancer results from the inhalation of radon (Rn-222) and radon decay products (various isotopes of polonium). Radon is an α emitting, chemically-inert gas that seeps out of the soil, which can sometimes accumulate in houses. It is an intermediate product in a naturally occurring decay chain that starts with uranium-238 and ends with a stable isotope of lead. Figure 9 diagrams this radioactive series. Three other similar series can be drawn. One begins with uranium-235 and ends with lead-207, another begins with thorium-232 and ends with lead-208, and the third begins with the artificial isotope, plutonium-241, and ends with bismuth-209.

An important parameter that characterizes a given radioactive isotope is its half-life, which is the time required for half of the atoms to spontaneously transform, or decay, into other elements. For example, if we start with 100 g of an isotope that has a half-life of 1 year, we would find 50 g remaining after 1 year, 25 g after 2 years, 12.5 g after 3 years, and so on. Although the half-life for a given isotope is constant, half-lives of radionuclides in general vary from fractions of a second to billions of years. Half-lives for the radon portion of the U-238 decay chain are given in Table 6.

There are a number of commonly used radiation units, which unfortunately can be easily confused. The *curie* (Ci) is the basic unit of decay rate; 1 Ci corresponds to the disintegration of 3.7×10^{10} atoms per second, which is approximately the decay rate of 1 g of radium. A more intuitive unit is the *becquerel* (Bq), which

TABLE 6

Half-Lives for the Radon Decay Chain

Isotope	Emission	Half-Life
Rn-222	α	3.8 day
Po-218	α	3.05 min
Pb-214	β	26.8 min
Bi-214	β	19.7 min
Po-214	α	160 μsec
Pb-210	β	19.4 yr
Ti-210	β	1.32 min
Bi-210	β	4.85 day
Po-210	α	1.38 day
Pb-206	—	stable

corresponds to 1 radioactive decay per second; hence, $1\,\text{Ci} = 3.7 \times 10^{10}\,\text{Bq}$. Although the curie is a measure of the rate at which radiation is emitted by a source, it tells us nothing about the radiation dose, which is actually absorbed by an object. The *roentgen* (R) is defined in terms of the number of ionizations produced in a given amount of air by x or γ rays. (X rays are electromagnetic radiation similar to gamma rays, but are generated by electron transitions rather than nuclear transformations.) Of more interest is the amount of energy actually absorbed by tissue, be it bone, fat, muscle, or whatever. The *rad* (*r*adiation *a*bsorbed *d*ose) corresponds to an absorption of 100 ergs of energy per gram of any substance. The rad has the further advantage that it may be used for any form of radiation, α, β, γ, or x. For water and soft tissue, the rad and roentgen are approximately equal.

Another unit, the *rem* (*r*oentgen *e*quivalent *m*an) has been introduced to take into account the different biological effects that various forms of radiation have on humans. Thus, for example, if a 10-rad dose of β particles produces the same biological effect as a 1-rad dose of α particles, both doses would have the same value when expressed in rems. This unit is rather loosely defined, making it difficult to convert from rads to rems. However, in many situations involving x-rays, γ-rays, and β radiation, rads and rems are approximately the same. The rem or millirem are units most often used when describing human radiation exposure.

Nuclear Fission

Nuclear reactors obtain their energy from the heat that is produced when uranium atoms *fission*. In a fission reaction, uranium-235 captures a neutron and becomes unstable uranium-236, which splits apart and discharges two *fission fragments,* two or three neutrons, and γ-rays (Figure 10). Most of the energy released is in the form of kinetic energy in the fission fragments. That energy is used to heat water, producing steam that powers a turbine and generator.

The fission fragments produced are always radioactive, and concerns for their proper disposal have created much of the controversy surrounding nuclear reactors. Typical fission fragments include cesium-137, which concentrates in muscles and has a half-life of 30 years; strontium-90, which concentrates in bone and has a half-life of 28 years; and iodine-131, which concentrates in the thyroid gland and has a

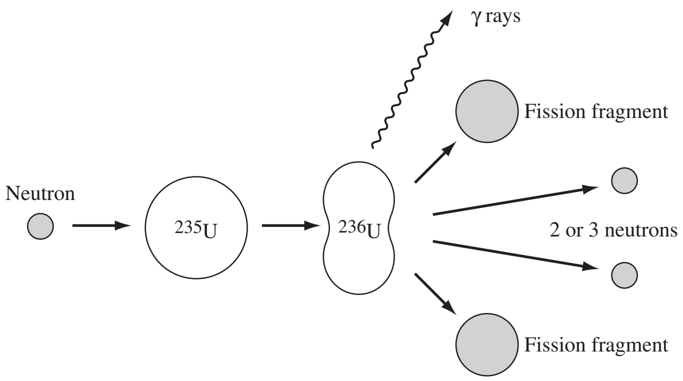


FIGURE 10 The fissioning of uranium-235 creates two radioactive fission fragments, neutrons, and gamma rays.

half-life of 8.1 days. The half lives of fission fragments tend to be no longer than a few tens of years so that after a period of several hundred years, their radioactivity will decline to relatively insignificant levels.

Reactor wastes also include some radionuclides with very long half-lives. Of major concern is plutonium, which has a half-life of 24,390 years. Only a few percent of the uranium atoms in reactor fuel is the fissile isotope ^{235}U, whereas essentially all of the rest is ^{238}U, which does not fission. Uranium-238 can, however, capture a neutron and be transformed into plutonium as the following reactions suggest:

$$^{238}_{92}\text{U} + \text{n} \rightarrow {}^{239}_{92}\text{U} \xrightarrow{\beta} {}^{239}_{93}\text{Np} \xrightarrow{\beta} {}^{239}_{94}\text{Pu} \qquad (47)$$

This plutonium, along with several other long-lived radionuclides, makes nuclear wastes dangerously radioactive for tens of thousands of years and greatly increases the difficulty of providing safe disposal. Removing the plutonium from nuclear wastes before disposal has been proposed as a way to shorten the decay period but that introduces another problem. Plutonium not only is radioactive and highly toxic, it is also the critical ingredient in the manufacture of nuclear weapons. A single nuclear reactor produces enough plutonium each year to make dozens of small atomic bombs, and some have argued that if the plutonium is separated from nuclear wastes, the risk of illicit diversions for such weapons would be unacceptable. Without reprocessing to reuse the plutonium and uranium in the spent fuel rods, the nuclear fuel cycle is as diagrammed in Figure 11.

If reprocessing is included, the nuclear fuel system more closely approaches a true cycle as shown in Figure 12. Reprocessing high-level wastes recovers unused uranium and newly created plutonium that can be used to refuel reactors. The plutonium, however, also can be used for nuclear weapons. In a very controversial decision, Japan has begun to ship its nuclear wastes to France for reprocessing, and in 1995, the first shipments of recovered plutonium were transported back to Japan for reuse in their nuclear program.

As Figures 11 and 12 suggest, the power reactor itself is only one of several sources in the complete nuclear fuel cycle that creates waste disposal problems. Uranium ores must first be mined and the uranium extracted. Mine tailings typically

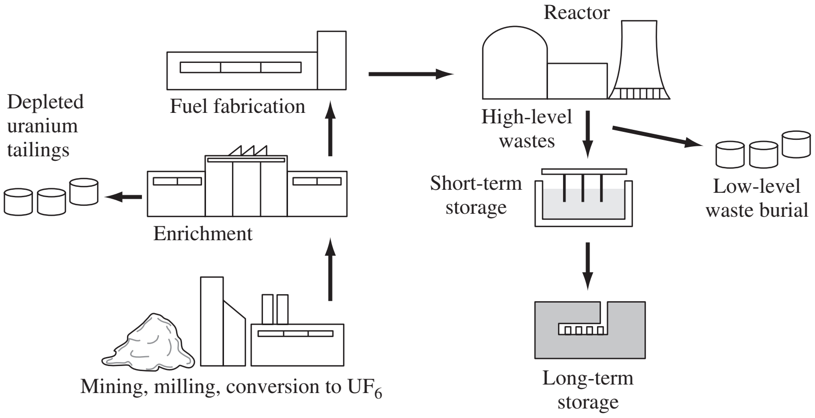


FIGURE 11 A once-through fuel system for nuclear reactors, typical for the United States.

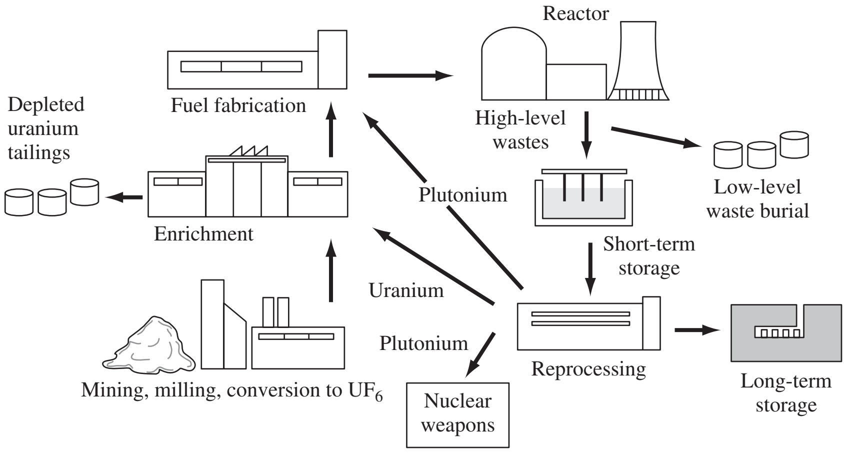


FIGURE 12 Nuclear fuel cycle with reprocessing.

contain toxic metals such as arsenic, cadmium, and mercury, as well as the radionuclides associated with the decay of ^{238}U shown earlier in Figure 9. Only about 0.72 percent of naturally occurring uranium is the desired isotope ^{235}U, and an enrichment facility is needed to increase that concentration to 2 or 3 percent for use in reactors. The enriched uranium is then fabricated into fuel pellets and fuel rods that are shipped to reactors. Highly radioactive wastes from reactors are temporarily stored on site until, eventually, they will be transported to their ultimate disposal in a federal repository. Such a repository, however, is not expected to become operational in the United States until well into the twenty-first century. Low-level wastes can be disposed of in specially designed landfills. When the reactor has reached the end of its lifetime, after roughly 30 years, it must be decommissioned, and radioactive components must be eventually transported to a secure disposal site.

Providing proper disposal of radioactive wastes at every stage in the nuclear fuel cycle is a challenging engineering task, but many argue that it is well within our capabilities.

PROBLEMS

1 Consider the following reaction representing the combustion of propane:

$$C_3H_8 + O_2 \rightarrow CO_2 + H_2O$$

- (a) Balance the equation.
- (b) How many moles of oxygen are required to burn 1 mol of propane?
- (c) How many grams of oxygen are required to burn 100 g of propane?
- (d) At standard temperature and pressure, what volume of oxygen would be required to burn 100 g of propane? If air is 21 percent oxygen, what volume of air at STP would be required?
- (e) At STP, what volume of CO_2 would be produced when 100 g of propane are burned?

2 Trinitrotoluene (TNT), $C_7H_5N_3O_6$, combines explosively with oxygen to produce CO_2, water, and N_2. Write a balanced chemical equation for the reaction and calculate the grams of oxygen required for each 100 g of TNT.

3 An unknown substance is empirically determined to be 40.00 percent carbon by weight, 6.67 percent hydrogen, and 53.33 percent oxygen. Its molecular weight is roughly 55 g/mol. Determine the molecular formula and its correct molecular weight.

4 What is the molarity of 10 g of glucose ($C_6H_{12}O_6$) dissolved in 1 L of water?

5 Eighty-six-proof whiskey is 43 percent ethyl alcohol, CH_3CH_2OH, by volume. If the density of ethyl alcohol is 0.79 kg/L, what is its molarity in whiskey?

6 The Earth's ozone layer is under attack in part by chlorine released when ultraviolet radiation breaks apart certain fluorocarbons. Consider three fluorocarbons known as CFC-11 (CCl_3F), CFC-12 (CCl_2F_2), and HCFC-22 (CHF_2Cl).

- (a) Calculate the percentage by weight of chlorine in each of these fluorocarbons.
- (b) If each kilogram of CFC-11 could be replaced by 1 kg of HCFC-22, by what percentage would the mass of chlorine emissions be reduced?

7 Suppose total world energy consumption of fossil fuels, equal to 3×10^{17} kJ/yr, were to be obtained entirely by combustion of petroleum with the approximate chemical formula C_2H_3. Combustion of petroleum releases about 43×10^3 kJ/kg.

- (a) Estimate the emmisions of CO_2 per year.
- (b) What is the ratio of grams of C emitted per unit of energy for petroleum vs. methane? (Use the results of Example 3.)

8 The standard enthalpy for selected organic compounds is given in this problem. For each, compute the gross and net heats of combustion.

- (a) Propane (C_3H_8); −103.8 kJ/mol
- (b) *n*-Butane (C_4H_{10}); −124.7 kJ/mol
- (c) Isobutane (C_4H_{10}); −131.6 kJ/mol
- (d) Ethanol (*l*), (C_2H_5OH); −277.6 kJ/mol
- (e) Methanol (*l*), (CH_3OH); −238.6 kJ/mol

9 The gross heat of combustion for ethyl alcohol (C_2H_5OH) is −1,370 kJ/mol. What is its standard enthalpy?

10 The standard enthalpy of propane (C_3H_8) is −103.8 kJ/mol. Find the gross heat released when 1 kg is burned.

11 For the following possible alternative automobile fuels, express their higher heating value (HHV) in Btu/gal:

(a) Methanol (CH_3OH), density 6.7 lb/gal, $H^0 = -238.6\,\text{kJ/mol}$

(b) Ethanol (C_2H_5OH), density 6.6 lb/gal, $H^0 = -277.6\,\text{kJ/mol}$

(c) Propane (C_3H_8), density 4.1 lb/gal, $H^0 = -103.8\,\text{kJ/mol}$

12 Repeat problem 11, but find the net heat of combustion for each fuel (Btu/gal).

13 In Example 6, photolysis of ozone (O_3) was described as producing an atomic oxygen with enthalpy 247.5 kJ/mol. More correctly, photolysis yields an oxygen that is temporarily in an excited state (O*) with an enthalpy of 438 kJ/mol.

$$\begin{array}{lccccc} & O_3 & + \; h\nu & \rightarrow & O_2 & + \quad O^* \\ H^0: & (142.9) & & & (0) & (438)\ \text{kJ/mol} \end{array}$$

What maximum wavelength could cause this photolysis?

14 What maximum wavelength could a photon have to cause the dissociation of gaseous nitrogen dioxide?

$$NO_2 + h\nu \rightarrow NO + O$$

15 Ethylene diamine tetra-acetic acid (EDTA) is a water-soluble compound that readily combines with metals, such as calcium, magnesium, and iron. The molecular formula for EDTA is $C_{10}N_2O_8H_{16}$. One EDTA molecule complexes (associates with) one metal atom. A factory produces an aqueous waste that contains 20 mg/L calcium and collects the waste in 44-gallon drums. What mass of EDTA would need to be added to each drum to completely complex all of the calcium in the barrel?

16 Elemental (metallic) aluminum (Al) is reacted with hydrochloric acid (HCl) to yield aluminum chloride ($AlCl_3$) and hydrogen gas (H_2). How many grams of metallic aluminum would be required to produce 1 gram of pure hydrogen gas?

17 Hydrochloric acid, HCl, completely ionizes when dissolved in water. Calculate the pH of a solution containing 25 mg/L of HCl.

18 What is the pH of a solution containing 3×10^{-4} mg/L of OH^- (25°C)?

19 Find the hydrogen ion concentration and the hydroxide ion concentration in baking soda with pH 8.5.

20 Find the theoretical oxygen demand for the following solutions:

(a) 200 mg/L of acetic acid, CH_3COOH

(b) 30 mg/L of ethanol, C_2H_5OH

(c) 50 mg/L of glucose, $C_6H_{12}O_6$

21 Water is frequently disinfected with chlorine gas, forming hypochlorous acid (HOCl), which partially ionizes to hypoclorite and hydrogen ions as follows:

$$HOCl \rightleftharpoons H^+ + OCl^- \quad \text{with equilibrium constant } K_{a1} = 2.9 \times 10^{-8}$$

The amount of [HOCl], which is the desired disinfectant, depends on the pH. Find the fraction that is hypochlorous acid—that is, $[HOCl]/\{[HOCl] + [OCl^-]\}$—as a function of pH. What would the hypochlorous fraction be for pH = 6, 8, and 10?

22 One-half gram of solid calcium sulfate, $CaSO_4(s)$, is added to 1.0 L of pure water. Immediately, the solid begins to dissolve according to the following reaction:

$$CaSO_4(s) \rightleftharpoons Ca^{2+} + SO_4^{2-}$$

What is the concentration of dissolved sulfate in the water once equilibrium is achieved?

23 A solution is initially supersaturated with Ca^{2+} and CO_3^{2-} ions by rapid addition of calcium carbonate, such that the concentrations of each are 2.10×10^{-3} M. The water's pH is held at 8.5. When equilibrium is finally reached, determine:

(a) The concentration of free Ca^{2+} in the water

(b) The concentration of $CO_2(aq)$ in the water

(c) The mass of $CaCO_3(s)$ per liter of water

24 Hydrogen sulfide (H_2S) is an odorous gas that can be stripped from solution in a process similar to that described in Example 11 for ammonia. The reaction is

$$H_2S \rightleftharpoons H^+ + HS^- \quad \text{with equilibrium constant } K_{a1} = 0.86 \times 10^{-7}$$

Find the fraction of hydrogen sulfide in H_2S form at pH 6 and pH 8.

25 Solid aluminum phosphate, $AlPO_4$, is in equilibrium with its ions in solution:

$$AlPO_4 \rightleftharpoons Al^{3+} + PO_4^{3-} \quad \text{with } K_{sp} = 10^{-22}$$

Find the equilibrium concentration of phosphate ions (in mg/L).

26 Calculate the equilibrium concentration of dissolved oxygen in 15°C water at 1 atm, and again at 2,000 m elevation.

27 Suppose the gas above the soda in a bottle of soft drink is pure CO_2 at a pressure of 2 atm. Calculate $[CO_2]$ at 25°C.

28 Calculate the pH of the soft drink in Problem 27. Start with the following chemical reaction and realize the solution will be somewhat acidic (negligible carbonate):

$$CO_2 + H_2O \rightleftharpoons H^+ + HCO_3^-$$

29 It has been estimated that the concentration of CO_2 in the atmosphere before the Industrial Revolution was about 275 ppm. If the accumulation of CO_2 in the atmosphere continues, then by the middle of this century, it will probably be around 600 ppm. Calculate the pH of rainwater (neglecting the effect of any other gases) at 25°C both before the Industrial Revolution and by the middle of this century.

30 One strategy for dealing with the acidification of lakes is periodically to add powdered limestone ($CaCO_3$) to them. Calculate the pH of a lake that has more than enough powdered limestone in it to saturate the water with its ions Ca^{2+} and CO_3^{2-}. *Suggestions:* Begin with the carbonate system equations (43), (44), and (45), and then add a charge balance equation:

$$[H^+] + 2[Ca^{2+}] = [HCO_3^-] + 2[CO_3^{2-}] + [OH^-] \approx [HCO_3^-] + [OH^-]$$

(The preceding holds for pH values that are less than about 10; when you work out the solution, check to see whether this was a valid assumption.) This calculation is the same as would be made to estimate the pH of the oceans, which are saturated with $CaCO_3$.

31 Draw the Lewis structure and the more common structural formulas for the following:

(a) Ethylene, C_2H_4

(b) 2-Chloropropane, $CH_3CHClCH_3$

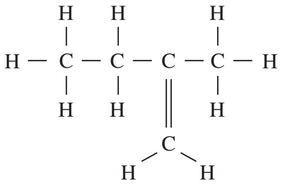
(c) Methanol, CH_3OH

32 Draw kinky diagrams for the following molecules:

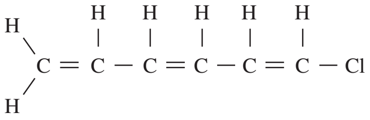
(a) $CH_3CH_2CH_2CH_2CH_3$

(b) $CH_3CH_2CHCH_2$

(c)

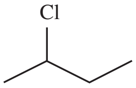


(d)

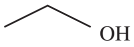


33 Write the chemical structures from the following kinky diagrams:

(*a*)



(*b*)



(*c*)

34 From the names alone, write the structures for the following organics:

(a) Dichloromethane

(b) Trichloromethane (chloroform)

(c) 1,1-Dichloroethylene

(d) Trichlorofluoromethane (CFC-11)

(e) 1,1,2,2-Tetrachlorethane

(f) *o*-Dichlorobenzene

(g) Tetrachloroethene (PCE)

(h) Dichlorofluoromethane (CFC-21)

35 What values of a and b would complete each of the following (X and Y are not meant to be any particular elements):

$$^{266}_{88}\mathrm{X} \rightarrow \alpha + {}^{a}_{b}\mathrm{Y}$$

$$^{a}_{15}\mathrm{X} \rightarrow \beta + {}^{32}_{b}\mathrm{Y}$$

36 The half-life of iodine-125 is about 60 days. If we were to start with 64 g of it, about how much would remain after 1 year?

REFERENCES

Baird, C., 1995, *Environmental Chemistry,* W. H. Freeman, New York.

Masters, G. M., 1974, *Introduction to Environmental Science and Technology,* Wiley, New York.

Sawyer, C. N., P. L. McCarthy, and G. F. Parkin, 1994, *Chemistry for Environmental Engineering,* 5th ed., McGraw-Hill, New York.

Thomann, R. V., and J. A. Mueller, 1987, *Principles of Surface Water Quality and Control,* Harper & Row, New York.

Mathematics of Growth

> Prediction is very difficult, especially if it's about the future.
>
> —*Niels Bohr (1885–1962)*

1 Introduction

A sense of the future is an essential component in the mix of factors that should be influencing the environmental decisions we make today. In some circumstances, only a few factors may need to be considered, and the time horizon may be relatively short. For example, a wastewater treatment plant operator may need to predict the growth rate of bacteria in a digester over a period of hours or days. The designer of the plant, on the other hand, probably needs to estimate the local population growth rate over the next decade or two in order to size the facility. At the other extreme, to make even the crudest estimate of future carbon emissions and their effect on global warming, scientists need to forecast population and economic growth rates, anticipated improvements in energy efficiency, the global fossil fuel resource base and the consumption rate of individual fuels, the rate of deforestation or reforestation that could be anticipated, and so on. These estimates need to be made for periods of time that are often measured in hundreds of years.

As Niels Bohr suggests in the preceding quote, we cannot expect to make accurate predictions of the future, especially when the required time horizon may be

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

extremely long. However, often simple estimates can be made that are robust enough that the insight they provide is most certainly valid. We can say, for example, with considerable certainty that world population growth at today's rates cannot continue for another hundred years. We can also use simple mathematical models to develop very useful "what if" scenarios: *If* population growth continues at a certain rate, and *if* energy demand is proportional to economic activity, and so forth, *then* the following would occur.

The purpose of this chapter is to develop some simple but powerful mathematical tools that can shed considerable light on the future of many environmental problems. We begin with what is probably the most useful and powerful mathematical function encountered in environmental studies: the *exponential function*. Other growth functions encountered often are also explored, including the *logistic* and the *Gaussian* functions. Applications that will be demonstrated using these functions include population growth, resource consumption, pollution accumulation, and radioactive decay.

2 Exponential Growth

Exponential growth occurs in any situation where the increase in some quantity is proportional to the amount currently present. This type of growth is quite common, and the mathematics required to represent it is relatively simple, yet extremely important. Exponential growth is a particular case of a first-order rate process. We will approach this sort of growth first as discrete, year-by-year increases and then in the more usual way as a continuous growth function.

Suppose something grows by a fixed percentage each year. For example, if we imagine our savings at a bank earning 5 percent interest each year, compounded once a year, then the amount of increase in savings over any given year is 5 percent of the amount available at the beginning of that year. If we start now with \$1,000, then at the end of one year we would have \$1,050 (1,000 + 0.05 × 1,000); at the end of two years, we would have \$1,102.50 (1,050 + 0.05 × 1,050); and so on. This can be represented mathematically as follows:

N_0 = initial amount
N_t = amount after t years
r = growth rate (fraction per year)

Then,

$$N_{t+1} = N_t + rN_t = N_t(1 + r)$$

For example,

$$N_1 = N_0(1 + r);\ N_2 = N_1(1 + r) = N_0(1 + r)^2$$

and, in general,

$$N_t = N_0(1 + r)^t \tag{1}$$

EXAMPLE 1 **U.S. Electricity Growth (Annual Compounding)**

In 2005, the United States produced 4.0×10^{12} kWhr/yr of electricity. The average annual growth rate of U.S. electricity demand in the previous 15 years was about 1.8 percent. (For comparison, it had been about 7 percent per year for many decades before the 1973 oil embargo.) Estimate the electricity consumption in 2050 if the 1.8 percent per year growth rate were to remain constant over those 45 years.

Solution From (1), we have

$$\begin{aligned} N_{45} &= N_0(1 + r)^{45} \\ &= 4.0 \times 10^{12} \times (1 + 0.018)^{45} \\ &= 8.93 \times 10^{12} \text{ kWhr/yr} \end{aligned}$$

Continuous Compounding

For most events of interest in the environment, it is usually assumed that the growth curve is a smooth, continuous function without the annual jumps that (1) is based on. In financial calculations, this is referred to as continuous compounding. With the use of a little bit of calculus, the growth curve becomes the true exponential function that we will want to use most often.

One way to state the condition that leads to exponential growth is that the quantity grows in proportion to itself; that is, the rate of change of the quantity N is proportional to N. The proportionality constant r is called the rate of growth and has units of (time^{-1}).

$$\frac{dN}{dt} = rN \tag{2}$$

The solution is

$$N = N_0 e^{rt} \tag{3}$$

which is plotted in Figure 1.

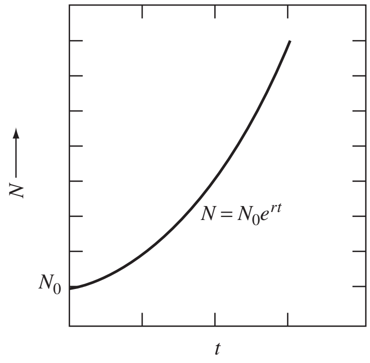


FIGURE 1 The exponential function.

EXAMPLE 2 **U.S. Electricity Growth (Continuous Compounding)**

Suppose we repeat Example 1, but now consider the 1.8 percent growth rate to be continuously compounded. Starting with the 2005 electricity consumption of 4.0×10^{12} kWhr/yr, what would consumption be in 2050 if the growth rate remains constant?

Solution Using (3),

$$\begin{aligned} N &= N_0 e^{rt} \\ &= 4.0 \times 10^{12} \times e^{0.018 \times 45} \\ &= 8.99 \times 10^{12} \text{ kWhr/yr} \end{aligned}$$

Example 1, in which increments were computed once each year, and Example 2, in which growth was continuously compounded, have produced nearly identical results. As either the period of time in question or the growth rate increases, the two approaches begin to diverge. At 12 percent growth, for example, those answers would differ by nearly 50 percent. In general, it is better to express growth rates as if they are continuously compounded so that (3) becomes the appropriate expression.

Doubling Time

Calculations involving exponential growth can sometimes be made without a calculator by taking advantage of the following special characteristic of the exponential function. A quantity that is growing exponentially requires a fixed amount of time to double in size, regardless of the starting point. That is, it takes the same amount of time to grow from N_0 to $2N_0$ as it does to grow from $2N_0$ to $4N_0$, and so on, as shown in Figure 2.

The *doubling time* (t_d) of a quantity that grows at a fixed exponential rate r is easily derived. From (3) the doubling time can be found by setting $N = 2N_0$ at $t = t_d$:

$$2N_0 = N_0 e^{rt_d}$$

Since N_0 appears on both sides of the equation, it can be canceled out, which is another way of saying the length of time required to double the quantity does not

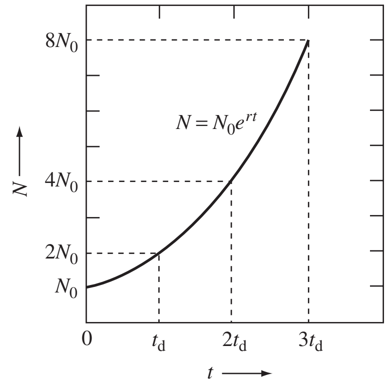


FIGURE 2 Illustrating the concept of doubling time.

depend on how much you start with. So canceling N_0 and taking the natural log of both sides gives

$$\ln 2 = rt_{\mathrm{d}}$$

or

$$t_{\mathrm{d}} = \frac{\ln 2}{r} \cong \frac{0.693}{r} \tag{4}$$

If the growth rate r is expressed as a percentage instead of as a fraction, we get the following important result:

$$t_{\mathrm{d}} \cong \frac{69.3}{r(\%)} \cong \frac{70}{r(\%)} \tag{5}$$

Equation (5) is well worth memorizing; that is,

> *The length of time required to double a quantity growing at r percent is about equal to 70 divided by r percent.*

If your savings earn 7 percent interest, it will take about 10 years to double the amount you have in the account. If the population of a country grows continuously at 2 percent, then it will double in size in 35 years, and so on.

EXAMPLE 3 Historical World Population Growth Rate

It took about 300 years for the world's population to increase from 0.5 billion to 4.0 billion. If we assume exponential growth at a constant rate over that period of time, what would that growth rate be? Do this example first with a calculator and (3), and then with the rule of thumb suggested in (5).

Solution Rearranging (3) and taking the natural log of both sides gives

$$\ln(e^{rt}) = rt = \ln\left(\frac{N}{N_0}\right)$$

$$r = \frac{1}{t}\ln\left(\frac{N}{N_0}\right) \tag{6}$$

$$r = \frac{1}{300}\ln\left(\frac{4.0}{0.5}\right) = 0.00693 = 0.693 \text{ percent}$$

Using the rule of thumb approach given by (5), three doublings would be required to have the population grow from 0.5 billion to 4.0 billion. Three doublings in 300 years means each doubling took 100 years.

$$t_{\mathrm{d}} = 100 \text{ yrs} \cong \frac{70}{r\%}$$

so

$$r(\%) \cong \frac{70}{100} = 0.7 \text{ percent}$$

Our answers would have been exactly the same if the rule of thumb in (5) had not been rounded off.

TABLE 1

Exponential Growth Factors for Various Numbers of Doubling Times

Number of Doublings (n)	Growth Factor (2^n)
1	2
2	4
3	8
4	16
5	32
10	1024
20	$\approx 1.05 \times 10^6$
30	$\approx 1.07 \times 10^9$

Quantities that grow exponentially very quickly also increase in size at a deceptively fast rate. As Table 1 suggests, quantities growing at only a few percent per year increase in size with incredible speed after just a few doubling times. For example, at the 1995 rate of world population growth ($r = 1.5$ percent), the doubling time was about 46 years. If that rate were to continue for just 4 doubling times, or 184 years, world population would increase by a factor of 16, from 5.7 billion to 91 billion. In 20 doubling times, there would be 6 million billion (quadrillion) people, or more than one person for each square foot of surface area of the earth. The absurdity of these figures simply points out the impossibility of the underlying assumption that exponential growth, even at this relatively low-sounding rate, could continue for such periods of time.

Half-Life

When the rate of decrease of a quantity is proportional to the amount present, exponential growth becomes exponential decay. Exponential decay can be described using either a *reaction rate coefficient* (k) or a *half-life* ($t_{1/2}$), and the two are easily related to each other. A reaction rate coefficient for an exponential decay plays the same role that r played for exponential growth.

Exponential decay can be expressed as

$$N = N_0 e^{-kt} \tag{7}$$

where k is a reaction rate coefficient (time^{-1}), N_0 is an initial amount, and N is an amount at time t.

A plot of (7) is shown in Figure 3 along with an indication of half-life. To relate half-life to reaction rate, we set $N = N_0/2$ at $t = t_{1/2}$, giving

$$\frac{N_0}{2} = N_0 e^{-kt_{1/2}}$$

Canceling the N_0 term and taking the natural log of both sides gives us the desired expression

$$t_{1/2} = \frac{\ln 2}{k} \cong \frac{0.693}{k} \tag{8}$$

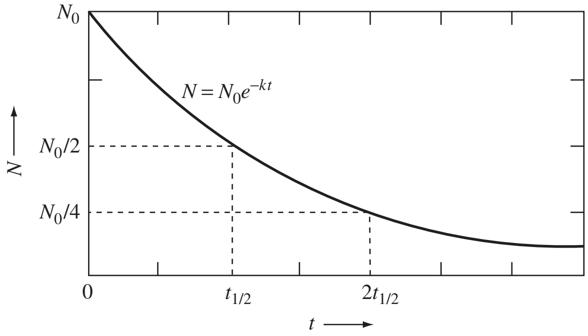


FIGURE 3 Exponential decay and half-life.

EXAMPLE 4 Radon Half-Life

If we start with a 1.0-Ci radon-222 source, what would its activity be after 5 days?

Solution Rearranging (8) to give a reaction rate coefficient results in

$$k = \frac{\ln 2}{t_{1/2}} = \frac{0.693}{3.8\text{ day}} = 0.182/\text{day}$$

Using (7) gives

$$N = N_0e^{-kt} = 1\text{ Ci} \times e^{-0.182/\text{day}\times 5\text{ days}} = 0.40\,\text{Ci}$$

Disaggregated Growth Rates

Often the quantity to be modeled can be considered to be the product of a number of individual factors. For example, to estimate future carbon emissions, we might *disaggregate* demand into the following three factors:

$$\text{Carbon emissions} = (\text{Population}) \times \left(\frac{\text{Energy}}{\text{Person}}\right) \times \left(\frac{\text{Carbon}}{\text{Energy}}\right) \tag{9}$$

By considering demand this way, we would hope to increase the accuracy of our forecast by estimating separately the rates of change of population, per capita energy consumption, and carbon emissions per unit of energy. Equation (9) is an example of the following simple conceptual model of factors that drive the environmental impacts of human activity:

$$\text{Impacts} = (\text{Population}) \times (\text{Affluence}) \times (\text{Technology}) \tag{10}$$

In (9), affluence is indicated by per capita energy demand, and technology is represented by the carbon emissions per unit of energy.

Expressing a quantity of interest as the product of individual factors, such as has been done in (9), leads to a very simple method of estimating growth. We begin with

$$P = P_1P_2\ldots P_n \tag{11}$$

If each factor is itself growing exponentially

$$P_i = p_i e^{r_i t}$$

then

$$\begin{aligned} P &= (p_1 e^{r_1 t})(p_2 e^{r_2 t}) \ldots (p_n e^{r_n t}) \\ &= (p_1 p_2 \ldots p_n) e^{(r_1 + r_2 + \ldots r_n)t} \end{aligned}$$

which can be written as

$$P = P_0 e^{rt}$$

where

$$P_0 = (p_1 p_2 \ldots p_n) \tag{12}$$

and

$$r = r_1 + r_2 + \cdots r_n \tag{13}$$

Equation (13) is a very simple, very useful result. That is,

> *If a quantity can be expressed as a product of factors, each growing exponentially, then the total rate of growth is just the sum of the individual growth rates.*

EXAMPLE 5 **Future U.S. Energy Demand**

One way to disaggregate energy consumption is with the following product:

$$\text{Energy demand} = (\text{Population}) \times \left(\frac{\text{GDP}}{\text{person}}\right) \times \left(\frac{\text{Energy}}{\text{GDP}}\right)$$

GDP is the gross domestic product. Suppose we project per capita GDP to grow at 2.3 percent and population at 0.6 percent. Assume that through energy-efficiency efforts, we expect energy required per dollar of GDP to *decrease* exponentially at the rate that it did between 1995 and 2005. In 1995, (energy/GDP) was 13.45 kBtu/$, and in 2005, it was 11.35 kBtu/$ (in constant 1987 dollars). Total U.S. energy demand in 2005 was 100.2 quads (quadrillion Btu). If the aforementioned rates of change continue, what would energy demand be in the year 2020?

Solution First, find the exponential rate of decrease of (energy/GDP) over the 10 years between 1995 and 2005. Using (6) gives

$$r_{\text{energy/GDP}} = \frac{1}{t} \ln\left(\frac{N}{N_0}\right) = \frac{1}{10} \ln\left(\frac{11.35}{13.45}\right) = -0.017 = -1.7 \text{ percent}$$

Notice that the equation automatically produces the negative sign. The overall energy growth rate is projected to be the sum of the three rates:

$$r = 0.6 + 2.3 - 1.7 = 1.2 \text{ percent}$$

(For comparison, energy growth rates in the United States before the 1973 oil embargo were typically 3 to 4 percent per year.) Projecting out 15 years to the year 2020, (3) gives

$$\begin{aligned} \text{Energy demand in 2020} &= 100.2 \times 10^{15} \text{ Btu/yr} \times e^{0.012 \times 15} \\ &= 120 \times 10^{15} \text{ Btu/yr} = 120 \text{ quads/yr} \end{aligned}$$

3 Resource Consumption

To maintain human life on Earth, we depend on a steady flow of energy and materials to meet our needs. Some of the energy used is *renewable* (e.g., hydroelectric power, windpower, solar heating systems, and even firewood if the trees are replanted), but most comes from *depletable* fossil fuels. The minerals extracted from the Earth's crust, such as copper, iron, and aluminum, are limited as well. The future of our way of life to a large extent depends on the availability of abundant supplies of inexpensive, depletable energy and materials.

How long will those resources last? The answer, of course, depends on how much there is and how quickly we use it. We will explore two different ways to model the rate of consumption of a resource: One is based on the simple exponential function that we have been working with, and the other is based on a more realistic bell-shaped curve.

Exponential Resource Production Rates

When a mineral is extracted from the Earth, geologists traditionally say the resource is being *produced* (rather than *consumed*). Plotting the rate of production of a resource versus time, as has been illustrated in Figure 4, the area under the curve between any two times will represent the total resource that has been produced during that time interval. That is, if P is the production rate (e.g., barrels of oil per day, tonnes of aluminum per year, etc.), and Q is the resource produced (total barrels, total tonnes) between times t_1 and t_2, we can write

$$Q = \int_{t_1}^{t_2} P\,dt \tag{14}$$

Assuming that the production rate of a resource grows exponentially, we can easily determine the total amount produced during any time interval; or, conversely, if the total amount to be produced is known, we can estimate the length of time that it will take to produce it. The basic assumption of exponential growth is probably not a good one if the time frame is very long, but nonetheless, it will give us some very useful insights. In the next section, we will work with a more reasonable production rate curve.

Figure 5 shows an exponential rate of growth in production. If the time interval of interest begins with $t = 0$, we can write

$$Q = \int_0^t P_0 e^{rt}\,dt = \frac{P_0}{r} e^{rt}\Big|_0^t$$

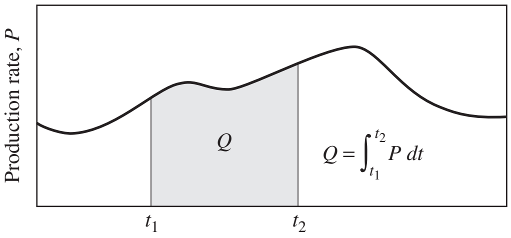


FIGURE 4 The total amount of a resource produced between times t_1 and t_2 is the shaded area under the curve.

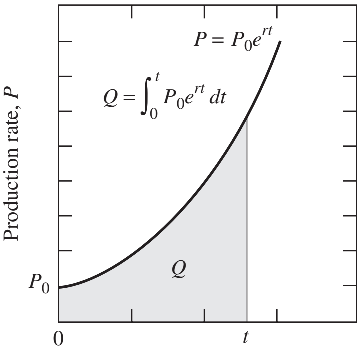


FIGURE 5 Total consumption of a resource experiencing exponential growth.

which has as a solution

$$Q = \frac{P_0}{r}(e^{rt} - 1) \tag{15}$$

where

Q = the total resource produced from time 0 to time t
P_0 = the initial production rate
r = the exponential rate of growth in production

Equation (15) tells us how much of the resource is produced in a given period of time if the production rate grows exponentially. If we want to know how long it will take to produce a given amount of the resource, rearrange (15) to give

$$t = \frac{1}{r}\ln\left(\frac{rQ}{P_0} + 1\right) \tag{16}$$

where t = the length of time required to produce an amount Q.

Applications of (16) often lead to startling results, as should be expected anytime we assume that exponential rates of growth will continue for a prolonged period of time.

EXAMPLE 6 World Coal Production

World coal production in 2005 was estimated to be 6.1 billion (short) tons per year, and the estimated total recoverable reserves of coal were estimated at 1.1 trillion tons. Growth in world coal production in the previous decade averaged 1.9 percent per year. How long would it take to use up those reserves at current production rates, and how long would it take if production continues to grow at 1.9 percent?

Solution At those rates, coal reserves would last

$$\frac{\text{Reserves}}{\text{Production}} = \frac{1.1 \times 10^{12}\ \text{tons}}{6.1 \times 10^{9}\ \text{tons/yr}} = 180\ \text{years}$$

If production grows exponentially, we need to use (16):

$$t = \frac{1}{r}\ln\left(\frac{rQ}{P_0} + 1\right)$$

$$= \frac{1}{0.019}\ln\left(\frac{0.019 \times 1.1 \times 10^{12}}{6.1 \times 10^{9}} + 1\right) = 78 \text{ years}$$

Even though the growth rate of 1.9 percent might seem modest, compared to a constant production rate, it cuts the length of time to deplete those reserves by more than half.

Example 6 makes an important point: By simply dividing the remaining amount of a resource by the current rate of production, we can get a misleading estimate of the remaining lifetime for that resource. If exponential growth is assumed, what would seem to be an abundant resource may actually be consumed surprisingly quickly. Continuing this example for world coal, let us perform a sensitivity analysis on the assumptions used. Table 2 presents the remaining lifetime of coal reserves, assuming an initial 2005 production rate of 6.1 billion tons per year, for various estimates of total available supply and for differing rates of production growth.

Notice how the lifetime becomes less and less sensitive to estimates of the total available resource as exponential growth rates increase. At a constant 3 percent growth rate, for example, a total supply of 500 billion tons of coal would last 41 years; if our supply is four times as large, 2,000 billion tons, the resource only lasts another 38 years. Having four times as much coal does not even double the projected lifetime.

This is a good time to mention the important distinction between the *reserves* of a mineral and the ultimately producible *resources*. As shown in Figure 6, *reserves* are quantities that can reasonably be assumed to exist and are producible with existing technology under present economic conditions. *Resources* include present reserves as well as deposits not yet discovered or deposits that have been identified but are not recoverable under present technological and economic conditions.

As existing reserves are consumed, further exploration, advances in extraction technology, and higher acceptable prices constantly shift mineral resources into the reserves category. Estimating the lifetime of a mineral based on the available reserves

TABLE 2

Years Required to Consume All of the World's Recoverable Coal Reserves, Assuming Various Reserve Estimates and Differing Production Growth Rates[a]

Growth rate (%)	500 Billion tons (yrs)	1,000 Billion tons (yrs)	2,000 Billion tons (yrs)
0	82	164	328
1	60	97	145
2	49	73	101
3	41	59	79
4	36	51	66
5	33	44	57

[a] Initial production rate 6.1 billion tons/year (2005); actual reserves estimated at 1,100 billion tons.

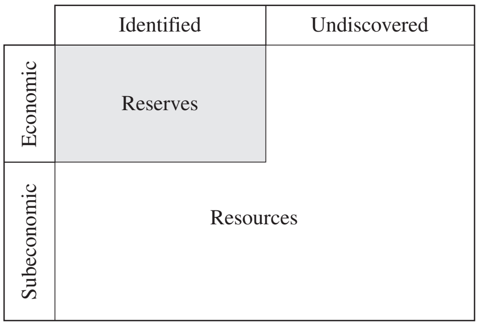


FIGURE 6 Classification of minerals. Reserves are a subcategory of resources.

rather than on the ultimately producible resources can therefore be misleading. World oil reserves in 1970, for example, were estimated at 550 billion barrels (1 barrel equals 42 gallons) while production was 17 billion barrels per year. The reserves-to-production ratio was about 32 years; that is, at 1970 production rates, those reserves would be depleted in 32 years. However, 35 years later (2005), instead of being out of reserves, they had grown to 1,300 billion barrels. Production in 2005 was 30.7 billion barrels per year, giving a reserves-to-production ratio of 42 years.

A Symmetrical Production Curve

We could imagine a complete production cycle to occur in many ways. The model of exponential growth until the resource is totally consumed that we just explored is just one example. It might seem an unlikely one since the day before the resource collapses, the industry is at full production, and the day after, it is totally finished. It is a useful model, however, to dispel any myths about the possibility of long-term exponential growth in the consumption rate of any finite resource.

A more reasonable approach to estimating resource cycles was suggested by M. King Hubbert (1969). Hubbert argued that consumption would more likely follow a course that might begin with exponential growth while the resource is abundant and relatively cheap. As new sources get harder to find, prices go up and substitutions would begin to take some of the market. Eventually, consumption rates would peak and then begin a downward trend as the combination of high prices and resource substitutions would prevail. The decline could be precipitous when the energy needed to extract and process the resource exceeds the energy derived from the resource itself. A graph of resource consumption rate versus time would therefore start at zero, rise, peak, and then decrease back to zero, with the area under the curve equaling the total resource consumed.

Hubbert suggested that a symmetrical production rate cycle that resembles a bell-shaped curve be used. One such curve is very common in probability theory, where it is called the *normal* or *Gaussian* function. Figure 7 shows a graph of the function and identifies some of the key parameters used to define it.

The equation for the complete production rate cycle of a resource corresponding to Figure 7 is

$$P = P_{\mathrm{m}} \exp\left[-\frac{1}{2}\left(\frac{t - t_{\mathrm{m}}}{\sigma}\right)^2\right] \tag{17}$$

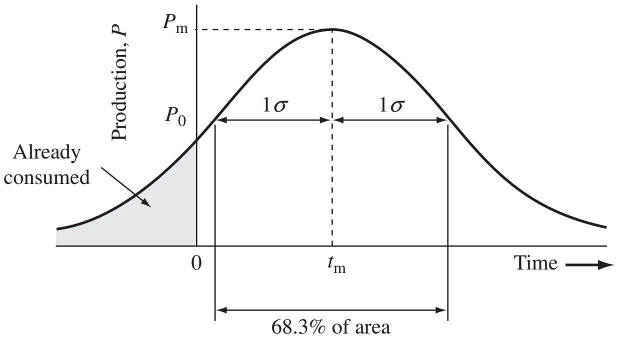


FIGURE 7 Resource production following a Gaussian distribution function.

where

P = the production rate of the resource
P_m = the maximum production rate
t_m = time at which the maximum production rate occurs
σ = standard deviation, a measure of the width of the bell-shaped curve
exp[ ] = the exponential function

The parameter σ is the standard deviation of a normal density function, and in this application, it has units of time. Within ±1 σ away from the time of maximum production, 68.3 percent of the production occurs; within ±2 σ, 95 percent of the production occurs. Notice that with this bell-shaped curve, it is not possible to talk about the resource ever being totally exhausted. It is more appropriate to specify the length of time required for some major fraction of it to be used up. Hubbert uses 80 percent as his criterion, which corresponds to 1.3 σ after the year of maximum production.

We can find the total amount of the resource ever produced, Q_∞, by integrating (17)

$$Q_\infty = \int_{-\infty}^{\infty} P\,dt = \int_{-\infty}^{\infty} P_m \exp\left[-\frac{1}{2}\left(\frac{t - t_m}{\sigma}\right)^2\right]dt$$

which works out to

$$Q_\infty = \sigma P_m \sqrt{2\pi} \tag{18}$$

Equation (18) can be used to find σ if the ultimate amount of the resource ever to be produced Q_∞ and the maximum production rate P_m can be estimated.

It is also interesting to find the length of time required to reach the maximum production rate. If we set $t = 0$ in (17), we find the following expression for the initial production rate, P_0

$$P_0 = P_m \exp\left[-\frac{1}{2}\left(\frac{t_m}{\sigma}\right)^2\right] \tag{19}$$

which leads to the following expression for the time required to reach maximum production:

$$t_m = \sigma\sqrt{2\ln\frac{P_m}{P_0}} \tag{20}$$

Let us demonstrate the use of these equations by fitting a Gaussian curve to data for U.S. coal resources.

EXAMPLE 7 **U.S. Coal Production**

Suppose ultimate total production of U.S. coal is double the 1995 recoverable reserves, which were estimated at 268×10^9 (short) tons. The U.S. coal production rate in 1995 was 1.0×10^9 tons/year. How long would it take to reach a peak production rate equal to four times the 1995 rate if a Gaussian production curve is followed?

Solution Equation (18) gives us the relationship we need to find an appropriate σ.

$$\sigma = \frac{Q_\infty}{P_m\sqrt{2\pi}} = \frac{2 \times 268 \times 10^9 \text{ ton}}{4 \times 1.0 \times 10^9 \text{ ton/yr}\sqrt{2\pi}} = 53.5 \text{ yr}$$

A standard deviation of 53.5 years says that in a period of 107 years ($2\,\sigma$), about 68 percent of the coal would be consumed.

To find the time required to reach peak production, use (20):

$$t_m = \sigma\sqrt{2\ln\frac{P_m}{P_0}} = 53.5\sqrt{2\ln 4} = 89 \text{ yr}$$

The complete production curve is given by (17)

$$P = P_m \exp\left[-\frac{1}{2}\left(\frac{t - t_m}{\sigma}\right)^2\right] = 4 \times 1.0 \times 10^9 \exp\left[-\frac{1}{2}\left(\frac{t - 89}{53.5}\right)^2\right] \text{ tons/yr}$$

which is plotted in Figure 8 along with the production curves that result from the maximum production rate reaching two, four, and eight times the current production rates. The tradeoffs between achieving high production rates versus making the resource last beyond the next century are readily apparent from this figure.

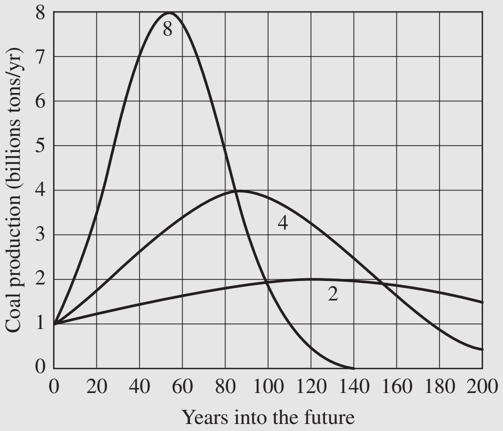


FIGURE 8 Gaussian curves fitted to U.S. production of coal. Ultimate production is set equal to twice the 1995 reserves. The plotted parameter is the ratio of peak production rate to initial (1995) production rate.

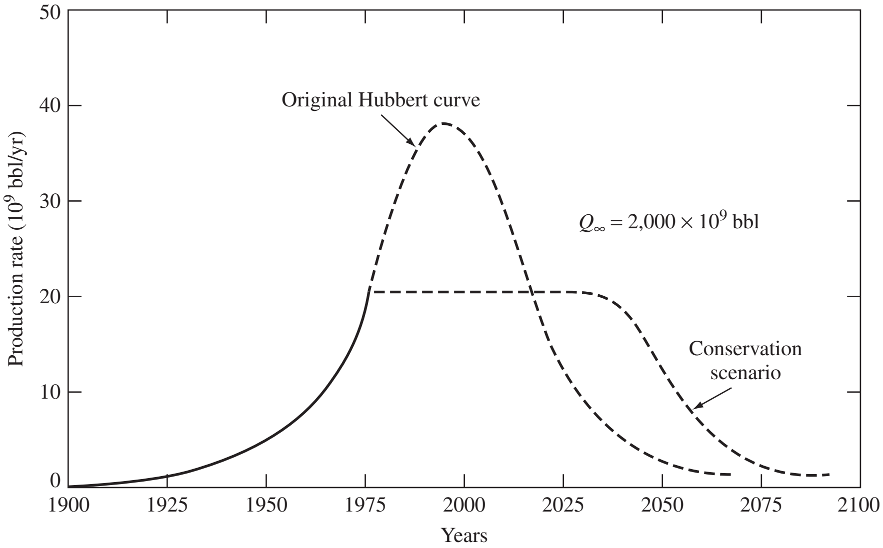


FIGURE 9 Two possible production rate curves for the world's ultimately recoverable crude oil resources, estimated by Hubbert in 1977 to be 2 trillion barrels. The original Hubbert curve indicates that peak production would have been reached before the year 2000, whereas the "conservation" scenario, which holds production to the 1975 level as long as possible, extends the resource for only a few decades more.
(*Source:* Hubbert, 1977.)

Hubbert's analysis of world oil production is presented in Figure 9 for two scenarios. One scenario is based on fitting a Gaussian curve to the historical production record before 1975; the other is based on a conservation scenario in which oil production follows the Gaussian curve until 1975 but then remains constant at the 1975 rate for as long as possible before a sudden decline takes place. With the bell-shaped curve, world oil consumption would have peaked just before the year 2000; with the conservation scenario, supplies would last until about 2040 before the decline would begin. The actual production curve between 1975 and 1995 fell between these two scenarios.

The curves drawn in Figure 9 were based on Hubbert's 1977 estimate of 2,000 billion barrels of oil ultimately produced. For perspective, by 2005, approximately 1,000 billion barrels had been produced (and consumed), and proved reserves stood at 1,300 billion barrels (two-thirds of which were in the Middle East). Amounts already produced plus proved reserves total 2,300 billion barrels, more than Hubbert's estimate for the total amount ever to be produced.

Will more oil ultimately be produced than the current 1.3 trillion barrels of proved reserves? Finding new sources, using enhanced oil recovery techniques to remove some of the oil that is now left behind, and eventually tapping into unconventional hydrocarbon resources such as extra heavy oil, bitumen, and shale oil could stretch petroleum supplies much farther into the future but at significantly higher environmental and economic costs.

4 Population Growth

The simple exponential growth model can serve us well for short time horizons, but obviously, even small growth rates cannot continue for long before environmental constraints limit further increases.

The typical growth curve for bacteria, shown in Figure 10, illustrates some of the complexities that natural biological systems often exhibit. The growth curve is divided into phases designated as lag, exponential, stationary, and death. The *lag phase,* characterized by little or no growth, corresponds to an initial period of time when bacteria are first inoculated into a fresh medium. After the bacteria have adjusted to their new environment, a period of rapid growth, the *exponential phase,* follows. During this time, conditions are optimal, and the population doubles with great regularity. (Notice that the vertical scale in Figure 10 is logarithmic so that exponential growth produces a straight line.) As the bacterial food supply begins to be depleted, or as toxic metabolic products accumulate, the population enters the no-growth, or *stationary phase*. Finally, as the environment becomes more and more hostile, the *death phase* is reached and the population declines.

Logistic Growth

Population projections are often mathematically modeled with a *logistic* or S-shaped (*sigmoidal*) growth curve such as the one shown in Figure 11. Such a curve has great intuitive appeal. It suggests an early exponential growth phase while conditions for growth are optimal, followed by slower and slower growth as the population nears the carrying capacity of its environment. Biologists have successfully used logistic curves to model populations of many organisms, including protozoa, yeast cells, water fleas, fruit flies, pond snails, worker ants, and sheep (Southwick, 1976).

Mathematically, the logistic curve is derived from the following differential equation:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \tag{21}$$

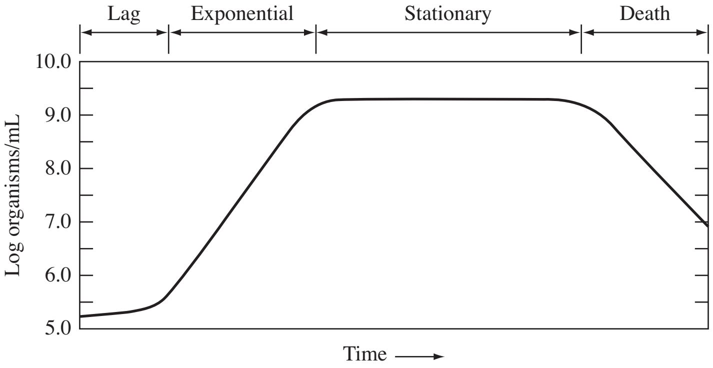


FIGURE 10 Typical growth curve for a bacterial population.
(*Source:* Brock, *Biology of Microorganisms,* 2nd ed., © 1974. Reprinted by permission of Prentice Hall, Inc., Englewood Cliffs, New Jersey.)

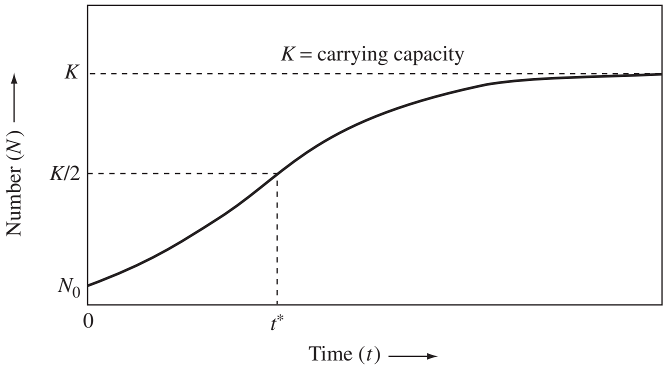


FIGURE 11 The logistic growth curve suggests a smooth transition from exponential growth to a steady-state population.

where N is population size, K is called the *carrying capacity* of the environment, and r is the exponential growth rate constant that would apply if the population size is far below the carrying capacity. Strictly speaking, r is the growth rate constant when $N = 0$; the population is zero. However, when population is much less than carrying capacity, population growth can be modeled with a rate constant r without significant error. As N increases, the rate of growth slows down, and eventually, as N approaches K, the growth stops altogether, and the population stabilizes at a level equal to the carrying capacity. The factor $(1 - N/K)$ is often called the *environmental resistance*. As population grows, the resistance to further population growth continuously increases.

The solution to (21) is

$$N = \frac{K}{1 + e^{-r(t-t^*)}} \tag{22}$$

Note that t^* corresponds to the time at which the population is half of the carrying capacity, $N = K/2$. Substituting $t = 0$ into (22) lets us solve for t^*. Applying the relationship into (22), we can derive the integrated expression of (21) as

$$N(t) = \frac{KN_0}{N_0 + (K - N_0)e^{-rt}} \tag{23}$$

where N_0 is the population at time $t = 0$. In the usual application of (23), the population growth rate is known at $t = 0$. However, the rate constant for this growth rate is not the same as the rate constant r. It cannot necessarily be assumed that the effect of environmental resistance on the growth rate is negligible. To find r from the growth rate measured at $t = 0$, the factor R_0 is introduced. Let R_0 = instantaneous rate constant at $t = 0$ (but not necessarily when $N \ll K$). If we characterize growth at $t = 0$ as exponential, then

$$\left.\frac{dN}{dt}\right|_{t=0} = R_0 N_0 \tag{24}$$

But, from (21)

$$\left.\frac{dN}{dt}\right|_{t=0} = rN_0\left(1 - \frac{N_0}{K}\right) \tag{25}$$

so that equating (24) with (25) yields

$$r = \frac{R_0}{1 - N_0/K} \tag{26}$$

Equation (26) lets us use quantities that are known at $t = 0$, namely the population size N_0 and the population growth constant R_0, to find the appropriate growth constant r for (23). The relationship between the exponential rate constants r and R_0 can be confusing. Remember, there is a different R_0 corresponding to each value of N_0, but r has a single value corresponding to the exponential growth that occurs only when $N_0 = 0$. It may help to think of r as the intrinsic rate constant for the particular logistic growth curve, as shown in (21), and R_0 as a provisional rate constant for the instantaneous growth rate at any time, if the growth is assumed to be exponential (and not logistic).

The following example demonstrates the application of the logistic growth model.

EXAMPLE 8 **Logistic Human Population Curve**

Suppose the human population follows a logistic curve until it stabilizes at 15.0 billion. In 2006, the world's population was 6.6 billion, and its growth rate was 1.2 percent. When would the population reach 7.5 billion—one half of its assumed carrying capacity?

Solution Find r using (26)

$$r = \frac{R_0}{1 - N_0/K} = \frac{0.012}{(1 - 6.6 \times 10^9/15 \times 10^9)} = 0.0214$$

To solve for the time required to reach a population of 7.5 billion, rearrange (23) to solve for t:

$$t = \frac{1}{r}\ln\left(\frac{N(K - N_0)}{N_0(K - N)}\right) \tag{27}$$

Now substitute $N_0 = 6.6$ billion and $N(t) = 7.5$ billion into (27)

$$t = \frac{1}{0.0214}\ln\left(\frac{7.5 \times 10^9(15 \times 10^9 - 6.6 \times 10^9)}{6.6 \times 10^9(15 \times 10^9 - 7.5 \times 10^9)}\right) = 11 \text{ yrs}$$

The Earth would reach one-half its carrying capacity in 2017.

Maximum Sustainable Yield

The logistic curve can also be used to introduce another useful concept in population biology called the *maximum sustainable yield* of an ecosystem. The maximum sustainable yield is the maximum rate that individuals can be harvested (removed) without reducing the population size. Imagine, for example, harvesting fish from a pond. If the pond is at its carrying capacity, there will be no population growth. Any fish removed will reduce the population. Therefore, the maximum sustainable yield will correspond to some population less than the carrying capacity. In fact, because the yield is the same as dN/dt, the maximum yield will correspond to the point on the logistic curve where the slope is a maximum (its inflection point). Setting the derivative of the slope equal to zero, we can find that point. The slope of the logistic curve is given by (21):

$$\text{Yield} = \text{Slope} = \frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \tag{21}$$

Setting the derivative of the slope equal to zero gives

$$\frac{d}{dt}\left(\frac{dN}{dt}\right) = r\frac{dN}{dt} - \frac{r}{K}\left(2N\frac{dN}{dt}\right) = 0$$

Letting N^* be the population at the maximum yield point gives

$$1 - \frac{2N^*}{K} = 0$$

so that

$$N^* = \frac{K}{2} \qquad \text{(for maximum sustainable yield)} \tag{28}$$

That is, if population growth is logistic, then the maximum sustainable yield will be obtained when the population is half the carrying capacity. The yield at that point can be determined by substituting (28) into (21):

$$\text{Maximum yield} = \left(\frac{dN}{dt}\right)_{\max} = r\frac{K}{2}\left(1 - \frac{K/2}{K}\right) = \frac{rK}{4} \tag{29}$$

Using (26) lets us express the maximum yield in terms of the current growth rate R_0 and current size N_0:

$$\left(\frac{dN}{dt}\right)_{\max} = \left(\frac{R_0}{1 - N_0/K}\right) \times \frac{K}{4} = \frac{R_0K^2}{4(K - N_0)} \tag{30}$$

EXAMPLE 9 Harvesting Fish

Observations of a pond newly stocked with 100 fish shows their population doubles in the first year, but after many years, their population stabilizes at 4,000 fish. Assuming a logistic growth curve, what would be the maximum sustainable yield from this pond?

Solution In the first year, with no harvesting, the population doubles. One approach is to estimate the initial growth rate constant, R_0, from the doubling time equation (5):

$$R_0 = \frac{\ln 2}{t_d} = \frac{\ln 2}{1 \text{ yr}} = 0.693 \text{ yr}^{-1}$$

Since R_0 is an instantaneous rate, it conceptually only applies for a single value of N, not for all values of N (100–200 fish) occurring during the first year. A good estimate of the appropriate population size corresponding to the value of R_0 calculated would be the midpoint population for the year, 150 fish. However, as is often done, N_0 could also be taken as 100 fish (the initial population for the period in which the growth rate was measured) without significant error. So from (30) the maximum yield would be

$$\left(\frac{dN}{dt}\right)_{\max} = \frac{R_0K^2}{4(K - N_0)} = \frac{0.693\ \text{yr}^{-1} \times (4{,}000\ \text{fish})^2}{4(4{,}000 - 150\ \text{fish})} = 720\ \text{fish/yr}$$

If we had used 100 fish for N_0, the maximum sustainable yield would have been calculated as 710 fish/yr, showing there will only be a small error in the final answer as long as the initial population is small relative to the carrying capacity. This suggests a more intuitive way we could have gone about this because the initial population, N_0, is much lower than the carrying capacity, K. In this case, the instantaneous growth rate constant, R_0, is approximately equal to the growth rate constant, r, when there is no environmental resistance. That is, $r \approx R_0 = 0.693/\text{yr}$. Using this approximation along with the fact that population size when yield is highest is $K/2 = 2{,}000$ fish, from (21) we have

$$\begin{aligned}\text{Maximum yield} &= rN\left(1 - \frac{N}{K}\right) \approx 0.693 \times 2{,}000\left(1 - \frac{2{,}000}{4{,}000}\right)\\ &= 693\ \text{fish/yr}\end{aligned}$$

5 Human Population Growth

The logistic growth equations just demonstrated are frequently used with some degree of accuracy and predictive capability for density-dependent populations, but they are not often used to predict human population growth. More detailed information on fertility and mortality rates and population age composition are usually available for humans, increasing our ability to make population projections. Use of such data enables us to ask such questions as: What if every couple had just two children? What if such replacement level fertility were to take 40 years to achieve? What would be the effect of reduced fertility rates on the ratios of retired people to workers in the population? The study of human population dynamics is known as *demography,* and what follows is but a brief glimpse at some of the key definitions and most useful mathematics involved.

The simplest measure of fertility is the *crude birth rate,* which is just the number of live births per 1,000 population in a given year. It is called crude because it does not take into account the fraction of the population that is physically capable of giving birth in any given year. The crude birth rate is typically in the range of 30 to 40 per 1,000 per year for the poorest, least developed countries of the world, whereas it is typically around 10 for the more developed ones. For the world as a whole, in 2006, the crude birth rate was 21 per 1,000. That means the 6.56 billion people alive then would have had about $6.56 \times 10^9 \times (21/1{,}000) = 138$ million live births in that year. (These statistics, and most others in this section, are taken from data supplied by the Population Reference Bureau in a most useful annual publication called the *2006 World Population Data Sheet*.)

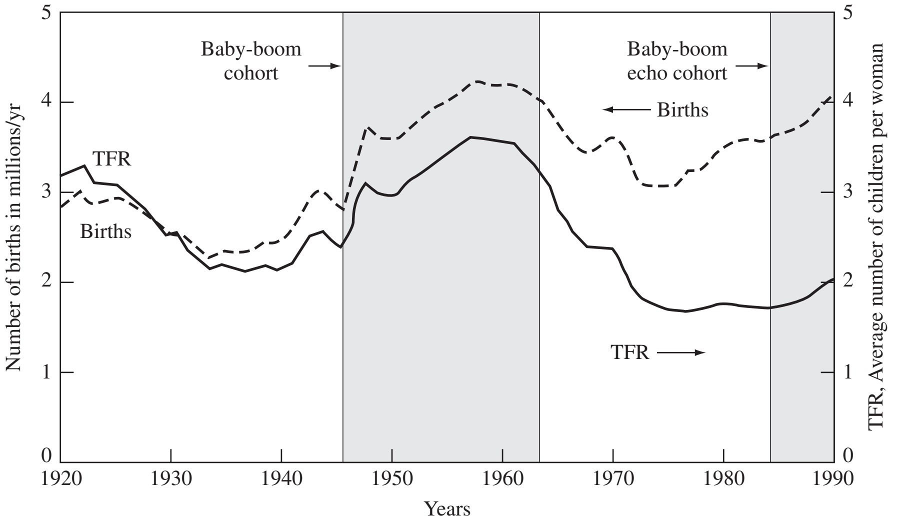


FIGURE 12 Annual births and total fertility rate in the United States.
(*Source:* Bouvier and De Vita, 1991.)

The *total fertility rate* (TFR) is the average number of children that would be born alive to a woman, assuming current age-specific birth rates remain constant through the woman's reproductive years. It is a good measure of the average number of children each woman is likely to have during her lifetime. In the developed countries of the world in 2006, the total fertility rate was 1.6 children per female, whereas in the less developed countries, the average was 2.9. In some of the poorest countries, the TFR is more than 6 children per woman. Figure 12 shows the history of the TFR for the United States. During the depression years of the 1930s, TFR was barely above 2; during the baby-boom era, from 1946 to 1964, it was well over 3 children per woman; in the mid-1970s, it dropped back to a level close to 2.

The number of children that a woman must have, on the average, to replace herself with one daughter in the next generation is called *replacement level fertility*. Replacement level fertility accounts for differences in the ratio of male to female births as well as child mortality rates. For example, in the United States, replacement level fertility is 2.11 children per female. At that fertility level, 100 women would bear 211 children on the average. Statistically, 108 of the children would be boys and 103 would be girls. About 3 percent of the girls would be expected to die before bearing children, leaving a net of 100 women in the next generation. Notice that differing mortality rates in other countries cause the replacement level of fertility to vary somewhat compared to the 2.11 just described. In many developing countries, for example, higher infant mortality rates raise the level of fertility needed for replacement to approximately 2.7 children.

As will be demonstrated later, having achieved replacement level fertility does not necessarily mean a population has stopped growing. When a large fraction of the population is young people, as they pass through their reproductive years, the total births will continue to grow despite replacement fertility. Figure 12 illustrates this concept for the United States. Even though replacement level fertility had been achieved by the mid-1970s, the total number of births rose as the baby boomers had their children. The cohorts born between 1981 and 1995 are often referred to as the baby boom echo. The continuation of population growth, despite achievement of replacement level fertility, is a phenomenon known as *population momentum,* and it will be described more carefully a bit later in the chapter.

The simplest measure of mortality is the *crude death rate,* which is the number of deaths per 1,000 population per year. Again, caution should be exercised in interpreting *crude* death rates because the age composition of the population is not accounted for. The United States, for example, has a higher crude death rate than Guatemala, but that in no way indicates equivalent risks of mortality. In Guatemala, only 4 percent of the population is over 65 years of age (and hence at greater risk of dying), whereas in the United States, a much large fraction of the population, 12 percent, is over 65.

An especially important measure of mortality is the *infant mortality rate,* which is the number of deaths to infants (under 1 year of age) per 1,000 live births in a given year. The infant mortality rate is one of the best indicators of poverty in a country. In some of the poorest countries of the world, infant mortality rates are over 140, which means 1 child in 7 will not live to see his or her first birthday. In the more developed countries, infant mortality rates are around 6 per 1,000 per year.

EXAMPLE 10 Birth and Death Statistics

In 2006, over 80 percent of the world's population, some 5.3 billion people, lived in the less developed countries of the world. In those countries, the average crude birth rate was 23, crude death rate was 8.5, and the infant mortality rate was 53. What fraction of the total deaths is due to infant mortality? If the less developed countries were able to care for their infants as well as they are cared for in most developed countries, resulting in an infant mortality rate of 6, how many infant deaths would be avoided each year?

Solution To find the number of infant deaths each year, we must first find the number of live births and then multiply that by the fraction of those births that die within the first year:

Now:

$$\begin{aligned}\text{Infant Deaths} &= \text{Population} \times \text{Crude Birth Rate} \times \text{Infant Mortality Rate}\\ &= 5.3 \times 10^9 \times (23/1{,}000) \times (53/1{,}000) = 6.5 \times 10^6 \text{ per year}\\ \text{Total Deaths} &= \text{Population} \times \text{Crude Death Rate}\\ &= 5.3 \times 10^9 \times (8/1{,}000) = 42.4 \times 10^6 \text{ per year}\\ \text{Fraction Infants} &= 6.5/42.4 = 0.15 = 15 \text{ percent}\end{aligned}$$

With lowered infant mortality rate:

$$\begin{aligned}\text{Infant Deaths} &= 5.3 \times 10^9 \times (23/1{,}000) \times (6/1{,}000) = 0.73 \times 10^6 \text{ per year}\\ \text{Avoided Deaths} &= (6.5 - 0.73) \times 10^6 = 5.8 \text{ million per year}\end{aligned}$$

It is often argued that reductions in the infant mortality rate would eventually result in reduced birth rates as people began to gain confidence that their offspring would be more likely to survive. Hence, the reduction of infant mortality rates through such measures as better nutrition, cleaner water, and better medical care is thought to be a key to population stabilization.

The difference between crude birth rate b and crude death rate d is called the *rate of natural increase* of the population. While it can be expressed as a rate per 1,000 of population, it is more common to express it either as a decimal fraction or as a percentage rate. As a simple, but important equation,

$$r = b - d \tag{31}$$

where r is the rate of natural increase. If r is treated as a constant, then the exponential relationships developed earlier can be used. For example, in 2006, the crude birth rate for the world was 21 per 1,000, and the crude death rate was 9 per 1,000, so the rate of natural increase was $(21 - 9)/1{,}000 = 12/1{,}000 = 0.012 =$ 1.2 percent. If this rate continues, then the world would double in population in about $70/1.2 = 58$ years.

Although it is reasonable to talk in terms of a rate of natural increase for the world, it is often necessary to include effects of migration when calculating growth rates for an individual country. Letting m be the *net migration rate,* which is the difference between immigration (in-migration) and emigration (out-migration), we can rewrite the preceding relationship as follows:

$$r = b - d + m \tag{32}$$

At the beginning of the industrial revolution, it is thought that crude birth rates were around 40 per 1,000, and death rates were around 35 per 1,000, yielding a rate of natural increase of about 0.5 percent for the world. As parts of the world began to achieve the benefits of a better and more assured food supply, improved sanitation, and modern medicines, death rates began to drop. As economic and social development has proceeded in the currently more developed countries, crude birth rates have fallen to about 12 per 1,000 and crude death rates to about 10 per 1,000. These countries have undergone what is referred to as the *demographic transition,* a transition from high birth and death rates to low birth and death rates, as shown in Figure 13.

The less developed countries of the world have also experienced a sizable drop in death rates, especially during the last half century. Imported medicines and better control of disease vectors have contributed to a rather sudden drop in death rates to their 2006 level of 8 per 1,000. In general, birth rates have not fallen nearly as fast, however. They are still up around 23 per 1,000, which yields a growth rate of 1.5 percent for over 80 percent of the world's population. The rapid (and historically speaking, extraordinary) population growth that the world is experiencing now is almost entirely due to the drop in death rates in developing countries without a significant corresponding drop in birth rates. Many argue that decreases in fertility depend on economic growth, and many countries today face the danger that economic growth may not be able to exceed population growth. Such countries may be

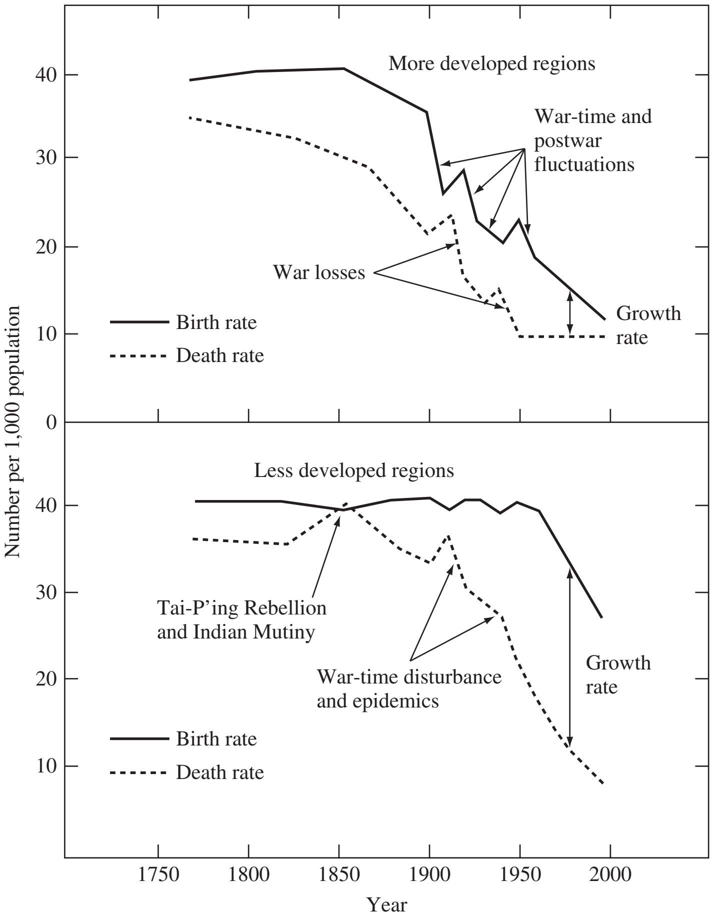


FIGURE 13 The demographic transition to low birth and death rates took over 100 years in the more developed nations. The less developed regions are in the middle of the process.
(*Source:* United Nations, 1971.)

temporarily stuck in the middle of the demographic transition and are facing the risk that population equilibrium may ultimately occur along the least desirable path, through rising death rates.

One developing country that stands out from the others is China. After experiencing a disastrous famine from 1958 to 1961, China instituted aggressive family planning programs coupled with improved health care and living standards that cut birth rates in half in less than 2 decades. Figure 14 shows China's rapid passage through the demographic transition, which was unprecedented in speed (the developed countries took more than 100 years to make the transition).

Some of these data on fertility and mortality are presented in Table 3 for the world, for more developed countries, for China, and for the less developed countries excluding China. China has been separated out because, in terms of population control, it is so different from the rest of the less developed countries and because its size so greatly influences the data.

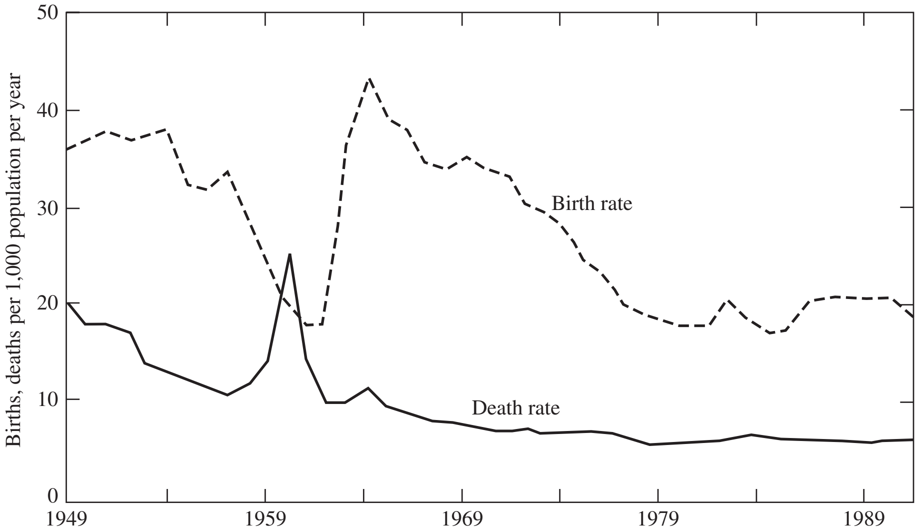


FIGURE 14 Birth and death rates in China, 1949 to 1991.
(*Source:* Tien et al., 1992.)

TABLE 3

Some Important Population Statistics (2006)

	World	More Developed Countries	Less Developed Countries (Excluding China)	China	USA
Population (millions)	6,555	1,216	4,028	1,311	299
% of world population	100	19	61	20	4.6
Crude birth rate, *b*	21	11	27	12	14
Crude death rate, *d*	9	10	9	7	8
Natural increase, *r* %	1.2	0.1	1.8	0.5	0.6
% Population under age 15	29	17	35	20	20
Total fertility rate	2.7	1.6	3.4	1.6	2.0
Infant mortality rate	52	6	61	27	6.7
% of total added 2006 to 2050	41	4	63	10	40
Per capita GNI[a] (US$)	9,190	27,790	4,410	6,600	41,950
% urban	48	77	42	37	79
Est. population 2025 (millions)	7,940	1,255	5,209	1,476	349
Added pop. 2006 to 2025 (millions)	1,385	39	1,181	165	50

[a] Gross National Income per capita 2005.
Source: Population Reference Bureau, 2006.

Age Structure

A great deal of insight and predictive power can be obtained from a table or diagram representing the age composition of a population. A graphical presentation of the data, indicating numbers of people (or percentages of the population) in each

FIGURE 15 A rapidly growing, expansive population (Morocco), compared with a population that is nearing stationary condition (Sweden).
(*Source:* Haupt and Kane, 1985.)

age category, is called an *age structure* or a *population pyramid*. An age structure diagram is a snapshot of a country's population trends that tells a lot about the recent past as well as the near future. Developing countries with rapidly expanding populations, for example, have pyramids that are triangular in shape, with each cohort larger than the cohort born before it. An example pyramid is shown in Figure 15 for Morocco. It is not uncommon in such countries to have nearly half of the population younger than 15 years old. The age structure shows that even if replacement fertility is achieved in the near future, there are so many young women already born who will be having their children that population size will continue to grow for many decades.

The second pyramid in Figure 15 is an example of a population that is not growing very rapidly, if at all. Notice that the sides of the structure are much more vertical. There is also a pinching down of the age structure in the younger years, suggesting a fertility rate below replacement level, so that if these trends continue, the population will eventually begin to decrease.

Two important terms are used by demographers to describe the shape of an age structure. A *stable population* has had constant age-specific birth and death rates for such a long time that the percentage of the population in any age category does not change. That is, the shape of the age structure is unchanging. A population that is stable does not have to be fixed in size; it can be growing or shrinking at a constant rate. When a population is both stable and unchanging in size it is called a *stationary population*. Thus all stationary populations are stable, but not all stable populations are stationary.

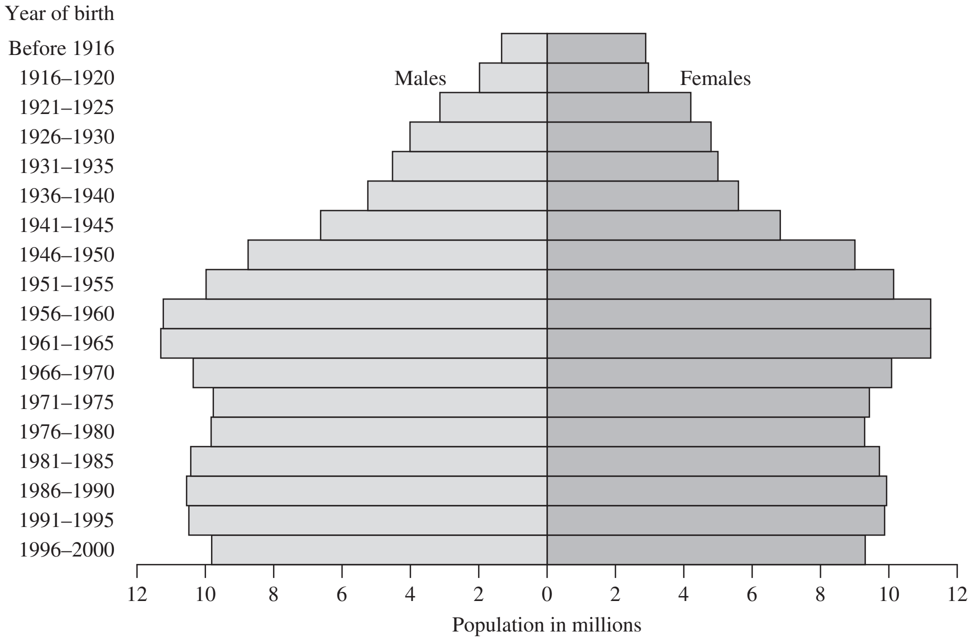


FIGURE 16 The age structure for the United States in 2000.
(*Source:* U.S. Census Bureau.)

The population pyramid for the United States in 2000 is shown in Figure 16. The nearly vertical sides on the lower part of the pyramid suggest that this is a population that is well underway toward a stationary condition. The dominant feature, however, is the bulge in the middle corresponding to the baby-boom cohorts born between 1946 and 1964. Thinking of the age structure as a dynamic phenomenon, the bulge moving upward year after year suggests an image of a python that has swallowed a pig, which slowly works its way through the snake. As the boomers pass through the age structure, resources are stretched to the limit. When they entered school, there were not enough classrooms and teachers, so new schools were built and teacher training programs expanded. When they left, schools closed, and there was an oversupply of teachers. Universities found themselves competing for smaller and smaller numbers of high school graduates. During their late teens and early twenties, arrest rates increased since those are the ages that are, statistically speaking, responsible for the majority of crimes. As they entered the workforce, competition was fierce and unemployment rates were persistently high, housing was scarce, and the price of real estate jumped. The baby-boom generation is now nearing retirement, and there is a boom in construction of retirement communities coupled with growing concern about the capability of the social security system to accommodate so many retirees.

By contrast, children born in the 1930s, just ahead of the baby-boom generation, are sometimes referred to as the "good times cohorts." Since there are so few of them, as they sought their educations, entered the job market, bought homes, and are now retiring, the competition for those resources has been relatively modest. Children born in the 1970s also benefit by being part of a narrow portion of the age

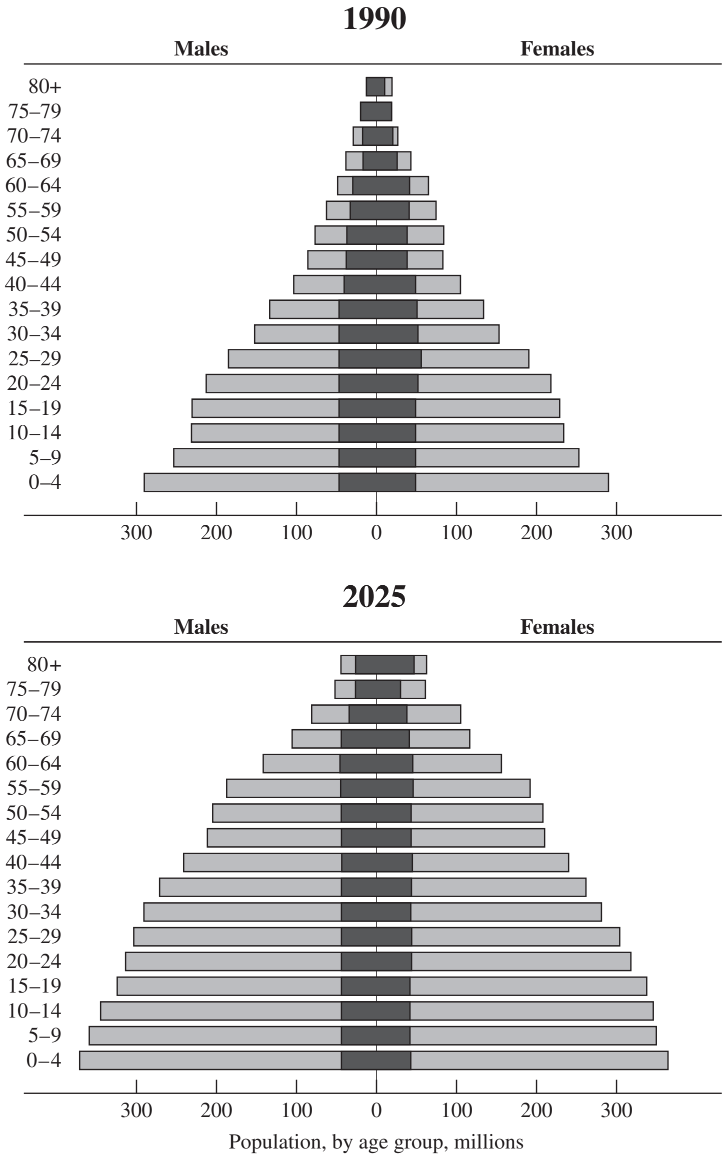


FIGURE 17 Age structures for the world in 1990 and projections for 2025, showing the developed countries in darker tones than the developing countries.
(*Source: The Economist,* January 27, 1996.)

structure. However, their future is less rosy since they will be part of the working population that will face the responsibility of supporting the more elderly baby-boom generation. At the bottom of the age structure are the baby-boom echo cohorts born in the 1980s and 1990s.

Figure 17 shows age structures for the world in 1990 and projections for 2025. The developed countries essentially have stationary populations in both years, but the developing countries are still growing rapidly.

Human Population Projections

If the age structure for a population is combined with data on age-specific birth and death rates, it is possible to make realistic projections of future population size and composition. The techniques that we will explore now are especially useful for predicting the effects and implications of various changes in fertility patterns that might be imagined or advocated. For example, suppose we want to determine the effects of replacement-level fertility. If replacement fertility were achieved today, how long would it take to achieve a stationary population? What would the age composition look like in the interim? What fraction of the population would be young and looking for work, and what fraction would be retired and expecting support? Or suppose a policy is established that sets a target for maximum size of the population (as is the case for China). How many children would each couple need to have to achieve that goal, and how would the number change with time?

Not long ago, the calculations required to develop these scenarios were tedious and best left to professional demographers. Now, however, with the widespread use of simple spreadsheet programs on personal computers, it is possible to work them out with relative ease.

The starting point in a population projection is the current age structure combined with mortality data obtained from *life tables*. Life tables are the mainstay of insurance companies for predicting the average number of years of life remaining as a function of the age of their clients. A life table is developed by applying a real population's age-specific death rates (the fraction of the people in a given age category who will die each year) to *hypothetical* stable and stationary populations having 100,000 live births per year, evenly distributed through the year, with no migration. As the 100,000 people added each year get older, their ranks are thinned in accordance with the age-specific death rates. It is then possible to calculate the numbers of people who would be alive within each age category in the following year.

Table 4 presents a life table for the United States that has been simplified by broadening the age intervals to 5-year increments rather than the 1-year categories used by insurance companies and demographers. These data, remember, are for a hypothetical population with 100,000 live births each year (and 100,000 deaths each year as well since this is a stationary population). The first column shows the age interval (e.g., 10–14 means people who have had their tenth birthday but not their fifteenth). The second column is the number of people who would be alive at any given time in the corresponding age interval and is designated L_x, where x is the age at the beginning of the interval. The third column, L_{x+5}/L_x, is the ratio of number of people in the next interval to the number in the current interval; it is the probability that individuals will live 5 more years (except in the case of those 80 and older, where the catch-all category of 85+ years old modifies the interpretation).

If we assume that the age-specific death rates that were used to produce Table 4 remain constant, we can use the table to make future population projections at five-year intervals for all but the 0–4 age category. The 0–4 age category will depend on fertility data, to be discussed later.

If we let $P_x(0)$ be the number of people in age category x to $x + 5$ at time $t = 0$, and $P_{x+5}(5)$ be the number in the next age category five years later, then

$$P_{x+5}(5) = P_x(0)\frac{L_{x+5}}{L_x} \tag{33}$$

TABLE 4

Age Distribution for a Hypothetical, Stationary Population with 100,000 Live Births Per Year in the United States[a]

Age Interval x to $x + 5$	Number in Interval L_x	$\frac{L_{x+5}}{L_x}$
0–4	494,285	0.9979
5–9	493,247	0.9989
10–14	492,704	0.9973
15–19	491,374	0.9951
20–24	488,966	0.9944
25–29	486,228	0.9940
30–34	483,311	0.9927
35–39	479,783	0.9898
40–44	474,889	0.9841
45–49	467,338	0.9745
50–54	455,421	0.9597
55–59	437,068	0.9381
60–64	410,013	0.9082
65–69	372,374	0.8658
70–74	322,401	0.8050
75–79	259,533	0.7163
80–84	185,904	0.9660
85 and over	179,583	0.0000

[a] Age-specific death rates are U.S. 1984 values.

Source: Abstracted from data in U.S. Dept. of Health and Human Services (1987).

That is, five years from now, the number of people in the next five-year age interval will be equal to the number in the interval now times the probability of surviving for the next five years (L_{x+5}/L_x).

For example, in the United States in 1985, there were 18.0 million people in the age group 0–4 years; that is, $P_0(1985) = 18.0$ million. We would expect that in 1990, the number of people alive in the 5–9 age category would be

$$\begin{aligned} P_5(1990) &= P_0(1985)(L_5/L_0) \\ &= 18.0 \times 10^6 \times 0.9979 = 17.98 \text{ million} \end{aligned}$$

Let's take what we have and apply it to the age composition of the United States in 1985 to predict as much of the structure as we can for 1990. This involves application of (33) to all of the categories, giving us a complete 1990 age distribution except for the 0–4 year olds. The result is presented in Table 5.

To find the number of children ages 0–4 to enter into the age structure, we need to know something about fertility patterns. Demographers use *age-specific fertility rates,* which are the number of live births per woman in each age category, to make these estimates. To do this calculation carefully, we would need to know the number of people in each age category who are women as well as statistical data on child mortality during the first five years of life. Since our purpose here is to develop a tool for asking questions of the "what if" sort, rather than making actual population projections, we can use the following simple approach to estimate the number

TABLE 5

1990 U.S. Population Projection Based on the 1985 Age Structure (Ignoring Immigration)

Age Interval x to $x + 5$	$\frac{L_{x+5}}{L_x}$	P_x (Thousands) 1985	P_x (Thousands) 1990
0–4	0.9979	18,020	$P_0(1990)$
5–9	0.9989	17,000	17,982
10–14	0.9973	16,068	16,981
15–19	0.9951	18,245	16,025
20–24	0.9944	20,491	18,156
25–29	0.9940	21,896	20,376
30–34	0.9927	20,178	21,765
35–39	0.9898	18,756	20,031
40–44	0.9841	14,362	18,564
45–49	0.9745	11,912	14,134
50–54	0.9597	10,748	11,609
55–59	0.9381	11,132	10,314
60–64	0.9082	10,948	10,443
65–69	0.8658	9,420	9,943
70–74	0.8050	7,616	8,156
75–79	0.7163	5,410	6,131
80–84	0.9660	3,312	3,875
85 and over	0.0000	2,113	3,199
Total:		237,627	$227,684 + P_0(1990)$

Source: 1985 data from Vu (1985).

of children ages 0–4 to put into our age composition table:

$$P_0(n + 5) = b_{15}P_{15}(n) + b_{20}P_{20}(n) + \cdots b_{45}P_{45}(n) \tag{34}$$

where b_x is the number of the surviving children born per person in age category x to $x + 5$. Equation (34) has been written with the assumption that no children are born to individuals under 15 or over 49 years of age, although it could obviously have been written with more terms to include them. Notice that we can interpret the sum of b_x's (Σb_x) to be the number of children each person would be likely to have. The total fertility rate, which is the expected number of children per woman, is therefore close to $2\Sigma b_x$.

For example, these fertility factors and the numbers of people in their reproductive years for the United States in 1985 are shown in Table 6. The total fertility rate can be estimated as $2\Sigma b_x$, which is $2 \times 1.004 = 2.01$ children per woman.

With this 20,385,000 added to the 227,684,000 found in Table 5, the total population in the United States in 1990 would have been 248,069,000. Notice that, in spite of the fact that fertility was essentially at the replacement level in 1985, the population would still be growing by about 2 million people per year. This is an example of *population momentum,* a term used to describe the fact that a youthful

TABLE 6

Calculating Births in Five-Year Increments for the United States, 1985

Age Category	$P_x(1985)$ (Thousands)	b_x (Births Per Person During Five-Year Period)	Births (Thousands in Five Years)
15–19	18,245	0.146	2,664
20–24	20,491	0.289	5,922
25–29	21,896	0.291	6,372
30–34	20,178	0.190	3,834
35–39	18,756	0.075	1,407
40–45	14,362	0.013	187
Totals		1.004	20,385

age structure causes a population to continue growing for several decades after achieving replacement fertility. To finish off the aforementioned estimate, we would have to add in net migration, which is currently 897,000 legal immigrants per year plus probably hundreds of thousands of immigrants entering the country illegally.

Population Momentum

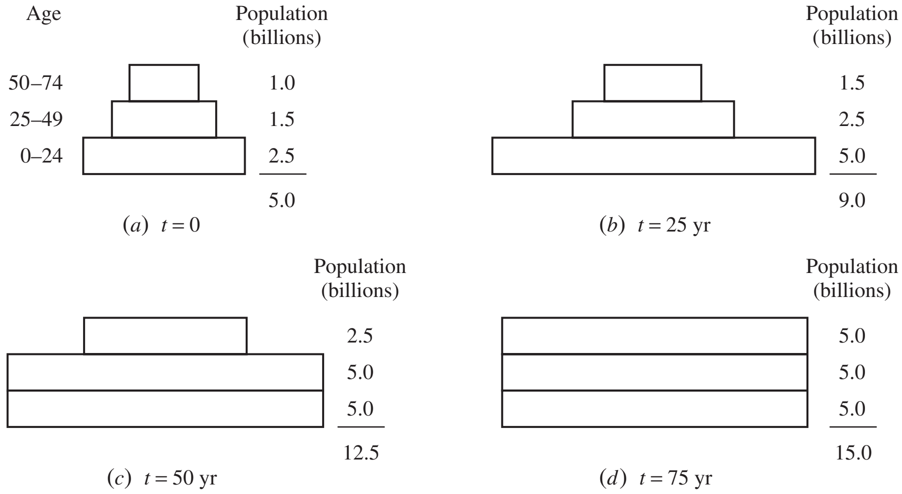
Let us begin with an extremely simplified example, which does not model reality very well, but which does keep the arithmetic manageable. We will then see the results of more carefully done population scenarios.

Suppose we have an age structure with only three categories: 0–24 years, 25–49 years, and 50–74 years, subject to the following fertility and mortality conditions:

1. All births occur to women as they leave the 0–24 age category.
2. All deaths occur at age 75.
3. The total fertility rate (TFR) is 4.0 for 25 years; then it drops instantaneously to 2.0.

Suppose a population of 5.0 billion people has an age structure at time $t = 0$, as shown in Figure 18*a*. Equal numbers of men and women are assumed so that the number of children per person is half the TFR. During the first 25 years, the 2.5 billion people who start in the 0–24 age category all pass their twenty-fifth birthday, and all of their children are born. Because TFR is 4.0, these 2.5 billion individuals bear 5.0 billion children (2 per person). Those 2.5 billion people now are in the 25–49 age category. Similarly, the 1.5 billion people who were 25–49 years old are now ages 50–74. Finally, the 1.0 billion 50–74 year olds have all passed their seventy-fifth birthdays and, by the rules, they are all dead. The total population has grown to 9.0 billion, and the age structure is as shown in Figure 18*b*.

After the first 25 years have passed, TFR drops to the replacement level of 2.0 (1 child per person). During the ensuing 25 years, the 5.0 billion 0–24 year olds have 5.0 billion children and, following the logic given previously, the age structure at $t = 50$ years is as shown in Figure 18*c*. With the replacement level continuing, at $t = 75$ years, the population stabilizes at 15 billion. A plot of population versus time is shown in Figure 19. Notice that the population stabilizes 50 years after



(a) $t = 0$

(b) $t = 25$ yr

(c) $t = 50$ yr

(d) $t = 75$ yr

FIGURE 18 Age structure diagrams illustrating population momentum. For the first 25 years, TFR = 4; thereafter, it is at the replacement level of 2.0, yet population continues to grow for another 50 years.

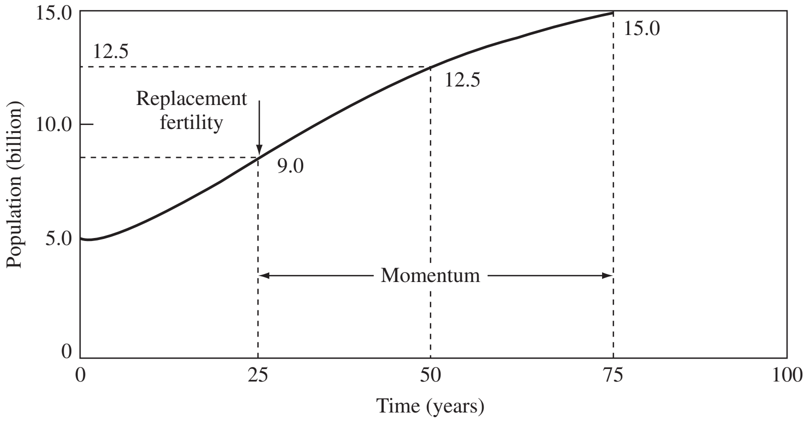


FIGURE 19 In the hypothetical example, it takes 50 years of replacement-level fertility before population stops growing.

replacement-level fertility is achieved, during which time it grows from 9 billion to 15 billion.

A more carefully done set of population scenarios is presented in Figure 20. In the constant-path scenario, population is assumed to continue growing at the 1980 growth rate into the indefinite future. The slow fertility-reduction path assumes that the world's fertility will decline to reach replacement level by 2065, which results in a stabilized population of about 14 billion by the end of the twenty-first century. The moderate fertility-reduction path assumes that replacement-level fertility will be reached by 2035, with the population stabilizing at about 10 billion. Finally, the

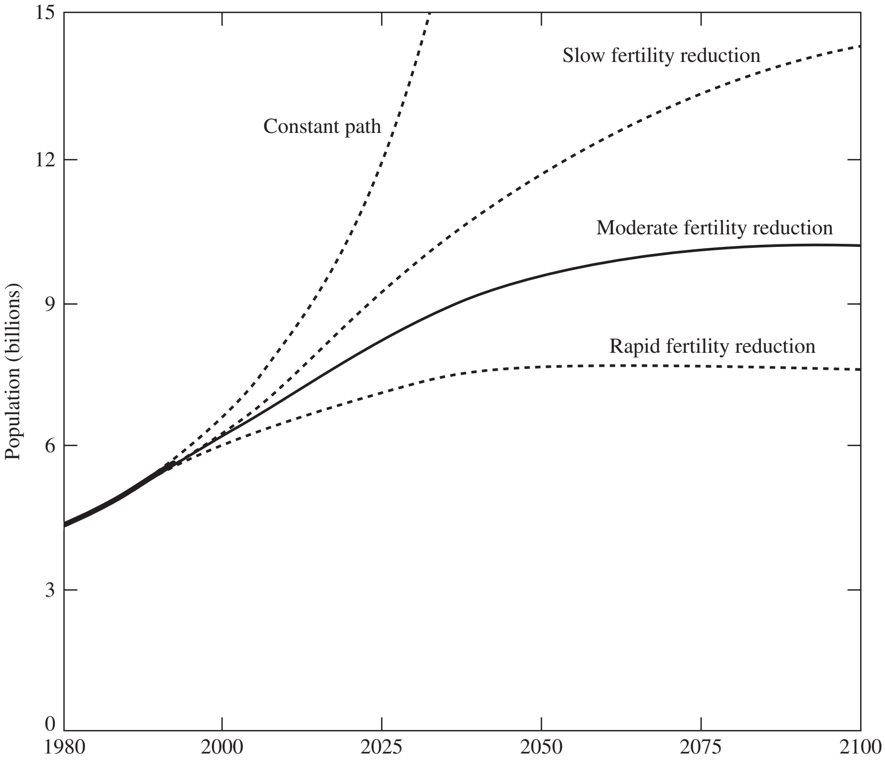


FIGURE 20 Four scenarios for the future of the world's population. The moderate path has fertility declining to replacement level by 2035, with momentum carrying the population to approximately 10 billion by the end of the twenty-first century. (*Source:* Haupt and Kane, 1985.)

rapid fertility-reduction path assumes that the world's fertility will decline to replacement level by 2010. Even under this most optimistic scenario, the world's population would still grow to approximately 7.5 billion. The implications of population momentum are obviously of tremendous importance.

PROBLEMS

1 World population in 1850 has been estimated at about 1 billion. World population reached 4 billion in 1975.

  (a) Use the doubling time approximation (5) to estimate the exponential rate of growth that would produce those numbers.

  (b) Use the exponential growth equation (6) to find the growth rate.

2 Tuition at a major university rose from \$1,500/yr in 1962 to \$20,000/yr in 1995.

  (a) What exponential rate of growth characterized that period of time?

  (b) If that growth rate were to continue for another 25 years (enough time for current students to have children in college!) what would the tuition be?

3 The world's population 10,000 years ago has been estimated at about 5 million. What exponential rate of growth would have resulted in the population in 1850, which is estimated at 1 billion? Had that rate continued, what would the population have been in the year 2000?

4 Suppose we express the amount of land under cultivation as the product of four factors:

$$\text{Land} = (\text{land/food}) \times (\text{food/kcal}) \times (\text{kcal/person}) \times (\text{population})$$

The annual growth rates for each factor are (1) the land required to grow a unit of food, −1 percent (due to greater productivity per unit of land); (2) the amount of food grown per calorie of food eaten by a human, +0.5 percent (because with affluence, people consume more animal products, which greatly reduces the efficiency of land use); (3) the per capita calorie consumption, +0.1 percent; and (4) the size of the population, +1.5 percent. At these rates, how long would it take to double the amount of cultivated land needed? At that time, how much less land would be required to grow a unit of food?

5 Suppose world carbon emissions are expressed as the following product:

$$\text{Carbon emissions} = (\text{energy/person}) \times (\text{carbon/energy}) \times (\text{population})$$

If per capita energy demand increases at 1.5 percent per year, fossil fuel emissions of carbon per unit of energy increase at 1 percent per year, and world population grows at 1.5 percent per year,

(a) How long would it take before we are emitting carbon at twice the current rate?

(b) At that point, by what fraction would per capita energy demand have increased?

(c) At that point, by what fraction would total energy demand have increased?

6 Under the assumptions stated in Problem 5, if our initial rate of carbon emission is 5×10^9 tonnes C/yr and if there are 700×10^9 tonnes of carbon in the atmosphere now,

(a) How long would it take to emit a total amount of carbon equal to the amount now contained in the atmosphere?

(b) If half of the carbon that we emit stays in the atmosphere (the other half being "absorbed" in other parts of the biosphere), how long would it take for fossil fuel combustion to double the atmospheric carbon concentration?

7 Consider the following disaggregation of carbon emissions:

$$\text{Carbon emissions (kg C/yr)} = \text{Population} \times \frac{\text{Energy (kJ/yr)}}{\text{Person}} \times \frac{\text{Carbon (kgC)}}{\text{Energy (kJ)}}$$

Using the following estimates for the United States and assuming that growth rates remain constant,

	Population	(kJ/yr)/Person	kg C/kJ
1990 amounts	250×10^6	320×10^6	15×10^{-6}
Growth, *r* (%/yr)	0.6	0.5	−0.3

(a) Find the carbon emission rate in 2020.

(b) Find the carbon emitted in those 30 years.

(c) Find total energy demand in 2020.

(d) Find the per capita carbon emission rate in 2020.

8 World reserves of chromium are about 800 million tons, and current usage is about 2 million tons per year. If growth in demand for chromium increases exponentially at a constant rate of 2.6 percent per year, how long would it take to use up current reserves? Suppose the total resource is five times current reserves; if the use rate continues to grow at 2.6 percent, how long would it take to use up the resource?

9 Suppose a Gaussian curve is used to approximate the production of chromium. If production peaks at six times its current rate of 2 million tons per year, how long would it take to reach that maximum if the total chromium ever mined is 4 billion tons (five times the current reserves)? How long would it take to consume about 80 percent of the total resource?

10 Suppose we assume the following:

(a) Any chlorofluorocarbons (CFCs) released into the atmosphere remain in the atmosphere indefinitely.

(b) At current rates of release, the atmospheric concentration of fluorocarbons would double in 100 years (what does that say about Q/P_0?).

(c) Atmospheric release rates are, however, not constant but growing at 2 percent per year. How long would it take to double atmospheric CFC concentrations?

11 Bismuth-210 has a half-life of 4.85 days. If we start with 10 g of it now, how much would we have left in 7 days?

12 Suppose some sewage drifting down a stream decomposes with a reaction rate coefficient k equal to 0.2/day. What would be the half-life of this sewage? How much would be left after 5 days?

13 Suppose human population grows from 6.3 billion in 2000 to an ultimate population of 10.3 billion following the logistic curve. Assuming a growth rate of 1.5 percent in 2000, when would the population reach 9 billion? What would the population be in 2050? Compare this to the moderate fertility reduction scenario of Figure 20.

14 Suppose a logistic growth curve had been used to project the world's population back in 1970, when there were 3.65 billion people, and the growth rate was 2.0 percent per year. If a steady-state population of 10.3 billion had been used (the moderate path in Figure 20), what would the projected population have been for 1995 (when it was actually 5.7 billion) and for 2025?

15 Suppose we stock a pond with 100 fish and note that the population doubles in the first year (with no harvesting), but after a number of years, the population stabilizes at what we think must be the carrying capacity of the pond, 2,000 fish. Growth seems to have followed a logistic curve.

(a) What population size should be maintained to achieve maximum yield, and what would be the maximum sustainable fish yield?

(b) If the population is maintained at 1,500 fish, what would be the sustainable yield?

16 What would be the sustainable yield from the pond in Example 9 if the population is maintained at 3,000 fish?

17 A lake has a carrying capacity of 10,000 fish. At the current level of fishing, 2,000 fish per year are taken with the catch uniformly distributed throughout the year. It is seen that the fish population holds fairly constant at about 4,000. If you wanted to maximize the sustainable yield, what would you suggest in terms of population size and yield?

18 Suppose we stock an island with 250 rabbits and find that the population is 450 after one year (with no harvesting). After a number of years, the population stabilizes at what we

think must be the carrying capacity of the island, 3,500 rabbits. Assume growth has followed a logistic curve.

(a) We now want to start harvesting the rabbits to sell, with the objective of harvesting the most rabbits possible per year from the island. What population size should be maintained to achieve the maximum sustainable yield and what would be the maximum sustainable rabbit yield?

(b) If the population is maintained at 1,200 rabbits, what would be the sustainable yield?

19 The Upset National Forest has a carrying capacity of 7,000 deer. At the current level of recreational hunting, 300 deer per year are taken during a 2-week hunting season. After the season, the deer population is always about 2,200 deer. You want to maximize the sustainable yield of deer from the forest.

(a) What should be the population of deer in the forest for maximum sustainable yield?

(b) What will be the maximum sustainable yield of deer from the forest?

(c) If hunting were stopped so no further deer were taken from the forest, how long would it take for the population size to reach the population (the population you calculated in part (a)) that is necessary for maximum sustainable yield?

20 The following statistics are for India in 1985: population, 762 million; crude birth rate, 34; crude death rate, 13; infant mortality rate, 118 (rates are per thousand per year). Find (a) the fraction of the total deaths that are infants less than 1 year old; (b) the avoidable deaths, assuming that any infant mortality above 10 could be avoided with better sanitation, food, and health care; and (c) the annual increase in the number of people in India.

21 The following statistics are for India in 1995: population, 931 million; crude birth rate, 29; crude death rate, 9; infant mortality rate, 74 (rates are per thousand per year). Find (a) the fraction of the total deaths that are infants less than 1 year old; (b) the avoidable deaths, assuming that any infant mortality above 10 could be avoided with better sanitation, food, and health care; and (c) the annual increase in the number of people in India. (For comparison, in 1985, the population was growing at 16 million per year, and the avoidable deaths were 2.8 million per year.)

22 Consider a simplified age structure that divides a population into three age groups: 0–24, with 3.0 million; 25–49, with 2.0 million; and 50–74, with 1.0 million. Suppose we impose the following simplified fertility and mortality constraints: All births occur just as the woman leaves the 0–24 age category, and no one dies until their seventy-fifth birthday, at which time they all die. Suppose we have replacement-level fertility starting now. Draw the age structure in 25 years, 50 years, and 75 years. What is the total population size at each of these times?

23 Use the same initial population structure given in Problem 22, with all births on the mothers twenty-fifth birthday. Draw the age structure in 25, 50, and 75 years under the following conditions: No deaths occur until the fiftieth birthday, at which time 20 percent die; the rest die on their seventy-fifth birthday. For the first 25 years, the total fertility rate is 4 (2 children/person), and thereafter it is 2.

24 Consider the following simplified age structure: All births are on the mothers twentieth birthday and all deaths are on the sixtieth birthday. Total population starts at 290,000 (half males, half females) and is growing at a constant rate of 3.5 percent per year (see Figure P24). (*Hint:* what does that say about the doubling time?)

Draw the age structure in 20 years. If the total fertility rate is a single, constant value during those 20 years, what is it?

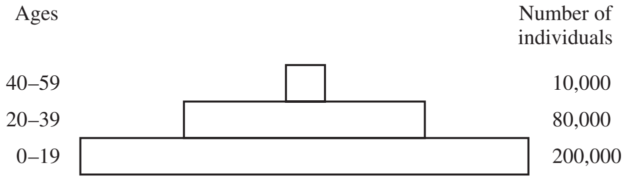


FIGURE P24

25 The following age structure and survival data are for China in 1980. Suppose the birth factors (corresponding to a total fertility rate of 1.0) are as shown. Estimate the population of China in 1990.

Age	Population (millions)	L_{x+10}/L_x	b_x
0–9	235	0.957	0
10–19	224	0.987	0.25
20–29	182	0.980	0.25
30–39	124	0.964	0
40–49	95	0.924	0
50–59	69	0.826	0
60–69	42	0.633	0
70–79	24	0.316	0
80–89	6	0	0
Total:	*1,001*		

26 Use a spreadsheet to project the China population data given in Problem 25 out to the year 2030. What is the population at that time?

27 Use a spreadsheet to project the China population data given in Problem 25 out to the year 2030 but delay the births by one 10-year interval (that is, $b_{10} = 0$, $b_{20} = 0.25$, and $b_{30} = 0.25$). Compare the peak population in Problem 25 to that obtained by postponing births by 10 years.

28 The birth-rate data for China in Problem 25 were for the very optimistic one-child per family scenario (TFR = 1). In reality, the TFR has been closer to 2. Assuming that each woman had 2 children while she was in the 20 to 30 age group (and none at any other age),

(a) Repeat this population projection for China for 1990.

(b) Continuing that birth rate, find the population in 2000.

REFERENCES

Brock, T. D., 1974, *Biology of Microorganisms,* 2nd ed., Prentice Hall, Englewood Cliffs, NJ.

Bouvier, L. F., and C. J. De Vita, 1991, *The Baby Boom—Entering Midlife,* Population Reference Bureau, Inc., Washington, DC.

Energy Information Administration, 1996, *Annual Energy Review, 1996,* U.S. Department of Energy, Washington, DC.

Haupt, A., and T. T. Kane, 1985, *Population Handbook,* 2nd ed., Population Reference Bureau, Washington, DC.

Hubbert, M. K., 1969, *Resources and Man,* National Academy of Sciences, Freeman, San Francisco.

Hubbert, M. K., 1977, *Project Independence: U.S. and World Energy Outlook through 1990,* Congressional Research Service, Washington, DC.

Hobbs, G. W., 1995, Oil, gas, coal, uranium, tar sand resource trends on rise, *Oil and Gas Journal,* September 4, 104–108.

Population Reference Bureau, 1996, *1995 World Population Data Sheet,* Washington, DC.

Population Reference Bureau, 2006, *2006 World Population Data Sheet,* Washington, DC.

Southwick, C. H., 1976, *Ecology and the Quality of our Environment,* 2nd ed., Van Nostrand, New York.

Tien, H. Y., Z. Tianhu, P. Yu, L. Jingnent, and L. Zhongtang, 1992, *China's Demographic Dilemmas,* Population Reference Bureau, Inc., Washington, DC.

United Nations, 1971, *A Concise Summary of the World Population Situation in 1970,* Population Studies No. 48, New York.

U.S. Dept. of Health and Human Services, 1987, *Vital Statistics of the United States, 1984, Life Tables,* National Center for Health Statistics, Hyattsville, MD.

Vu, M. T., 1985, *World Population Projections, 1985, Short- and Long-Term Estimates by Age and Sex with Related Demographic Statistics,* World Bank, Johns Hopkins Press, Baltimore.

Weeks, J. R., 1994, *Population: An Introduction to Concepts and Issues,* Wadsworth, Belmont, CA.

This page intentionally left blank

Risk Assessment

> All substances are poisons; there is none which is not a poison. The right dose differentiates a poison and a remedy.
>
> —*Paracelsus (1493–1541)*

1 Introduction

One of the most important changes in environmental policy in the 1980s was the acceptance of the role of risk assessment and risk management in environmental decision making. In early environmental legislation, such as the Clean Air and Clean Water Acts, the concept of risk is hardly mentioned; instead, these acts required that pollution standards be set that would allow adequate margins of safety to protect public health. Intrinsic to these standards was the assumption that pollutants have thresholds and that exposure to concentrations below these thresholds would produce no harm. All of that changed when the problems of toxic waste were finally recognized and addressed. Many toxic substances are suspected carcinogens; that is, they may cause cancer, and for carcinogens, the usual assumption is that even the smallest exposure creates some risk.

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

If any exposure to a substance causes some risk, how can air quality and water quality standards be set? When cleaning up a hazardous waste site, at what point is the project completed; that is, how clean is clean? At some point in the cleanup, the remaining health and environmental risks may not justify the continued costs and, from a risk perspective, society might be better off spending the money elsewhere. Almost by definition, achieving zero risk would cost an infinite amount of money, so policy makers have had to grapple with the tradeoff between acceptable risk and acceptable cost. Complicating those decisions is our very limited understanding of diseases such as cancer coupled with a paucity of data on the tens of thousands of synthetic chemicals that are in widespread use today. Unfortunately, those who have responsibility for creating and administering environmental regulations have to take action even if definitive answers from the scientific community on the relationship between exposure and risk are not available.

The result has been the emergence of the controversial field of environmental risk assessment. Hardly anyone is comfortable with it. Scientists often deplore the notion of condensing masses of frequently conflicting, highly uncertain, often ambiguous data that have been extrapolated well beyond anything actually measured down to a single number or two. Regulatory officials are battered by the public when they propose a level of risk that they think a community living next to a toxic waste site should tolerate. Critics of government spending think risks are being systematically overestimated, resulting in too much money being spent for too little real improvement in public health. Others think risks are underestimated since risk assessments are based on data obtained for exposure to individual chemicals, ignoring the synergistic effects that are likely to occur when we are exposed to thousands of them in our daily lives.

Some of the aforementioned conflicts can best be dealt with if we make the distinction between risk assessment and risk management. *Risk assessment* is the scientific side of the story. It is the gathering of data that are used to relate response to dose. Such dose-response data can then be combined with estimates of likely human exposure to produce overall assessments of risk. *Risk management,* on the other hand, is the process of deciding what to do. It is decision making, under extreme uncertainty, about how to allocate national resources to protect public health and the environment. Enormous political and social judgment is required to make those decisions. Is a one-in-a-million lifetime risk of getting cancer acceptable and, if it is, how do we go about trying to achieve it?

2 Perspectives on Risks

The usual starting point for an explanation of risk is to point out that there is some risk in everything we do, and since we will all die someday, our lifetime risk of death from all causes is 1.0, or 100 percent. It is easy to gather good data on the causes of death, such as are shown in Table 1, which gives us one approach to talking about risk. For example, of the total 2.393 million predicted deaths in the United States in 2006, 564,800 were predicted to be caused by cancer. Neglecting age structure complications, we could say that, on the average, the risk, or probability, of someone in the United States dying of cancer is about 24 percent (564,800/2,393,000 = 0.24).

TABLE 1

Leading Causes of Death in the United States, 1992

Cause	Annual Deaths (thousands)	Percent
Cardiovascular (heart) disease	720	33
Cancer (malignant neoplasms)	521	24
Cerebrovascular diseases (strokes)	144	7
Pulmonary diseases (bronchitis, emphysema, asthma)	91	4
Pneumonia and influenza	76	3
Diabetes mellitus	50	2
Nonmotor vehicle accidents	48	2
Motor vehicle accidents	42	2
HIV/AIDS	34	1.6
Suicides	30	1.4
Homicides	27	1.2
All other causes	394	18
Total annual deaths (rounded)	2,177	100

Source: Kolluru et al., 1996.

Notice that there are no units associated with risk, although other clarifiers may be needed, such as whether the risk is a lifetime risk or an annual risk, whether it is an average risk to the general public or a risk faced by individuals who engage in some activity, or whether it is being expressed as a percentage or as a decimal fraction. For example, in the United States, smoking is thought to be causing approximately 400,000 deaths per year. On the average, the probability of death caused by smoking is therefore about 18 percent (400,000/2,177,000 = 0.18). Obviously, however, the risk that an individual faces from smoking depends on how much that person smokes and how much exposure is caused by smoke from other people's cigarettes. The probability that a pack-a-day smoker will die of lung cancer, heart disease, or emphysema brought on by smoking is about 0.25, or 25 percent (Wilson and Crouch, 1987). Statistically, smokers shorten their life expectancy by about 5 minutes for each cigarette smoked, which is roughly the time it takes to smoke that cigarette.

Environmental risk assessments deal with incremental probabilities of some harm occurring. For example, the Environmental Protection Agency (EPA) attempts to control our exposure to toxics to levels that will pose incremental lifetime cancer risks to the most exposed members of the public of roughly 10^{-6} (1 additional cancer in 1 million people) to 10^{-4} (100 additional cancers per million people). For perspective, suppose all 260 million Americans faced a 10^{-6} lifetime risk of cancer from exposure to a particular toxic chemical. That would mean 260 extra cancers during their lifetimes. Suppose we assume a typical lifetime of 70 years. Then spreading those 260 cancers out over 70 years suggests roughly 4 extra cancers per year in the United States. Table 1 tells us there are 521,000 cancer deaths per year, so the 4 extra cancers caused by toxic exposure would be less than 0.001 percent of the nominal rate.

Presenting risks as an annual probability of death to individuals who engage in some activity is a much more specific way to express risks than simply looking at the

TABLE 2

Annual Risks of Death Associated with Certain Activities

Activity/Exposure	Annual Risk (Deaths per 100,000 Persons at Risk)
Motorcycling	2,000
Smoking, all causes	300
Smoking (cancer)	120
Hang gliding	80
Coal mining	63
Farming	36
Motor vehicles	24
Chlorinated drinking water (chloroform)	0.8
4 tbsp peanut butter per day (aflatoxin)	0.8
3 oz charcoal broiled steak per day (PAHs)	0.5
1-in-a-million lifetime risk	0.0014

Source: Based on Wilson and Crouch, 1987.

population as a whole. Table 2 shows risk data for some common activities. For example, the probability of dying in a motorcycle accident among those who ride motorcycles is 2,000 deaths per year per 100,000 motorcyclists. Another example is the risk associated with consuming 4 tablespoons of peanut butter per day. It turns out that mold on peanuts creates a group of chemicals called *aflatoxins* that are known to cause malignant tumors in a number of animals, such as rats, mice, guinea pigs, and monkeys. The Food and Drug Administration (FDA) restricts the aflatoxin concentration in peanut products to 20 ppb (a risk-based decision), and at that level, eating 4 tablespoons of peanut butter per day may cause an estimated 0.8 cancer deaths per year per 100,000. Another interesting item on the list is associated with the polycyclic aromatic hydrocarbons (PAHs) created when food is charbroiled. PAHs are formed when carbon-containing materials are not completely oxidized during combustion, so consumption of burned steak poses some cancer risk. In Table 2 that risk is estimated at 0.5 deaths per year per 100,000 people who eat 3 ounces of broiled steak per day. Notice, for comparison, that the annual risk associated with a one-in-a-million lifetime risk is about 0.0014 per 100,000.

The data in Table 1 are based on actuarial data, so they may be considered accurate, but the data in Tables 2 and 3 are a mix of actuarial values and estimates based on various risk models. Always keep in mind that when risks are based on models, there are generally very large uncertainties in the estimates.

Wilson (1979) provides some perspectives on risk by comparing various activities on the basis of equal, one-in-one-million (10^{-6}) risks. For example, aircraft statistics indicate that 100 billion passenger miles of travel each year in the United States result in about 100 deaths per year; that is, 1 death per billion passenger miles. A trip of 1,000 miles would therefore result in a risk of about 10^{-6}. As another example, Wilson cites statistics on death rates caused by sulfur emissions from coal plants east of the Mississippi. At 20,000 deaths per year among 100 million people exposed to this dirty air, the average risk would be 20,000/100,000,000, or 0.0002 per year of exposure. Two days of breathing this polluted air would pose a

TABLE 3

Activities That Increase Mortality Risk by One in a Million

Activity	Type of Risk
Smoking 1.4 cigarettes	Cancer, heart disease
Drinking 1/2 liter of wine	Cirrhosis of the liver
Spending 1 hour in a coal mine	Black lung disease
Living 2 days in New York or Boston	Air pollution
Traveling 300 miles by car	Accident
Flying 1,000 miles by jet	Accident
Flying 6,000 miles by jet	Cancer by cosmic radiation
Traveling 10 miles by bicycle	Accident
Traveling 6 minutes by canoe	Accident
Living 2 summer months in Denver (vs. sea level)	Cancer by cosmic radiation
Living 2 months with a cigarette smoker	Cancer, heart disease
Eating 40 tablespoons of peanut butter	Liver cancer caused by aflatoxin
Eating 100 charcoal-broiled steaks	Cancer from benzopyrene
Living 50 years within 5 miles of a nuclear reactor	Accident releasing radiation

Source: Wilson, 1979.

TABLE 4

Estimated Expenditures per Life-Year Saved for Selected Programs

Program	1990 U.S.$
Childhood immunizations	Direct savings
Eliminating lead in gasoline	Direct savings
Safety rules at underground construction sites	52,000
Hemodialysis at a dialysis center	56,000
Coronary artery bypass surgery	68,000
Front seat air bags in new cars	109,000
Dioxin effluent controls at paper mills	5,570,000

Source: Kolluru et al., 1996, based on data from the Harvard School of Public Health.

risk of $2/365 \times 0.0002 = 10^{-6}$. Other examples of one-in-a-million risks are given in Table 3. As suggested there, for example, smoking 1.4 cigarettes is equivalent in risk terms to living 50 years within 5 miles of a nuclear power plant. Again, bear in mind that these values are at best rough approximations.

One of the purposes of risk assessment is to provide a starting point in balancing the tradeoffs between an acceptable incremental risk and the cost of controlling risk to that level. Table 4 shows some estimated expenditures to prevent a life from being shortened by one year. Immunizations and phasing out leaded gasoline are indicated to have no cost to them because the direct savings in health care far exceed their cost. Pollution control in the case of lead emissions is very cost-effective, but the table suggests that saving lives by controlling dioxin at a paper mill is very costly indeed.

3 Perception of Risk

Data such as those given in the preceding tables are often used to try to put the health risk associated with pollution into perspective. It usually turns out, however, that the perceptions of risk as seen by an engineer or scientist familiar with the numbers are very different from those of an individual who lives next to a toxic waste site. Social scientists have studied this phenomenon and conclude that there are a number of attributes of risk that can increase the anxiety level of someone evaluating his or her own personal exposures.

For example, people are more likely to be outraged when they have no control of the risks they are exposed to, and they are more fearful of unknown risks than ones they are familiar with. We quite readily take on the risk of crashing into a tree when skiing because it is a voluntary activity, and we are familiar with and understand the risks. We put ourselves at great risk by driving cars, but we feel somewhat in control and believe the risk is worth the benefits. We also accept natural risks such as earthquakes and hurricanes much more readily than unnatural ones, such as the 1984 accidental release of methyl isocyanate from a storage tank in Bhopal, India, which killed 3,400 people. We are also probably more comfortable living next to a gas station, despite the exposure to the carcinogen benzene, than living anywhere near a nuclear power plant, with its perceived unknown and uncertain risks. Table 5 illustrates this notion by comparing attributes that seem to be associated with elevated perceptions of risk.

That these risk attributes can be so important to the public can be a source of frustration in the technical community, but they are real and must be acknowledged by anyone who needs to communicate risk concepts to the public. Finding a way to help people act on the real risks in life and worry less about the minor ones is a difficult challenge.

TABLE 5

Some Characteristics That Change Our Perception of Risk

Attributes That Elevate the Perception	Attributes That Lower the Perception
Involuntary	Voluntary
Exotic	Familiar
Uncontrollable	Controllable
Controlled by others	Controlled by self
Dread	Accept
Catastrophic	Chronic
Caused by humans	Natural
Inequitable	Equitable
Permanent effect	Temporary effect
No apparent benefits	Visible benefits
Unknown	Known
Uncertainty	Certainty
Untrusted source	Trusted source

Source: Based on Slovic, 1987, and Slovic et al., 1980.

4 Risk Assessment

Our concern is with the probability that exposure of some number of people to some combination of chemicals will cause some amount of response, such as cancer, reproductive failure, neurological damage, developmental problems, or birth defects. That is, we want to begin to develop the notions of risk assessment. The National Academy of Sciences (1983) suggests that risk assessment be divided into the following four steps: hazard identification, dose-response assessment, exposure assessment, and risk characterization. After a risk assessment has been completed, the important stage of risk management follows, as shown in Figure 1. The four steps of risk assessment are described here:

- *Hazard identification* is the process of determining whether or not a particular chemical is causally linked to particular health effects, such as cancer or birth defects. Because human data are so often difficult to obtain, this step usually focuses on whether a chemical is toxic in animals or other test organisms.
- *Dose-response assessment* is the process of characterizing the relationship between the dose of an agent administered or received and the incidence of an adverse health effect. Many different dose-response relationships are possible for any given agent depending on such conditions as whether the response is carcinogenic (cancer causing) or noncarcinogenic and whether the experiment is a one-time acute test or a long-term chronic test. Because most tests are performed with high doses, the dose-response assessment must include a consideration for the proper method of extrapolating data to low exposure rates that humans are likely to experience. Part of the assessment must also include a method of extrapolating animal data to humans.

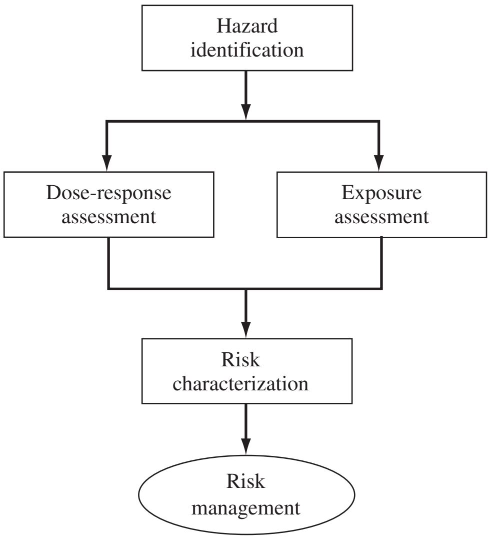


FIGURE 1 Risk assessment is usually considered a four-step process, followed by risk management.

- *Exposure assessment* involves determining the size and nature of the population that has been exposed to the toxicant under consideration and the length of time and toxicant concentration to which they have been exposed. Consideration must be given to such factors as the age and health of the exposed population, smoking history, the likelihood that members of the population might be pregnant, and whether or not synergistic effects might occur due to exposure to multiple toxicants.
- *Risk characterization* is the integration of the foregoing three steps, which results in an estimate of the magnitude of the public-health problem.

5 Hazard Identification

The first step in a risk analysis is to determine whether or not the chemicals that a population has been exposed to are likely to have any adverse health effects. This is the work of toxicologists who study both the nature of the adverse effects caused by toxic agents as well as the probability of their occurrence. We start our description of this hazard identification process by summarizing the pathways that a chemical may take as it passes through a human body and the kinds of damage that may result. A simple diagram of the human circulatory system is shown in Figure 2, which identifies some of the principal organs and the nomenclature for toxic effects.

A toxicant can enter the body using any of three pathways: by ingestion with food or drink, through inhalation, or by contact with the skin (dermal) or other exterior surfaces, such as the eyes. Once in the body, it can be absorbed by the blood and distributed to various organs and systems. The toxicant may then be stored (for example, in fat, as in the case of DDT), or it may be eliminated from the body by excretion or by transformation into something else. The biotransformation process usually yields metabolites that can be more readily eliminated from the body than the original chemicals; however, metabolism can also convert chemicals to more toxic forms. Figure 3 presents the most important movements of chemical toxicants in the body, showing absorption, distribution, storage, and excretion. Although these are shown as separate operations, they all occur simultaneously.

Several organs are especially vulnerable to toxicants. The liver, for example, which filters the blood before it is pumped through the lungs, is often the target. Since toxics are transported by the bloodstream, and the liver is exposed to so much of the blood supply, it can be directly damaged by toxics. Moreover, since a major function of the liver is to metabolize substances, converting them into forms that can be excreted more easily from the body, it is also susceptible to chemical attack by toxic chemicals formed during the biotransformation process itself. Chemicals that can cause liver damage are called *hepatotoxins*. Examples of hepatotoxic agents include a number of synthetic organic compounds, such as carbon tetrachloride (CCl_4), chloroform ($CHCl_3$), and trichloroethylene (C_2HCl_3); pesticides, such as DDT and paraquat; heavy metals, such as arsenic, iron, and manganese; and drugs, such as acetaminophen and anabolic steroids. The kidneys also filter the blood, and they too are frequently susceptible to damage.

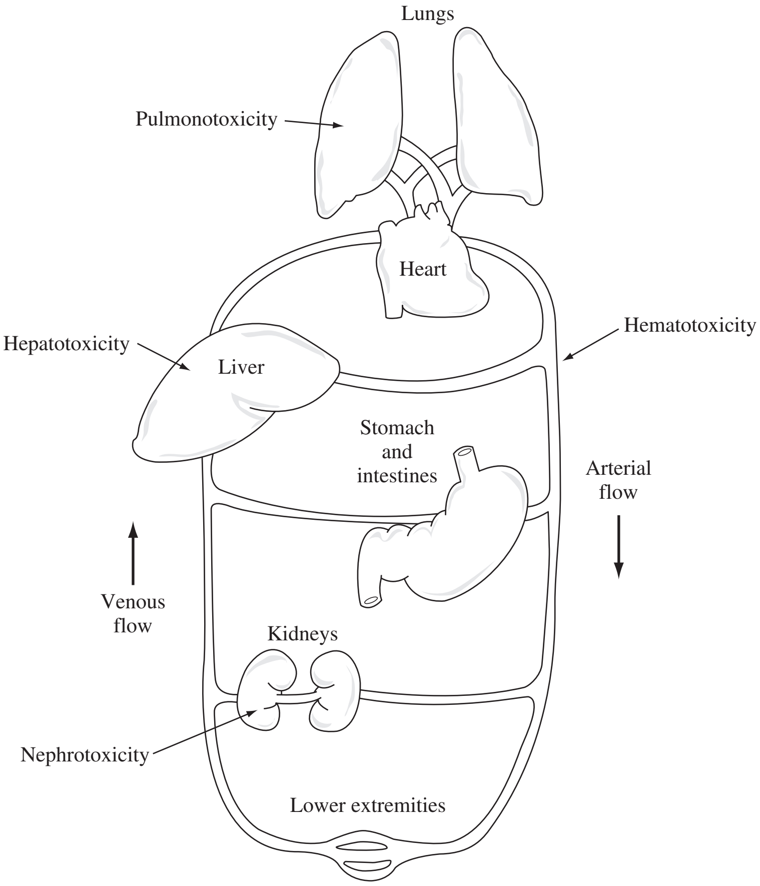


FIGURE 2 The circulatory system and nomenclature for toxic effects: hepatotoxicity (liver), nephrotoxicity (kidneys), pulmonotoxicity (lungs), hematotoxicity (blood). (*Source:* Based on James, 1985.)

Toxic chemicals often injure other organs and organ systems as well. The function of the kidneys is to filter blood to remove wastes that will be excreted in the form of urine. Toxicants that damage the kidneys, called *nephrotoxics,* include metals such as cadmium, mercury, and lead, as well as a number of chlorinated hydrocarbons. Excessive kidney damage can decrease or stop the flow of urine, causing death by poisoning from the body's own waste products. *Hematotoxicity* is the term used to describe the toxic effects of substances on the blood. Some hematotoxins, such as carbon monoxide in polluted air and nitrates in groundwater, affect the ability of blood to transport oxygen to the tissues. Other toxicants, such as benzene, affect the formation of platelets, which are necessary for blood clotting. The lungs and skin, due to their proximity to pollutants, are also often affected by chemical toxicants. Lung function can be impaired by such substances as cigarette smoke, ozone, asbestos, and quartz rock dust. The skin reacts in a variety of ways to chemical toxicants, but the most common and serious environmentally related skin problem is cancer induced by excessive ultraviolet radiation.

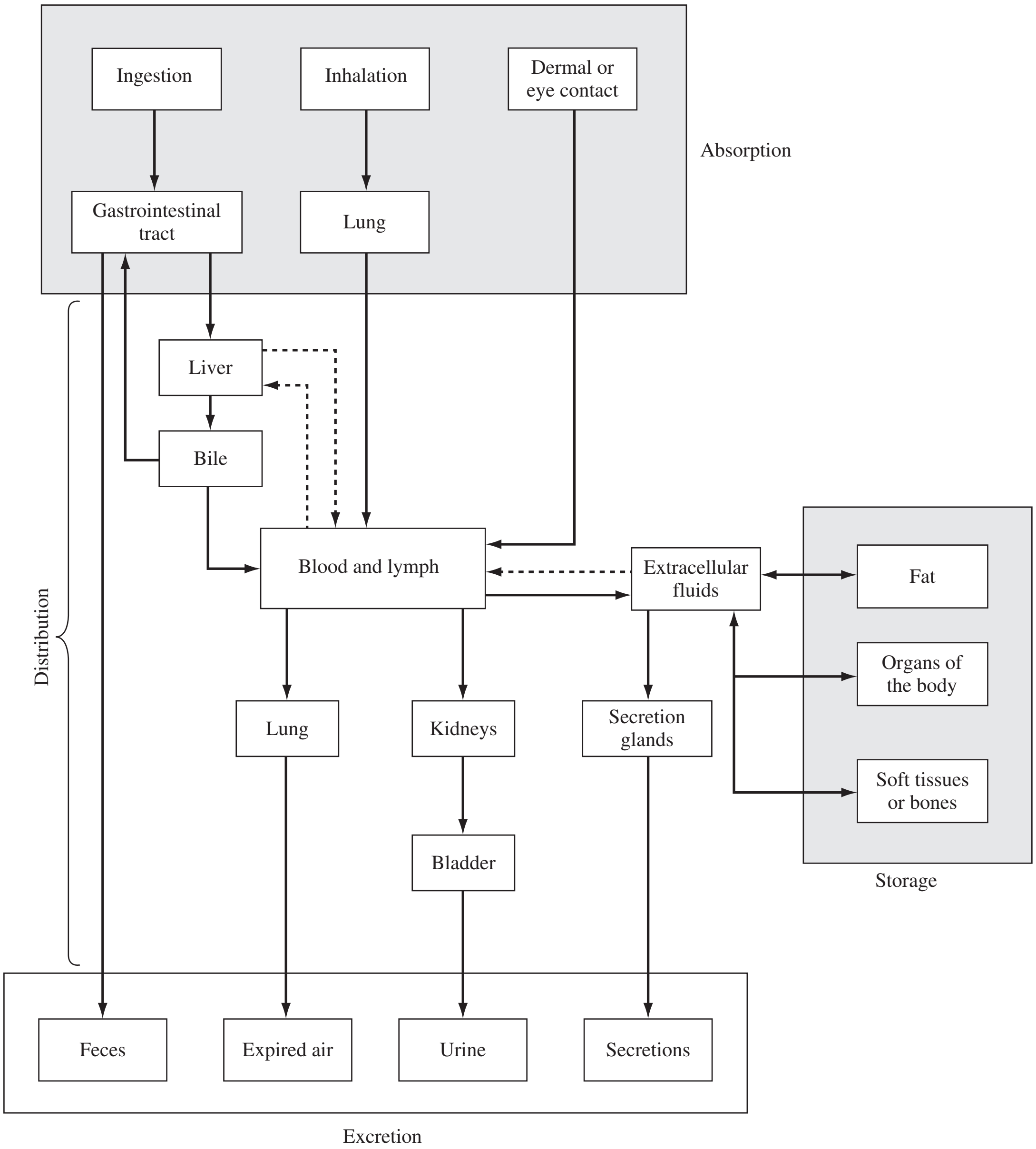


FIGURE 3 Fate of chemical toxicants in the body.
(*Source:* Environ, 1988.)

Acute Toxicity

This chapter began with a quote by Philippus Aureolus Theophrastus Bombastus von Hohenheim-Paracelsus to the effect that it is the dose that makes the poison. One measure of the toxicity of something is the amount needed to cause some acute response, such as organ injury, coma, or even death. Acute toxicity refers to effects that are caused within a short period of time after a single exposure to the chemical;

later we will discuss chronic toxicity effects that take place after prolonged exposure periods.

One way to describe the toxicity of chemicals is by the amount that is required to kill the organism. Table 6 shows a conventional toxicity rating scheme that expresses the dose in terms of milligrams of chemical ingested per kilogram of body weight. That is, ingestion of a given amount of toxin will be more dangerous for a small person, such as a child, than a larger adult. Normalizing the dose using body weight is the first step in trying to relate a lethal dose to a laboratory animal to what might be expected in a human. For example, it takes on the order of 20,000 mg of ordinary sucrose per kilogram to kill a rat. Using the rating system in Table 6, sucrose would be considered practically nontoxic. If we scale up that dose to a 70-kg human (without any other adjustments), it might take something like 1.4 kg of sucrose (3 pounds) ingested all at one time to be lethal. At the other extreme, the bacteria *Clostridium botulinum,* responsible for botulism (food poisoning), is lethal with a single dose of only 0.00001 mg/kg, so it is supertoxic (Rodricks, 1992).

Not every member of an exposed population will react the same way to a toxin, so one way to illustrate the variation is with a dose-response curve that shows the percentage of a population that is affected as a function of the dose received. In the dose-response curves of Figure 4, a logarithmic scale for dose is shown, which

TABLE 6

A Conventional Rating System for the Acute Toxicity of Chemicals in Humans

	Probable Lethal Oral Dose for Humans	
Toxicity Rating	Dose (mg/kg of Body Weight)	For Average Adult
1. Practically nontoxic	More than 15,000	More than 1 quart
2. Slightly toxic	5,000–15,000	1 pint to 1 quart
3. Moderately toxic	500–5,000	1 ounce to 1 pint
4. Very toxic	50–500	1 teaspoon to 1 ounce
5. Extremely toxic	5–50	7 drops to 1 teaspoon
6. Supertoxic	Less than 5	Less than 7 drops

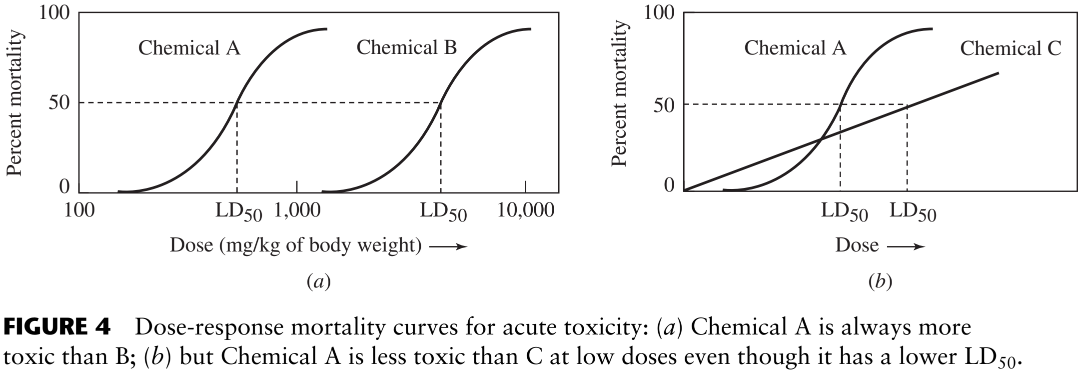


FIGURE 4 Dose-response mortality curves for acute toxicity: (*a*) Chemical A is always more toxic than B; (*b*) but Chemical A is less toxic than C at low doses even though it has a lower LD_{50}.

tends to yield the familiar S-shaped curve. Also notice that the dose is expressed as milligrams of chemical ingested per kilogram of body weight. Normalizing with body weight allows the dose to be extrapolated to individuals of different sizes, such as a child versus an adult. Also, it provides a first cut at extrapolating the likely effects on a human when the dose-response curve has been generated using animal tests.

The curves of Figure 4 show the response to chemical exposure as a mortality rate. The dose that will kill 50 percent of a population is designated LD_{50}, where LD stands for lethal dose. In Figure 4*a*, the dose-response curves for two chemicals are shown. Chemical A has a lower LD_{50} than Chemical B, and it is always more toxic. Figure 4*b* warns us to be aware that just because one chemical has a lower LD_{50} than another does not necessarily mean it is always more toxic. Chemical A has a lower LD_{50}, which would normally suggest it is more toxic than C, but notice it is not as toxic as C at low doses. So the dose-response curve does provide more information than a simple table of LD_{50} doses.

Mutagenesis

In contrast to the short-term responses associated with acute toxicity, most risk assessments are focused on responses that may take years to develop. Measuring the ability of specific chemicals to cause cancer, reproductive failure, and birth defects is much more difficult than the acute toxicity testing just described.

Deoxyribonucleic acid (DNA) is an essential component of all living things and a basic material in the chromosomes of the cell nucleus. It contains the genetic code that determines the overall character and appearance of every organism. Each molecule of DNA has the capability to replicate itself exactly, transmitting that genetic information to new cells. Our interest here in DNA results from the fact that certain chemical agents, as well as ionizing radiation, are *genotoxic;* that is, they are capable of altering DNA. Such changes, or *mutations,* in the genetic material of an organism can cause cells to misfunction, leading in some cases to cell death, cancer, reproductive failure, or abnormal offspring. Chemicals that are capable of causing cancer are called *carcinogens;* chemicals that can cause birth defects are *teratogens.*

Mutations may affect somatic cells, which are the cells that make up the tissues and organs of the body itself, or they may cause changes in germ cells (sperm or ovum) that may be transmitted to future offspring. As is suggested in Figure 5, one possible outcome of a mutagenic event is the death of the cell itself. If the mutation is in a somatic cell and it survives, the change may be such that the cell no longer responds to signals that normally control cell reproduction. If that occurs, the cell may undergo rapid and uncontrolled cellular division, forming a tumor. Mutations in somatic cells may damage or kill the affected individual, and if the individual is a pregnant female, the embryo may be damaged, leading to a birth defect. Germ cell mutations, on the other hand, have the potential to become established in the gene pool and be transmitted to future generations.

Carcinogenesis

Cancer is second only to heart disease in terms of the number of Americans killed every year. In 2006, nearly 1.4 million new cancer cases were diagnosed in the

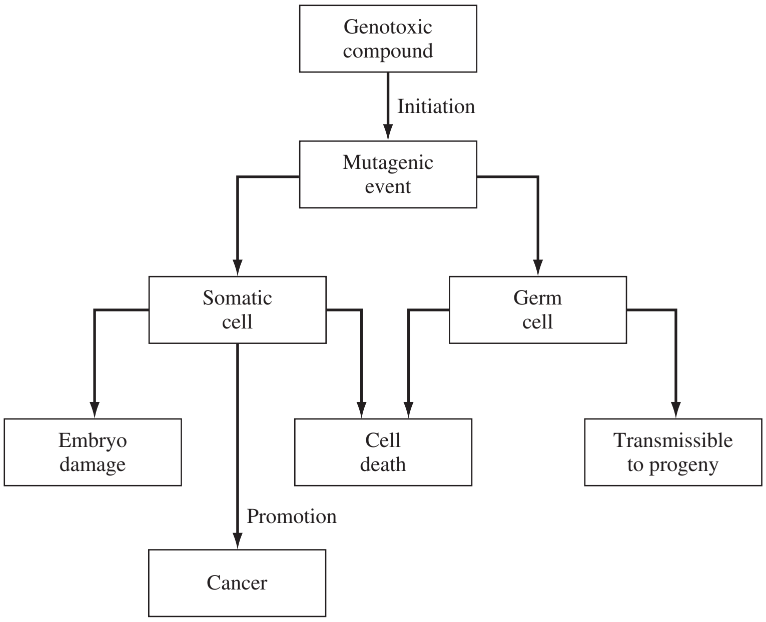


FIGURE 5 Possible consequences of a mutagenic event in somatic and germinal cells.

United States, and over one-half million people die each year from cancer. Cancer is truly one of the most dreaded diseases.

Chemically induced carcinogenesis is thought to involve two distinct stages, referred to as *initiation* and *promotion.* In the initiation stage, a mutation alters a cell's genetic material in a way that may or may not result in the uncontrolled growth of cells that characterizes cancer. In the second, or promotion, stage of development, affected cells no longer recognize growth constraints that normally apply, and a tumor develops. Promoters can increase the incidence rate of tumors among cells that have already undergone initiation, or they can shorten the latency period between initiation and the full carcinogenic response. The model of initiation followed by promotion suggests that some carcinogens may be initiators, others may be promoters, and some may be complete carcinogens capable of causing both stages to occur. Current regulations do not make this distinction, however, and any substance capable of increasing the incidence of tumors is considered a carcinogen, subject to the same risk assessment techniques. Tumors, in turn, may be *benign* or *malignant* depending on whether or not the tumor is contained within its own boundaries. If a tumor undergoes *metastasis*—that is, it breaks apart and portions of it enter other areas of the body—it is said to be malignant. After a tumor has metastasized, it is obviously much harder to treat or remove.

The theoretical possibility that a single genotoxic event can lead to a tumor is referred to as the *one-hit hypothesis.* Based on this hypothesis, exposure to even the smallest amount of a carcinogen leads to some nonzero probability that a malignancy will result. That is, in a conservative, worst-case risk assessment for carcinogens, it is assumed that there is no threshold dose below which the risk is zero.

A brief glossary of carcinogenesis terminology is presented in Table 7.

TABLE 7

Glossary of Carcinogenesis Terminology

Acute toxicity	Adverse effects caused by a toxic agent occurring within a short period of time following exposure
Benign tumor	A new tumor composed of cells that, though proliferating in an abnormal manner, do not spread to surrounding, normal tissue
Cancer	An abnormal process in which cells begin a phase of uncontrolled growth and spread
Carcinogen	Any cancer-producing substance
Carcinoma	A malignant tumor in the tissue that covers internal or external surfaces of the body such as the stomach, liver, or skin
Chronic toxicity	Adverse effects caused by a toxic agent after a long period of exposure
Initiator	A chemical that initiates the change in a cell that irreversibly converts the cell into a cancerous or precancerous state
Malignant tumor	Relatively autonomous growth of cells or tissue that invade surrounding tissue and have the capability to metastasize
Mutagenesis	Alteration of DNA in either somatic or germinal cells not associated with the normal process of recombination
Mutation	A permanent, transmissible change in DNA that changes the function or behavior of the cell
Neoplasm	Literally, new growth, usually of an abnormally fast-growing tissue
Oncogenic	Giving rise to tumors or causing tumor formation
Pharmacokinetics	The study of how a chemical is absorbed, distributed, metabolized, and excreted
Promoter	A chemical that can increase the incidence of response to a carcinogen previously administered
Sarcoma	A cancer that arises from mesodermal tissue (e.g., fat, muscle, bone)
Teratogen	Any substance capable of causing malformation during development of the fetus
Toxicity	A relative term generally used in comparing the harmful effect of one chemical on some biological mechanism with the effect of another chemical

Source: Based on Williams and Burson, 1985.

Toxicity Testing in Animals

With several thousand new chemicals coming onto the market each year, a backlog of tens of thousands of relatively untested chemicals already in commerce, and a limited number of facilities capable of providing the complex testing that might be desired, it is just not possible to test each and every chemical for its toxicity. As a result, a hierarchy of testing procedures has been developed that can be used to help select those chemicals that are most likely to pose serious risks.

The starting point is the relatively straightforward acute toxicity testing already described. The next step may be to compare the structure of the chemical in question with other chemicals that are known or suspected to be human carcinogens, such as those shown in Figure 6. New chemicals that are similar to these, and other suspected carcinogens, would be potential candidates for further testing.

The prevailing carcinogenesis theory, that human cancers are initiated by gene mutations, has led to the development of short-term, in vitro (in glassware) screening procedures, which are one of the first steps taken to determine whether a chemical is

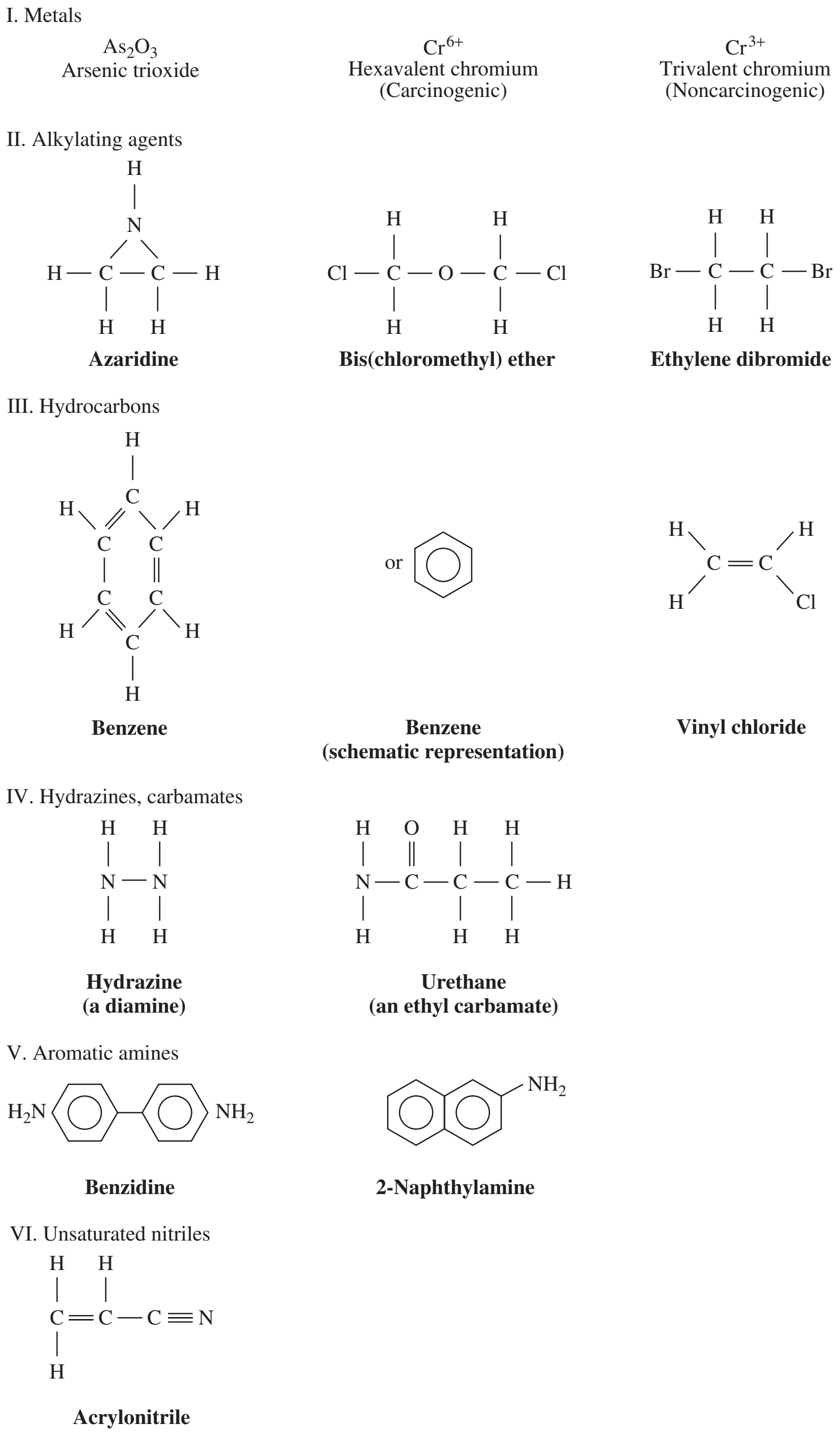


FIGURE 6 Selected structural formulae for some classes of carcinogenic chemicals. (*Source:* Williams and Burson, 1985.)

carcinogenic. It is thought that if a chemical can be shown to be mutagenic, then it *may* be carcinogenic, and further testing may be called for. The most widely used short-term test, called the *Ames mutagenicity assay,* subjects special tester strains of bacteria to the chemical in question. These tester strains have previously been rendered incapable of normal bacterial division, so unless they mutate back to a form that is capable of division, they will die. Bacteria that survive and form colonies do so through mutation; therefore, the greater the survival rate of these special bacteria, the more mutagenic is the chemical.

Intermediate testing procedures involve relatively short-term (several months duration) carcinogenesis bioassays in which specific organs in mice and rats are subjected to known mutagens to determine whether tumors develop.

Finally, the most costly, complex, and long-lasting test, called a *chronic carcinogenesis bioassay,* involves hundreds or thousands of animals over a time period of several years. To assure comparable test results and verifiable data, the National Toxicology Program in the United States has established minimum test requirements for an acceptable chronic bioassay, which include the following:

- *Two species of rodents must be tested.* Mice and rats, using specially inbred strains for consistency, are most often used. They have relatively short lifetimes, and their small size makes them easier to test in large numbers.
- *At least 50 males and 50 females of each species for each dose must be tested.* Many more animals are required if the test is to be sensitive enough to detect risks of less than a few percent.
- *At least two doses must be administered (plus a no-dose control).* One dose is traditionally set at the maximum tolerated dose (MTD), a level that can be administered for a major portion of an animal's lifetime without significantly impairing growth or shortening the lifetime. The second dose is usually one-half or one-fourth the MTD.

Exposure begins at 6 weeks of age and ends when the animal reaches 24 months of age. At the end of the test, all animals are killed and their remains are subjected to detailed pathological examinations. These tests are expensive as well as time consuming. Testing a typical new chemical costs between \$500,000 and \$1.5 million, takes up to 2 or 3 years, and may entail the sacrifice of thousands of animals (Goldberg and Frazier, 1989).

Notice that, following the aforementioned protocol, the minimum number of animals required for a bioassay is 600 (2 species × 100 animals × 3 doses), and at that number, it is still only relatively high risks that can be detected. With this number of animals, for the test to show a statistically significant effect, the exposed animals must have at least 5 or 10 percent more tumors than the controls in order to conclude that the extra tumors were caused by the chemical being tested. That is, the risk associated with this chemical can be measured only down to roughly 0.05 or 0.10 unless we test a lot more animals.

A simple example may help to clarify this statistical phenomenon. Suppose we test 100 rats at a particular dose and find one tumor. To keep it easy, let's say the control group never gets tumors. Can the actual probability (risk) of tumors caused by this chemical at this dose be 1 percent? Yes, definitely. If the risk is 1 percent, we would expect to get one tumor, which is what we got. Could the actual probability be 2 percent? Well, if the actual risk is 2 percent, and *if we were able to run the test over* *and over again* on sets of 100 rats each, some of those groups would have no tumors, some would certainly have 1 tumor, and some would have more. So our actual test of only 1 group of 100, which found 1 tumor, is not at all inconsistent with an actual risk of 2 percent. Could the actual risk be 3 percent? Running many sets of 100 rats through the test would likely result in at least 1 of those groups having only 1 tumor. So it would not be out of the question to find 1 tumor in a single group of 100 rats even if the actual risk is 3 percent. Getting back to the original test of 100 rats and finding 1 tumor, we have just argued that the actual risk could be anything from 0 percent to, say, 2 or 3 percent, maybe even more, and still be consistent with finding just 1 tumor. We certainly cannot conclude that the risk is only 1 percent. In other words, with 100 animals, we cannot perform a statistically significant test and be justified in concluding that the risk is anything less than a few percent. Bioassays designed to detect lower risks require many thousands of animals. In fact, the largest experiment ever performed involved more than 24,000 mice and yet was still insufficiently sensitive to measure a risk of less than 1 percent (Environ, 1988).

The inability of a bioassay to detect small risks presents one of the greatest difficulties in applying the data so obtained to human risk assessment. Regulators try to restrict human risks due to exposure to carcinogens to levels of about 10^{-6} (one in a million), yet animal studies are only capable of detecting risks of down to 0.01 to 0.1. It is necessary, therefore, to find some way to extrapolate the data taken for animals exposed to high doses to humans who will be exposed to doses that are several orders of magnitude lower.

Human Studies

Another shortcoming in the animal testing methods just described, besides the necessity to extrapolate the data toward zero risk, is the obvious difficulty in interpreting the data for humans. How does the fact that some substance causes tumors in mice relate to the likelihood that it will cause cancer in humans as well? Animal testing can always be criticized in this way, but since we are not inclined to perform the same sorts of tests directly on humans, other methods must be used to gather evidence of human toxicity.

Sometimes human data can be obtained by studying victims of tragedies, such as the chemical plant leak that killed and injured thousands in Bhopal, India, and the atomic bombing of Hiroshima and Nagasaki, Japan. The most important source of human risk information, however, comes from epidemiologic studies. Epidemiology is the study of the incidence rate of diseases in real populations. By attempting to find correlations between disease rates and various environmental factors, an epidemiologist attempts to show in a quantitative way the relationship between exposure and risk. Such data can be used to complement animal data, clinical data, and scientific analyses of the characteristics of the substances in question.

Epidemiologists have a number of strategies for gathering useful information, but they share the common feature of trying to identify two populations of people having different exposures to the risk factor being studied. Preliminary data analysis usually involves setting up a simple 2×2 matrix such as the one shown in Figure 7. The rows divide the populations according to those who have, and those who have not, been exposed to the risk factor. The columns are based on the numbers of individuals who have acquired the disease being studied and those who have not.

	With disease	Without disease
Exposed	*a*	*b*
Not exposed	*c*	*d*

FIGURE 7 A 2 × 2 matrix for an epidemiologic rate comparison. Rows divide people by exposure; columns divide them by disease.

Various measures can be applied to the data given in Figure 7 to see whether they *suggest* an association between exposure and disease.

- The *relative risk* is defined as

$$\text{Relative risk} = \frac{a/(a + b)}{c/(c + d)} \tag{1}$$

  Notice that the numerator is the fraction of those exposed who have the disease, and the denominator is the fraction of those not exposed who have the disease. If those two ratios are the same, the odds of having the disease would not depend on whether an individual had been exposed to the risk factor, and the relative risk would be 1.0. Above 1.0, the higher the relative risk, and the more the data suggests an association between exposure and risk.

- The *attributable risk* is defined as

$$\text{Attributable risk} = \frac{a}{a + b} - \frac{c}{c + d} \tag{2}$$

  The attributable risk is the difference between the odds of having the disease with exposure and the odds of having the disease without exposure. An attributable risk of 0.0 suggests no relationship between exposure and risk.

- The *odds ratio* is defined as the cross product of the entries in the matrix:

$$\text{Odds ratio} = \frac{ad}{bc} \tag{3}$$

  The odds ratio is similar to the relative risk. Numbers above 1.0 suggest a relationship between exposure and risk.

EXAMPLE 1 Epidemiologic Data Analysis

An evaluation of personnel records for employees of a plant that manufactures vinyl chloride finds that out of 200 workers, 15 developed liver cancer. A control group consisting of individuals with smoking histories similar to the exposed workers, and who were unlikely to have encountered vinyl chloride, had 24 with liver cancers and 450 who did not develop liver cancer. Find the relative risk, attributable risk, and odds ratio for these data.

Solution Putting the data into the 2 × 2 matrix gives

$$\begin{array}{c|c|c|}
 & D & \overline{D} \\
\hline
E & 15 & 185 \\
\hline
\overline{E} & 24 & 450 \\
\hline
\end{array}$$

Solving for each measure yields:

$$\text{Relative risk} = \frac{15/(15 + 185)}{24/(24 + 450)} = \frac{.075}{0.05} = 1.48$$

$$\text{Attributable risk} = \frac{15}{200} - \frac{24}{474} = 0.024$$

$$\text{Odds ratio} = \frac{15 \times 450}{185 \times 24} = 1.52$$

The relative risk and the odds ratio both are above 1.0, so they suggest a relationship between exposure and risk. For those who were exposed, the risk of cancer has increased by 0.024 (the attributable risk) over that of their cohorts who were not exposed. All three measures indicate that further study of the relationship between vinyl chloride and liver cancer would be warranted.

Caution must be exercised in interpreting every epidemiologic study since any number of confounding variables may lead to invalid conclusions. For example, the study may be biased because workers are compared with nonworkers (workers are generally healthier), or because relative rates of smoking have not been accounted for, or there may be other variables that are not even hypothesized in the study that may be the actual causal agent. As an example of the latter, consider an attempt to compare lung cancer rates in a city having high ambient air pollution levels with rates in a city having less pollution. Suppose the rates are higher in the more polluted city, even after accounting for smoking history, age distribution, and working background. To conclude that ambient air pollution is causing those differences may be totally invalid. Instead, it might well be different levels of radon in homes, or differences in other indoor air pollutants associated with the type of fuel used for cooking and heating, that are causing the cancer variations.

Weight-of-Evidence Categories for Potential Carcinogens

Based on the accumulated evidence from clinical studies, epidemiologic evidence, in vitro studies, and animal data, the EPA uses the following categories to describe the likelihood that a chemical substance is carcinogenic (U.S. EPA, 1986a). Using both human and animal data, five categories, A through E, have been established as follows:

Group A: Human carcinogen. A substance is put into this category only when sufficient epidemiologic evidence supports a causal association between exposure to the agent and cancer.

TABLE 8

Weight-of-Evidence Categories for Human Carcinogenicity

	Animal Evidence				
Human Evidence	Sufficient	Limited	Inadequate	No Data	No Evidence
Sufficient	A	A	A	A	A
Limited	B1	B1	B1	B1	B1
Inadequate	B2	C	D	D	D
No data	B2	C	D	D	E
No evidence	B2	C	D	D	E

Source: U.S. EPA, 1986a.

Group B: Probable human carcinogen. This group is actually made up of two subgroups. An agent is categorized as B1 if there is limited epidemiologic evidence; and an agent is put into B2 if there is inadequate human data but sufficient evidence of carcinogenicity in animals.

Group C: Possible human carcinogen. This group is used for agents with limited evidence of carcinogenicity in animals and an absence of human data.

Group D: Not classified. This group is for agents with inadequate human and animal evidence or for which no data are available.

Group E: Evidence of noncarcinogenicity. This group is used for agents that show no evidence for carcinogenicity in at least two adequate animal tests in different species or in both adequate epidemiologic and animal studies.

Table 8 summarizes this categorization scheme.

6 Dose-Response Assessment

As the name suggests, the fundamental goal of a dose-response assessment is to obtain a mathematical relationship between the amount of a toxicant that a human is exposed to and the risk that there will be an unhealthy response to that dose. We have seen dose-response curves for acute toxicity, in which the dose is measured in milligrams per kilogram of body weight. The dose-response curves that we are interested in here are the result of chronic toxicity; that is, the organism is subjected to a prolonged exposure over a considerable fraction of its life. For these curves, the abscissa is dose, which is usually expressed as the average milligrams of substance per kilogram of body weight per day (mg/kg-day). The dose is an exposure averaged over an entire lifetime (for humans, assumed to be 70 years). The ordinate is the response, which is the risk that there will be some adverse health effect. As usual, response (risk) has no units; it is a probability that there will be some adverse health effect. For example, if prolonged exposure to some chemical would be expected to produce 700 cancers in a population of 1 million, the response could be expressed as 0.0007, 7×10^{-4}, or 0.07 percent. The annual risk would be obtained by spreading that risk over an assumed 70-year lifetime, giving a risk of 0.00001 or 1×10^{-5} per year.

For substances that induce a carcinogenic response, it is always conventional practice to assume that exposure to any amount of the carcinogen will create some

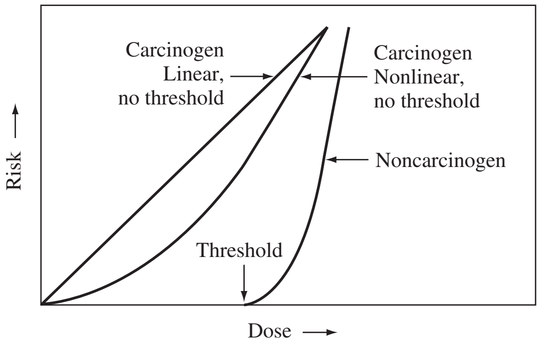


FIGURE 8 Hypothetical dose-response curves. Dose-response curves for carcinogens are assumed to have no threshold; that is, any exposure produces some chance of causing cancer.

likelihood of cancer. That is, a plot of response versus dose is required to go through the origin. For noncarcinogenic responses, it is usually assumed that there is some *threshold dose,* below which there will be no response. As a result of these two assumptions, the dose-response curves and the methods used to apply them are quite different for carcinogenic and noncarcinogenic effects, as suggested in Figure 8. The same chemical, by the way, may be capable of causing both kinds of response.

To apply dose-response data obtained from animal bioassays to humans, a *scaling factor* must be introduced. Sometimes the scaling factor is based on the assumption that doses are equivalent if the dose per unit of body weight in the animal and human is the same. Sometimes, especially if the exposure is dermal, equivalent doses are normalized to body surface area rather than body weight when scaling up from animal to human. In either case, the resulting human dose-response curve is specified with the standard mg/kg-day units for dose. Adjustments between animal response and human response may also have to be made to account for differences in the rates of chemical absorption. If enough is known about the differences between the absorption rates in test animals and in humans for the particular substance in question, it is possible to account for those differences later in the risk assessment. Usually, however, there is insufficient data, and it is simply assumed that the absorption rates are the same.

Extrapolations from High Doses to Low Doses

The most controversial aspect of dose-response curves for carcinogens is the method chosen to extrapolate from the high doses actually administered to test animals to the low doses to which humans are likely to be exposed. Recall that even with extremely large numbers of animals in a bioassay, the lowest risks that can be measured are usually a few percent. Since regulators attempt to control human risk to several orders of magnitude less than that, there will be no actual animal data anywhere near the range of most interest.

Many mathematical models have been proposed for the extrapolation to low doses. Unfortunately, no model can be proved or disproved from the data, so there is no way to know which model is the most accurate. That means the choice of models is strictly a policy decision. One commonly used model is called the *one-hit* *model,* in which the relationship between dose (d) and lifetime risk (probability) of cancer, $P(d)$, is given in the following form (Crump, 1984):

$$P(d) = 1 - e^{-(q_0 + q_1 d)} \tag{4}$$

where q_0 and q_1 are parameters picked to fit the data. The one-hit model corresponds to the simplest mechanistic model of carcinogenesis, in which it is assumed that a single chemical hit is capable of inducing a tumor.

If we substitute $d = 0$ into (1), the result will be an expression for the background rate of cancer incidence, $P(0)$. Using the mathematical expansion for an exponential

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} \cong 1 + x \quad \text{(for small } x\text{)} \tag{5}$$

and assuming that the background cancer rate is small allows us to write

$$P(0) = 1 - e^{-q_0} \cong 1 - [1 + (-q_0)] = q_0 \tag{6}$$

That is, the background rate for cancer incidence corresponds to the parameter q_0. Using the exponential expansion again, the one-hit model suggests that the lifetime probability of cancer for small dose rates can be expressed as

$$P(d) \cong 1 - [1 - (q_0 + q_1 d)] = q_0 + q_1 d = P(0) + q_1 d \tag{7}$$

For low doses, the additional risk of cancer above the background rate is

$$\text{Additional risk} = A(d) = P(d) - P(0) \tag{8}$$

Substituting (7) into (8) yields the following equation for the additional cancer risk incurred when the organism in question is exposed to a dose d:

$$\text{Additional risk} = A(d) \cong q_1 d \tag{9}$$

That is, the one-hit model predicts that for low doses, the extra lifetime probability of cancer is linearly related to dose.

The one-hit model relating risk to dose is not the only one possible. Another mathematical model that has been proposed has its roots in the multistage model of tumor formation; that is, that tumors are the result of a sequence of biological events (Crump, 1984). The *multistage* model expresses the relationship between risk and dose as

$$P(d) = 1 - e^{-(q_0 + q_1 d + q_2 d^2 + \cdots q_n d^n)} \tag{10}$$

where the individual parameters q_i are positive constants picked to best fit the dose-response data. Again, it is easy to show that for small values of dose d, the multistage model also has the simplifying feature of producing a linear relationship between additional risk and dose. Figure 9 illustrates the use of a one-hit model and a multistage model to fit experimental data. The multistage model will always fit the data better since it includes the one-hit model as a special case.

Since the choice of an appropriate low-dose model is not based on experimental data, no model can be proved to be more correct than another. To protect public health, EPA chooses to err on the side of safety and overemphasize risk. The EPA's model of choice is a modified multistage model, called the *linearized multistage model.* It is linear at low doses with the constant of proportionality picked in a way that the probability of overestimating the risk is 95 percent.

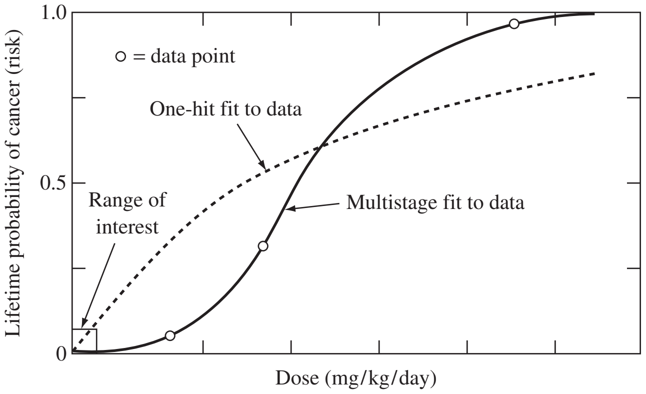


FIGURE 9 Dose-response curves showing two methods of fitting an equation to the data. The range of interest is well below the point where any data actually exist.
(*Source:* Based on Crump, 1984.)

Potency Factor for Carcinogens

For chronic toxicity studies, a low dose is administered over a significant portion of the animal's lifetime. The resulting dose-response curve has the incremental risk of cancer (above the background rate) on the *y*-axis, and the lifetime average daily dose of toxicant along the *x*-axis. At low doses, where the dose-response curve is assumed to be linear, the slope of the dose-response curve is called the *potency factor* (PF), or *slope factor*.

$$\text{Potency factor} = \frac{\text{Incremental lifetime cancer risk}}{\text{Chronic daily intake (mg/kg-day)}} \tag{11}$$

The denominator in (11) is the dose averaged over an entire lifetime; it has units of average milligrams of toxicant absorbed per kilogram of body weight per day, which is usually written as (mg/kg-day) or (mg/kg/day). Since risk has no units, the units for potency factor are therefore $(\text{mg/kg-day})^{-1}$.

If we have a dose-response curve, we can find the potency factor from the slope. In fact, one interpretation of the potency factor is that it is the risk produced by a chronic daily intake of 1 mg/kg-day, as shown in Figure 10.

Rearranging (11) shows us where we are headed. If we know the chronic daily intake CDI (based on exposure data) and the PF (from EPA), we can find the lifetime, incremental cancer risk from

$$\text{Incremental lifetime cancer risk} = \text{CDI} \times \text{PF} \tag{12}$$

The linearized multistage risk-response model assumptions built into (12) should make this value an upper-bound estimate of the actual risk. Moreover, (12) estimates the risk of getting cancer, which is not necessarily the same as the risk of dying of cancer, so it should be even more conservative as an upper-bound estimate of cancer death rates.

Potency factors needed for (12) can be found in an EPA database on toxic substances called the Integrated Risk Information System (IRIS). Included in the

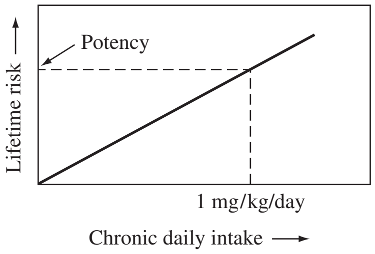


FIGURE 10 The potency factor is the slope of the dose-response curve. It can also be thought of as the risk that corresponds to a chronic daily intake of 1 mg/kg-day.

TABLE 9

Toxicity Data for Selected Potential Carcinogens

Chemical	Category	Potency Factor Oral Route $(\text{mg/kg-day})^{-1}$	Potency Factor Inhalation Route $(\text{mg/kg-day})^{-1}$
Arsenic	A	1.75	50
Benzene	A	2.9×10^{-2}	2.9×10^{-2}
Benzol(a)pyrene	B2	11.5	6.11
Cadmium	B1	—	6.1
Carbon tetrachloride	B2	0.13	—
Chloroform	B2	6.1×10^{-3}	8.1×10^{-2}
Chromium VI	A	—	41
DDT	B2	0.34	—
1,1-Dichloroethylene	C	0.58	1.16
Dieldrin	B2	30	—
Heptachlor	B2	3.4	—
Hexachloroethane	C	1.4×10^{-2}	—
Methylene chloride	B2	7.5×10^{-3}	1.4×10^{-2}
Nickel and compounds	A	—	1.19
Polychlorinated biphenyls (PCBs)	B2	7.7	—
2,3,7,8-TCDD (dioxin)	B2	1.56×10^{5}	—
Tetrachloroethylene	B2	5.1×10^{-2}	$1.0 - 3.3 \times 10^{-3}$
1,1,1-Trichloroethane (1,1,1-TCA)	D	—	—
Trichloroethylene (TCE)	B2	1.1×10^{-2}	1.3×10^{-2}
Vinyl chloride	A	2.3	0.295

Source: U.S. EPA, www.epa.gov/iris.

rather extensive background information on each potential carcinogen in IRIS is the potency factor and the weight-of-evidence category (recall Table 8). A short list of some of these chemicals, PFs (for both oral and inhalation exposure routes), and cancer categories is given in Table 9.

The other factor we need to develop more fully in order to use the basic risk equation (12) is the concept of chronic daily intake. The CDI is, by definition,

$$\text{CDI (mg/kg-day)} = \frac{\text{Average daily dose (mg/day)}}{\text{Body weight (kg)}} \tag{13}$$

The numerator in (13) is the total lifetime dose averaged over an assumed 70-year lifetime. The next example shows how to combine CDI and potency to find risk.

EXAMPLE 2 Risk Assessment for Chloroform in Drinking Water

When drinking water is disinfected with chlorine, an undesired byproduct, chloroform ($CHCl_3$), may be formed. Suppose a 70-kg person drinks 2 L of water every day for 70 years with a chloroform concentration of 0.10 mg/L (the drinking water standard).

a. Find the upper-bound cancer risk for this individual.
b. If a city with 500,000 people in it also drinks the same amount of this water, how many extra cancers per year would be expected? Assume the standard 70-year lifetime.
c. Compare the extra cancers per year caused by chloroform in the drinking water with the expected number of cancer deaths from all causes. The cancer death rate in the United States is 189 per 100,000 per year.

Solution

a. From Table 9, we see that chloroform is a Class B2 probable human carcinogen with a potency factor of 6.1×10^{-3} (mg/kg-day)$^{-1}$. Using (13), the chronic daily intake is

$$\text{CDI (mg/kg-day)} = \frac{\text{Average daily dose (mg/day)}}{\text{Body weight (kg)}}$$

$$= \frac{0.10 \text{ mg/L} \times 2 \text{ L/day}}{70 \text{ kg}} = 0.00286 \text{ mg/kg-day}$$

From (12), the incremental lifetime cancer risk is

$$\text{Risk} = \text{CDI} \times \text{PF}$$

$$= 0.00286 \text{ (mg/kg-day)} \times 6.1 \times 10^{-3} \text{ (mg/kg-day)}^{-1} = 17.4 \times 10^{-6}$$

So, over a 70-year period, the upper-bound estimate of the probability that a person will get cancer from this drinking water is about 17 in 1 million.

b. If there are 17.4 cancers per million people over a 70-year period, then in any given year in a population of 500,000, the number of cancers caused by chloroform would be

$$500{,}000 \text{ people} \times \frac{17.4 \text{ cancer}}{10^6 \text{ people}} \times \frac{1}{70 \text{ yr}} = 0.12 \text{ cancers/yr}$$

c. The total number of cancer deaths that would be expected in a city of 500,000 would be

$$500{,}000 \text{ people} \times \frac{189 \text{ cancer/yr}}{100{,}000 \text{ people}} = 945 \text{ cancer deaths/yr}$$

It would seem that an additional 0.12 new cancers per year would not be detectable.

Once again, it is necessary to emphasize that the science behind a risk assessment calculation of the sort demonstrated in Example 2 is primitive, and enormous uncertainties are associated with any particular answer so computed. There is still great value, however, to this sort of procedure since it does organize a mass of data into a format that can be communicated to a much wider audience, and it can greatly help that audience find legitimate perspectives based on that data. For example, it matters little whether the annual extra cancers associated with chloroform in the preceding example are found to be 0.12 or even ten times as much, 1.2; the conclusion that the extra cancers would be undetectable would not change.

Another use for these risk calculations is to estimate the concentration of a contaminant in drinking water that would result in a politically acceptable risk level. Often that risk goal is 10^{-6}, and the concentration that will produce that risk is called the *drinking water equivalent level* (DWEL). To find the DWEL, it is usually assumed that a 70-kg adult consumes 2 L of water per day. As Example 3 shows, we can find the DWEL from the potency factor using (12).

EXAMPLE 3 Drinking Water Concentration of Chloroform for a 10^{-6} Risk

Find the concentration of chloroform in drinking water that would result in a 10^{-6} risk for a 70-kg person who drinks 2 L/day throughout his or her entire lifetime.

Solution Rearranging (4.12) and using the PF from Table 9 gives

$$\text{CDI} = \frac{\text{Risk}}{\text{Potency factor}} = \frac{10^{-6}}{6.1 \times 10^{-3}\ (\text{kg-day/mg})} = 1.64 \times 10^{-4}\ (\text{mg/kg-day})$$

Since CDI is just the average daily intake divided by body mass, we can write

$$\text{CDI} = \frac{C(\text{mg/L}) \times 2\ \text{L/day}}{70\ \text{kg}} = 1.64 \times 10^{-4}\ (\text{mg/kg-day})$$

where C(mg/L) is the allowable concentration of chloroform. Solving for C gives

$$C = 70 \times 1.64 \times 10^{-4}/2 = 0.0057 \cong 6 \times 10^{-3}\ \text{mg/L} = 6\ \mu\text{g/L}$$

So a DWEL of 6 μg/L for chloroform would result in an upper-bound risk of 10^{-6}.

In Examples 2 and 3 it was assumed that everyone drinks 2 L of contaminated water every day for 70 years. When a risk assessment is made for exposures that do not last the entire lifetime, we need to develop the chronic daily intake a little more carefully.

If the contaminant is in drinking water, the CDI can be expressed as

$$\text{CDI} = \frac{\text{Concentration (mg/L)} \times \text{Intake rate (L/day)} \times \text{Exposure (days/life)}}{\text{Body weight (kg)} \times 70\ \text{(yr/life)} \times 365\ \text{(days/yr)}} \tag{14}$$

where *Concentration* refers to the contaminant concentration, *Intake rate* is the amount of water ingested each day, and *Exposure* is the number of days in a lifetime that the person drinks contaminated water.

If the exposure route is inhalation of a contaminant, the chronic daily intake can be expressed as

$$\text{CDI} = \frac{\text{Concentration (mg/m}^3\text{)} \times \text{Intake rate (m}^3\text{/day)} \times \text{Exposure (days/life)}}{\text{Body weight (kg)} \times 70 \text{ (yr/life)} \times 365 \text{ (days/yr)}} \qquad (15)$$

where *Concentration* is the contaminant concentration in air, and the *Intake rate* is the amount of air inhaled during each day that the person is exposed to the contamination. Similar expressions can be used for consumption of contaminated food or soil and for dermal contact with contaminated soil.

EXAMPLE 4 **An Occupational Exposure**

Estimate the incremental cancer risk for a 60-kg worker exposed to a particular carcinogen under the following circumstances. Exposure time is 5 days per week, 50 weeks per year, over a 25-year period of time. The worker is assumed to breathe 20 m^3 of air per day. The carcinogen has a potency factor of 0.02 $(\text{mg/kg-day})^{-1}$, and its average concentration is 0.05 mg/m^3.

Solution Since this is an inhalation exposure, we will use (15).

$$\text{CDI} = \frac{0.05 \text{ mg/m}^3 \times 20 \text{ m}^3\text{/day} \times 5 \text{ days/wk} \times 50 \text{ wk/yr} \times 25 \text{ yr}}{60 \text{ kg} \times 70 \text{ yr/life} \times 365 \text{ days/yr}}$$

$$= 0.0041 \text{ mg/kg-day}$$

Using (12), the upper-bound, incremental cancer risk is CDI × potency:

$$\text{Incremental risk} = 0.0041 \text{ mg/kg-day} \times 0.02 \text{ (mg/kg-day)}^{-1} = 81 \times 10^{-6}$$

which is considerably higher than the usual goal of 10^{-6} risk.

The EPA has developed a set of recommended default values for daily intakes, exposures, and body weights to be used in risk calculations when more site-specific information is not available. Table 10 shows some of these default factors, and the next example illustrates their use.

EXAMPLE 5 **A Proposed Source of Benzene in Your Neighborhood**

Suppose an industrial facility that emits benzene into the atmosphere is being proposed for a site near a residential neighborhood. Air quality models predict that 60 percent of the time, prevailing winds will blow the benzene away from the neighborhood, but 40 percent of the time the benzene concentration will be 0.01 mg/m^3. Use standard exposure factors from Table 10 to assess the incremental risk to adults in the neighborhood if the facility is allowed to be built. If the acceptable risk is 10^{-6}, should this plant be allowed to be built?

Solution Using factors from Table 10, the chronic daily intake will be

$$\text{CDI} = \frac{0.01 \text{ mg/m}^3 \times 20 \text{ m}^3\text{/day} \times 350 \text{ day/yr} \times 30 \text{ yr}}{70 \text{ kg} \times 365 \text{ day/yr} \times 70 \text{ yr}} \times 0.40$$

$$= 0.00047 \text{ mg/kg-day}$$

The potency factor from Table 9 for benzene is 2.9×10^{-2} (mg/kg-day)$^{-1}$, so the incremental risk would be

$$\begin{aligned}\text{Incremental risk} &= 0.00047 \text{ mg/kg-day} \times 2.9 \times 10^{-2} (\text{mg/kg-day})^{-1} \\ &= 1.3 \times 10^{-5}\end{aligned}$$

The risk is higher than the acceptable level, so the facility should not be built as it is being proposed.

The Reference Dose for Noncarcinogenic Effects

The key assumption for noncarcinogens is that there is an exposure threshold; that is, any exposure less than the threshold would be expected to show no increase in adverse effects above natural background rates. One of the principal goals of toxicant testing is therefore to identify and quantify such thresholds. Unfortunately, for the usual case, inadequate data are available to establish such thresholds with any degree of certainty and, as a result, it has been necessary to introduce a number of special assumptions and definitions.

Suppose a precise threshold exists for some particular toxicant for some particular animal species. To determine the threshold experimentally, we might imagine a testing program in which animals would be exposed to a range of doses. Doses below the threshold would elicit no response; doses above the threshold would produce responses. The lowest dose administered that results in a response is given a special name: the *lowest-observed-effect level* (LOEL). Conversely, the highest dose

TABLE 10

Example EPA Exposure Factors Recommended for Risk Assessments

Land Use	Exposure Pathway	Daily Intake	Exposure Frequency, Days/Year	Exposure Duration, Years	Body Weight, kg
Residential	Ingestion of potable water	2 L (adult) 1 L (child)	350	30	70 (adult) 15 (child)
	Ingestion of soil and dust	200 mg (child) 100 mg (adult)	350	6 24	15 (child) 70 (adult)
	Inhalation of contaminants	20 m^3 (adult) 12 m^3 (child)	350	30	70
Industrial and commercial	Ingestion of potable water	1 L	250	25	70
	Ingestion of soil and dust	50 mg	250	25	70
	Inhalation of contaminants	20 m^3 (workday)	250	25	70
Agricultural	Consumption of homegrown produce	42 g (fruit) 80 g (veg.)	350	30	70
Recreational	Consumption of locally caught fish	54 g	350	30	70

Source: U.S. EPA, 1991.

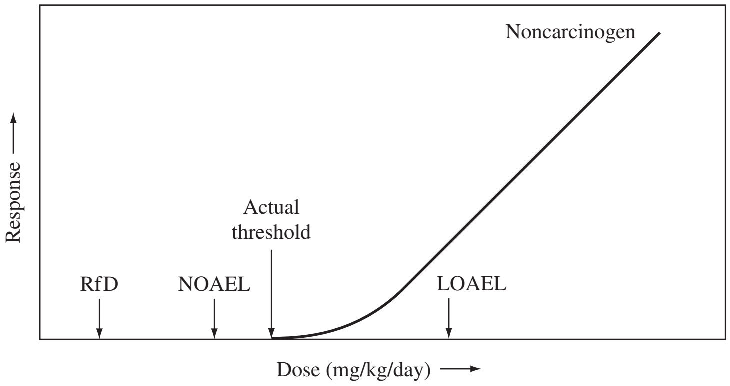


FIGURE 11 Nomenclature used to characterize response to different doses of a hypothetical noncarcinogen. The reference dose RfD is the no-observed-adverse-effects-level (NOAEL) divided by an uncertainty factor typically between 10 and 1,000.

administered that does not create a response is called the *no-observed-effect level* (NOEL). NOELs and LOELs are often further refined by noting a distinction between effects that are *adverse* to health and effects that are not. Thus, there are also *no-observed-adverse-effect levels* (NOAELs) and *lowest-observed-adverse-effect levels* (LOAELs).

Figure 11 illustrates these levels and introduces another exposure called the *reference dose,* or RfD. The RfD used to be called the *acceptable daily intake* (ADI), and as that name implies, it is intended to give an indication of a level of human exposure that is likely to be without appreciable risk. The units of RfD are mg/kg-day averaged over a lifetime, just as they were for the chronic daily intake CDI. The RfD is obtained by dividing the NOAEL by an appropriate *uncertainty factor* (sometimes called a safety factor). A 10-fold uncertainty factor is used to account for differences in sensitivity between the most sensitive individuals in an exposed human population, such as pregnant women, babies, and the elderly, and "normal, healthy" people. Another factor of 10 is introduced when the NOAEL is based on animal data that is to be extrapolated to humans. Finally, another factor of 10 is sometimes applied when there are no good human data, and the animal data available are limited. Thus, depending on the strength of the available data, human RfD levels are established at doses that are anywhere from one-tenth to one-thousandth of the NOAEL, which is itself somewhat below the actual threshold. Table 11 gives a short list of some commonly encountered toxicants and their RfDs.

The Hazard Index for Noncarcinogenic Effects

Since the reference dose RfD is established at what is intended to be a safe level, well below the level at which any adverse health effects have been observed, it makes sense to compare the actual exposure to the RfD to see whether the actual dose is supposedly safe. The hazard quotient is based on that concept:

$$\text{Hazard quotient} = \frac{\text{Average daily dose during exposure period (mg/kg-day)}}{\text{RfD}} \tag{16}$$

TABLE 11

Oral RfDs for Chronic Noncarcinogenic Effects of Selected Chemicals

Chemical	RfD (mg/kg-day)
Acetone	0.100
Arsenic	0.0003
Cadmium	0.0005
Chloroform	0.010
1,1-Dichloroethylene	0.009
cis-1,2-Dichloroethylene	0.010
Fluoride	0.120
Mercury (inorganic)	0.0003
Methylene chloride	0.060
Phenol	0.600
Tetrachloroethylene	0.010
Toluene	0.200
1,1,1-Trichloroethane	0.035
Xylene	2.000

Source: U.S. EPA, www.epa.gov/iris.

Notice that the daily dose is averaged only over the period of exposure, which is different from the average daily dose used in risk calculations for carcinogens. For noncarcinogens, the toxicity is important only during the time of exposure. Recall that for a cancer risk calculation (e.g., Eq. 13), the dose is averaged over an assumed 70-year lifetime.

The hazard quotient has been defined so that if it is less than 1.0, there should be no significant risk of systemic toxicity. Ratios above 1.0 could represent a potential risk, but there is no way to establish that risk with any certainty.

When exposure involves more than one chemical, the sum of the individual hazard quotients for each chemical is used as a measure of the potential for harm. This sum is called the *hazard index:*

$$\text{Hazard index} = \text{Sum of the hazard quotients} \tag{17}$$

EXAMPLE 6 Hazard Index

Suppose drinking water contains 1.0 mg/L of toluene and 0.01 mg/L of tetrachloroethylene (C_2Cl_4). A 70-kg adult drinks 2 L per day of this water for 10 years.

a. Does the hazard index suggest that this was a safe level of exposure?

b. Tetrachloroethylene is a B2 carcinogen. What is the carcinogenic risk faced by someone drinking this water? Is it less than a goal of 10^{-6}?

Solution

a. First we need to find the average daily doses (ADDs) for each of the chemicals and then their individual hazard quotients.

For toluene, the RfD is given in Table 11 as 0.200 mg/kg-day, so

$$\text{ADD (toluene)} = \frac{1.0\text{ mg/L} \times 2\text{ L/day}}{70\text{ kg}} = 0.029\text{ mg/kg-day}$$

$$\text{Hazard quotient (toluene)} = \frac{0.029\text{ mg/kg-day}}{0.200\text{ mg/kg-day}} = 0.14$$

The RfD for tetrachloroethylene is 0.01 mg/kg-day, so

$$\text{ADD } (C_2Cl_4) = \frac{0.01\text{ mg/L} \times 2\text{ L/day}}{70\text{ kg}} = 0.00029\text{ mg/kg-day}$$

$$\text{Hazard quotient } (C_2Cl_4) = \frac{0.00029\text{ mg/kg-day}}{0.01\text{ mg/kg-day}} = 0.029$$

So

$$\text{Hazard index} = 0.14 + 0.029 = 0.17 < 1.0$$

The hazard index suggests that this water is safe. By the way, notice that we did not need to know that the person drank this water for 10 years.

b. The incremental carcinogenic risk associated with the C_2Cl_4 is

$$\text{Risk} = \text{CDI} \times \text{PF}$$

$$\text{CDI} = \frac{0.01\text{ mg/L} \times 2\text{ L/day} \times 365\text{ days/yr} \times 10\text{ yrs}}{70\text{ kg} \times 365\text{ days/yr} \times 70\text{ yrs}}$$

$$= 4.0 \times 10^{-5}\text{ mg/kg-day}$$

From Table 9, the oral potency is 5.1×10^{-2} (mg/kg-day)$^{-1}$, so the risk is

$$\begin{aligned}\text{Risk} &= \text{CDI} \times \text{PF} \\ &= 4.0 \times 10^{-5}\text{ mg/kg-day} \times 5.1 \times 10^{-2}\text{ (mg/kg-day)}^{-1} = 2 \times 10^{-6}\end{aligned}$$

So, from a cancer risk standpoint, this water does not meet the 10^{-6} risk goal. Notice how the tetrachloroethylene was way below the RfD but was above the desired risk goal. This is not uncommon when the hazard index is computed for carcinogens.

7 Human Exposure Assessment

One of the most elementary concepts of risk assessment is all too often overlooked in public discussions: that risk has two components—the toxicity of the substance involved, and the amount of exposure to that substance. Unless individuals are exposed to the toxicants, there is no human risk.

A human exposure assessment is itself a two-part process. First, pathways that allow toxic agents to be transported from the source to the point of contact with people must be evaluated. Second, an estimate must be made of the amount of contact that is likely to occur between people and those contaminants. Figure 12 suggests some of the transport mechanisms that are common at a toxic waste site.

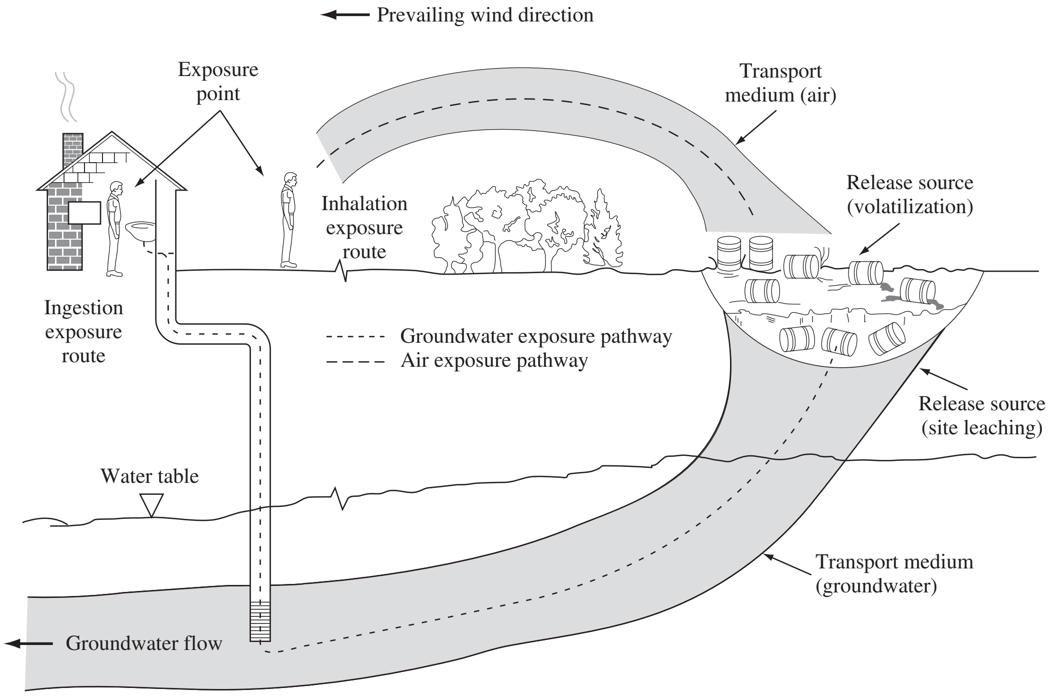


FIGURE 12 Illustration of exposure pathways.
(*Source:* U.S. EPA, 1986b.)

Substances that are exposed to the atmosphere may volatilize and be transported with the prevailing winds (in which case, plume models are often used). Substances in contact with soil may leach into groundwater and eventually be transported to local drinking water wells. As pollutants are transported from one place to another, they may undergo various transformations that can change their toxicity and/or concentration. Many of these fate and transport pathways for pollutants will be covered later in this book. A useful summary of exposure pathway models that the EPA uses is given in the *Superfund Exposure Assessment Manual* (U.S. EPA, 1988).

After the exposure pathways have been analyzed, an estimate of the concentrations of toxicants in the air, water, soil, and food at a particular exposure point can be made. With the concentrations of various toxic agents established, the second half of an exposure assessment begins. Human contact with those contaminants must be estimated. Necessary information includes number of people exposed, duration of exposure, and amounts of contaminated air, water, food, and soil that find their way into each exposed person's body. Often, the human intake estimates are based on a lifetime of exposure, assuming standard, recommended daily values of amounts of air breathed, water consumed, and body weight, such as given earlier in Table 10. In some circumstances, the exposure may be intermittent, and adjustments might need to be made for various body weights, rates of absorption, and exposure periods, as illustrated previously in Example 4.

Bioconcentration

One potentially important exposure route is human consumption of contaminated fish. It is relatively straightforward to estimate concentrations of contaminants in water, and it also is reasonable to make estimates of consumption of fish that individuals may consume (for example, the standard intake values given in Table 10). What is more difficult is the task of estimating the concentration of a contaminant in fish, given only the chemical concentration in water. The *bioconcentration factor* (BCF) provides the key link measuring the tendency for a substance to accumulate in fish tissue. The equilibrium concentration of a chemical in fish can be estimated by multiplying the chemical concentration in water by the bioconcentration factor:

$$\text{Concentration in fish} = (\text{concentration in water}) \times (\text{bioconcentration factor}) \quad (18)$$

The units of BCF (L/kg) are picked to allow the concentration of substance in water to be the usual mg/L, and the concentration in fish to be milligrams of substance per kilogram of fish. Some example values of BCF are given in Table 12. Note the high bioconcentration factors for chlorinated hydrocarbon pesticides, such as chlordane, DDT, and heptachlor, and the especially high concentration factor for polychlorinated biphenyls (PCBs). These high bioconcentration factors played an important role in the decision to reduce or eliminate the use of these chemicals in the United States.

TABLE 12

Bioconcentration Factors (BCFs) for a Selected List of Chemicals

Chemical	Bioconcentration Factor (L/kg)
Aldrin	28
Arsenic and compounds	44
Benzene	5.2
Cadmium and compounds	81
Carbon tetrachloride	19
Chlordane	14,000
Chloroform	3.75
Chromium III, VI, and compounds	16
Copper	200
DDE	51,000
DDT	54,000
1,1-Dichloroethylene	5.6
Dieldrin	4760
Formaldehyde	0
Heptachlor	15,700
Hexachloroethane	87
Nickel and compounds	47
Polychlorinated biphenyls (PCBs)	100,000
2,3,7,8-TCDD (dioxin)	5,000
Tetrachloroethylene	31
1,1,1-Trichloroethane	5.6
Trichloroethylene (TCE)	10.6
Vinyl chloride	1.17

Source: U.S. EPA, 1986b.

Example 7 illustrates the use of bioconcentration factors in a carcinogenic risk assessment.

EXAMPLE 7 Bioconcentration of TCE

Using the standard exposure factors in Table 10 for a person eating locally caught fish, estimate the lifetime cancer risk from fish taken from waters containing a concentration of trichloroethylene (TCE) equal to 100 ppb (0.1 mg/L).

Solution In Table 12, the bioconcentration factor for TCE is given as 10.6 L/kg. From (18) the expected concentration of TCE in fish is therefore

$$\text{TCE concentration} = 0.1\text{ mg/L} \times 10.6\text{ L/kg} = 1.06\text{ mg TCE/kg fish}$$

From Table 10, standard exposure factors include a 70-kg person consuming 54 g of fish, 350 days per year for 30 years. The chronic daily intake CDI is thus

$$\text{CDI} = \frac{0.054\text{ kg/day} \times 1.06\text{ mg TCE/kg} \times 350\text{ days/yr} \times 30\text{ yrs}}{70\text{ kg} \times 365\text{ days/yr} \times 70\text{ yrs}}$$

$$= 3.36 \times 10^{-4}\text{ mg/kg-day}$$

From Table 9, the cancer potency factor for an oral dose of TCE is 1.1×10^{-2} (mg/kg-day)$^{-1}$. Using (12), the upper-bound, incremental lifetime risk of cancer is

$$\text{Risk} = \text{CDI} \times \text{PF}$$

$$= 3.36 \times 10^{-4}\text{ mg/kg-day} \times 1.1 \times 10^{-2}\text{ (mg/kg-day)}^{-1} = 3.6 \times 10^{-6}$$

or about 4 in 1 million.

Contaminant Degradation

Many toxic chemicals of concern are nonconservative; that is, they degrade with time. Degradation may be the result of a number of processes that remove pollutants from the medium in which they reside. There may be phase transfer as a chemical volatilizes, chemical transformation if it reacts with other substances, or biological transformation if it is consumed by microorganisms. The persistence of a chemical as it moves through various environmental media may be affected by some combination of these mechanisms. A convenient way to deal with such complexity is simply to combine the degradation processes into a single, overall *half-life*. The half-life of a given substance will depend on whether it appears in soil, air, surface water, or groundwater. Some representative half-lives are given in Table 13.

If the concentration of a substance is modeled with a simple exponential decay relationship,

$$C(t) = C(0)e^{-kt} \tag{19}$$

TABLE 13

Range of Half-Lives (days) of Various Contaminants in Air and Surface Water

	Air		Surface Water	
Chemical	Low	High	Low	High
Benzene	6	—	1	6
Benzo(a)pyrene	1	6	0.4	—
Carbon tetrachloride	8,030	—	0.3	300
Chlordane	40	—	420	500
Chloroform	80	—	0.3	30
DDT	—	—	56	110
1,1-Dichloroethane	45	—	1	5
Formaldehyde	0.8	—	0.9	3.5
Heptachlor	40	—	0.96	—
Hexachloroethane	7,900	—	1.1	9.5
Polychlorinated biphenyls (PCBs)	58	—	2	12.9
2,3,7,8-TCDD (dioxin)	—	—	365	730
1,1,1-Trichloroethane	803	1,752	0.14	7
Trichloroethylene	3.7	—	1	90
Vinyl chloride	1.2	—	1	5

Source: U.S. EPA, 1986b.

then the time required for the concentration to be decreased by 50 percent is the half-life, given by

$$t_{1/2} = \frac{\ln 2}{k} = \frac{0.693}{k} \tag{20}$$

An example of how to use half-lives is given in the following example.

EXAMPLE 8 A Leaking Underground Storage Tank Exposure Assessment

Suppose an underground storage tank has been leaking for many years, contaminating the groundwater and causing a contaminant concentration directly beneath the site of 0.30 mg/L. The contamination is flowing at the rate of 0.5 ft per day toward a public drinking water well 1 mile away. The half-life of the contaminant is 10 years.

a. Estimate the steady-state pollutant concentration expected at the well.
b. If the potency factor for the contaminant is 0.02 $(\text{mg/kg-day})^{-1}$, estimate the cancer risk if a 70-kg person drank 2 L of this water per day for 10 years.

Solution

a. The time required to travel to the well 1 mile away is

$$\text{Time to well} = \frac{5{,}280 \text{ ft}}{0.5 \text{ ft/day}} = 10{,}560 \text{ days}$$

The pollutant is assumed to degrade exponentially, so the reaction rate coefficient *k* can be found using (20):

$$k = \frac{0.693}{t_{1/2}} = \frac{0.693}{10\text{ yr} \times 365\text{ days/yr}} = 1.9 \times 10^{-4}\text{/day}$$

In the 10,560 days required to travel to the drinking water well, (4.19) suggests that the initial 0.30 mg/L is reduced to

$$C(t) = C(0)e^{-kt} = 0.30e^{-(1.9\times10^{-4}/d\times10{,}560\ d)} = 0.040\text{ mg/L}$$

b. The CDI for someone drinking this water for 10 years out of a 70-year lifetime is

$$\text{CDI} = \frac{0.040\text{ mg/L} \times 2\text{ L/day} \times 10\text{ yr}}{70\text{ kg} \times 70\text{ yr}} = 1.6 \times 10^{-4}\text{ mg/kg-day}$$

So the lifetime cancer risk would be

$$\begin{aligned}\text{Risk} &= \text{CDI} \times \text{PF}\\ &= 1.6 \times 10^{-4}\text{ mg/kg-day} \times 0.020\text{ (mg/kg-day)}^{-1} = 3.2 \times 10^{-6}\end{aligned}$$

This is probably an upper-bound estimate of the individual risk and is subject to all of the uncertainties that currently characterize all risk assessments.

8 Risk Characterization

The final step in a risk assessment is to bring the various studies together into an overall risk characterization. In its most primitive sense, this step could be interpreted to mean simply multiplying the exposure (dose) by the potency to get individual risk, and then multiplying that by the number of people exposed to get an estimate of overall risk to some specific population.

Although there are obvious advantages to presenting a simple, single number for extra cancers, or some other risk measure, a proper characterization of risk should be much more comprehensive. The final expressions of risk derived in this step will be used by regulatory decision makers in the process of weighing health risks against other societal costs and benefits. The public will also use these expressions to help them decide on the adequacy of proposed measures to manage the risks. Both groups need to appreciate the extraordinary leaps of faith that, by necessity, have had to be used to determine these simple quantitative estimates. It must always be emphasized that these estimates are preliminary, subject to change, and extremely uncertain.

The National Academy of Sciences (1983) suggests a number of questions that should be addressed in a final characterization of risk, including the following:

- What are the statistical uncertainties in estimating the extent of health effects? How are these uncertainties to be computed and presented?

- What are the biological uncertainties? What are their origins? How will they be estimated? What effect do they have on quantitative estimates? How will the uncertainties be described to agency decision makers?
- Which dose-response assessments and exposure assessments should be used?
- Which population groups should be the primary targets for protection, and which provide the most meaningful expression of the health risk?

Rodricks (page 181, 1992) offers the following example of the sort of qualifying statement that ought to accompany all risk assessments (in this case for a hypothetical contaminant difluoromuckone, DFM):

> *Difluoromuckone (DFM) has been found to increase the risk of cancer in several studies involving experimental animals. Investigations involving groups of individuals exposed in the past to relatively high levels of DFM have not revealed that the chemical increases cancer risk in humans. Because these human studies could not detect a small increase in risk, and because there is a scientific basis for assuming results from animal experiments are relevant to humans, exposure to low levels of DFM may create an increase in risk of cancer for people. The magnitude of this risk is unknown, but probably does not exceed one in 50,000. This figure is the lifetime chance of developing cancer from a daily exposure to the highest levels of DFM detected in the environment. Average levels, which are more likely to be experienced over the course of a lifetime, suggest a lifetime risk more like one in 200,000. These risk figures were derived using scientific assumptions that are not recognized as plausible by all scientists, but which are consistently used by regulatory scientists when attempting to portray the risks of environmental chemicals. It is quite plausible that actual risks are lower than the ones cited above; higher risks are not likely but cannot be ruled out. Regulators typically seek to reduce risks that exceed a range of one in 100,000 to one in 1,000,000. Note that the lifetime cancer risk we face from all sources of these diseases is about 1 in 5 (1 in 10 for nonsmokers), so that, even if correct, the DFM risk is a minor contributor to the overall cancer problem. Prudence may dictate the need for some small degree of risk reduction for DFM in the environment.*

9 Comparative Risk Analysis

In 1987, the EPA released a report entitled *Unfinished Business: A Comparative Assessment of Environmental Problems* (U.S. EPA, 1987), in which the concepts of risk assessment were applied to a variety of pressing environmental problems. The goal of the study was to attempt to use risk as a policy tool for ranking major environmental problems in order to help the agency establish broad, long-term priorities.

At the outset, it was realized that direct comparisons of different environmental problems would be next to impossible. Not only are the data usually insufficient to quantify risks, but the kinds of risk associated with some problems, such as global warming, are virtually incomparable with risks of others, such as hazardous waste. In most cases, considerable professional judgment rather than hard data was required to finalize their rankings. In spite of difficulties such as these, the report is noteworthy both in terms of its methodology and its conclusions.

The study was organized around a list of 31 environmental problems including topics as diverse as conventional (criteria) air pollutants, indoor radon, stratospheric ozone depletion, global warming, active (RCRA) and inactive (Superfund) hazardous waste sites, damage to wetlands, mining wastes, and pesticide residues on foods. Each of these 31 problems was analyzed by 4 separate working groups from the perspective of 4 different types of risk: cancer risks, noncancer health risks, ecological effects, and welfare effects (visibility impairment, materials damage, etc.).

The ranking of cancer risks was perhaps the most straightforward part of the study because EPA already has established risk assessment procedures, and there are considerable data already available from which to work. Rankings were based primarily on overall cancer risk to the entire U.S. population, although high risks to specific groups of individuals such as farm workers were noted. Ordinal rankings were given, but it was emphasized that these should not be interpreted as being precise, especially when similarly ranked problems are being compared. Given all of the uncertainties, in the cancer working group's final judgment, two problems were tied at the top of the list: (1) worker exposure to chemicals, which doesn't involve a large number of individuals but does result in high individual risks to those exposed; and (2) indoor radon exposure, which is causing significant risk to a large number of people. Inactive (Superfund) hazardous waste sites ranked 8th and active (RCRA) hazardous waste sites were 13th. Interestingly, it was noted that with the exception of pesticide residues on food, the major route of exposure for carcinogens is inhalation. Their final ranking of the top 10 carcinogenic risks is reproduced in Table 14.

The other working groups had considerably greater difficulty ranking the 31 environmental problem areas since there are no accepted guidelines for quantitatively assessing relative risks. As noted in *Unfinished Business,* a perusal of the rankings of the 31 problem areas for each of the 4 types of risk (cancer, noncancer health effects, ecological, and welfare effects) produced the following general results:

- No problems rank relatively high in all four types of risk, or relatively low in all four.
- Problems that rank relatively high in three of the four risk types, or at least medium in all four, include criteria air pollutants, stratospheric ozone depletion, pesticide residues on food, and other pesticide risks (runoff and air deposition of pesticides).
- Problems that rank relatively high in cancer and noncancer health risks, but low in ecological and welfare risks include hazardous air pollutants, indoor radon, indoor air pollution other than radon, pesticide application, exposure to consumer products, and worker exposures to chemicals.
- Problems that rank relatively high in ecological and welfare risks, but low in both health risks, include global warming, point and nonpoint sources of surface water pollution, and physical alteration of aquatic habitats (including estuaries and wetlands) and mining wastes.
- Areas related to groundwater consistently rank medium or low.

TABLE 14

Consensus Ranking of Environmental Problem Areas on the Basis of Population Cancer Risk

Rank	Problem Area	Selected Comments
1 (tied)	Worker exposure to chemicals	About 250 cancer cases per year estimated based on exposure to 4 chemicals; but workers face potential exposures to more than 20,000 substances. Very high individual risk possible.
1 (tied)	Indoor radon	Estimated 5,000 to 20,000 lung cancers annually from exposure in homes.
3	Pesticide residues on foods	Estimated 6,000 cancers annually, based on exposure to 200 potential oncogens.
4 (tied)	Indoor air pollutants (nonradon)	Estimated 3,500–6,500 cancers annually, mostly due to tobacco smoke.
4 (tied)	Consumer exposure to chemicals	Risk from 4 chemicals investigated is about 100–135 cancers annually; an estimated 10,000 chemicals in consumer products. Cleaning fluids, pesticides, particleboard, and asbestos-containing products especially noted.
6	Hazardous/toxic air pollutants	Estimated 2,000 cancers annually based on an assessment of 20 substances.
7	Depletion of stratospheric ozone	Ozone depletion projected to result in 10,000 additional annual deaths in the year 2100. Not ranked higher because of the uncertainties in future risk.
8	Hazardous waste sites, inactive	Cancer incidence of 1,000 annually from 6 chemicals assessed. Considerable uncertainty since risk is based on extrapolation from 35 sites to about 25,000 sites.
9	Drinking water	Estimated 400–1,000 annual cancers, mostly from radon and trihalomethanes.
10	Application of pesticides	Approximately 100 cancers annually; small population exposed but high individual risks.

Note: Not ranked: Biotechnology; global warming; other air pollutants; discharges to estuaries, coastal waters, and oceans; discharges to wetlands.
Source: Based on data from U.S. EPA, 1987.

In spite of the great uncertainties involved in making their assessments, the divergence between the EPA effort in the 1980s and relative risks is noteworthy. As concluded in the study, areas of relatively high risk but low EPA effort include indoor radon; indoor air pollution; stratospheric ozone depletion; global warming; nonpoint sources; discharges to estuaries, coastal waters, and oceans; other pesticide risks; accidental releases of toxics; consumer products; and worker exposures. Areas of high EPA effort but relatively medium or low risks include RCRA sites, Superfund sites, underground storage tanks, and municipal nonhazardous waste sites.

The *Unfinished Business* report was the first major example of what has come to be known as *comparative risk analysis*. Comparative risk analysis differs from conventional risk assessment because its purpose is not to establish absolute values of risk but rather to provide a process for ranking environmental problems by their seriousness. A subsequent 1990 report, *Reducing Risks*, by the EPA's Science Advisory Board recommended that the EPA reorder its priorities on the basis of reducing the most serious risks. The combination of these two reports has had considerable influence on the way that the EPA perceives its role in environmental protection. The EPA's Office of Research and Development (U.S. EPA, 1996) has incorporated these recommendations in setting forth its strategic principles, which include the following:

- Focus research and development on the greatest risks to people and the environment, taking into account their potential severity, magnitude, and uncertainty.
- Focus research on reducing uncertainty in risk assessment and on cost-effective approaches for preventing and managing risks.
- Balance human health and ecological research.

Based on those strategic principles, EPA has defined its six highest priority research topics for the next few years (U.S. EPA, 1996):

- *Drinking water disinfection.* Some microorganisms, especially the protozoan *Cryptosporidium,* are able to survive conventional disinfection processes, and some carcinogens, such as chloroform, are created during chlorination of drinking water. Questions to be addressed include the comparative risk between waterborne microbial disease and the disinfection byproducts formed during drinking water disinfection.
- *Particulate matter.* Inhalation of particulate matter in the atmosphere poses a high potential human health risk. The relationship between morbidity/mortality and low ambient levels of particulate matter, and cost-effective methods to reduce particulate matter emissions are principal areas of interest.
- *Endocrine disruptors.* Declines in the quality and quantity of human sperm production and increased incidence of certain cancers that may have an endocrine-related basis form the basis of concern for this high-priority research topic.
- *Improved ecosystem risk assessment.* Understanding the impacts of human activities on ecosystems has not developed as rapidly as human health impacts. Topics such as forest decline, toxic microorganisms in estuaries, reproductive failure of wildlife, and the reappearance of vector-borne epidemic diseases need to be addressed.
- *Improved health risk assessment.* Continued focus on reducing the uncertainty in source-exposure-dose relationships, including the impacts of mixtures of chemical insults, is needed.
- *Pollution prevention and new technologies.* Avoiding the creation of environmental problems is the most cost-effective risk-management strategy, but it is not clear how best to integrate pollution prevention into government and private-sector decision making.

PROBLEMS

1 Consider a carcinogenic VOC with the dose-response curve shown in Figure P1. If 70-kg people breath 20 m^3/day of air containing 10^{-3} mg/m^3 of this VOC throughout their entire 70-year lifetime, find the cancer risk (you first need to find the potency).

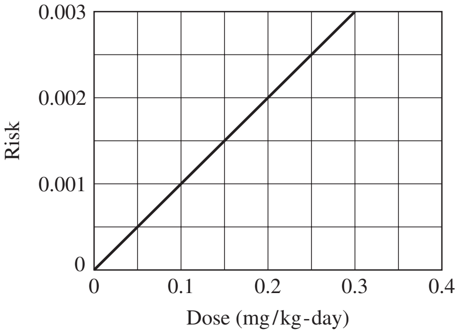


FIGURE P1

2 Suppose a city's water supply has 0.2 ppb (1 ppb = 10^{-3} mg/L) of polychlorinated biphenyls (PCBs) in it. Using the PCB oral potency factor (Table 9) and the EPA recommended exposure factors given in Table 10,

(a) What would be the CDI?

(b) What would be the individual lifetime cancer risk for an adult residential consumer?

(c) In a city of 1 million people, use this risk assessment to estimate the number of extra cancers per year caused by the PCBs in the water supply.

(d) Assume the average cancer death rate in the United States is 193 per 100,000 per year. How many cancer deaths would be expected in a city of 1 million? Do you think the incremental cancers caused by PCBs in the drinking water would be detectable?

3 Suppose 30 out of 500 rats exposed to a potential carcinogen develop tumors. A control group of 300 rats not exposed to the carcinogen develops only 10 tumors. Based on these data, compute (a) the relative risk, (b) the attributable risk, and (c) the odds ratio. Do these indicators suggest that there might be a relationship between exposure and tumor risk?

4 Suppose 5 percent of individuals exposed to a chemical get a tumor, and 2 percent of those not exposed get the same kind of tumor. Find (a) the relative risk, (b) the attributable risk, and (c) the odds ratio.

5 Suppose a 70-kg individual drinks 2 L/day of water containing 0.1 mg/L of 1,1-dichloroethylene for 20 years.

(a) Find the hazard quotient for this exposure.

(b) Find the cancer risk.

(c) If the individual drinks this water for 30 years instead of just 20, recompute the hazard quotient and the cancer risk.

6 Compute the drinking water equivalent level (DWEL) for methylene chloride based on a 10^{-6} risk for a lifetime consumption of 2 L of water per day for a 70-kg individual.

7 Based on a 10^{-6} risk for a 70-kg individual consuming 2 L of water per day, the DWEL for a contaminant is 10 μg/L. What potency would produce this value?

8 The drinking water standard for 2,3,7,8-TCDD (dioxin) is 3×10^{-8} mg/L. Using EPA exposure factors for residential consumption, what lifetime risk would this pose?

9 The drinking water standard for tetrachloroethylene is 0.005 mg/L. Using EPA exposure factors for residential consumption, what lifetime risk would this pose?

10 Human exposure to radiation is often measured in rems (roentgen-equivalent man), or millirems (mrem). The cancer risk caused by exposure to radiation is thought to be approximately 1 fatal cancer per 8,000 person-rems of exposure (e.g., 1 cancer death if 8,000 people are exposed to 1 rem each, or 10,000 people exposed to 0.8 rems each, etc.).

(a) Natural radioactivity in the environment is thought to expose us to roughly 130 mrem/yr. How many cancer deaths in the United States (population 260 million) would be expected per year from this exposure?

(b) A single 3,000-mile, cross-country jet flight exposes an individual to about 4 mrem. How many cross-country flights would be required to elevate your cancer risk by 1 in 1 million? How does this answer compare with the value given in Table 3?

11 Exposure to cosmic radiation increases with increasing altitude. At sea level, it is about 40 mrem/yr, whereas at the elevation of Denver, it is about 120 mrem/yr. Using the radiation potency factor given in Problem 10,

(a) Compare the lifetime (70-yr) probability of dying of cancer induced by cosmic radiation for a person living at sea level with that of a person living in Denver.

(b) Estimate the incremental cancer deaths per year caused by the elevated cosmic radiation exposure in Denver (population 0.57 million in 2006). Compare that incremental death rate with the expected cancer deaths per year at the typical U.S. rate of 189 per 100,000 per year.

(c) How long would an individual have to live in Denver to cause an incremental cancer risk of 10^{-6} (compared with sea level)?

(d) If all 300 million Americans lived at sea level, estimate the total cancer deaths per year caused by cosmic radiation.

12 Living in a home with 1.5 pCi/L of radon is thought to cause a cancer risk equivalent to that caused by approximately 400 mrem/yr of radiation. Using the radiation potency factor given in Problem 10,

(a) Estimate the annual cancers in the United States (population 300 million) caused by radon in homes.

(b) Estimate the lifetime risk for an individual living in a home with that amount of radon.

13 It has been estimated that about 75 million people in the Ukraine and Byelorussia were exposed to an average of 0.4 rem of radiation as a result of the Chernobyl nuclear accident. Using potency from Problem 10,

(a) How many extra cancer deaths might eventually be expected from this exposure?

(b) If the normal probability of dying of cancer from all causes is 0.22, how many cancer deaths would you normally expect among those 75 million people?

14 In Table 3, it was estimated that living 50 years within 5 miles of a nuclear reactor would increase risk by 10^{-6}. Using radiation potency given in Problem 10, what mrem/yr exposure rate would yield this risk?

15 You are given a job adjusting the flow meters in gasoline pumps. The job exposes you to benzene fumes.

   (a) What is the maximum concentration of benzene you can tolerate in your workplace air and keep your incremental risk below the normal acceptable risk set by EPA?

   (b) You have the option of working part-time at the gasoline pump job. Having taken this class, you have vowed not to accept an incremental risk greater than 10^{-7} from workplace exposure to carcinogens. How many days a year could you work on the gasoline pump job and maintain your promise, if the benzene concentration is what you calculated in part a?

16 A man works in an aluminum smelter for 10 years. The drinking water in the smelter contains 0.0700 mg/L arsenic and 0.560 mg/L methylene chloride. His only exposure to these chemicals in water is at work.

   (a) What is the hazard index (HI) associated with this exposure?

   (b) Does the HI indicate this is a safe level of exposure?

   (c) What is the incremental lifetime cancer risk for the man due solely to the water he drinks at work *and* does it seem to be an acceptable risk according to EPA?

17 You have just taken a job at a metal plating plant, and some of the metals you work with contain arsenic as an impurity. The heated metals give off arsenic vapors, which create an average concentration of arsenic of 0.00002 mg/m^3 in the air in the workplace.

   (a) What is your increased risk of cancer if you continue to work in this job for your entire career?

   (b) There are about 220,000 workers in the United States in metal casting plants similar to the one in which you work. How many worker deaths above the background cancer rate would be expected due to their occupation?

   (c) Of the 220,000 workers in metal casting plants, how many would be expected to die of cancer?

   (d) Suppose your workplace air also contains mercury in the air with an average concentration of 0.0001 mg/m^3. Would the HI suggest that this is a safe level of exposure?

18 One way to estimate maximum acceptable concentrations of toxicants in drinking water or air is to pick an acceptable lifetime risk and calculate the concentration that would give that risk assuming agreed-on exposures such as the residential factors given in Table 10. Find the acceptable concentrations of the following substances:

   (a) Benzene in drinking water (mg/L), at a lifetime acceptable risk of 1×10^{-5}

   (b) Trichloroethylene in air (mg/m^3), at a lifetime acceptable risk of 1×10^{-6}

   (c) Benzene in air (mg/m^3), at a lifetime acceptable risk of 1×10^{-5}

   (d) Vinyl chloride in drinking water (mg/L), at a lifetime acceptable risk of 1×10^{-4}

19 Using exposure factors in Table 10, what would be an acceptable concentration of trichloroethylene in the air of an industrial facility if worker risk is to be less than 10^{-4}? Express the answer in mg/m^3 and ppm.

20 Suppose an individual eats fish from a river contaminated by benzene. What concentration of benzene (mg/L) in the water would produce a lifetime risk of 1×10^{-6} to an individual who eats the amount of fish suggested by the EPA exposure factors in Table 10? Use the oral PF for benzene and the BCF given in Table 12.

21 Estimate the cancer risk for a 70-kg individual consuming 2 g of fish every day for 70 years from a stream with 20 ppb of DDT.

22 Suppose a 50-kg individual drinks 1 L/day of water containing 2 mg/L of 1,1,1-trichloroethane, 0.04 mg/L of tetrachloroethylene, and 0.1 mg/L of 1,1-dichloroethylene. What is the hazard index? Is there cause for concern?

23 Suppose 1.0 g/day of heptachlor leaks into a 30,000 m^3 pond. If heptachlor has a reaction rate coefficient of 0.35/day, and complete mixing occurs,

(a) What would be the steady-state concentration in the pond?

(b) Suppose a 70-kg individual drank 2 L/day of that water for 5 years. Estimate the maximum risk of cancer due to that exposure to heptachlor.

24 Mainstream smoke inhaled by a 70-kg smoker contains roughly 0.03 mg per cigarette of the class B2 carcinogen, benzo(a)pyrene. For an individual who smokes 20 cigarettes per day for 40 years, estimate the lifetime risk of cancer caused by that benzo(a)pyrene (there are other carcinogens in cigarettes as well).

25 Consider the problem of indoor air pollution caused by sidestream smoke (unfiltered smoke from an idling cigarette). Suppose the sidestream smoke from one cigarette in a small apartment will produce an airborne benzo(a)pyrene (BaP) concentration that will average 6×10^{-4} mg/m^3 for 1 hour. How many cigarettes would need to be smoked in the presence of a nonsmoking roommate to raise the nonsmoker's cancer risk by 10^{-6} just due to this BaP? Assume an inhalation rate of 0.83 m^3/hr (20 m^3/day). At eight cigarettes of exposure per day, how many days would it take to create this risk?

26 The sidestream smoke from one cigarette releases about 0.1 mg of benzo(a)pyrene (BaP). In an apartment with fresh air entering through holes and cracks (infiltration) at an average rate of 120 m^3/hr ,

(a) What would be the steady-state indoor concentration of BaP if one cigarette per hour is smoked? (Assume that BaP is a conservative pollutant.)

(b) What would be the incremental cancer risk to a nonsmoking roommate who spends 8 hr/day for 1 year in this apartment? (Assume an inhalation rate of 20 m^3/day.)

27 For the following carcinogens, the U.S. drinking water standards are given. For each, find the lifetime individual cancer risk and the incremental cancers per year in a population of 260 million as computed using a standard risk assessment based on residential exposure factors recommended by the EPA.

(a) Trichloroethylene (TCE), 0.005 mg/L

(b) Benzene, 0.005 mg/L

(c) Arsenic, 0.01 mg/L

(d) Carbon tetrachloride, 0.005 mg/L

(e) Vinyl chloride, 0.002 mg/L

(f) Polychlorinated biphenyls (PCBs), 0.0005 mg/L

28 One way to express cancer potency for substances that are inhaled is in terms of risk caused by a lifetime (20 m^3/day for 70 years) of breathing air with a concentration of 1.0 μg/m^3 of carcinogen. The potency for formaldehyde in these terms is 1.3×10^{-5} cancer/μg/m^3. What is the cancer risk caused by a lifetime of breathing formaldehyde at the not unusual (in smoggy cities) concentration of 50 μg/m^3 (the threshold of eye irritation)?

29 Trichloroethylene (TCE) is a common groundwater contaminant. In terms of cancer risk, which would be better: (1) to drink unchlorinated groundwater with 10 ppb (0.010 mg/L) of TCE; or (2) to switch to a surface water supply that, as a result of chlorination, has a chloroform concentration of 50 ppb?

30 Suppose a 70-kg man is exposed to 0.1 mg/m^3 of tetrachloroethylene in the air at his workplace. If he inhales 1 m^3/hr, 8 hours per day, 5 days per week, 50 weeks per year, for 30 years, and if tetrachloroethylene has an absorption factor of 90 percent and an inhalation potency of 2×10^{-3} (mg/kg-day)$^{-1}$, what would be his lifetime cancer risk? What would the risk be to a 50-kg woman similarly exposed?

31 Suppose a factory releases a continuous flow of wastewater into a local stream, resulting in an in-stream carcinogen concentration of 0.1 mg/L just below the outfall. Suppose this carcinogen has an oral potency factor of 0.30 (mg/kg-day)$^{-1}$ and that it is degradable with a reaction rate coefficient k of 0.10/day. To keep the problem simple, assume that the stream is uniform in cross section, flowing at the rate of 1 mph, and that there are no other sources or sinks for this carcinogen. At a distance of 100 miles downstream, a town uses this stream as its only source of water. Estimate the individual residential lifetime cancer risk caused by drinking this water.

32 The following tumor data were collected for rats exposed to ethylene-thiourea (ETU) (data from Crump, 1984):

Dietary Concentration	Animals with Tumors
125 ppm	3%
250 ppm	23%
500 ppm	88%

A one-hit model fitted to the data has coefficients $q_0 = 0.01209$ and $q_1 = 0.001852$/ppm. A multistage model has coefficients

$$q_0 = 0.02077$$

$$q_1 = q_2 = 0.0$$

$$q_3 = 1.101 \times 10^{-8}/(\text{ppm})^3$$

$$q_4 = 1.276 \times 10^{-11}/(\text{ppm})^4$$

(a) For each of the three concentrations given, compare the measured data with the values derived from each of these two models.

(b) For a concentration of 1 ppm, compare the values that each of the two models would predict for percent tumors.

33 Suppose 10 million people are exposed to a carcinogen that poses an individual lifetime (70-yr) cancer risk of 10^{-4}.

(a) How many cancers per year might be caused by this carcinogen?

(b) If spending \$1 per year per person (for an indefinitely long time) to reduce exposure to that carcinogen reduces that risk to 10^{-5}, what would be the cost of each cancer avoided?

REFERENCES

American Cancer Society, 2006, *Cancer Facts and Figures,* American Cancer Society, Atlanta.

Crump, K. S., 1984, An improved procedure for low-dose carcinogenic risk assessment from animal data, *Journal of Environmental Pathology, Toxicology, and Oncology,* (5)4/5:339–349.

Environ, 1988, *Elements of Toxicology and Chemical Risk Assessment,* Environ Corporation, Washington, DC.

Goldberg, A. M., and J. M. Frazier, 1989, Alternatives to animals in toxicity testing, *Scientific American,* August, 24–30.

James, R. C., 1985, Hepatotoxicity: Toxic Effects in the Liver, in *Industrial Toxicology,* P. L. Williams and J. L. Burson (eds.), Van Nostrand Reinhold, New York.

Kolluru, R. V., S. M. Bartell, R. M. Pitblado, and R. S. Stricoff, 1996, *Risk Assessment and Management Handbook,* McGraw-Hill, New York.

National Academy of Sciences, 1983, *Risk Assessment in the Federal Government: Managing the Process,* National Academy Press, Washington, DC.

Rodricks, J. V., 1992, *Calculated Risks: The Toxicity and Human Health Risks of Chemicals in Our Environment,* Cambridge University Press, Cambridge, UK.

Slovic, P., B. Fischhoff, and S. Lichtenstein, 1980, Facts and Fears: Understanding Perceived Risk, in *Societal Risk Assessment: How Safe Is Safe Enough?,* R. Schwing and W. Albers (eds.), Plenum Press, New York.

Slovic, P., 1987, Perception of risk, *Science,* April 17, 236:280–285.

U.S. EPA, 1986a, *Guidelines for Carcinogen Risk Assessment,* Federal Register, Vol. 51, No. 185, pp. 33992–34003, September 24, 1986.

U.S. EPA, 1986b, *Superfund Public Health Evaluation Manual,* Office of Emergency and Remedial Response, Washington, DC.

U.S. EPA, 1987, *Unfinished Business: A Comparative Assessment of Environmental Problems,* Office of Policy, Planning and Evaluation, EPA/2302-87/025, Washington, DC.

U.S. EPA, 1988, *Superfund Exposure Assessment Manual,* Environmental Protection Agency, Office of Remedial Response, EPA/540/1-88/001, Washington, DC.

U.S. EPA, 1990, *Reducing Risks: Setting Priorities and Strategies for Environmental Protection,* Science Advisory Board, U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA, 1991, *Human Health Evaluation Manual,* Supplemental Guidance: Standard Default Exposure Factors, OSWER Directive 9285.6-03, U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA, 1996, *Strategic Plan for the Office of Research and Development,* EPA/600/R-96/059, U.S. Environmental Protection Agency, Washington, DC.

Williams, P. L., and J. L. Burson, 1985, *Industrial Toxicology, Safety and Health Applications in the Workplace,* Van Nostrand Reinhold, New York.

Wilson, R., 1979, Analyzing the daily risks of life, *Technology Review,* 81(4):41–46.

Wilson, R., and E. A. C. Crouch, 1987, Risk assessment and comparisons: An introduction, *Science,* April 17, 267–270.

Water Pollution

> When the well's dry, we know the worth of water.
>
> —*Ben Franklin,* Poor Richard's Almanac

1 Introduction

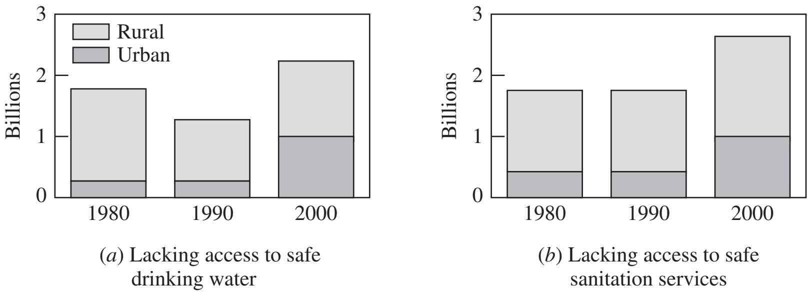
In the late 1960s, the Cuyahoga River in Ohio was clogged with debris and trash, floating in layers of black, heavy oil several inches thick. When it caught fire in 1969, it became a symbol of the environmental degradation that can result from uncontrolled industrialization. Along with images of thick mats of algae along the

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

shores of Lake Erie and dead fish floating on its surface, visible manifestations of pollution quickly galvanized the public, created the modern environmental movement, and led to the enactment of some of our most important environmental legislation, including the Clean Water Act of 1972, which has jurisdiction over water quality in rivers, lakes, estuaries, and wetlands, and the Safe Drinking Water Act of 1974, which regulates tap water quality.

In many ways, the quality of our surface waters has improved significantly as a result of the Clean Water Act, but problems remain. The level of municipal sewage treatment and the fraction of the population served by these facilities has risen sharply as a result of almost $130 billion of federal, state, and local expenditures on wastewater treatment plants. Municipal treatment plant upgrades, coupled with a pollution discharge permit system, have dramatically reduced point-source discharges of raw sewage and industrial wastes. But diffuse, "nonpoint-source" pollution, such as runoff from agricultural lands and urban streets, has proven to be much more difficult to regulate and control, resulting in surface water quality that still does not meet the Clean Water Act's goals of "fishable and swimmable" waters throughout the nation by 1983. In fact, by 2006, EPA estimated that 40 percent of the waters in the United States did not meet the water quality standards set for them and that over two-thirds of the population—218 million people—live within 10 miles of these "impaired waters" (U.S. EPA, 2006).

Water pollution problems in many parts of the world are far worse. It has been estimated that in the year 2002, 1.1 billion people lacked access to safe drinking water, and 2.6 billion lacked access to improved sanitation (WHO, 2004). As shown in Figure 1, growth in numbers lacking access to safe water and sanitation will be driven in large part by the growth rate of people living in urban areas (Gleick, 1993). Most urban centers in Africa and Asia have no sewage system at all, including many cities with populations over 1 million people. The result is a tragic rate of morbidity and mortality in the less developed parts of the world. In 2004, it was estimated that 88 percent of the world's diarrheal disease (including cholera) is due to unsafe water and sanitation (WHO, 2004). Worldwide, 1.8 million people die every year from diarrheal disease, including over 6,000 children under the age of 5 every day (WHO/UNICEF, 2005). In addition, recent studies have shown a clear correlation in developing countries between chronic malnutrition and a lack of safe drinking water due to the ingestion of waterborne human parasites (Braghetta, 2006).



(*a*) Lacking access to safe drinking water

(*b*) Lacking access to safe sanitation services

FIGURE 1 Population in developing countries lacking access to safe drinking water (*a*); and lacking access to sanitation services (*b*).
(*Source:* Based on Gleick, 1993.)

2 Water Resources

Water is so common that we take it for granted. After all, it covers nearly three-fourths of the surface of the Earth. And we probably think it is much like any other liquid, but it is not. In fact, nearly every physical and chemical property of water is unusual when compared with other liquids, and these differences are essential to life as we know it.

Unusual Properties of Water

Consider a simple molecule of water, H_2O. As shown in Figure 2, the two - hydrogen-to-oxygen chemical bonds form a 105° angle with each other, resulting in a molecule that has a slightly positive charge at one end and a slightly negative charge at the other. This *dipolar* character means water molecules are attracted to each other, which helps explain why water boils at such a high temperature and why it takes an unusually large amount of energy to cause it to vaporize. It also helps explain water's high surface tension, which allows it to support relatively heavy objects, such as insects that skate along its surface.

The dipolar nature of water molecules causes them to adhere to other surfaces easily, and the combination of surface tension and adhesion lets water crawl up the sides of objects. This *capillary* action is crucial to life since, for example, it causes sap to rise in trees, water to rise in soil, and food to move through organisms. The dipolar property also makes water a very effective solvent since water molecules tend to surround charged ions and effectively neutralize them.

Density. Water is the only common liquid that expands when it freezes. In fact, a plot of density versus temperature shows a maximum density at 4°C, which means that as temperatures move away from this point, water continuously becomes lighter and more buoyant. As a result, ice floats. If it did not, ice that would form on the surface of bodies of water would sink to the bottom, making it possible for rivers and lakes to freeze solid from the bottom up. The expansion of water as it freezes also contributes to the weathering of rocks by literally breaking them apart when water freezes in the cracks. When water is warmed beyond 4°C, it becomes buoyant once again, so warm water floats on top of cold water in lakes. This thermal stratification affects aquatic life and pollutant distribution in unusual ways, as will be described later in this chapter.

Melting and Boiling Points. Water has unusually high boiling and freezing temperatures for a compound having such a low molecular weight. If water were

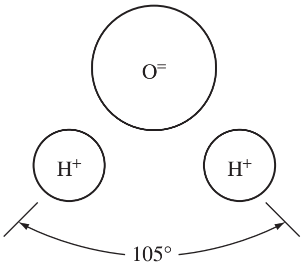


FIGURE 2 A water molecule is dipolar; that is, it appears to have a positive charge at one end and a negative charge at the other. This dipolar character helps explain a number of water's unusual properties.

similar to other "H_2X" substances, such as H_2S, H_2Se, and H_2Te, it would boil at normal Earth temperatures, so it would exist mostly as a gas rather than a liquid or solid. It also has an unusually high difference in temperature between the melting point and boiling point, thus remaining a liquid over most of the globe. It is the only substance that appears in all three states, gaseous, liquid, and solid (ice), within the normal range of temperatures on Earth. With only slightly different phase change temperatures, life on Earth would be very different, if it could exist at all.

Specific Heat. Water has a higher heat capacity (4,184 J/kg°C) than any other known liquid except ammonia. It is five times higher than the specific heat of most common heavy solids, such as rock and concrete. As a result, it takes water longer to heat up and to cool down than almost anything else. This high heat capacity helps make the oceans the major moderating factor in maintaining the temperature of the surface of the Earth. It also serves the important function of protecting life from rapid thermal fluctuations, which are often lethal.

Heat of Vaporization. The heat required to vaporize water (2,258 kJ/kg) is one of the highest of all liquids. This high heat of vaporization means that water vapor stores an unusually large amount of energy, which is released when the water vapor condenses. This property is important in distributing heat from one place on the globe to another and is a major factor affecting the Earth's climate.

Water as a Solvent. Water dissolves more substances than any other common solvent. As a result, it serves as an effective medium for both transporting dissolved nutrients to tissues and organs in living things as well as eliminating their wastes. Water also transports dissolved substances throughout the biosphere.

Greenhouse Effect. The 105° bond angle for water is actually an average value. The hydrogen atoms vibrate back and forth, causing the bond angle to oscillate. This H—O—H *bending vibration* resonates with certain wavelengths of electromagnetic radiation coming from the sun, allowing water vapor in the atmosphere to absorb solar energy. Other vibrations in water molecules also cause absorption of infrared radiation leaving the Earth's surface. Earth's temperature depends to a significant degree on the absorption of incoming solar and outgoing infrared radiation. As will be described later, water vapor is in fact the most important greenhouse gas in our atmosphere.

The Hydrologic Cycle

Almost all of the world's water (97 percent) is located in the oceans, but as might be expected, the high concentration of salts renders the oceans virtually unusable as a source of water for municipal, agricultural, or most industrial needs. We do use the oceans, however, for thermal cooling of power plants and as a sink for much of our pollution. Although desalination technologies exist, the capital and energy requirements to produce significant quantities of fresh water are generally prohibitive, although some small regions of the world do rely heavily on this approach. Fortunately, the sun performs that desalination service for us when it provides the energy needed to evaporate water, leaving the salts behind. In fact, close to one-half

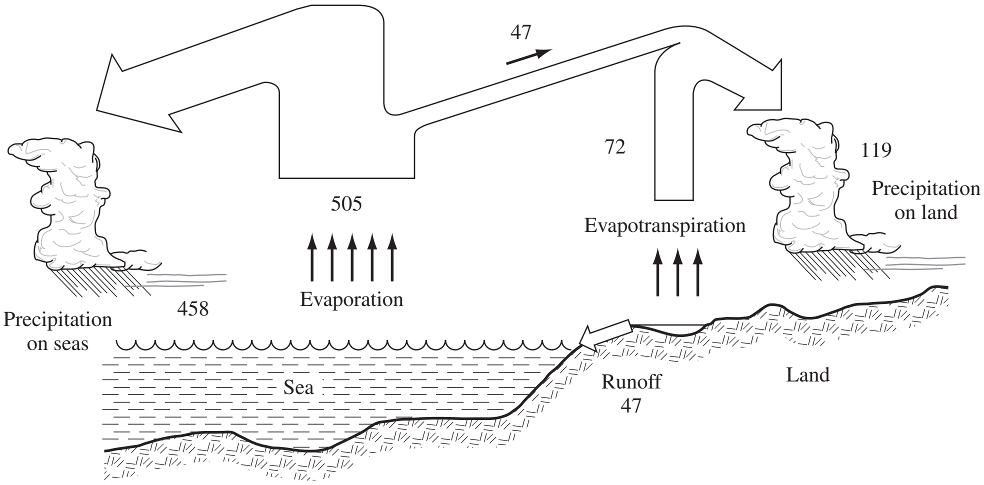


FIGURE 3 The hydrologic cycle. Units are 10^3 km^3/yr.
(*Source:* Based on Shiklomanov, 1993.)

of the sun's energy that is absorbed on the Earth's surface is converted to latent heat, removing water from wet surfaces by evaporation and from the leaves of plants by transpiration. The combination of processes, called *evapotranspiration,* requires an enormous amount of energy, equivalent to roughly 4,000 times the rate at which we use energy resources to power our societies.

Evapotranspiration removes an amount of water equivalent to a layer about 1 m thick around the globe each year. About 88 percent of that is evaporation from the oceans, whereas the remaining 12 percent is evapotranspiration from the land. The resulting water vapor is transported by moving air masses and eventually condenses and returns to the Earth's surface as precipitation. There is more evaporation than precipitation over the oceans, and there is more precipitation than evapotranspiration over the land. The difference between the amount of precipitation and evapotranspiration on land is water that is returned to the oceans as both stream flow and groundwater flow. This is termed *runoff.* Figure 3 illustrates this simple concept of evapotranspiration, precipitation, and runoff as components of the *hydrologic cycle.*

This representation of the hydrologic cycle is highly simplified, masking many of the complexities of timing and distribution. Snowfall may remain locked in polar ice for thousands of years; groundwater may emerge at the surface, contributing to surface water flow and vice versa; droughts and floods attest to the erratic rates of precipitation; and our own activities, most importantly those contributing to global climate change, are probably beginning to alter the relative balances in these hydrologic processes.

As the data in Figure 3 indicate, 60 percent of the precipitation falling on the Earth's land masses is eventually returned to the atmosphere as evapotranspiration. Of the 40 percent that does not evaporate, most collects on the surface, flows into streams and rivers and empties into the oceans, while some seeps into the soil to become underground water that slowly moves toward the seas. This combined groundwater and surface water runoff, 47,000 km^3/yr, is a renewable supply of

TABLE 1

Stocks of Water on Earth

Location	Amount (10^6 km$^{3^{-}}$)	Percentage of World Supply
Oceans	1,338.0	96.5
Glaciers and permanent snow	24.1	1.74
Groundwater	23.4	1.7
Ground ice/permafrost	0.30	0.022
Freshwater lakes	0.091	0.007
Saline lakes	0.085	0.006
Swamp water	0.011	0.008
Atmosphere	0.013	0.001
Average in stream channels	0.002	0.0002
Water in living biomass	0.001	0.0001

Source: Shiklomanov, 1993.

freshwater that can potentially be used year after year without ever depleting the freshwater resources of the world.

Although the rates of evaporation, precipitation, and runoff are obviously important, the amounts of water stored in various locations and forms are also critical. It has already been mentioned that almost all of the world's water is contained in the oceans. The remainder is distributed as shown in Table 1. Freshwater lakes, rivers, and streams, which are the main source of water for human use, account for just 0.0072 percent of the world's stock of water, or about 93,000 km^3.

Water Usage

Roughly 10 percent of the world's annual runoff is withdrawn for human use each year. While that small figure may suggest ample supplies for the future, that is not at all the case. Some areas of the world are inundated with water, whereas others have so little rainfall that human existence is barely possible. Even areas with adequate average precipitation are vulnerable to chaotic variations from one year to the next. Unless major water storage and conveyance facilities are constructed, a region may have plenty of water on the average but not enough to cover needs during dry spells. As population grows, the demands for water will rise, and the amount available for each person will drop, as Figure 4 suggests.

The geographic distribution of water does not match well the distribution of people on the planet. Asia, with 60 percent of the world's population, has only 36 percent of global runoff, whereas South America, with only 5 percent of the world's population, has 25 percent of the runoff. Variations within regions or continents can be extreme. For example, the per capita water availability in North Africa is less than 7 percent of the African average, which is already low.

As populations grow and development proceeds, rising demands for water increase the potential for internal disruption within countries and external conflict with other countries. Many countries depend on local rivers for their water supply, but their upstream neighbors control the flow, examples of which are given in Table 2. Egypt, for example, gets 32.3 times as much surface water from precipitation that falls outside of Egypt as it does from its own rainfall. Moreover, Egypt depends on

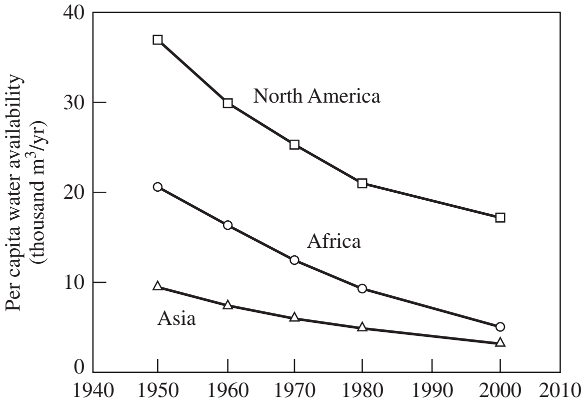


FIGURE 4 Per capita water availability for North America, Africa, and Asia, showing the implications of growing population.
(*Source:* Shiklomanov, 1993. Reprinted by permission of Oxford University Press.)

TABLE 2

Dependence on Imported Surface Water for Selected Countries

Country	Percent of Total Flow Originating Outside of Border	Ratio of External Water Supply to Internal Supply
Egypt	97	32.3
Hungary	95	17.9
Mauritania	95	17.5
Gambia	86	6.4
Syria	79	3.7
Sudan	77	3.3
Iraq	66	1.9
Bangladesh	42	0.7

Source: Gleick, 1993.

the Nile for 97 percent of its surface water supplies, while its neighbor, Ethiopia, controls almost all of the Nile's total flow. Similar circumstances exist all around the globe. Turkey, Syria, and Iraq, for example, share the Euphrates; Bangladesh relies on the Ganges, which is controlled by India; Israel and Jordan share an uneasy dependence on the Jordan River. As these and other such countries increase their demands for water, stresses between neighbors will no doubt intensify.

When describing how water is used, and especially if comparisons are to be made with available supplies, we have to be careful about the terminology. For example, what does it mean to say that a sector of society "uses" a certain amount of water? For example, almost all of the water that is withdrawn from a river for power plant cooling is returned to the river (at a slightly higher temperature) and can be used again. The rest is lost to the atmosphere by evaporation. Downriver, that same water may sprinkle someone's lawn. Again, evaporation takes its toll. To keep track of water usage, it is important to distinguish between consumptive uses of water, in which water is made unavailable for future use (lost mostly to evaporation), and nonconsumptive uses, in which water serves its purpose for one user and

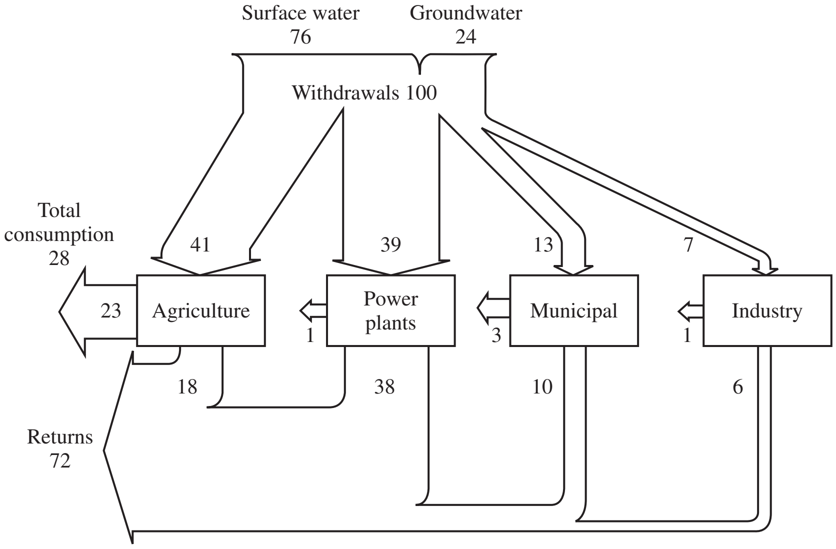


FIGURE 5 Fresh water use in the United States, 1990. The proportion of withdrawals by category of use was unchanged in 2000. Annual 2000 freshwater withdrawals are about 476 km^3.

(*Sources:* From Gleick, 1993; U.S. Water Resources Council, 1978; and Hutson et al., 2004.)

Notes: Industry includes mining, *agriculture* includes irrigation and livestock, and *municipal* includes public, domestic, and commercial uses.

is made available for the next. A simple equation relates the key terms: water *withdrawals,* water *returns,* and water *consumption.*

$$\text{Withdrawals} = \text{Consumption} + \text{Returns} \tag{1}$$

Figure 5 provides data on fresh water withdrawals, consumption, and returns in the United States. Total freshwater withdrawals are about 476 km^3/yr, which is about one-third of the annual runoff. For the whole country, groundwater provides one-fourth of all withdrawals and surface water provides the other three fourths. Saline water withdrawals bring the total to 564 km^3/yr. However, the distribution of freshwater withdrawals is very uneven across the country. In Kansas, for example, close to 60 percent of withdrawals are from groundwater, whereas in Montana, only 2 percent of withdrawals are from groundwater (Hutson et al., 2004).

Although water to cool power plants accounts for almost 40 percent of freshwater withdrawals, essentially all of that water is returned, and it is returned in a relatively unpolluted condition. If those withdrawals are removed from Figure 5, the importance of agriculture stands out. Excluding power plant cooling, agriculture accounts for 67 percent of freshwater withdrawals in the United States and 85 percent of consumption. In some areas, the predominant role of agriculture is even more dramatic. In drought-prone California, for example, irrigation accounts for over 80 percent of both withdrawals and consumption (excluding cooling water), whereas municipal withdrawals account for approximately 15 percent of total withdrawals. As the state continues to grow, water transfers from agricultural uses to cities would be one way to meet increasing urban needs. A reduction of only 20 percent in agriculture, for example, could double the water supply in cities.

TABLE 3

Examples of Water Withdrawals to Supply Various End Uses

Household Water Usage[a]	**Liters**	**Gallons**
Standard toilet, per flush	10–30	3–8
Ultralow volume toilet, per flush	6 or less	1.6 or less
Dishwasher, per load	50–120	13–30
Water saver dishwasher, per load	40–100	10–25
Washing dishes in a filled sink	20–40	5–10
Shower head, per minute	20–30	5–8
Low-flow shower head, per minute	6–11	1.5–3
Slowly dripping faucet, per day	40–80	10–20
Washing car with running water, 20 min	400–800	100–200
Washing car with pistol-grip faucet, 20 min	60 or more	15 or more
Uncovered 60-m^2 pool, per day	100–400	25–100
Covered pool, per day	10–40	2.5–10
Agricultural Items[b]		
One egg	150	40
Glass of milk	380	100
One pound of rice	2,120	560
One pound of grain-fed beef	3,030	800
One pound of cotton	7,730	2,040

Sources: [a]From Gleick, 1993, and [b]from USGS, 1984.

Average per-capita household water use in the United States is about 400 liters per day (110 gal/person-day). If water for landscaping is not included, it is not uncommon for one-third of the remaining water to be used for toilet flushings, another one-third is used for bathing, and the remaining third is for cooking, laundry, and everything else. Table 3 shows some examples of water use for various household activities, giving standard and water-conserving estimates.

3 Water Pollutants

Water that has been withdrawn, used for some purpose, and then returned will be polluted in one way or another. Agricultural return water contains pesticides, fertilizers, and salts; municipal return water carries human sewage, pharmaceuticals, and surfactants; power plants discharge water that is elevated in temperature; and industry contributes a wide range of chemical pollutants and organic wastes. To aggravate the problem, pollutants also enter water from natural sources and from human sources via nonaqueous routes. Arsenic, antimony, and fluoride often come from natural mineral deposits through which groundwater flows. Much of the mercury in water is deposited from the air after being emitted from coal combustion. The polybrominated biphenyl ethers (PBDEs), discussed later in this chapter, are now found in water throughout the world and are believed to be transported largely on dust in the air. The list of pollutants that contaminate water is lengthy, so it helps to organize the list into a smaller number of major categories, as has been done in the following sections.

Pathogens

It has long been known that contaminated water is responsible for the spread of many contagious diseases. In a famous study in 1849, Dr. John Snow provided some of the earliest evidence of the relationship between human waste, drinking water, and disease. He noted that individuals who drank from a particular well on Broad Street in London were much more likely to become victims of a local cholera epidemic than those from the same neighborhood who drank from a different well. He not only found a likely source of the contamination, sewage from the home of a cholera patient, but he was able to effectively end the epidemic by simply removing the handle from the pump on the Broad Street well. It wasn't until later in the nineteenth century, however, when Pasteur and others convincingly established the germ theory of disease, that the role of pathogenic microorganisms in such epidemics was understood.

Pathogens are disease-causing organisms that grow and multiply within the host. The resulting growth of microorganisms in a host is called an infection. Examples of pathogens associated with water include *bacteria,* responsible for cholera, bacillary dysentery (shigellosis), typhoid, and paratyphoid fever; *viruses,* responsible for infectious hepatitis and poliomyelitis; *protozoa,* which cause amebic dysentery, giardiasis, and cryptosporidiosis; and *helminths,* or parasitic worms, that cause diseases such as schistosomiasis and dracunculiasis (guinea-worm disease). Although it is not the only source of pathogens in water, contamination by human feces is the most important. Table 4 provides a more complete list of pathogens excreted in human feces. The intestinal discharges of an infected individual, a carrier, may contain billions of these pathogens. If they are allowed to enter the water supply, they can cause epidemics of immense proportions. Carriers may not even necessarily exhibit symptoms of their disease, which makes it even more important to carefully protect all water supplies from any human waste contamination. In developing countries, where resources are scarce, even simple measures can be quite effective. Household water, for example, is often taken from open wells or streams that are easily contaminated. By enclosing a well and replacing the dirty rope and bucket with a less-easily-contaminated handpump, the incidence rate of these diseases can be greatly reduced.

Epidemics of infectious diseases periodically emerge in areas where crowded conditions and poor sanitation enable the microbes to reach new victims at a rapid rate. With international travel now commonplace, local epidemics can become global pandemics. A pandemic of cholera, for example, that began in 1961 in Indonesia was tracked by the World Health Organization (WHO) as the disease spread around the globe (Figure 6). By 1970, it had reached Africa, where it spread quickly and remained for another two decades. It finally jumped to Peru in 1991, where it killed almost 4,000 people. Within a year, 400,000 new cases of cholera were reported in South America where there had been almost none before.

There are many ways that contaminated water is associated with infectious diseases. *Waterborne* diseases, such as cholera and typhoid, are spread by ingestion of contaminated water; *water-washed* diseases, such as trachoma and scabia, are associated with lack of sufficient water to maintain cleanliness; *water-based* diseases, such as schistosomiasis and dracunculiasis, involve water contact but don't require ingestion; and *water-related* diseases, such as malaria and dengue, involve a host

TABLE 4

Typical Pathogens Excreted in Human Feces

Pathogen Group and Name	Associated Diseases
Virus	
Adenoviruses	Respiratory, eye infections
Enteroviruses	
Polioviruses	Aseptic meningitis, poliomyelitis
Echoviruses	Aseptic meningitis, diarrhea, respiratory infections
Coxsackie viruses	Aseptic meningitis, herpangina, myocarditis
Hepatitis A virus	Infectious hepatitis
Reoviruses	Not well known
Other viruses	Gastroenteritis, diarrhea
Bacterium	
Salmonella typhi	Typhoid fever
Salmonella paratyphi	Paratyphoid fever
Other salmonellae	Gastroenteritis
Shigella species	Bacillary dysentery
Vibrio cholerae	Cholera
Other vibrios	Diarrhea
Yersinia enterocolitica	Gastroenteritis
Protozoan	
Entamoeba histolytica	Amoebic dysentery
Giardia lamblia	Diarrhea
Cryptosporidium species	Diarrhea
Helminth	
Ancylostoma duodenale (Hookworm)	Hookworm
Ascaris lumbricoides (Roundworm)	Ascariasis
Hymenolepis nana (Dwarf tapeworm)	Hymenolepiasis
Necator americanus (Hookworm)	Hookworm
Strongyloides stercoralis (Threadworm)	Strongyloidiasis
Trichuris trichiura (Whipworm)	Trichuriasis

Source: Hammer & Hammer, 1996.

that depends on water for its habitat (e.g., mosquitoes), but human contact with the water is not required. Table 5 shows estimates of global morbidity, mortality, and numbers of people living in areas where these diseases are common, for a number of the most prevalent diseases associated with water.

One disease, schistosomiasis (bilharzia), is particularly insidious since mere contact with contaminated water is sufficient to cause infection, and its incidence rate is to a large extent the result of water-resource development projects. Schistosomiasis is one of the most common water-associated diseases in the world, affecting approximately 200 million people, 200,000 of whom die each year. It is spread by free-swimming larvae in the water, called *cercaria,* that attach themselves to human skin, penetrate it, and enter the bloodstream. Cercaria mature in the liver into worms that lay masses of eggs on the walls of the intestine. When these eggs are

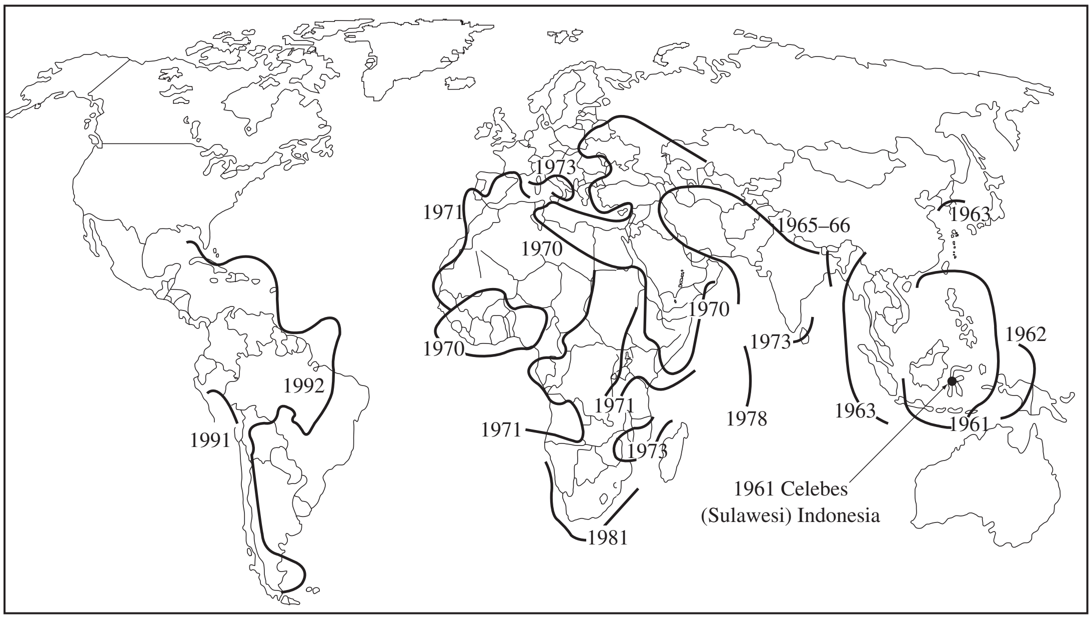


FIGURE 6 The progressive spread of cholera from Indonesia in 1961 to the known global distribution in 1992.
(*Source:* UNEP, 1993.)

TABLE 5

Selected Examples of Global Morbidity, Mortality, and Populations at Risk for Infectious Diseases Associated with Water

Disease	Vector	Morbidity	Mortality	Population at Risk
Diarrheal diseases	microorganisms	>1.5 billion	4 million	>2 billion
Schistosomiasis	water snails	200 million	200,000	500–600 million
Malaria	mosquitoes	267 million	1–2 million	2.1 billion
Onchocerciasis	blackflies	18 million	20 to 50,000	90 million

Source: UNEP, 1993.

excreted into water, they hatch and have only a few hours to find a snail host in which they develop into new cercaria. Cercaria excreted by the snails then have a few days to find another human host, continuing the cycle. Continuation of the cycle requires continued contamination by schistosomiasis carriers in waters that are still enough to allow snails to thrive. Unfortunately, development projects such as dams and irrigation canals, built in countries with poor sanitation, often lead to an increase in schistosomiasis by creating the still water conditions needed by the intermediate hosts, the snails.

Not that long ago, even developed countries such as the United States experienced numerous epidemics of waterborne diseases such as typhoid and cholera. At the start of the twentieth century, typhoid, for example, was killing approximately 28,000 Americans each year. It was only after the advent of chlorination, which began in the United States in 1908, that outbreaks of waterborne diseases such as these became rare. Even now, however, inadequate sanitation in developing countries continues to contribute to high rates of disease and death. The WHO, for example, estimates that 80 to 100 percent of the illness caused by cholera, typhoid, and guinea-worm infection could be eliminated in developing countries with improved water supply and sanitary disposal of excreta (UNEP, 1993).

In the United States, most waterborne diseases are adequately controlled, and we usually don't need to worry about water coming from the tap. There are times, however, when even disinfection is not totally effective, as was the case in Milwaukee, in 1993, when an outbreak of the gastrointestinal illness, cryptosporidiosis, occurred. *Cryptosporidium parvum* is a protozoan parasite that causes diarrhea, abdominal pain, nausea, and vomiting in its host. For individuals whose immune systems are compromised, it can be life threatening. Most incidents of cryptosporidiosis have been traced to water supplies that were contaminated with agricultural runoff containing cattle feces. An infected cow sheds billions of oocysts (eggs) in feces per day and these oocysts are resistant to the usual chlorine, or chlorine dioxide, disinfection processes used in water treatment plants. At present, the most suitable method of removing them at water treatment plants is with microfiltration or with combined coagulation, flocculation, and depth-filtration processes.

Another protozoan, *Giardia lamblia,* is less deadly than *Cryptosporidium,* but it shares a number of similar characteristics. *Giardia* cysts can be carried by wild animals as well as humans; they survive for months in the environment; and they are not easily destroyed by chlorination. As every backpacker knows these days, it is no longer safe to drink surface water from even the most sparkling stream because of the risk of giardiasis caused by the *Giardia lamblia*.

Oxygen-Demanding Wastes

One of the most important measures of the quality of a water source is the amount of dissolved oxygen (DO) present. The saturated value of dissolved oxygen in water is modest, on the order of 8 to 15 mg of oxygen per liter of water, depending on temperature and salinity. The minimum recommended amount of DO for a healthy fish population has often been set at 5 mg/L, but a more careful analysis of the needs of different species of fish at different times in their lifecycle yields a range of oxygen requirements. For example, EPA recommends at least 8 mg/L for coldwater species, such as trout and salmon, during their embryonic and larval stages, and the first 30 days after hatching, but 5 mg/L is the recommendation for early life stages of warmwater species such as bluegill and bass. For older fish, EPA recommends a 7-day mean minimum of 5 mg/L for coldwater fish and 4 mg/L for warmwater fish.

Oxygen-demanding wastes are substances that oxidize in the receiving body of water. As bacteria decompose these wastes, they utilize oxygen dissolved in the water, which reduces the remaining amount of DO. As DO drops, fish and other aquatic life are threatened and, in the extreme case, killed. In addition, as dissolved oxygen levels fall, undesirable odors, tastes, and colors reduce the acceptability of that water as a domestic supply and reduce its attractiveness for recreational uses. Oxygen-demanding wastes are usually biodegradable organic substances contained in municipal wastewaters or in effluents from certain industries, such as food processing and paper production. In addition, the oxidation of certain inorganic compounds may contribute to the oxygen demand. Even naturally occurring organic matter, such as leaves and animal droppings, that finds its way into surface water contributes to oxygen depletion.

There are several measures of oxygen demand commonly used. The *chemical oxygen demand* (COD) is the amount of oxygen needed to chemically oxidize the wastes, whereas the *biochemical oxygen demand* (BOD) is the amount of oxygen required by microorganisms to biologically degrade the wastes. BOD has traditionally been the most important measure of the strength of organic pollution, and the amount of BOD reduction in a wastewater treatment plant is a key indicator of process performance.

Nutrients

Nutrients are chemicals, such as nitrogen, phosphorus, carbon, sulfur, calcium, potassium, iron, manganese, boron, and cobalt, that are essential to the growth of living things. In terms of water quality, nutrients can be considered as pollutants when their concentrations are sufficient to allow excessive growth of aquatic plants, particularly algae.

Nutrient enrichment can lead to blooms of algae, which eventually die and decompose. Their decomposition removes oxygen from the water, potentially leading to levels of DO that are insufficient to sustain normal life forms. Algae and decaying organic matter add color, turbidity, odors, and objectionable tastes to water that are difficult to remove and that may greatly reduce its acceptability as a domestic water source. The process of nutrient enrichment, called *eutrophication,* is especially important in lakes, and it is described more fully in a later section of this chapter.

Aquatic species require a long list of nutrients for growth and reproduction, but from a water-quality perspective, the three most important ones are carbon, nitrogen, and phosphorus. Plants, including algae, require relatively large amounts of each of these three nutrients, and unless all three are available, growth will be limited. The nutrient that is least available relative to the plant's needs is called the *limiting nutrient*. Carbon is usually available from a number of natural sources, including alkalinity, dissolved carbon dioxide from the atmosphere, and decaying organic matter, so it is not often the limiting nutrient. Rather, it is usually either nitrogen or phosphorus that controls algal growth rates. In general, seawater is most often limited by nitrogen, whereas freshwater lakes are most often limited by phosphorus.

Nutrients as well as other pollution may come from either *point-sources* or *nonpoint-sources*. The most common point-sources are the discharge pipes from industries and wastewater treatment plants but may also include discharges from waste disposal sites, mines, animal feedlots, and large construction sites where the polluted water is collected in networks of pipes or channels and taken to a central point before discharge. Nonpoint-sources are usually more intermittent than point-sources and are characterized by many points of entry to the receiving water. Runoff from agricultural lands, pastures and ranges, small construction sites, urban areas, abandoned mines, and logging sites are typical nonpoint-sources of pollution. Point sources, which have been the focus of most remedial effort in the past 30 years, are easier to treat and monitor because they are both more continuous and centralized. However, nonpoint-sources are now the largest contributor to surface water pollution in the United States, and agriculture is the largest contributor to nonpoint-source pollution (Carpenter, 1998; U.S. EPA, 2002).

Major sources of nitrogen include municipal wastewater discharges, runoff from animal feedlots, chemical fertilizers, and nitrogen-deposition from the atmosphere, especially in the vicinity of coal-fired power plants. In addition, certain bacteria and blue-green algae can obtain their nitrogen directly from the atmosphere. These life forms are usually abundant in lakes that have high rates of biological productivity, making the control of nitrogen in such places extremely difficult.

Not only is nitrogen capable of contributing to eutrophication problems, but when found in drinking water in a particular form, nitrate (NO_3), it can pose a serious public health threat. For most people, nitrate in drinking water does not cause a particular toxicity problem. However, certain bacteria that convert nitrate to nitrite (NO_2) are found in the alkaline digestive tract of infants but not later in adults as the digestive system becomes more acidic. Hemoglobin in the bloodstream is oxidized by nitrite to methemoglobin, which cannot carry oxygen. The oxygen starvation characteristic of *methemoglobinemia* causes a bluish discoloration of the infant; hence, it is commonly referred to as the blue baby syndrome. In extreme cases, the victim may die from suffocation. Usually after the age of about six months, the digestive system of a child is sufficiently developed that this syndrome does not occur.

While there are usually enough natural sources of nitrogen to allow algae and aquatic weeds to grow, there isn't much phosphorus available from nature, so it tends to become the controlling nutrient in rivers and lakes. Human activities, however, often provide enough phosphorus to allow excessive growth of aquatic weeds and algae. Human sources of phosphorus include agricultural runoff in fertilized areas, discharges from animal feedlots, and domestic sewage. In sewage, part of the phosphorus is from human feces and part is from detergents.

Detergent formulations have evolved in response to environmental concerns since their introduction just after World War II to replace soaps that tended to form a scum of insoluble precipitates (the ring around the bathtub) in hard water. When they were first introduced, detergents were nonbiodegradable, which led to mountains of foam on rivers, lakes, and sewage treatment plants, and, in some areas, foamy water came out of the tap. The problem was caused by the choice of surfactant, which is the ingredient in detergents that lowers the surface tension in water and allows dirt particles to be lifted or floated from the soiled material during washing. By 1965, the choice of surfactants had changed so that all detergents became biodegradable, but that caused a new problem. Detergents contained large amounts of phosphorus, and when that was released during degradation, the phosphorus acted as a stimulant to algal growth resulting in enormous blooms. Concern for the environmental effects of phosphorus has now led to reduction in its use in detergents. In many countries, including the United States, phosphorus-containing surfactants have now been largely replaced, but there have been concerns for the safety of some of the new nonphosphorus-containing surfactants.

Despite the decrease in phosphorus in detergents and the implementation of phosphorus removal processes at some wastewater treatment plants, phosphorus-induced eutrophication is still the largest water quality problem in lakes, reservoirs, and ponds in the United States (U.S. EPA, 2002). Most of phosphorus is from nonpoint-sources. It has been estimated that even with no point-source phosphorus inputs, almost 90 percent of the impaired lakes would still not meet water quality standards due to the nonpoint-source inputs (Carpenter, 1998). Addition of phosphorus to agricultural land, both in industrial fertilizers and manure, has built up a phosphorus accumulation in farm and pasture lands that is about 25 percent higher than background levels (Carpenter, 1998). This excess is steadily being augmented by application of phosphorus above the needs of the crops being grown and a fraction of the excess finds its way to adjacent rivers, streams, and water bodies, both dissolved in runoff water and attached to particles carried by the run-off.

Salts

Water naturally accumulates a variety of dissolved solids, or *salts,* as it passes through soils and rocks on its way to the sea. These salts typically include such cations as sodium, calcium, magnesium, and potassium, and anions such as chloride, sulfate, and bicarbonate. A careful analysis of salinity would result in a list of the concentrations of the primary cations and anions, but a simpler, more commonly used measure of salinity is the concentration of *total dissolved solids* (TDS). As a rough approximation, *fresh* water can be considered to be water with less than 1,500 mg/L TDS; *brackish* waters may have TDS values up to 5,000 mg/L; and *saline* waters are those with concentrations above 5,000 mg/L (Tchobanoglous & Shroeder, 1985). Seawater contains 30,000 to 34,000 mg/L TDS.

The concentration of dissolved solids is an important indicator of the usefulness of water for various applications. Drinking water, for example, has a recommended maximum TDS concentration of 500 mg/L. Many people will begin to notice water tastes salty at about 1,000 mg/L of TDS, although this is very dependent on the particular dissolved constituents. Livestock can tolerate higher concentrations. Upper limits for stock water concentrations quoted by the U.S. Geological Survey (1985) include poultry at 2,860 mg/L, pigs at 4,290 mg/L, and beef cattle at 10,100 mg/L. Of greater importance, however, is the salt tolerance of crops. As the concentration of salts in irrigation water increases above 500 mg/L, the need for careful water management to maintain crop yields becomes increasingly important. With sufficient drainage to keep salts from accumulating in the soil, up to 1,500 mg/L TDS can be tolerated by most crops with little loss of yield (Frederick & Hanson, 1982), but at concentrations above 2,100 mg/L, water is generally unsuitable for irrigation except for the most salt tolerant of crops.

All naturally occurring water has some amount of salt in it. In addition, many industries discharge high concentrations of salts, and urban runoff may contain large amounts in areas where salt is used to keep ice from forming on roads in the winter. Although such human activities may increase salinity by adding salts to a given volume of water, it is more often the opposite process, the removal of fresh water by evaporation, that causes salinity problems. When water evaporates, the salts are left behind, and since there is less remaining fresh water to dilute them, their concentration increases.

Irrigated agriculture, especially in arid areas, is always vulnerable to an accumulation of salts due to this evapotranspiration on the cropland itself. The salinity is enhanced by the increased evaporation in reservoirs that typically accompany

TABLE 6

Mean Annual Flow and TDS Levels in the Rio Grande as It Travels Through New Mexico and Texas

Station Flow	Flow (10^6 m^3/yr)	Dissolved Solids (mg/L)
Otowi Bridge, NM	1.33	221
San Marcial, NM	1.05	449
Elephant Butte Outlet, NM	0.97	478
Caballo Dam, NM	0.96	515
Leasburg Dam, NM	0.92	551
El Paso, TX	0.65	787
Fort Quitman, TX	0.25	1691

Source: Skogerboe & Law, 1971.

irrigation projects. In addition, irrigation drainage water may pick up additional salt as it passes over and through soils. As a result, irrigation drainage water is always higher in salinity than the supply water and, with every reuse, its salt concentration increases even more. In rivers that are heavily used for irrigation, the salt concentration progressively increases downstream as the volume of water available to dilute salts decreases due to evaporation, and as the salt load increases due to salty drainage water returning from irrigated lands. As an example, Table 6 shows decreasing flows and increasing TDS for the Rio Grande as it travels from New Mexico to Texas.

It has been estimated that roughly one-third of the irrigated lands in the western part of the United States have a salinity problem that is increasing with time, including regions in the Lower Colorado River Basin and the west side of the San Joaquin Valley in California. An estimated 100,000 tons of salt are imported annually into southern Arizona through its consumption of Colorado River water via the 300-mile long, Central Arizona Project canal. Salinity problems are also having major impacts on irrigated lands in Iraq, Pakistan, India, Mexico, Argentina, Mali, and North Africa, among others. The collapse of ancient civilizations, such as those that once flourished in the Fertile Crescent in what is now Iraq, is thought to have been precipitated by the demise of irrigated agriculture caused by accumulating salt (Reisner, 1986).

Salt accumulation in soils is often controlled by flushing the salts away with additional amounts of irrigation water. This increases costs; wastes water, which may not be abundantly available in the first place; and unless adequate drainage is available, increases the likelihood that a rising water table will drown plant roots in salt-laden water. Providing adequate drainage can be an expensive and challenging task involving extensive on-farm subsurface drainage systems coupled with a central drain and disposal system. Even when salt is removed from the agricultural land by good drainage, it can have later repercussions. In the mid-1980s, it was found that birdlife in the natural freshwater marshes of the Kesterson Wildlife Refuge in central California was being poisoned by selenium draining from the region's agricultural fields. Since irrigation return water contains not only salts but fertilizers and pesticides as well, finding an acceptable method of disposal is difficult. These issues with salts highlight how important it is to not only deal with the immediate impacts of pollution but to develop the remedies so that further downstream impacts are not created.

Thermal Pollution

A large steam-electric power plant requires an enormous amount of cooling water. A typical nuclear plant, for example, warms about 150,000 m^3/hr of cooling water by 10°C as it passes through the plant's condenser. If that heat is released into a local river or lake, the resulting rise in temperature can adversely affect life in the vicinity of the thermal plume. For some species, such as trout and salmon, any increase in temperature is life threatening, although for others warmed water might be considered beneficial. Within certain limits, thermal additions can promote fish growth, and fishing may actually be improved in the vicinity of a power plant. On the other hand, sudden changes in temperature caused by periodic plant outages, both planned and unanticipated, can make it difficult for the local ecology to ever acclimate.

As water temperature increases, two factors combine to make it more difficult for aquatic life to get sufficient oxygen from the water. The first results from the fact that metabolic rates tend to increase with temperature, generally by a factor of 2 for each 10°C rise in temperature. This causes an increase in the amount of oxygen required by organisms. At the same time, the available supplies of dissolved oxygen (DO) are reduced both because waste assimilation is quicker, drawing down DO at a faster rate, and because the amount of DO that the water can hold decreases with temperature. Thus, as temperatures increase, the demand for oxygen goes up, and the amount of DO available goes down.

Heavy Metals

In some contexts, the definition of a *metal* is based on physical properties. Metals are characterized by high thermal and electrical conductivity, high reflectivity and metallic luster, strength, and ductility. From a chemical perspective, however, it is more common to use a broader definition that says a metal is an element that will give up one or more electrons to form a cation in an aqueous solution. These definitions are imprecise, as there are some elements that exhibit both metallic and nonmetallic characteristics and form anions in aqueous solution. These elements are sometimes referred to as *metalloids* and include selenium, arsenic, phosphorus, silicon, and boron. Including the metalloids, about 80 elements can be called metals. The term *heavy metal* is less precisely defined. It is often used to refer to metals with specific gravity greater than about 4 or 5. In terms of their environmental impacts, the most important heavy metals are mercury (Hg), lead (Pb), cadmium (Cd), and arsenic (As).

Most metals are toxic, including aluminum, arsenic, beryllium, bismuth, cadmium, chromium, cobalt, copper, iron, lead, manganese, mercury, nickel, selenium, strontium, thallium, tin, titanium, and zinc. Some of these metals, such as chromium and iron, are essential nutrients in our diets, but in higher doses, they can cause a range of adverse impacts on the body, including nervous system and kidney damage, creation of mutations, and induction of tumors. The most important route for the elimination of metals after they are inside a person is via the kidneys. In fact, kidneys can be considered to be complex filters whose primary purpose is to eliminate toxic substances from the body. The kidneys contain millions of excretory units called nephrons, and chemicals that are toxic to the kidneys are called *nephrotoxins*. Cadmium, lead, and mercury are examples of nephrotoxic metals.

Metals may be inhaled, as is often the case with lead, for example, and they may be ingested. How well they are absorbed in the body depends somewhat on the particular metal in question and the particular form that it exists in. For example, liquid mercury is not very toxic and most of what is ingested is excreted from the body. Mercury vapor, on the other hand, is highly toxic. As a vapor, it enters the lungs where it diffuses into the bloodstream. When blood containing mercury reaches the brain, the mercury can pass into the brain, where it causes serious damage to the central nervous system. By contrast, lead does not pose much of a threat as a vapor since it has such a low vapor pressure, and is most dangerous when it is dissolved into its ionic form, Pb^{2+}. Lead dissolved in blood is transferred to vital organs, including the kidneys and brain, and it readily passes from a pregnant woman to her fetus. Children and fetuses are the most at risk since their brains are growing rapidly, and exposure to lead can cause severe and permanent brain damage.

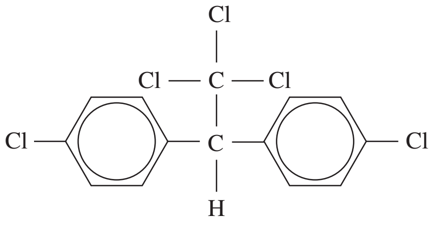
Metals differ from other toxic substances in that they are totally nondegradable, which means they are virtually indestructible in the environment. Thus, it is imperative when metal pollution is remediated that the waste products be disposed of in such a way as to minimize further environmental and human impacts. For instance, in 2001, a more stringent arsenic standard was passed that will cause an estimated 30,000 pounds of arsenic to be removed from drinking water annually. Most of the arsenic-bearing residuals from the water treatment will be sent to municipal solid waste landfills, where research suggests it may be mobilized into the landfill leachate—the liquids that drain out of the disposed solid waste. The questions then become what happens to the leachate, and, if the arsenic moves from the leachate into another process, for instance, via a sewer to a wastewater treatment plant, what are the risks and fate of the arsenic in this next environmental compartment.

Pesticides

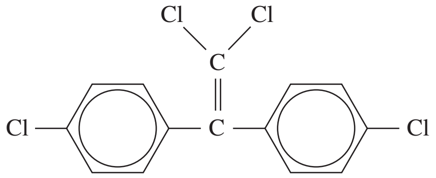
The term *pesticide* is used to cover a range of chemicals that kill organisms that humans consider undesirable. Pesticides can be delineated as insecticides, herbicides, rodenticides, and fungicides.

There are three main groups of synthetic organic insecticides: *organochlorines* (also known as *chlorinated hydrocarbons*), *organophosphates,* and *carbamates*. The most widely known organochlorine pesticide is DDT (dichlorodiphenyltrichloroethane), which has been widely used to control insects that carry diseases such as malaria (mosquitoes), typhus (body lice), and plague (fleas). By contributing to the control of these diseases, DDT is credited with saving literally millions of lives worldwide. In spite of its more recent reputation as a dangerous pesticide, in terms of human toxicity, DDT is considered to be relatively safe. In fact organochlorine insecticides in general are highly toxic to insects, but their acute human toxicity is relatively low. It was DDT's impact on food chains, rather than its toxicity to humans, that led to its ban in the developed countries of the world (it is still used in developing countries).

Organochlorine pesticides, such as DDT, have two properties that cause them to be particularly disruptive to food chains. They are very *persistent,* which means they last a long time in the environment before being broken down into other



DDT: *p*,*p*′-dichlorodiphenyltrichloroethane



DDE

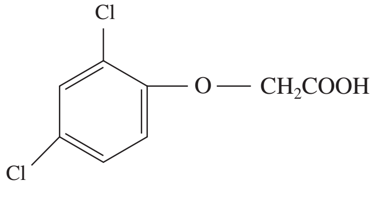
substances, and they are *soluble* in hydrocarbon solvents, which means they easily accumulate in fatty tissue. The accumulation of organochlorine pesticides in fatty tissue means that organisms at successively higher trophic levels in a food chain are consuming food that has successively higher concentrations of pesticide. At the top of the food chain, body concentrations of these pesticides are the highest, and it is there that organochlorine toxicity has been most recognizable. Birds, for example, are high on the food chain, and it was the adverse effect of DDT on their reproductive success that focused attention on this particular pesticide. DDT, and its metabolite, DDE (dichlorodiphenyldichloroethene), interferes with the enzyme that regulates the distribution of calcium in birds, resulting in eggs with shells that are too thin to support the weight of the parent. The resulting difficulty to reproduce has been shown to affect a number of species, including peregrine falcons, bald eagles, ospreys, and brown pelicans. This is an example of endocrine disruption by pollutants that will be discussed shortly. As environmental levels of DDT and DDE subside, some of these species are enjoying a resurgence in numbers. The bald eagle has been taken off the endangered species list in some areas, such as the Pacific Northwest and around the Great Lakes.

Other widely used organochlorines included methoxychlor, chlordane, heptachlor, aldrin, dieldrin, endrin, endosulfan, and Kepone. Animal studies have shown that dieldrin, heptachlor, and chlordane produce liver cancers, and aldrin, dieldrin, and endrin cause birth defects in mice and hamsters. Workers' exposure to Kepone in a manufacturing plant in Virginia resulted in severe neurological damage, and the plant was ultimately closed. Given the ecosystem disruption, their potential long-term health effects in humans (e.g., cancer), and the biological resistance to these pesticides that many insect species have developed, organochlorines have largely been replaced with organophosphates and carbamates.

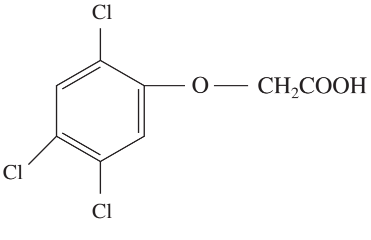
The organophosphates, such as parathion, malathion, diazinon, TEPP (tetraethyl prophosphate), and dimethoate, are effective against a wide range of insects, and they are not persistent. However, they are much more acutely toxic to humans than the organochlorines that they have replaced. They are rapidly absorbed through the skin, lungs, and gastrointestinal tract, so unless proper precautions are taken, they are very hazardous to those who use them. Humans exposed to excessive amounts have shown a range of symptoms, including tremor, confusion, slurred speech, muscle twitching, and convulsions.

The third category of insecticides, carbamates, is derived from carbamic acid, H_2NCOOH. They are similar to organophosphates in that they are short-lived in the environment. They, therefore, share that advantage of not being around long enough to bioaccumulate in food chains, but they also share the disadvantage of high human toxicity. Popular carbamate pesticides include propoxur, carbaryl, and aldicarb.

Acute human exposure to carbamates has led to a range of symptoms such as nausea, vomiting, blurred vision, and, in extreme cases, convulsions and death.

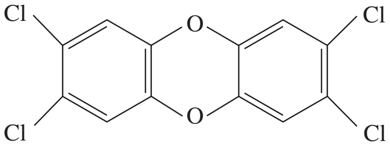


2,4-Dichlorophenoxyacetic acid
(2,4-D)



2,4,5-Trichlorophenoxyacetic acid
(2,4,5-T)

Chlorinated hydrocarbons are also used as herbicides. The chlorophenoxy compounds 2,4,5-T and 2,4,5-D are among the most well known because they were used as defoliants in the Vietnam War. Mixed together, they were called Agent Orange. They can kill broad-leafed plants without harming grasses, and have found use in controlling excessive growth of aquatic plants in lakes and reservoirs. The herbicide 2,4,5-T has been banned in part because the manufacturing process that produces 2,4,5-T also produces a highly toxic side-product, dioxin. Dioxins also enter the environment as products of combustion from incinerators. Other herbicides include paraquat, which acquired some fame as the pesticide of choice for destroying marijuana, and metolachlor, which is commonly used on soybeans and corn.



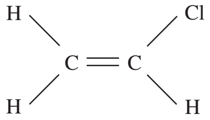
2,3,7,8-dioxin

There has been notable success in the past 25 years in producing less persistent, more degradable pesticides. Even though the pesticide may be degraded, its degradation products may move readily throughout the environment and, in themselves, be toxic to humans or other biota. In recent years, the transformation products of pesticides have been identified in both surface waters and groundwaters and, in a few cases, in treated drinking water (Coupe & Blomquist, 2004). There is growing concern that not only must we carefully manage and monitor the synthetic chemical products we manufacture but also the risks associated with their natural transformation products (Sinclair et al., 2006).

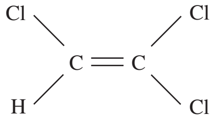
Volatile Organic Chemicals

Volatile organic chemicals (VOCs) are among the most commonly found contaminants in groundwater. They are often used as solvents in industrial processes and a number of them are either known or suspected carcinogens or mutagens. Their volatility means they are not often found in concentrations above a few μg/L in surface waters, but in groundwater, their concentrations can be hundreds or thousands of times higher. Their volatility also suggests the most common method of treatment, which is to aerate the water to encourage them to vaporize and disperse in the atmosphere. Since volatilization does not destroy the VOC but simply transfers it to the air, it is important that the resulting atmospheric concentration does not reach a level of concern.

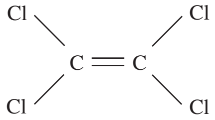
Five VOCs are especially toxic, and their presence in drinking water is cause for special concern: vinyl chloride, tetrachloroethylene, trichloroethylene, 1,2-dichloroethane, and carbon tetrachloride.



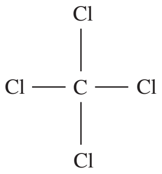
Vinyl chloride



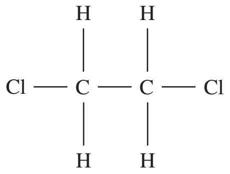
Trichloroethylene



Tetrachloroethylene



Carbon tetrachloride



1,2-Dichloroethane

The most toxic of the five is *vinyl chloride* (chloroethylene). It is a known human carcinogen used primarily in the production of polyvinyl chloride resins. *Tetrachloroethylene* is used as a solvent, as a heat transfer medium, and in the manufacture of chlorofluorocarbons. It causes tumors in animals, but there is inadequate evidence to call it a human carcinogen. Of the five, it is the one most commonly found in groundwater. *Trichloroethylene* (TCE) is a solvent that was quite commonly used to clean everything from electronics parts to jet engines and septic tanks. It is a suspected carcinogen and is among the most frequently found contaminants in groundwater. *1,2-dichloroethane* is a metal degreaser that is also used in the manufacture of a number of products, including vinyl chloride, tetraethyllead, fumigants, varnish removers, and soap compounds. Though it is not a known carcinogen, high levels of exposure are known to cause injury to the central nervous system, liver, and kidneys. It is also a common groundwater contaminant that is soluble, making it one of the more difficult to remove by air stripping. *Carbon tetrachloride* was a common household cleaning agent that is now more often used in grain fumigants, fire extinguishers, and solvents. It is very toxic if ingested; only a few milliliters can produce death. It is relatively insoluble in water, and therefore only occasionally found in contaminated groundwater.

Emerging Contaminants

As of January 2006, there were more than 239,000 substances on the CAS (Chemical Abstracts Service) list of regulated chemicals (CAS, 2006), and there are 4,800 chemicals whose production exceeds 1,000 tonnes/year (OECD, 2004). Not all potential pollutants are synthetic chemicals. Many are naturally occurring organisms or materials, such as pathogens and metals. This said, how is it possible to identify and prioritize those potential pollutants, or classes of pollutants, of most environmental concern? The answer is not straightforward but has to do with such things as the substance's environmental persistence; relative toxicity or deleterious impact (both to humans and other biota); occurrence frequency and concentration; and immediacy of impact. Since the 1990s, many previously little-recognized pollutants have become classified as *emerging contaminants*—contaminants that by meeting some combination of the foregoing attributes warrant particular interest and concern. Much of this trend in identifying and characterizing emerging contaminants impacts depends on improvements in instrumentation, sampling, and analytic techniques (Muir & Howard, 2006).

Foremost among the emerging contaminants are *endocrine disrupting chemicals* (EDCs). Since *Our Stolen Future* (Colborn) drew widespread attention to how EDCs may be interfering with the hormonal control of development in humans and wildlife, a wave of literature has been published documenting how numerous chemicals adversely affect normal endocrine system functioning. After the genetic code is implanted, an organism's development, behavior, and metabolism are largely dictated by an intricate system of chemical signaling and response controlled by the endocrine system. EDCs interfere with the natural functioning of this system either by being or acting like a natural hormone, blocking or counteracting the action of a natural hormone, or increasing or reducing the production of natural hormones. Hermaphroditism (the presence of both male and female reproductive organs in a single organism) and impaired offspring survival and development have been reported in frogs, salmon, oysters, newts, trout, and turtles exposed to EDCs. For instance, 84 percent of the female Chinook salmon in one Columbia River study were found to have undergone sex reversal early in development and were, in fact, chromosomally males (Nagler et al., 2001). Estrogenic compounds, like 17β-estradiol (the natural female sex hormone) and ethinylestradiol (the analog of 17β-estradiol that is in birth control pills), have received the most attention among EDCs and have been shown to cause reproductive and developmental effects in fish and amphibians at levels measured in the environment (Kavlock & Daston, 1996). At present, there is no firm evidence that the environmentally measured concentrations of EDCs are causing human health effects, although it has been suggested that EDCs are a factor in reduced sperm counts and increased incidences of breast and reproductive organ cancers (Snyder, Pleus, and Snyder, 2005). Hundreds of chemicals, including flame retardants, metals (e.g., arsenic, cadmium), pesticides, natural plant material, detergents, combustion byproducts, and pharmaceuticals are considered to be possible EDCs.

Polybrominated biphenyl ethers (PBDEs), which are used widely as flame retardant additives, and perfluorocarboxylates (PFCAs), which are waste and transformation products from the manufacture of fluorine containing surface treatments and lubricants, are now found in environments throughout the globe, including snow and mammal in the Arctic (Streets et al., 2006, and Scott et al., 2006). Both classes of compounds have been observed to bioaccumulate and possibly exhibit toxicity or endocrine disruption in wildlife and humans. With improved analytic techniques, many new pathogenic organisms have also been identified. Human adenoviruses have now been found around the world in surface waters (Jiang, 2006). Rapidly increasing incidences of antibiotic-resistant pathogens are a concern as they threaten to make medical remedies, such as penicillin, obsolete (Pruden et al., 2006). Bacterial pathogens have been found to grow in amoebas, which inhabit cooling towers, and are possibly spread through aqueous aerosols from the towers (Berk et al., 2006). In addition, manufacture and application of nanoparticles (materials with functionalities at the near-atomic scale) is growing exponentially along with the exposure of humans and the environment to these materials. Nanoparticles are now commonly found in paints, surface coatings, food additives, polishing compounds, industrial catalysts, and personal care products. However, basic study of the environmental fate and toxic impacts of these materials has only recently begun (Brunner et al., 2006). The foregoing list of emerging contaminants is not an exhaustive list but merely indicates the breadth of pollutants involved and the nature of some of the challenges for future environmental engineers and scientists.

4 Status of Surface Water Quality

As a way to help monitor progress toward the goals of the Clean Water Act, states are required to submit water quality assessments of their rivers, lakes, and estuaries. The starting point for these assessments is a designation of the *beneficial uses* that individual bodies of water shall be required to support. The EPA provides guidance on these beneficial uses, as shown in Table 7. The Clean Water Act authorizes the states to set their own water quality standards that accompany these beneficial uses, as long as those standards comply with the "fishable and swimmable" goals of the Act.

The water quality assessments that the states must submit identify sources of pollution and the fraction of the resource that is "impaired" by pollutants. A body of water is said to be impaired when at least one of the designated beneficial uses, such as recreational swimming or fish consumption, is not being supported by the quality of water. The water quality assessments submitted in 2000 showed that

TABLE 7

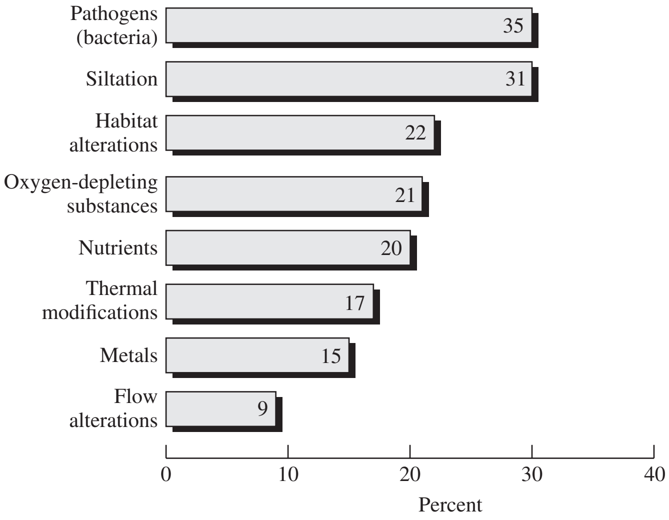
Beneficial Uses of Surface Water

Beneficial Use	Descriptor
Aquatic life support	The waterbody provides suitable habitat for survival and reproduction of desirable fish, shellfish, and other aquatic organisms.
Fish consumption	The waterbody supports a population of fish free from contamination that could pose a human health risk to consumers.
Shellfish harvesting	The waterbody supports a population of shellfish free from toxicants and pathogens that could pose a human health risk to consumers.
Drinking water supply	The waterbody can supply safe drinking water with conventional treatment.
Primary contact recreation	People can swim in the waterbody without risk of adverse human health effects (such as catching waterborne disease from raw sewage contamination).
Secondary contact recreation	People can perform activities on the water (such as kayaking) without risk of adverse human health effects from occasional contact with the water.
Agriculture	The water quality is suitable for irrigating fields or watering livestock.

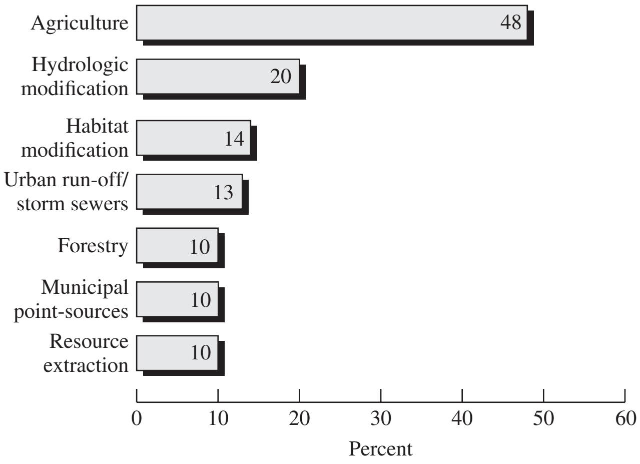
Source: U.S. EPA, 1994.

39 percent of the assessed miles of river, and 45 percent of the assessed area of lakes, were impaired for one reason or another (U.S. EPA, 2002).

A pollutant or a stressor may cause surface water quality impairment. Stressors are physical changes to the water body. For instance, a habitat alteration stressor encompasses activities that alter aquatic habitat (but not the water flow), such as removal of rocks from the bottom of a stream, whereas a flow alteration stressor includes activities such as damming a river or withdrawing irrigation water from a stream. The most important pollutants and stressors that cause impairment of water quality in rivers and streams are shown in Figure 7*a*, and the



(*a*) By pollutant/stressor

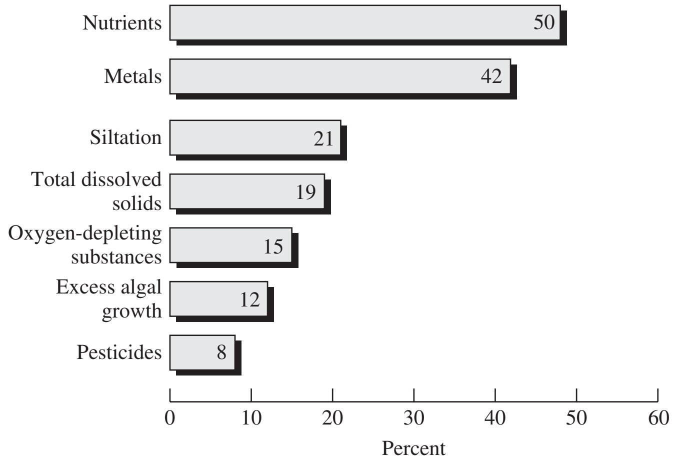


(*b*) By source

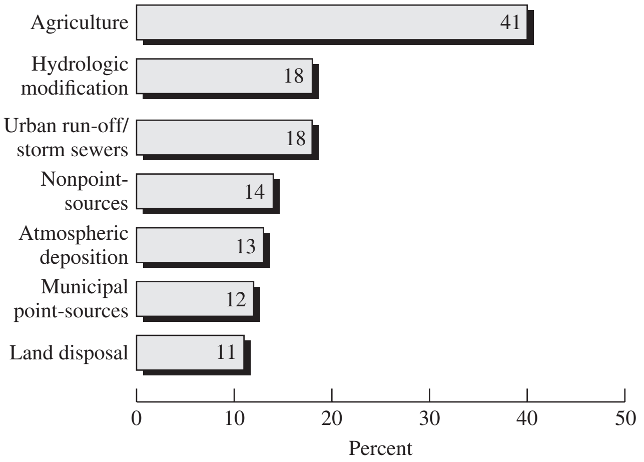
FIGURE 7 Fraction of impaired river and stream miles broken down by (*a*) pollutant/stressor and (*b*) sources of pollution.
(*Source:* U.S. EPA, 2002.)

leading sources of those problems are given in Figure 7*b*. The most common problems are pathogens (measured by bacterial counts) and siltation. Pathogens enter rivers from poorly treated sewage discharges, wildlife fecal matter, and run-off from farms and feedlots. The presence of pathogens suggests use of the water without further treatment may cause human illness. Siltation is caused by solids washed off of plowed fields, logging sites, urban areas, and stream banks when it rains. The resulting sediment load can smother habitats needed by fish and other aquatic organisms. In the United States, the leading source of river and stream impairment is agricultural activity, including crop production and animal husbandry.

The pollutants and sources of impairment for lakes and reservoirs are shown in Figure 8. Lakes and reservoirs are most often impaired by nutrients and metals coming from nonpoint-sources. Agricultural run-off causes the most problems, but



(*a*) By pollutant



(*a*) By source

FIGURE 8 Extent of impaired lake acres broken down by (*a*) pollutant and (*b*) sources of pollution.
(*Source:* U.S. EPA, 2002.)

urban run-off and hydrologic modification (which includes such things as flow regulation and dredging) are also very important. Because more than one pollutant or process may impact a given lake or river, the percentages in Figures 7 and 8 add up to more than 100 percent.

5 Biochemical Oxygen Demand

Surface water is obviously highly susceptible to contamination. It has historically been the most convenient sewer for industry and municipalities alike, while at the same time, it is the source of the majority of our water for all purposes. One particular category of pollutants, oxygen-demanding wastes, has been such a pervasive surface-water problem, affecting both moving water and still water, that it will be given special attention.

When biodegradable organic matter is released into a body of water, microorganisms, especially bacteria, feed on the wastes, breaking them down into simpler organic and inorganic substances. When this decomposition takes place in an *aerobic* environment—that is, in the presence of oxygen—the process produces nonobjectionable, stable end products such as carbon dioxide (CO_2), sulfate (SO_4), orthophosphate (PO_4), and nitrate (NO_3). A simplified representation of *aerobic decomposition* is given by the following:

$$\text{Organic matter} + O_2 \xrightarrow{\text{Microorganisms}} CO_2 + H_2O + \text{New cells} + \text{Stable products } (NO_3, PO_4, SO_4, \ldots)$$

When insufficient oxygen is available, the resulting *anaerobic* decomposition is performed by completely different microorganisms. They produce end products that can be highly objectionable, including hydrogen sulfide (H_2S), ammonia (NH_3), and methane (CH_4). *Anaerobic decomposition* can be represented by the following:

$$\text{Organic matter} \xrightarrow{\text{Microorganisms}} CO_2 + H_2O + \text{New cells} + \text{Unstable products } (H_2S, NH_3, CH_4, \ldots)$$

The methane produced is physically stable, biologically degradable, and a potent greenhouse gas. When emitted from bodies of water, it is often called swamp gas. It is also generated in the anaerobic environment of landfills, where it is sometimes collected and used as an energy source.

The amount of oxygen required by microorganisms to oxidize organic wastes aerobically is called the *biochemical oxygen demand* (BOD). BOD may have various units, but most often, it is expressed in milligrams of oxygen required per liter of wastewater (mg/L). The BOD, in turn, is made up of two parts: the *carbonaceous oxygen demand* (CBOD) and the *nitrogenous oxygen demand* (NBOD). Those distinctions will be clarified later.

Five-Day BOD Test

The total amount of oxygen that will be required for biodegradation is an important measure of the impact that a given waste will have on the receiving body of water. While we could imagine a test in which the oxygen required to degrade *completely* a sample of waste would be measured, for routine purposes, such a test would take too long to be practical (at least several weeks would be required). As a result, it has become standard practice simply to measure and report the oxygen demand over a shorter, restricted period of five days, realizing that the ultimate demand may be considerably higher.

The five-day BOD, or BOD_5, is the total amount of oxygen consumed by microorganisms during the first five days of biodegradation. In its simplest form, a BOD_5 test would involve putting a sample of waste into a stoppered bottle and measuring the concentration of dissolved oxygen (DO) in the sample at the beginning of the test and again five days later. The difference in DO divided by the volume of waste would be the five-day BOD. Light must be kept out of the bottle to keep algae from adding oxygen by photosynthesis, and the bottle is sealed to keep air from replenishing DO that has been removed by biodegradation. To standardize the procedure, the test is run at a fixed temperature of 20°C. Since the oxygen demand of typical waste is several hundred milligrams per liter, and the saturated value of DO for water at 20°C is only 9.1 mg/L, it is usually necessary to dilute the sample to keep the final DO above zero. If during the five days, the DO drops to zero, the test is invalid because more oxygen would have been removed had more been available.

The five-day BOD of a diluted sample is given by

$$\mathrm{BOD}_5 = \frac{\mathrm{DO_i} - \mathrm{DO_f}}{P} \tag{2}$$

where

$\mathrm{DO_i}$ = the initial dissolved oxygen (DO) of the diluted wastewater
$\mathrm{DO_f}$ = the DO of the diluted wastewater, 5 days later
P = the dilution fraction = $\dfrac{\text{volume of wastewater}}{\text{volume of wastewater plus dilution water}}$

A standard BOD bottle holds 300 mL, so P is just the volume of wastewater divided by 300 mL.

EXAMPLE 1 Unseeded Five-Day BOD Test

A 10.0-mL sample of sewage mixed with enough water to fill a 300-mL bottle has an initial DO of 9.0 mg/L. To help assure an accurate test, it is desirable to have at least a 2.0-mg/L drop in DO during the five-day run, and the final DO should be at least 2.0 mg/L. For what range of BOD_5 would this dilution produce the desired results?

Solution The dilution fraction is $P = 10/300$. To get at least a 2.0-mg/L drop in DO, the minimum BOD needs to be

$$\mathrm{BOD}_5 \geq \frac{\mathrm{DO_i} - \mathrm{DO_f}}{P} = \frac{2.0\ \mathrm{mg/L}}{(10/300)} = 60\ \mathrm{mg/L}$$

To assure at least 2.0 mg/L of DO remains after five days requires that

$$\mathrm{BOD}_5 \leq \frac{(9.0 - 2.0)\ \mathrm{mg/L}}{(10/300)} = 210\ \mathrm{mg/L}$$

This dilution will be satisfactory for BOD_5 values between 60 and 210 mg/L.

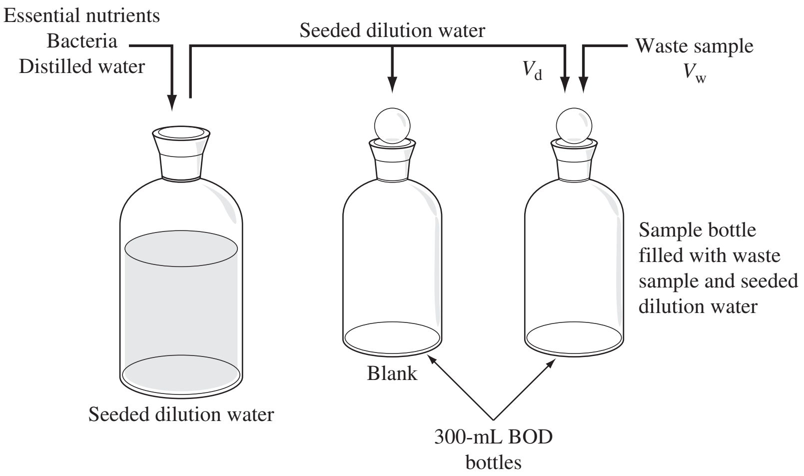


FIGURE 9 Laboratory test for BOD using seeded dilution water.

So far, we have assumed that the dilution water added to the waste sample has no BOD of its own, which would be the case if pure water were added. In some cases it is necessary to seed the dilution water with microorganisms to assure that there is an adequate bacterial population to carry out the biodegradation. In such cases, to find the BOD of the waste itself, it is necessary to subtract the oxygen demand caused by the seed from the demand in the mixed sample of waste and dilution water.

To sort out the effect of seeded dilution water from the waste itself, two BOD bottles must be prepared, one containing just the seeded dilution water and the other containing the mixture of both the wastewater and seeded dilution water (Figure 9). The change in DO in the bottle containing just seeded dilution water (called the "blank") as well as the change in DO in the mixture are then noted. The oxygen demand of the waste itself (BOD_w) can then be determined as follows:

$$\text{BOD}_m V_m = \text{BOD}_w V_w + \text{BOD}_d V_d \tag{3}$$

where

BOD_m = BOD of the mixture of wastewater and seeded dilution water
BOD_w = BOD of the wastewater alone
BOD_d = BOD of the seeded dilution water alone (the blank)
V_w = the volume of wastewater in the mixture
V_d = the volume of seeded dilution water in the mixture
V_m = the volume of the mixture = $V_d + V_w$

Let P = the fraction of the mixture that is wastewater = V_w/V_m so that $(1 - P)$ = the fraction of the mixture that is seeded dilution water = V_d/V_m. Rearranging (3) gives

$$\text{BOD}_w = \text{BOD}_m\left(\frac{V_m}{V_w}\right) - \text{BOD}_d\left(\frac{V_d}{V_w} \times \frac{V_m}{V_m}\right) \tag{4}$$

where the last term has been multiplied by unity (V_m/V_m). A slight rearrangement of (4) yields

$$\text{BOD}_w = \frac{\text{BOD}_m}{(V_w/V_m)} - \text{BOD}_d \frac{(V_d/V_m)}{(V_w/V_m)} \tag{5}$$

Substituting the definitions of P and $(1 - P)$ into (5) gives

$$\text{BOD}_w = \frac{\text{BOD}_m - \text{BOD}_d(1 - P)}{P} \tag{6}$$

Because

$$\text{BOD}_m = \text{DO}_i - \text{DO}_f \quad \text{and} \quad \text{BOD}_d = B_i - B_f$$

where

B_i = initial DO in the seeded dilution water (blank)
B_f = final DO in the seeded dilution water

our final expression for the BOD of the waste itself is thus

$$\text{BOD}_w = \frac{(\text{DO}_i - \text{DO}_f) - (B_i - B_f)(1 - P)}{P} \tag{7}$$

EXAMPLE 2 A Seeded BOD Test

A test bottle containing just seeded dilution water has its DO level drop by 1.0 mg/L in a five-day test. A 300-mL BOD bottle filled with 15 mL of wastewater and the rest seeded dilution water (sometimes expressed as a dilution of 1:20) experiences a drop of 7.2 mg/L in the same time period. What would be the five-day BOD of the waste?

Solution The dilution factor P is

$$P = 15/300 = 0.05$$

Using (7), the five-day BOD of the waste would be

$$\text{BOD}_5 = \frac{7.2 - 1.0(1 - 0.05)}{0.05} = 125 \text{ mg/L}$$

Modeling BOD as a First-Order Reaction

Suppose we imagine a flask with some biodegradable organic waste in it. As bacteria oxidize the waste, the amount of organic matter remaining in the flask will decrease with time until eventually it all disappears. Another way to describe the organic matter in the flask is to say as time goes on, the amount of organic matter already oxidized goes up until finally all of the original organic matter has been oxidized. Figure 10 shows these two equivalent ways to describe the organic matter. We can also describe oxygen demand from those same two perspectives. We could say that the remaining demand for oxygen to decompose the wastes decreases with time until there is no more demand, or we could say the amount of oxygen

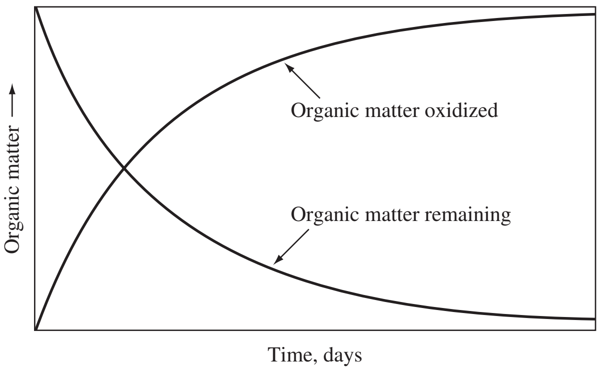


FIGURE 10 Two equivalent ways to describe the time dependence of organic matter in a flask.

demand already exerted, or utilized, starts at zero and rises until all of the original oxygen demand has been satisfied.

Translating Figure 10 into a mathematical description is straightforward. To do so, it is assumed that the rate of decomposition of organic wastes is proportional to the amount of waste that is left in the flask. If we let L_t represent the amount of oxygen demand left after time t, then, assuming a first-order reaction, we can write

$$\frac{dL_t}{dt} = -kL_t \tag{8}$$

where $k =$ the BOD reaction constant (time^{-1}).

The solution to (8) is

$$L_t = L_0 e^{-kt} \tag{9}$$

where L_0 is the *ultimate carbonaceous oxygen demand.* It is the total amount of oxygen required by microorganisms to oxidize the carbonaceous portion of the waste to simple carbon dioxide and water. (Later, we will see that there is an additional demand for oxygen associated with the oxidation of nitrogen compounds.) The ultimate carbonaceous oxygen demand is the sum of the amount of oxygen already consumed by the waste in the first t days (BOD_t), plus the amount of oxygen remaining to be consumed after time t. That is,

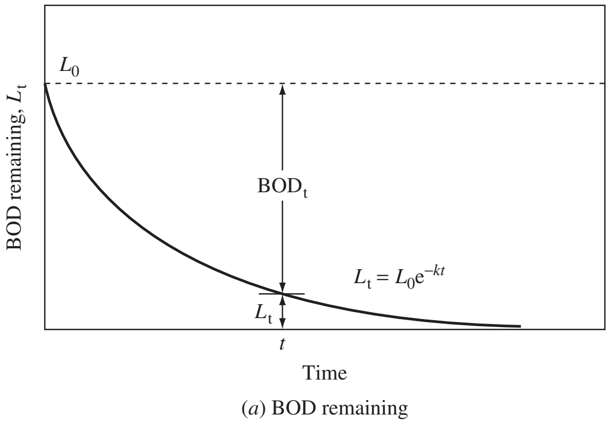
$$L_0 = \mathrm{BOD}_t + L_t \tag{10}$$

Combining (9) and (10) gives us

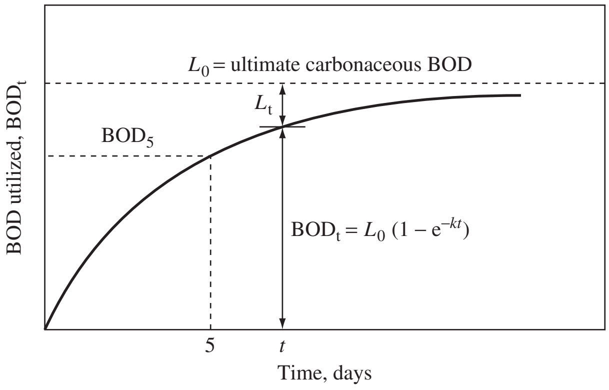
$$\mathrm{BOD}_t = L_0(1 - e^{-kt}) \tag{11}$$

A graph of Eqs. (9) and (11) is presented in Figure 11. If these two figures are combined, the result would look exactly like Figure 10. Notice that oxygen demand can be described by the BOD remaining (you might want to think of L_t as how much oxygen demand is *L*eft at time *t*), as in Figure 11*a*, or equivalently as oxygen demand already satisfied (or utilized, or exerted), BOD_t, as in Figure 11*b*. Also notice how the five-day BOD is more easily described using the BOD utilized curve.

Sometimes the analysis leading to (11) is made using logarithms to the base 10 rather than the base *e*, as they were here. The relationship equivalent to (11), but in



(*a*) BOD remaining



(*b*) BOD utilized

FIGURE 11 Idealized carbonaceous oxygen demand: (*a*) The BOD remaining as a function of time, and (*b*) the oxygen already consumed as a function of time.

base 10, is

$$\text{BOD}_t = L_0(1 - 10^{-Kt}) \tag{12}$$

where uppercase K is the reaction rate coefficient to the base 10. It is easy to show that

$$k = K \ln 10 = 2.303K \tag{13}$$

EXAMPLE 3 Estimating L_0 from BOD_5

The dilution factor P for an unseeded mixture of waste and water is 0.030. The DO of the mixture is initially 9.0 mg/L, and after five days, it has dropped to 3.0 mg/L. The reaction rate constant k has been found to be $0.22\ \text{day}^{-1}$.

a. What is the five-day BOD of the waste?
b. What would be the ultimate carbonaceous BOD?
c. What would be the remaining oxygen demand after five days?

Solution

a. From (2), the oxygen consumed in the first five days is

$$\mathrm{BOD}_5 = \frac{\mathrm{DO_i} - \mathrm{DO_f}}{P} = \frac{9.0 - 3.0}{0.030} = 200 \text{ mg/L}$$

b. The total amount of oxygen needed to decompose the carbonaceous portion of the waste can be found by rearranging (11):

$$L_0 = \frac{\mathrm{BOD}_5}{(1 - e^{-kt})} = \frac{200}{(1 - e^{-0.22\times 5})} = 300 \text{ mg/L}$$

c. After five days, 200 mg/L of oxygen demand out of the total 300 mg/L would have already been used. The remaining oxygen demand would therefore be (300 − 200) mg/L = 100 mg/L.

The BOD Reaction Rate Constant *k*

The BOD reaction rate constant k is a factor that indicates the rate of biodegradation of wastes. As k increases, the rate at which dissolved oxygen is used increases, although the ultimate amount required, L_0, does not change. The reaction rate will depend on a number of factors, including the nature of the waste itself (for example, simple sugars and starches degrade easily while cellulose does not), the ability of the available microorganisms to degrade the wastes in question (it may take some time for a healthy population of organisms to be able to thrive on the particular waste in question), and the temperature (as temperatures increase, so does the rate of biodegradation).

Some typical values of the BOD reaction rate constant, at 20°C, are given in Table 8. Notice that raw sewage has a higher rate constant than either well-treated sewage or polluted river water. This is because raw sewage contains a larger proportion of easily degradable organics that exert their oxygen demand quite quickly, leaving a remainder that decays more slowly.

The rate of biodegradation of wastes increases with increasing temperature. To account for these changes, the reaction rate constant k is often modified using the following equation:

$$k_{\mathrm{T}} = k_{20}\theta^{(T-20)} \tag{14}$$

TABLE 8

Typical Values for the BOD Rate Constant at 20°C

Sample	$k(\text{day}^{-1})^a$	$K(\text{day}^{-1})^b$
Raw sewage	0.35–0.70	0.15–0.30
Well-treated sewage	0.12–0.23	0.05–0.10
Polluted river water	0.12–0.23	0.05–0.10

[a] Lowercase k reaction rates to the base e.

[b] Uppercase K reaction rates to the base 10.

Source: Davis and Cornwell, 1991.

where k_{20} is the reaction rate constant at the standard 20°C laboratory reference temperature, and k_T is the reaction rate at a different temperature T (expressed in °C). The most commonly used value for θ is 1.047, and although θ is somewhat temperature dependent, that single value will suffice for our purposes.

EXAMPLE 4 Temperature Dependence of BOD_5

In Example 3, the wastes had an ultimate BOD equal to 300 mg/L. At 20°C, the five-day BOD was 200 mg/L, and the reaction rate constant was 0.22/day. What would the five-day BOD of this waste be at 25°C?

Solution First we will adjust the reaction rate constant with (14) using a value of θ equal to 1.047:

$$k_{25} = k_{20}\theta^{(T-20)} = 0.22 \times (1.047)^{(25-20)} = 0.277/\text{day}$$

So, from (12),

$$\text{BOD}_5 = L_0(1 - e^{-k5}) = 300(1 - e^{-0.277\times 5}) = 225 \text{ mg/L}$$

Notice that the five-day BOD at 25°C is somewhat higher than the 20°C value of 200 mg/L. The same total amount of oxygen is required at either temperature, but as temperature increases, it gets used sooner.

Nitrification

So far, it has been assumed that the only oxygen demand is associated with the biodegradation of the carbonaceous portion of the wastes. There is a significant additional demand, however, caused by the oxidation of nitrogen compounds.

Nitrogen is the critical element required for protein synthesis and, hence, is essential to life. When living things die or excrete waste products, nitrogen that was tied to complex organic molecules is converted to ammonia (NH_3) by bacteria and fungi. Then, in aerobic environments, nitrite bacteria (*Nitrosomonas*) convert ammonia to nitrite (NO_2^-), and nitrate bacteria (*Nitrobacter*) convert nitrite to nitrate (NO_3^-). This process, called nitrification, can be represented with the following two reactions:

$$2\,NH_3 + 3\,O_2 \xrightarrow{\text{Nitrosomonas}} 2\,NO_2^- + 2\,H^+ + 2\,H_2O \tag{15}$$

$$2\,NO_2^- + O_2 \xrightarrow{\text{Nitrobacter}} 2\,NO_3^- \tag{16}$$

This conversion of ammonia to nitrate requires oxygen, so nitrification exerts its own oxygen demand. Thus, we have the combination of oxygen requirements. The oxygen needed to oxidize organic carbon to carbon dioxide is called the *carbonaceous oxygen demand* (CBOD), and the oxygen needed to convert ammonia to nitrate is called the *nitrogenous oxygen demand* (NBOD).

Nitrification is just one part of the biogeochemical cycle for nitrogen. In the atmosphere nitrogen is principally in the form of molecular nitrogen (N_2) with a

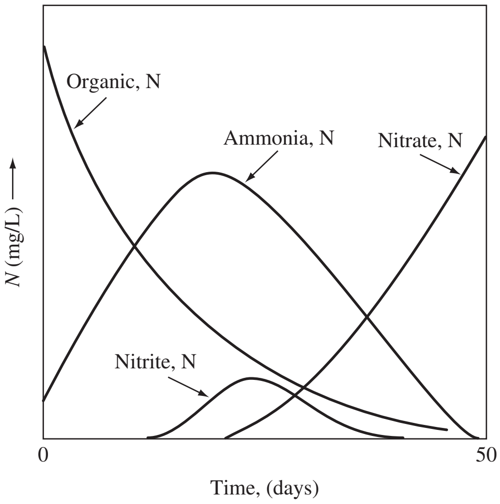


FIGURE 12 Changes in nitrogen forms in polluted water under aerobic conditions. (*Source:* Sawyer and McCarty, 1994. Reprinted by permission of McGraw-Hill, Inc.)

small but important fraction being nitrous oxide (N_2O). (Nitrous oxide is a greenhouse gas.) Nitrogen in the form of N_2 is unusable by plants and must first be transformed into either ammonia (NH_3) or nitrate (NO_3^-) in the process called *nitrogen fixation.* Nitrogen fixation occurs during electrical storms when N_2 oxidizes, combines with water, and is rained out as HNO_3. Certain bacteria and blue-green algae are also capable of fixing nitrogen. Under anaerobic conditions, certain denitrifying bacteria are capable of reducing NO_3^- back into NO_2 and N_2, completing the nitrogen cycle.

The entire nitrogen cycle obviously is important, but our concern in this section is with the nitrification process itself, in which organic-nitrogen in waste is converted to ammonia, ammonia to nitrite, and nitrite to nitrate. Figure 12 shows this sequential process, starting with all of the nitrogen bound up in organic form and weeks later ending with all of the nitrogen in the form of nitrate. Notice that the conversion of ammonia to nitrite does not begin right away, which means the nitrogenous biochemical oxygen demand does not begin to be exerted until a number of days have passed.

Figure 13 illustrates the carbonaceous and nitrogenous oxygen demands as they might be exerted for typical municipal wastes. Notice that the NBOD does not normally begin to exert itself for at least five to eight days, so most five-day tests are not affected by nitrification. In fact, the potential for nitrification to interfere with the standard measurement for CBOD was an important consideration in choosing the standard five-day period for BOD tests. To avoid further nitrification complications, it is now an accepted practice to modify wastes in a way that will inhibit nitrification during that five-day period.

A stoichiometric analysis of (15) and (16) allows us to quantify the oxygen demand associated with nitrification, as the following example illustrates.

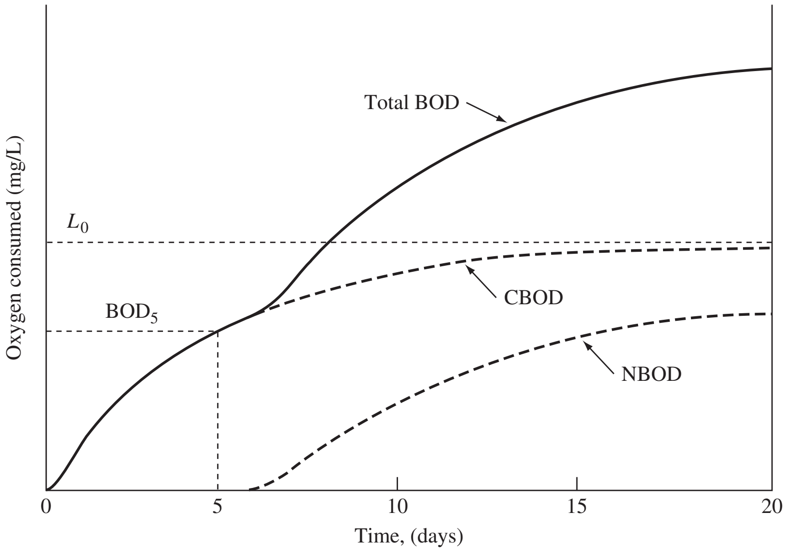


FIGURE 13 Illustrating the carbonaceous and nitrogenous biochemical oxygen demand. Total BOD is the sum of the two.

EXAMPLE 5 Nitrogenous Oxygen Demand

Some domestic wastewater has 30 mg/L of nitrogen in the form of either organic nitrogen or ammonia. Assuming that very few new cells of bacteria are formed during the nitrification of the waste (that is, the oxygen demand can be found from a simple stoichiometric analysis of the nitrification reactions given above), find

a. The ultimate nitrogenous oxygen demand
b. The ratio of the ultimate NBOD to the concentration of nitrogen in the waste.

Solution

a. Combining the two nitrification reactions (15) and (16) yields

$$NH_3 + 2O_2 \rightarrow NO_3^- + H^+ + H_2O \tag{17}$$

The molecular weight of NH_3 is 17, and the molecular weight of O_2 is 32. The foregoing reaction indicates that 1 g-mol of NH_3 (17 g) requires 2 g-mole of O_2 ($2 \times 32 = 64$ g). Since 17 g of NH_3 contains 14 g of N, and the concentration of N is 30 mg/L, we can find the final, or ultimate, NBOD:

$$\text{NBOD} = 30 \text{ mg N/L} \times \frac{17\text{g NH}_3}{14 \text{ g N}} \times \frac{64 \text{ g O}_2}{17 \text{ g NH}_3} = 137 \text{ mg O}_2/\text{L}$$

b. The oxygen demand due to nitrification divided by the concentration of nitrogen in the waste is

$$\frac{137 \text{ mg O}_2/\text{L}}{30 \text{ mg N/L}} = 4.57 \text{ mg O}_2/\text{mg N}$$

The total concentration of organic and ammonia nitrogen in wastewater is known as the *total Kjeldahl nitrogen* (TKN). As was demonstrated in the preceding example, the nitrogenous oxygen demand can be estimated by multiplying the TKN by 4.57. This is a result worth noting:

$$\text{Ultimate NBOD} \approx 4.57 \times \text{TKN} \tag{18}$$

Since untreated domestic wastewaters typically contain approximately 15–50 mg/L of TKN, the oxygen demand caused by nitrification is considerable, ranging from roughly 70 to 230 mg/L. For comparison, typical raw sewage has an ultimate carbonaceous oxygen demand of 250–350 mg/L.

Other Measures of Oxygen Demand

In addition to the CBOD and NBOD measures already presented, two other indicators are sometimes used to describe the oxygen demand of wastes. These are the *theoretical oxygen demand* (ThOD) and the *chemical oxygen demand* (COD).

The theoretical oxygen demand is the amount of oxygen required to oxidize completely a particular organic substance, as calculated from simple stoichiometric considerations. Stoichiometric analysis, however, for both the carbonaceous and nitrogenous components, tends to overestimate the amount of oxygen actually consumed during decomposition. The explanation for this discrepancy is based on a more careful understanding of how microorganisms actually decompose waste. While there is plenty of food for bacteria, they rapidly consume waste, and in the process, convert some of it to cell tissue. As the amount of remaining wastes diminishes, bacteria begin to draw on their own tissue for the energy they need to survive, a process called endogenous respiration. Eventually, as bacteria die, they become the food supply for other bacteria; all the while, protozoa act as predators, consuming both living and dead bacteria. Throughout this sequence, more and more of the original waste is consumed until finally all that remains is some organic matter, called humus, that stubbornly resists degradation. The discrepancy between theoretical and actual oxygen demands is explained by carbon still bound up in humus. The calculation of theoretical oxygen demand is of limited usefulness in practice because it presupposes a particular, single pollutant with a known chemical formula. Even if that is the case, the demand is overestimated.

Some organic matter, such as cellulose, phenols, benzene, and tannic acid, resist biodegradation. Other types of organic matter, such as pesticides and various industrial chemicals, are nonbiodegradable because they are toxic to microorganisms. The chemical oxygen demand (COD) is a measured quantity that does not depend either on the ability of microorganisms to degrade the waste or on knowledge of the particular substances in question. In a COD test, a strong chemical oxidizing agent is used to oxidize the organics rather than relying on microorganisms to do the job. The COD test is much quicker than a BOD test, taking only a matter of hours. However, it does not distinguish between the oxygen demand that will actually be felt in a natural environment due to biodegradation and the chemical oxidation of inert organic matter. It also does not provide any information on the rate at which actual biodegradation will take place. The measured value of COD is higher than BOD, though for easily biodegradable matter, the two will be similar. In fact, the COD test is sometimes used as a way to estimate the ultimate BOD.

6 The Effect of Oxygen-Demanding Wastes on Rivers

The amount of dissolved oxygen in water is one of the most commonly used indicators of a river's health. As DO drops below 4 or 5 mg/L, the forms of life that can survive begin to be reduced. In the extreme case, when anaerobic conditions exist, most higher forms of life are killed or driven off. Noxious conditions then prevail, including floating sludges; bubbling, odorous gases; and slimy fungal growths.

A number of factors affect the amount of DO available in a river. Oxygen-demanding wastes remove DO; photosynthesis adds DO during the day, but those plants remove oxygen at night; and the respiration of organisms living in the water as well as in sediments removes oxygen. In addition, tributaries bring their own oxygen supplies, which mix with those of the main river. In the summer, rising temperatures reduce the solubility of oxygen, while lower flows reduce the rate at which oxygen enters the water from the atmosphere. In the winter, ice may form, blocking access to new atmospheric oxygen. To model properly all of these effects and their interactions is a difficult task. A simple analysis, however, can provide insight into the most important parameters that affect DO. We should remember, however, that our results are only a first approximation to reality.

The simplest model of the oxygen resources in a river focuses on two key processes: the removal of oxygen by microorganisms during biodegradation, and the replenishment of oxygen through reaeration at the interface between the river and the atmosphere. In this simple model, it is assumed that there is a continuous discharge of waste at a given location on the river. As the water and wastes flow downriver, it is assumed that they are uniformly mixed at any given cross section of river, and it is assumed that there is no dispersion of wastes in the direction of flow. These assumptions are part of what is referred to as the *point-source, plug flow* model, illustrated in Figure 14.

Deoxygenation

The rate of deoxygenation at any point in the river is assumed to be proportional to the BOD remaining at that point. That is,

$$\text{Rate of deoxygenation} = k_d L_t \tag{19}$$

where

k_d = the deoxygenation rate constant (day^{-1})
L_t = the BOD remaining t (days) after the wastes enter the river, (mg/L)

The deoxygenation rate constant k_d is often assumed to be the same as the (temperature adjusted) BOD rate constant k obtained in a standard laboratory BOD test. For deep, slowly moving rivers, this seems to be a reasonable approximation, but for turbulent, shallow, rapidly moving streams, the approximation is less valid. Such streams have deoxygenation constants that can be significantly higher than the values determined in the laboratory.

Substituting (9), which gives BOD remaining after time t, into (19) gives

$$\text{Rate of deoxygenation} = k_d L_0 e^{-k_d t} \tag{20}$$

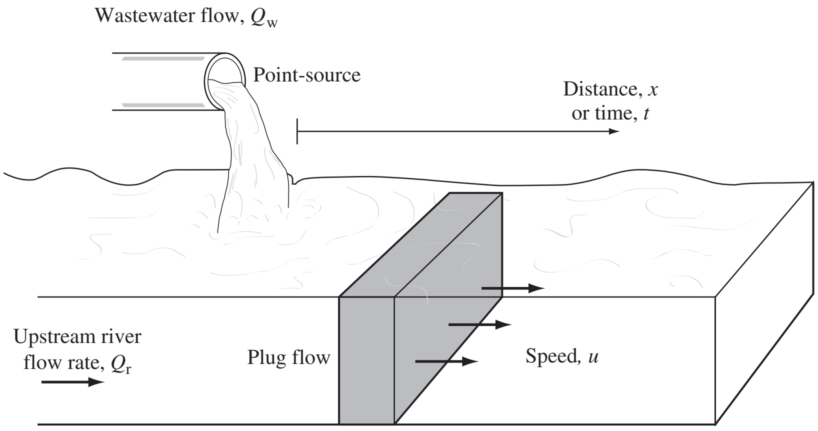


FIGURE 14 The point-source, plug flow model for dissolved-oxygen calculations.

where L_0 is the BOD of the mixture of streamwater and wastewater at the point of discharge. Assuming complete and instantaneous mixing,

$$L_0 = \frac{Q_w L_w + Q_r L_r}{Q_w + Q_r} \tag{21}$$

where

L_0 = ultimate BOD of the mixture of streamwater and wastewater (mg/L)
L_r = ultimate BOD of the river just upstream of the point of discharge (mg/L)
L_w = ultimate BOD of the wastewater (mg/L)
Q_r = volumetric flow rate of the river just upstream of the discharge point (m^3/s)
Q_w = volumetric flow rate of wastewater (m^3/s)

EXAMPLE 6 Downstream BOD

A wastewater treatment plant serving a city of 200,000 discharges 1.10 m^3/s of treated effluent having an ultimate BOD of 50.0 mg/L into a stream that has a flow of 8.70 m^3/s and a BOD of its own equal to 6.0 mg/L. The deoxygenation constant, k_d, is 0.20/day.

a. Assuming complete and instantaneous mixing, estimate the ultimate BOD of the river just downstream from the outfall.
b. If the stream has a constant cross section, so that it flows at a fixed speed equal to 0.30 m/s, estimate the BOD remaining in the stream at a distance 30,000 m downstream.

Solution

a. The BOD of the mixture of effluent and streamwater can be found using (21):

$$L_0 = \frac{1.10\ \text{m}^3/\text{s} \times 50.0\ \text{mg/L} + 8.70\ \text{m}^3/\text{s} \times 6.0\ \text{mg/L}}{(1.10 + 8.70)\ \text{m}^3/\text{s}} = 10.9\ \text{mg/L}$$

b. At a speed of 0.30 m/s, the time required for the waste to reach a distance 30,000 m downstream would be

$$t = \frac{30{,}000\ \text{m}}{0.30\ \text{m/s}} \times \frac{\text{hr}}{3600\ \text{s}} \times \frac{\text{day}}{24\ \text{hr}} = 1.16\ \text{days}$$

So the BOD remaining at that point, 30 km downstream, would be

$$L_t = L_0 e^{-k_d t} = 10.9\ e^{-(0.2/\text{d} \times 1.16\text{d})} = 8.7\ \text{mg/L}$$

Reaeration

The rate at which oxygen is replenished is assumed to be proportional to the difference between the actual DO in the river at any given location and the saturated value of dissolved oxygen. This difference is called the oxygen deficit, D:

$$\text{Rate of reaeration} = k_r D \tag{22}$$

where

k_r = reaeration constant (time^{-1})
D = dissolved oxygen deficit = $(\text{DO}_s - \text{DO})$ (23)
DO_s = saturated value of dissolved oxygen
DO = actual dissolved oxygen at a given location downstream

The reaeration constant, k_r, is very much dependent on the particular conditions in the river. A fast-moving, whitewater river will have a much higher reaeration constant than a sluggish stream or a pond. Many attempts have been made empirically to relate key stream parameters to the reaeration constant, with one of the most commonly used formulations being the following (O'Connor and Dobbins, 1958):

$$k_r = \frac{3.9\ u^{1/2}}{H^{3/2}} \tag{24}$$

where

k_r = reaeration coefficient at 20°C (day^{-1})
u = average stream velocity (m/s)
H = average stream depth (m)

Typical values of the reaeration constant k_r for various bodies of water are given in Table 9. Adjustments to the reaeration rate constant for temperatures other than 20°C can be made using (14) but with a temperature coefficient θ equal to 1.024.

The saturated value of dissolved oxygen varies with temperature, atmospheric pressure, and salinity. Table 10 gives representative values of the solubility of oxygen in water at various temperatures and chloride concentrations.

Both the wastewater that is being discharged into a stream and the stream itself are likely to have some oxygen deficit. If we assume complete mixing of the two, we can calculate the initial deficit of the polluted river using a weighted average based on their individual concentrations of dissolved oxygen:

$$D_0 = \text{DO}_s - \frac{Q_w\,\text{DO}_w + Q_r\,\text{DO}_r}{Q_w + Q_r} \tag{25}$$

TABLE 9

Typical Reaeration Constants for Various Bodies of Water

Water Body	Range of k_r at 20°C (day^{-1})[a]
Small ponds and backwaters	0.10–0.23
Sluggish streams and large lakes	0.23–0.35
Large streams of low velocity	0.35–0.46
Large streams of normal velocity	0.46–0.69
Swift streams	0.69–1.15
Rapids and waterfalls	>1.15

[a] Base *e*.

Source: Tchobanoglous and Schroeder, 1985.

TABLE 10

Solubility of Oxygen in Water (mg/L) at 1 atm Pressure

	Chloride Concentration in Water (mg/L)			
Temperature (°C)	0	5,000	10,000	15,000
0	14.62	13.73	12.89	12.10
5	12.77	12.02	11.32	10.66
10	11.29	10.66	10.06	9.49
15	10.08	9.54	9.03	8.54
20	9.09	8.62	8.17	7.75
25	8.26	7.85	7.46	7.08
30	7.56	7.19	6.85	6.51

Source: Thomann and Mueller, 1987.

where

D_0 = initial oxygen deficit of the mixture of river and wastewater

DO_s = saturated value of DO in water at the temperature of the river

DO_w = DO in the wastewater

DO_r = DO in the river just upstream of the wastewater discharge point

EXAMPLE 7 Initial Oxygen Deficit

The waste water in Example 6 has a dissolved oxygen concentration of 2.0 mg/L and a discharge rate of 1.10 m^3/s. The river that is receiving this waste has DO equal to 8.3 mg/L, a flow rate of 8.70 m^3/s, and a temperature of 20°C. Assuming complete and instantaneous mixing, estimate the initial dissolved oxygen deficit of the mixture of wastewater and river water just downstream from the discharge point.

Solution The initial amount of dissolved oxygen in the mixture of waste and river water would be

$$\text{DO} = \frac{1.10\ \text{m}^3/\text{s} \times 2.0\ \text{mg/L} + 8.70\ \text{m}^3/\text{s} \times 8.3\ \text{mg/L}}{(1.10 + 8.70)\ \text{m}^3/\text{s}} = 7.6\ \text{mg/L}$$

The saturated value of dissolved oxygen DO_s at 20°C is given in Table 10 as 9.09 mg/L, so the initial deficit would be

$$D_0 = 9.09 \text{ mg/L} - 7.6 \text{ mg/L} = 1.5 \text{ mg/L}$$

The Oxygen Sag Curve

The deoxygenation caused by microbial decomposition of wastes and oxygenation by reaeration are competing processes that are simultaneously removing and adding oxygen to a stream. Combining the two equations (20) and (22) yields the following expression for the rate of increase of the oxygen deficit:

Rate of increase of the deficit = Rate of deoxygenation − Rate of reaeration

$$\frac{dD}{dt} = k_d L_0 e^{-k_d t} - k_r D \tag{26}$$

which has the solution

$$D = \frac{k_d L_0}{k_r - k_d}\left(e^{-k_d t} - e^{-k_r t}\right) + D_0 e^{-k_r t} \tag{27}$$

Since the deficit D is the difference between the saturation value of dissolved oxygen DO_s and the actual value DO, we can write the equation for the DO as

$$DO = DO_s - \left[\frac{k_d L_0}{k_r - k_d}\left(e^{-k_d t} - e^{-k_r t}\right) + D_0 e^{-k_r t}\right] \tag{28}$$

Equation (28) is the classic *Streeter-Phelps oxygen sag equation* first described in 1925. A plot of DO following Streeter-Phelps behavior is given in Figure 15. As can be seen, there is a stretch of river immediately downstream of the discharge point where the DO drops rapidly. At the *critical point* downstream, dissolved oxygen reaches its minimum value and river conditions are at their worst. Beyond the critical point, the remaining organic matter in the river has diminished to the point where

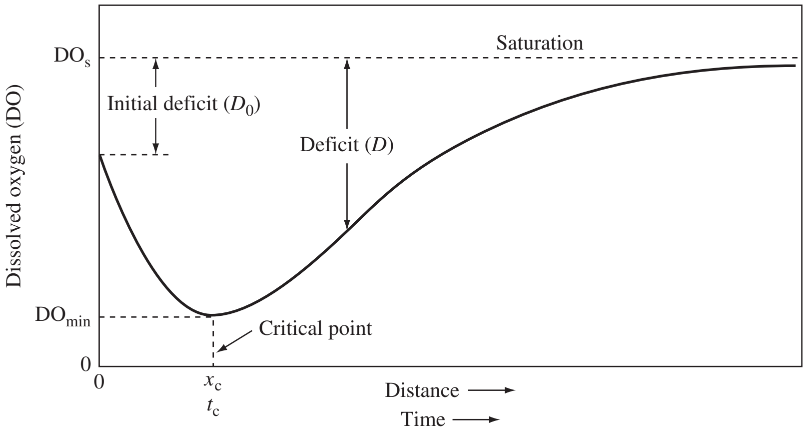


FIGURE 15 Streeter-Phelps oxygen sag curve.

oxygen is being added to the river by reaeration faster than it is being withdrawn by decomposition, and the river begins to recover.

We can add a few more things to this analysis of oxygen deficit. For the special case where $k_r = k_d$, the denominator in (27) goes to zero, which is mathematically unacceptable, so the equation needs to be rederived under those conditions. Under these circumstances, the solution to (26) becomes

$$D = (k_d L_0 t + D_0)e^{-k_d t} \tag{29}$$

If the stream has a constant cross-sectional area and is traveling at a speed u, then time and distance downstream are related by

$$x = ut \tag{30}$$

where

x = distance downstream
u = stream speed
t = elapsed time between discharge point and distance x downstream

Equation (27) can be rewritten as

$$D = \frac{k_d L_0}{k_r - k_d}\left(e^{-k_d x/u} - e^{-k_r x/u}\right) + D_0 e^{-k_r x/u} \tag{31}$$

The location of the critical point and the corresponding minimum value of DO is of obvious importance. At this point stream conditions are at their worst. Setting the derivative of the oxygen deficit equal to zero and solving for the critical time yields

$$t_c = \frac{1}{k_r - k_d}\ln\left\{\frac{k_r}{k_d}\left[1 - \frac{D_0(k_r - k_d)}{k_d L_0}\right]\right\} \tag{32}$$

The maximum deficit can then be found by substituting the value obtained for the critical time, t_c, into (28).

The oxygen sag curve should make some intuitive sense, even without the mathematical analysis. Near the outfall, there is so much organic matter being degraded that the rate of removal of oxygen from the water is higher than the rate that it can be returned by reaeration, so the dissolved oxygen drops. As we move further downstream, less and less organic matter remains, so the rate of removal of oxygen keeps dropping as well. At the critical point, the rate of removal of oxygen equals the rate of addition of oxygen by reaeration. Beyond the critical point, reaeration begins to dominate, returning oxygen to the river at a faster rate than the bacteria remove it, so the dissolved oxygen begins to climb back to the saturation value. Figure 16 shows the rate of deoxygenation, the rate of reaeration, and the oxygen sag curve.

EXAMPLE 8 Streeter-Phelps Oxygen Sag Curve

Just below the point where a continuous discharge of pollution mixes with a river, the BOD is 10.9 mg/L, and DO is 7.6 mg/L. The river and waste mixture has a temperature of 20°C, a deoxygenation constant k_d of 0.20/day, an average flow speed of 0.30 m/s, and an average depth of 3.0 m. (In other words, this is just a continuation of the problem started in Examples 6 and 7.)

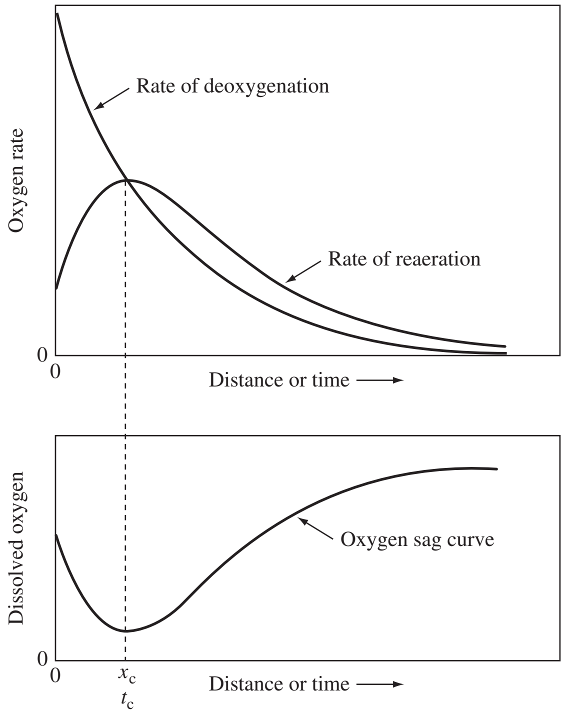


FIGURE 16 When the rate of deoxygenation exceeds the rate of reaeration, the DO in the river drops. At the critical point, those rates are equal. Beyond the critical point, reaeration exceeds decomposition, the DO curve climbs toward saturation, and the river recovers.

a. Find the time and distance downstream at which the oxygen deficit is at a maximum.

b. Find the minimum value of DO.

Solution From Table 10, the saturation value of DO at 20°C is 9.1 mg/L, so the initial deficit is

$$D_0 = 9.1 - 7.6 = 1.5 \text{ mg/L}$$

To estimate the reaeration constant, we can use the O'Connor and Dobbins relationship given in (24):

$$k_r = \frac{3.9\, u^{1/2}}{H^{3/2}} = \frac{3.9\,(0.30)^{1/2}}{(3.0)^{3/2}} = 0.41/\text{day}$$

a. Using (32), we can find the time at which the deficit is a maximum:

$$t_c = \frac{1}{k_r - k_d}\ln\left\{\frac{k_r}{k_d}\left[1 - \frac{D_0(k_r - k_d)}{k_d L_0}\right]\right\}$$

$$= \frac{1}{(0.41 - 0.20)}\ln\left\{\frac{0.41}{0.20}\left[1 - \frac{1.5(0.41 - 0.20)}{0.20 \times 10.9}\right]\right\} = 2.67 \text{ days}$$

so the critical distance downstream would be

$$x_c = ut_c = 0.30 \text{ m/s} \times 3{,}600 \text{ s/hr} \times 24 \text{ hr/d} \times 2.67 \text{ d}$$
$$= 69{,}300 \text{ m} = 69.3 \text{ km}$$

which is about 43 miles.

b. The maximum deficit can be found from (27):

$$D = \frac{k_d L_0}{k_r - k_d}(e^{-k_d t} - e^{-k_r t}) + D_0\, e^{-k_r t}$$
$$= \frac{0.20 \times 10.9}{(0.41 - 0.20)}(e^{-0.20 \times 2.67} - e^{-0.41 \times 2.67}) + 1.5e^{-0.41 \times 2.67} = 3.1 \text{ mg/L}$$

so the minimum value of DO will be the saturation value minus this maximum deficit:

$$\text{DO}_{\text{min}} = (9.1 - 3.1) \text{ mg/L} = 6.0 \text{ mg/L}$$

In the preceding example, the lowest value of DO was found to be 6.0 mg/L, an amount sufficient for most aquatic life. If the amount of BOD added to the river is excessive, the oxygen sag curve may drop below a minimum acceptable level, leading to a stretch of river with unhealthful conditions, as shown in Figure 17. Fish that cannot tolerate water with such low oxygen will be driven elsewhere, or they will die. The variety of animals that can inhabit the unhealthy region is diminished, and less desirable forms take over, forming thick mats of fungi, filamentous bacteria, sludge, and blood worms that blanket the bottom. The extreme case of excessive pollution is one in which the dissolved oxygen is driven to zero, creating a nearly lifeless, anearobic stretch of river. Decomposition continues at a much slower rate by anaerobic microbes, releasing noxious and toxic gases such as hydrogen sulfide and ammonia.

Figure 17 may seem to suggest that the cause of an unhealthful drop in DO is an increase in the BOD from a waste source. That may be the case, but the oxygen

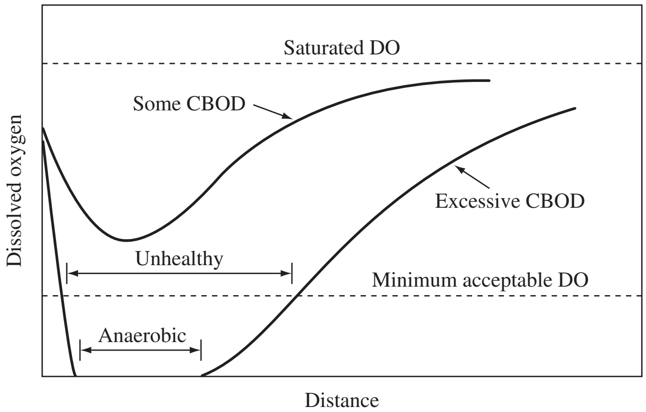


FIGURE 17 As a river gets more polluted, the oxygen sag curve drops below an acceptable level. In the extreme case, anaerobic conditions can occur.

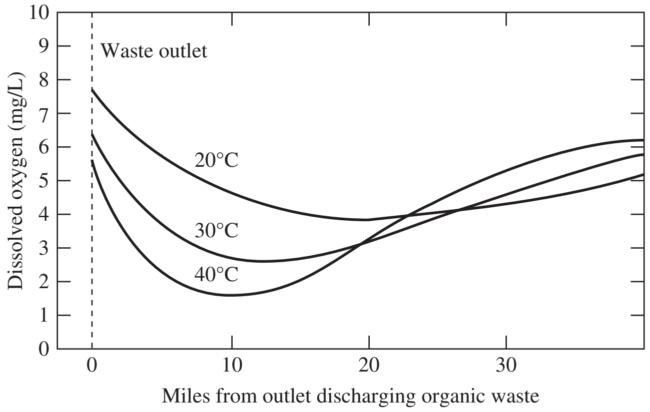


FIGURE 18 Changes in the oxygen sag curve as temperature increases. At higher temperatures, the minimum DO is lower.

sag curve is also sensitive to other sorts of changes. The DO curve changes with the seasons, with the temperature, and with the time of day, even if the pollutant load is constant. In summer months, for example, river flows are usually diminished, so the amount of waste dilution decreases, the BOD of the river/waste mix goes up, and the minimum DO drops.

The effect of temperature on the oxygen sag curve is also important (Figure 18). As temperatures rise, wastes decompose faster, and the rate of deoxygenation increases. At the same time, the saturated value of DO drops, so reaeration slows down. The combination of these effects causes the critical point downstream to be reached sooner and the minimum value of DO to be lower as well. Thus, a stream that may have sufficient DO in colder months may have an unacceptable deficit in the warmer months of summer. It also illustrates the potential adverse impact caused by thermal pollution from power plants. A river that might have been able to accept a certain sewage load without adverse effects could have unacceptably low oxygen levels when a power plant is added.

Photosynthesis also affects DO. Algae and other aquatic plants add DO during the daytime hours while photosynthesis is occurring, but at night, their continued respiration draws it down again. The net effect is a diurnal variation that can lead to elevated levels of DO in the late afternoon and depressed concentrations at night. For a lake or a slow-moving stream that is already overloaded with BOD and choked with algae, it is not unusual for respiration to cause offensive, anaerobic conditions late at night, even though the river seems fine during the day.

There are other factors not included in our simple oxygen sag model. Accumulated sludge along the bottom contributes to the oxygen demand; tributaries contribute their own BOD and DO; multiple sources cause multiple dips in the sag curve; nonpoint-sources contribute pulses of BOD and other contaminants when it rains; and this model has not included the effects of nitrification. Nitrification causes a second dip in the oxygen sag curve a little further downstream as ammonia and organic nitrogen convert to nitrite and nitrate. If there is not enough DO downstream for nitrification to proceed, nitrogen remains as ammonia, which is toxic.

Modeling the impacts of BOD on the oxygen resources in a stream or river is an important part of the permitting process for new sources. Stream models can help determine the maximum amount of additional BOD that will be allowed, which, in turn, affects facility siting decisions and the extent of on-site waste treatment that will be required.

7 Water Quality in Lakes and Reservoirs

All lakes gradually accumulate silt and organic matter as they undergo a natural aging process known as *eutrophication.* A young lake is characterized by a low nutrient content and low plant productivity. Such *oligotrophic* ("few foods") lakes gradually acquire nutrients from their drainage basins, which enables increased aquatic growth. Over time, the increased biological productivity causes the water to become murky with phytoplankton, while decaying organic matter contributes to the depletion of available dissolved oxygen. The lake becomes *eutrophic* ("well fed"). As the accumulating silt and organic debris cause the lake to get shallower and warmer, more plants take root along the shallow edges, and the lake slowly transforms into a marsh or bog.

While such eutrophication is a natural process that may take thousands of years, it is possible to accelerate greatly the rate of change through human activities. This is called *cultural eutrophication.* Municipal wastewater, industrial wastes, and runoff from fertilized agricultural lands add nutrients that stimulate algal growth and degrade water quality. Algal blooms die and decay, causing unsightly, odorous clumps of rotting debris along the shoreline and thick mats of dead organic matter in the lake. The decomposition of dead algae uses up available oxygen, resulting in the same sort of oxygen depletion problems already mentioned for streams. Among the first casualties are coldwater fish, whose temperature sensitivity forces them to stay in the colder bottom waters of a lake where the least amount of oxygen is available. In some lakes, there are periods of time when anaerobic conditions prevail near the bottom. Not only are organisms at risk for lack of oxygen, but the toxicity of the water increases as hydrogen sulfide and metals, such as iron and manganese, which are normally tied up as precipitates in sediments, are dissolved and released into the lake.

The EPA lake and reservoir assessments mentioned in Section 4 indicate that half of the U.S. lakes not meeting water quality standards are degraded by eutrophication caused by nutrient enrichment from human activities. As Figure 8 points out earlier in this chapter, the most common sources of lake and reservoir pollution are runoff from agricultural and urban nonpoint-sources. Unfortunately, these are the most difficult sources to control.

Lakes and reservoirs are susceptible to a host of other pollution problems besides eutrophication. Unlike streams, which can transport pollutants "away," the slow dispersal rates in lakes let contamination accumulate quickly, which would make them seem to be less likely to be used as chemical dumping grounds. Unfortunately, that does not seem to be the case. For example, 78 percent of the shoreline along the Great Lakes is not meeting the EPA's criteria for unimpaired beneficial uses because of toxic organic chemical pollution (U.S. EPA, 2002). Lakes and reservoirs are also vulnerable to effects of sulfuric and nitric acids that are rained out of the sky.

Controlling Factors in Eutrophication

Many factors control the rate of production of algae, including the availability of sunlight to power the photosynthetic reactions, and the concentration of nutrients required for growth.

The amount of light available is related to the transparency of the water, which is in turn a function of the level of eutrophication. An oligotrophic lake, such as Lake Tahoe, may have enough sunlight to allow significant rates of photosynthesis to take place at a depth of 100 m or more, whereas eutrophic lakes may be so murky that photosynthesis is restricted to a thin layer of water very near the surface. The top layer of water in a lake, where plants produce more oxygen by photosynthesis than they remove by respiration, is called the *euphotic zone*. Below that lies the *profundal zone*. The transition between the two zones is designated the *light compensation level*. The light compensation level corresponds roughly to a depth at which light intensity is about 1 percent of full sunlight.

While the amount of sunlight available can be a limiting factor in algal growth, it is not something that we could imagine controlling as a way to slow eutrophication. The more obvious approach is to try to reduce the supply of nutrients. The list of nutrients that we might consider controlling, however, is long because it could include every nutrient known to be essential to plant growth. That list would include carbon, nitrogen, phosphorus, sulfur, calcium, magnesium, potassium, sodium, iron, manganese, zinc, copper, boron, plus some other essential nutrients. The problem is greatly simplified by focusing on the two that most often limit algal growth: phosphorus and nitrogen.

Justus Liebig, in 1840, first formulated the idea that "growth of a plant is dependent on the amount of foodstuff that is presented to it in minimum quantity." This has come to be known as *Liebig's law of the minimum*. In essence, Liebig's law states that algal growth will be limited by the nutrient which is least available relative to its needs. Therefore, the quickest way to control eutrophication would be to identify the *limiting nutrient* and reduce its concentration.

Liebig's law also implies that reductions in a nonlimiting nutrient will not provide effective control unless its concentration can be reduced to the point where it becomes the limiting nutrient. Thus, for example, reducing phosphorus loading by eliminating phosphates in detergents will have little effect in a region with nitrogen-limited surface water, but the same reductions could be very effective where phosphorus is the limiting nutrient. As lakes eutrophy, the dominant species of algae are often blue-green *Cyanophyta,* which have the unusual characteristic of being able to obtain their nitrogen directly from the atmosphere. In addition, nitrogen enters water as "fallout" from combustion sources, particularly fossil-fuel-fired power plants. With the atmosphere providing a rather unlimited nitrogen supply, most freshwater systems are phosphorus limited.

To help illustrate the relative amounts of nitrogen and phosphorus that are required for algal growth, consider the following frequently used representation of algal photosynthesis (Stumm and Morgan, 1981):

$$106\ CO_2 + 16\ NO_3^- + HPO_4^{2-} + 122\ H_2O + 18\ H^+ \longrightarrow C_{106}H_{263}O_{110}N_{16}P + 138\ O_2 \quad (33)$$

Using a simple stoichiometric analysis and remembering the atomic weights of nitrogen (14) and phosphorus (31), the ratio of the mass of nitrogen to phosphorus in algae is

$$\frac{\mathrm{N}}{\mathrm{P}} = \frac{16 \times 14}{1 \times 31} = 7.2$$

As a first approximation, then, it takes about 7 times more nitrogen than phosphorus to produce a given mass of algae. As a rough guideline, when the concentration (mg/L) of nitrogen in water is more than 10 times the concentration of phosphorus, the body of water will probably be phosphorus limited. When it is less than 10:1, it will probably be nitrogen limited (Thomann and Mueller, 1987). Unlike freshwater systems, most marine waters have an N/P ratio less than 5 and are nitrogen limited.

Sawyer (1947) suggests that phosphorus concentrations in excess of 0.015 mg/L and nitrogen concentrations above 0.3 mg/L are sufficient to cause blooms of algae. These are in line with more recent estimates that suggest that 0.010 mg/L phosphorus is "acceptable," while 0.020 mg/L is "excessive" (Vollenweider, 1975).

A Simple Phosphorus Model

Suppose we want to estimate the phosphorus concentration that would be expected in a completely mixed lake under steady state conditions given some combination of phosphorus sources and sinks. By comparing the calculated phosphorus level with phosphorus concentrations that are generally considered "acceptable," we would be able to estimate the amount of phosphorus control needed to prevent a eutrophication problem. In phosphorus limited lakes, 0.020 mg/L of phosphorus results in excessive algae blooms, whereas 0.010 mg/L seems to be low enough to prevent excessive algal production, while 0.020 mg/L is excessive. Following the analysis presented in Thomann and Mueller (1987), we can model the phosphorus behavior in an idealized lake by making a number of simplifying assumptions and then performing a mass balance.

Consider the simple model shown in Figure 19. Phosphorus is shown entering the lake from a point-source as well as from the incoming stream flow. Phosphorus is assumed to leave the lake either as part of the exiting stream flow or by settling into sediments. It would be easy to add other sources, including nonpoint-sources, if their phosphorus loading rates could be estimated, but that will not be done here.

If we assume equal incoming and outgoing stream flow rates, negligible volumetric flow input from the point-source, a well-mixed lake, and steady state conditions representing a seasonal or annual average, then we can write the following phosphorus balance:

$$\text{Rate of addition of } P = \text{Rate of removal of } P$$

$$QC_{\text{in}} + S = QC + \nu_s AC \tag{34}$$

where

S = the rate of addition of phosphorus from the point-source (g/s)
C = the concentration of phosphorus in the lake (g/m^3)
C_{in} = the concentration of phosphorus in the incoming stream flow (g/m^3)
Q = the stream inflow and outflow rate (m^3/s)
ν_s = the phosphorus settling rate (m/s)
A = the surface area of the lake (m^2)

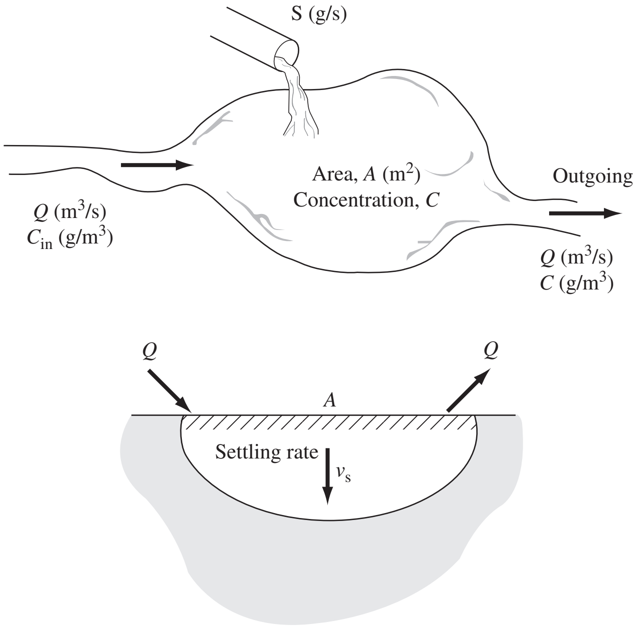


FIGURE 19 Well-mixed lake phosphorus mass balance.

which results in a steady state concentration of

$$C = \frac{QC_{in} + S}{Q + v_s A} \tag{35}$$

The settling rate, v_s, is an empirically determined quantity that is difficult to predict with any confidence. Thomann and Mueller cite studies that suggest settling rates are around 10 to 16 m/yr.

EXAMPLE 9 Phosphorus Loading in a Lake

A phosphorus limited lake with surface area equal to 80×10^6 m^2 is fed by a 15.0 m^3/s stream that has a phosphorus concentration of 0.010 mg/L. In addition, effluent from a point-source adds 1 g/s of phosphorus. The phosphorus settling rate is estimated at 10 m/yr.

a. Estimate the average total phosphorus concentration.
b. What rate of phosphorus removal at the wastewater treatment plant would be required to keep the concentration of phosphorus in the lake at an acceptable level of 0.010 mg/L?

Solution

a. The phosphorus loading from the incoming stream is

$$QC_{in} = 15.0 \text{ m}^3\text{/s} \times 0.010 \text{ mg/L} \times \frac{1 \text{ g/m}^3}{\text{mg/L}} = 0.15 \text{ g/s}$$

Adding the 1.0 g/s from the point-source gives a total phosphorus input rate of 1.15 g/s. The estimated settling rate is

$$\nu_s = \frac{10\ \text{m/yr}}{365\ \text{d/yr} \times 24\ \text{hr/d} \times 3600\ \text{s/hr}} = 3.17 \times 10^{-7}\ \text{m/s}$$

Using (35), the steady state concentration of total phosphorus would be

$$C = \frac{QC_{\text{in}} + S}{Q + \nu_s A} = \frac{1.15\ \text{g/s}}{15\ \text{m}^3/\text{s} + 3.17 \times 10^{-7}\ \text{m/s} \times 80 \times 10^6\ \text{m}^2}$$

$$= 0.028\ \text{g/m}^3 = 0.028\ \text{mg/L}$$

which is above the 0.010 mg/L needed to control eutrophication.

b. To reach 0.010 mg/L, the phosphorus loading from the point-source must be

$$\begin{aligned} S &= C(Q + \nu_s A) - QC_{\text{in}} \\ &= 0.010\ \text{g/m}^3(15\ \text{m}^3/\text{s} + 3.17 \times 10^{-7}\ \text{m/s} \times 80 \times 10^6\ \text{m}^2) \\ &\quad - 15\ \text{m}^3/\text{s} \times 0.010\ \text{g/m}^3 \\ &= 0.25\ \text{g/s} \end{aligned}$$

The point-source effluent currently supplies 1.0 g/s, so 75 percent removal of phosphorus is needed.

This model is based on a number of unrealistic assumptions. For example, due to a phenomenon called *thermal stratification,* which will be described in the next section, the assumption that the lake is well-mixed is usually reasonable only during certain times of the year and in certain parts of the lake. The assumption of steady state ignores the dynamic behavior of lakes as weather and seasons change. We have also assumed that phosphorus is the controlling nutrient and that a simple measure of its concentration will provide an adequate indication of the potential for eutrophication. Finally, the model assumes a constant phosphorus settling rate that does not depend on such factors as the fraction of incoming phosphorus that is in particulate form, the movement of phosphorus both to sediments and from sediments, and physical parameters of the lake such as the ratio of lake volume to stream flow (the hydraulic detention time). Despite these unrealistic assumptions, the model has been shown in practice to produce very useful results (Thomann and Mueller, 1987).

Thermal Stratification

As we have seen, nutrients stimulate algal growth and the subsequent death and decay of that algae can lead to oxygen depletion. This oxygen depletion problem is worsened by certain physical characteristics of lakes, which we will now consider.

As mentioned earlier, one of the most unusual properties of water is the fact that its density does not monotonically increase as the temperature drops. Instead, it has a maximum point at 4°C, as shown in Figure 20. One result of this density maximum is that ice floats because the water surrounding it is slightly warmer and denser. If water were like other liquids, ice would sink, and it would be possible for lakes to freeze solid from the bottom up. Fortunately, this is not the case.

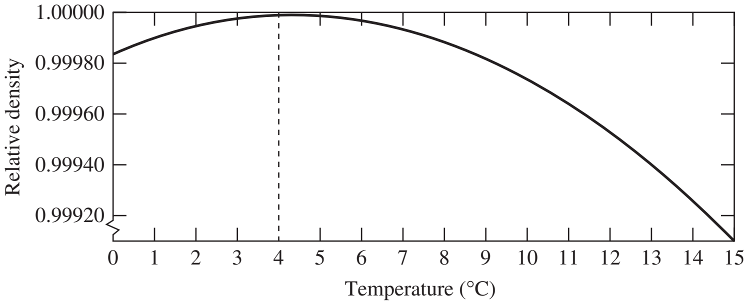


FIGURE 20 The density of water reaches a maximum at 4°C.

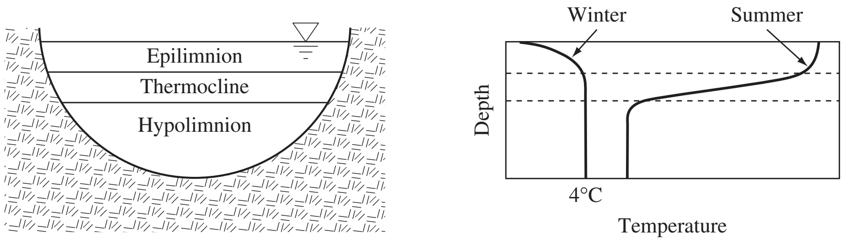


FIGURE 21 Thermal stratification of a lake showing winter and summer stratification temperature profiles.

Above 4°C, the density of water decreases with temperature. As a result, a lake warmed by the sun during the summer will tend to have a layer of warm water floating on the top of the denser, colder water below. Conversely, in the winter, if the lake's surface drops below 4°C, it will create a layer of cold water that floats on top of the more dense, 4°C water below. These density differences between surface water and the water nearer to the bottom inhibit vertical mixing in the lake, causing a very stable layering effect known as *thermal stratification.*

Figure 21 shows the stratification that typically occurs in a deep lake, in the temperate zone, during the summer. In the upper layer, known as the *epilimnion,* the warm water is completely mixed by the action of wind and waves, causing an almost uniform temperature profile. The thickness of the epilimnion varies from lake to lake and from month to month. In a small lake, it may be only a meter or so deep, whereas in large lakes, it may extend down 20 m or more. Below the epilimnion is a transition layer called the *thermocline,* or *metalimnion,* in which the temperature drops rather quickly. Most swimmers have experienced a thermocline's sudden drop in temperature when diving down into a lake. Below the thermocline is a region of cold water called the *hypolimnion.*

In terms of mixing, summer stratification creates essentially two separate lakes—a warm lake (epilimnion) floating on top of a cold lake (hypolimnion). The separation is quite stable and becomes increasingly so as the summer progresses. Once summer stratification begins, the lack of mixing between the layers causes the epilimnion, which is absorbing solar energy, to warm even faster, creating an even greater density difference. This difference is enhanced in a eutrophic lake since in such lakes most, if not all, of the absorption of solar energy occurs in the epilimnion.

As the seasons progress and winter approaches, the temperature of the epilimnion begins to drop, and the marked stratification of summer begins to disappear. Sometime in the fall, perhaps along with a passing storm that stirs things up, the stratification will disappear, the temperature will become uniform with depth, and complete mixing of the lake becomes possible. This is called the *fall overturn.* Similarly, in climates that are cold enough for the surface to drop below 4°C, there will be a winter stratification, followed by a *spring overturn* when the surface warms up enough to allow complete mixing once again.

Stratification and Dissolved Oxygen

Dissolved oxygen, one of the most important water quality parameters, is greatly affected by both eutrophication and thermal stratification. Consider, for example, two different stratified lakes, one oligotrophic and one eutrophic. In both lakes, the waters of the epilimnion can be expected to be rich in DO since oxygen is readily available from reaeration and photosynthesis. The hypolimnion, on the other hand, is cut off from the oxygen-rich epilimnion by stratification. The only source of new oxygen in the hypolimnion will be photosynthesis, but that will only happen if the water is clear enough to allow the euphotic zone to extend below the thermocline. That is, the hypolimnion of a clear, oligotrophic lake at least has the possibility of having a source of oxygen while that of a murky eutrophic lake does not.

In addition, the eutrophic lake is rich in nutrients and organic matter. Algal blooms suddenly appear and die off, leaving rotting algae that washes onto the beaches or sinks to the bottom. The rain of organic debris into the hypolimnion leads to increased oxygen demand there. Thus, not only is there inherently less oxygen available in the hypolimnion, but there is also more demand for oxygen due to decomposition, especially if the lake is eutrophic. Once summer stratification sets in, DO in the hypolimnion will begin dropping, driving fish out of the colder bottom regions of the lake and into the warmer, more oxygen-rich surface waters. As lakes eutrophy, fish that require cold water for survival are the first to die. In the extreme case, the hypolimnion of a eutrophic lake can become anaerobic during the summer, as is suggested in Figure 22.

During the fall and spring overturns, which may last for several weeks, the lake's waters become completely mixed. Nutrients from the bottom are distributed throughout the lake and oxygen from the epilimnion becomes mixed with the oxygen-poor hypolimnion. The lake, in essence, takes a deep breath of air.

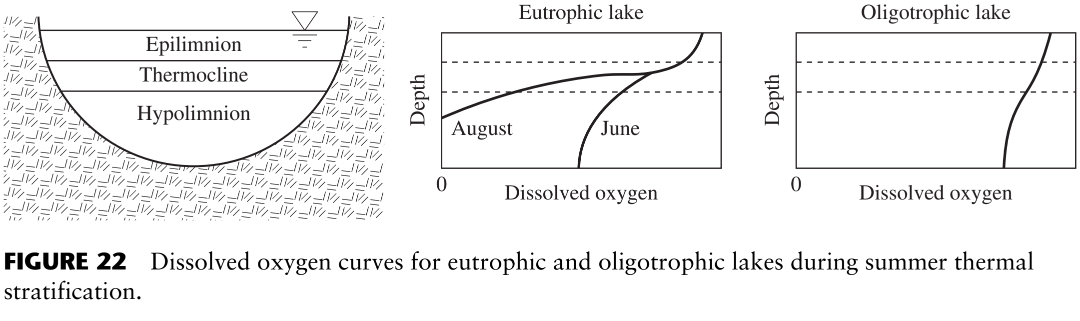


FIGURE 22 Dissolved oxygen curves for eutrophic and oligotrophic lakes during summer thermal stratification.

In winter, demands for oxygen decrease as metabolic rates decrease, while at the same time, the ability of water to hold oxygen increases. Thus, even though winter stratification may occur, its effects tend not to be as severe as those in the summer. If ice forms, however, both reaeration and photosynthesis may cease to provide oxygen, and fish may die.

Tropical lakes may remain stratified indefinitely since they do not experience the degree of seasonal variation that allows lakes to overturn. When a tropical lake eutrophies, the hypolimnion goes anaerobic and tends to stay that way. Decomposition in the hypolimnion produces gases such as hydrogen sulfide (H_2S), methane (CH_4), and carbon dioxide (CO_2), which tend to remain dissolved in the bottom waters, especially if the lake is deep and bottom pressure is high. In volcanically formed crater lakes, underground springs of carbonated groundwater can also contribute CO_2 to the hypolimnion. When lake bottoms become saturated with these dissolved gases, there is danger that some event, such as an earthquake or landslide, can trigger a sudden and dangerous gas bubble to erupt from the surface. Such an event occurred in Lake Monoun in Cameroon, Africa, in 1984, and another gaseous eruption occurred in Lake Nyos, Cameroon, two years later. Each eruption belched a cloud of heavier-than-air carbon dioxide that hugged the ground, rolled into valleys, and, by displacing oxygen, asphyxiated farm animals and people. The Lake Monoun event killed 37 people, and the Lake Nyos event killed more than 1,700 people, some as far away as 25 km (Monastersky, 1994). The bubble of CO_2 emerging from Lake Nyos is said to have created a toxic fountain that shot 150 m into the air. Both events occurred in August, the coolest and rainiest time of the year, when thermal stratification is the least stable.

Acidification of Lakes

All rainfall is naturally somewhat acidic. As was demonstrated, pure water in equilibrium with atmospheric carbon dioxide forms a weak solution of carbonic acid (H_2CO_3) with a pH of about 5.6. As a result, only rainfall with pH less than 5.6 is considered to be "acid rain." It is not unusual for rainfall in the northeastern United States to have pH between 4.0 and 5.0, and for acid fogs in southern California to have pH less than 3.0. There is little question that such pH values are caused by anthropogenic emissions of sulfur and nitrogen oxides formed during the combustion of fossil fuels. Since some sulfur oxides are actually particles that can settle out of the atmosphere without precipitation, the popular expression "acid rain" is more correctly described as *acid deposition*.

The effects of acid deposition on materials, terrestrial ecosystems, and aquatic ecosystems are still only partially understood, but some features are emerging quite clearly. Acids degrade building materials, especially limestone, marble (a form of limestone), various commonly used metals such as galvanized steel, and certain paints. In fact, the increased rate of weathering and erosion of building surfaces and monuments was one of the first indications of adverse impacts from acid rain. Terrestrial ecosystems, especially forests, seem to be experiencing considerable stress due to acid deposition, with reductions in growth and increased mortality being especially severe in portions of the eastern United States, eastern Canada, and northern Europe. It is the impact of acidification on aquatic ecosystems, however, that will be described here.

Bicarbonate Buffering

Aquatic organisms are very sensitive to pH. Most are severely stressed if pH drops below 5.5, and very few are able to survive when pH falls below 5.0. Moreover, as pH drops, certain toxic minerals, such as aluminum, lead, and mercury, which are normally insoluble and hence relatively harmless, enter solution and can be lethal to fish and other organisms.

It is important to note, however, that adding acid to a solution may have little or no effect on pH, depending on whether or not the solution has *buffers*. Buffers are substances capable of neutralizing added hydrogen ions. The available buffering of an aquatic ecosystem is not only a function of the chemical characteristics of the lake itself, but also of nearby soils through which water percolates as it travels from the land to the lake. Thus, information on the pH of precipitation alone, without taking into account the chemical characteristics of the receiving body of water and surrounding soils, is a poor indicator of the potential effect of acid rain on an aquatic ecosystem.

Most lakes are buffered by bicarbonate (HCO_3^-), which is related to carbonic acid (H_2CO_3) by the following reaction:

$$\underset{\text{Carbonic acid}}{H_2CO_3} \rightleftarrows H^+ + \underset{\text{Bicarbonate}}{HCO_3^-} \tag{36}$$

Some bicarbonate results from the dissociation of carbonic acid, as is suggested in (36), and some comes from soils. Consider what happens to a lake containing bicarbonate when hydrogen ions (acid) are added. As reaction (36) suggests, some of the added hydrogen ions will react with bicarbonate to form neutral carbonic acid. To the extent that this occurs, the addition of hydrogen ions does not show up as an increase in hydrogen ion concentration, so the pH may only change slightly. That is, bicarbonate is a buffer.

Notice that the reaction of hydrogen ions with bicarbonate removes bicarbonate from solution so, unless there is a source of new bicarbonate, its concentration will decrease as more acid is added. At some point, there may be so little bicarbonate left that relatively small additional inputs of acid will cause pH to decrease rapidly. This phenomenon leads to one way to classify lakes in terms of their acidification. As shown in Figure 23, a *bicarbonate* lake shows little decrease in pH as hydrogen ions are added, until the pH drops to about 6.3. As pH drops below this point, the bicarbonate buffering is rapidly depleted, and the lake enters a transitional phase. As shown in the figure, transitional lakes with pH between about 5.0 and 6.0 are very sensitive to small changes in acid. Below pH 5.0, the lakes are unbuffered and chronically acidic (Wright, 1984).

The implications of Figure 23 are worth repeating. Acid precipitation may have little or no effect on a lake's pH up until the point where the natural buffering is exhausted. During that stage, there may be no evidence of harm to aquatic life, and the lake may appear to be perfectly healthy. Continued exposure to acidification beyond that point, however, can result in a rapid drop in pH that can be disastrous to the ecosystem.

In a Norwegian study, 684 lakes were categorized by their water chemistry into bicarbonate lakes, transition lakes, and acid lakes, and for each category, observations of fish populations were made (Henriksen, 1980; Wright, 1984). Four

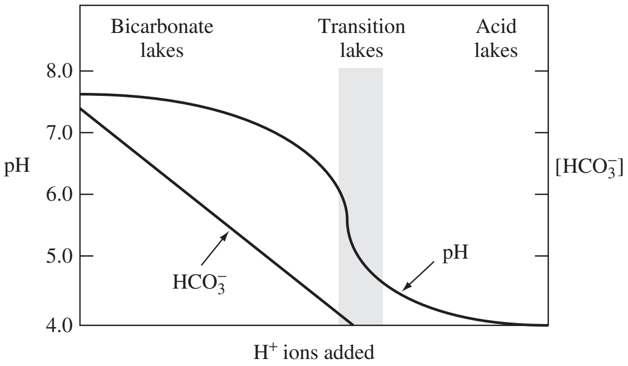


FIGURE 23 Bicarbonate buffering strongly resists acidification until pH drops below 6.3. As more H^+ ions are added, pH decreases rapidly.
(*Source:* Henriksen, 1980.)

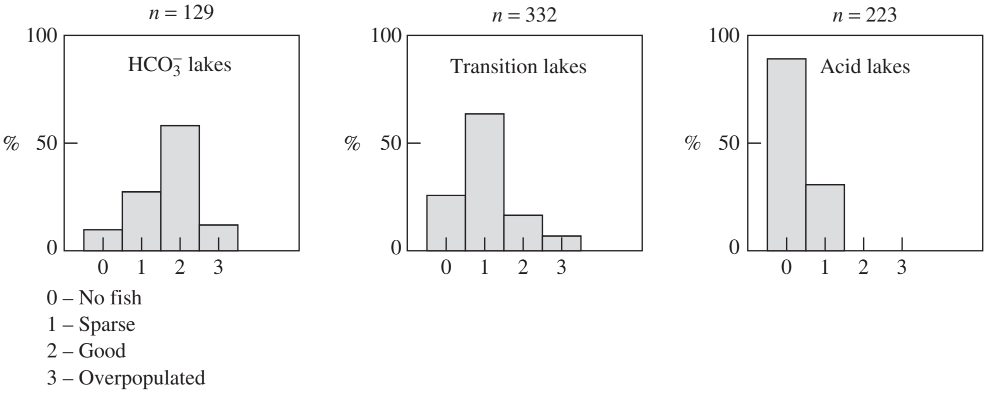


FIGURE 24 Frequency histograms of fish status for 684 Norwegian lakes categorized as bicarbonate, transition, or acid lakes.
(*Source:* Wright, 1984.)

fish-status categories were chosen: barren (long-term, chronic failures in fish reproduction), sparse (occasional fish reproduction failures), good, and overpopulated. The results, shown in Figure 24, clearly show the dependency of fish populations on lake acidification.

Importance of the Local Watershed

If there is a source of bicarbonate to replace that which is removed during acidification, then the buffering ability of bicarbonate can be extensive. Consider the ability of limestone ($CaCO_3$) to provide buffering:

$$H^+ + \underset{\text{Limestone}}{CaCO_3} \longrightarrow Ca^{2+} + \underset{\text{Bicarbonate}}{HCO_3^-} \tag{37}$$

$$\underset{\text{Bicarbonate}}{HCO_3^-} + H^+ \longrightarrow \underset{\text{Carbonic acid}}{H_2CO_3} \longrightarrow CO_2 + H_2O \tag{38}$$

As this pair of reactions suggests, limestone reacts with acids (hydrogen ions) to form bicarbonate, and the bicarbonate neutralizes even more hydrogen ions as it converts eventually to carbon dioxide. These reactions show why limestone is an effective buffer to acidification. They also show why limestone and marble monuments and building facades are deteriorating under the attack of acid deposition.

The bicarbonate that is formed in (37) can replenish the lake's natural bicarbonate buffers. Calcareous lakes, which have an abundance of calcium carbonate, are thus invulnerable to acidification. Copying this natural phenomenon, one way to temporarily mitigate the effects of acidification is to artificially treat vulnerable lakes with limestone ($CaCO_3$). This approach is being vigorously pursued in a number of countries that are most impacted by acid rain, including Sweden, Norway, Canada, and the United States (Shepard, 1986).

The ability of nearby soils to buffer acid deposition is an extremely important determinant of whether or not a lake will be subject to acidification. Soil systems derived from calcareous (limestone) rock, for example, are better able to assimilate acid deposition than soils derived from granite bedrock. The depth and permeability of soils are also important characteristics. Thin, relatively impermeable soils provide little contact between soil and runoff, which reduces the ability of natural soil buffers to affect acidity. The size and shape of the watershed itself also affect a lake's vulnerability. Steep slopes and a small watershed create conditions in which the runoff has little time to interact with soil buffers, increasing the likelihood of lake acidification. Even the type of vegetation growing in the watershed can affect acidification. Rainwater that falls through a forest canopy interacts with natural materials in the trees, such as sap, and its pH is affected. Deciduous foliage tends to decrease acidity, while water dripping from evergreen foliage tends to be more acidic than the rain itself.

In other words, to a large extent, the characteristics of the watershed itself will determine the vulnerability of a lake to acid. The most vulnerable lakes will be in areas with shallow soil of low permeability, granite bedrock, a steep watershed, and a predominance of conifers. Using factors such as these enables scientists to predict areas where lakes are potentially most sensitive to acid rain. One such prediction, using bedrock geology as the criterion, produced the map shown in Figure 25*a*. Figure 25*b* shows the average pH of rainfall across North America. Large areas in the northeastern portion of the continent, especially those near the Canadian border, have the unfortunate combination of vulnerable soils and very acidic precipitation.

8 Groundwater

Groundwater is the source of about one-third of this country's drinking water, and, excluding water for power plant cooling, it supplies almost 40 percent of our total water withdrawals. In rural areas, almost all of the water supply comes from groundwater, and more than one-third of our 100 largest cities depend on groundwater for at least part of their supply. Historically, the natural filtering resulting from water working its way through the subsurface was believed to provide sufficient protection from contamination to allow untreated groundwater to be delivered to customers. Then Love Canal, New York, and other dramatic incidents in the 1970s made us realize that groundwater was contaminated with hazardous substances from hundreds of

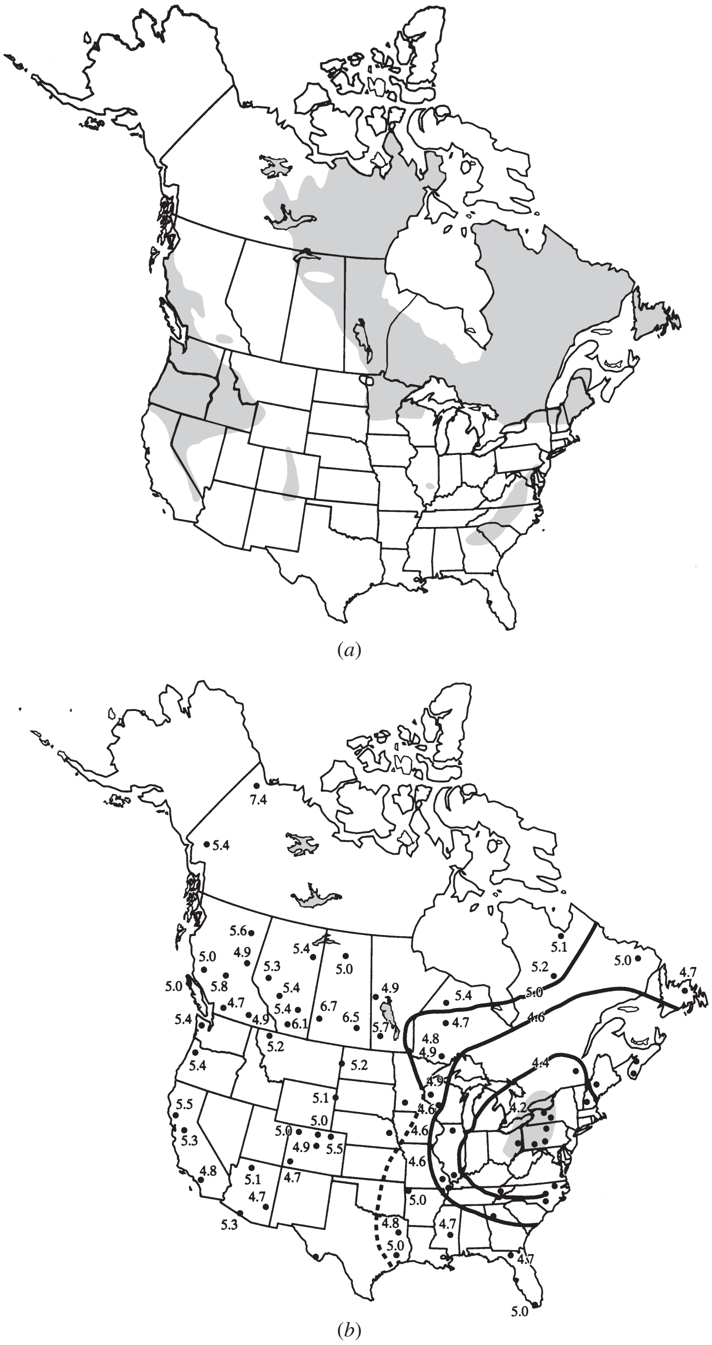


FIGURE 25 (*a*) Regions in North America containing lakes that would be sensitive to potential acidification by acid precipitation (shaded areas), based on bedrock geology (EPA, 1984, based on Galloway and Cowling, 1978); (*b*) pH of wet deposition in 1982 (Interagency Task Force on Acid Precipitation, 1983).

thousands of leaking underground storage tanks, industrial waste pits, home septic systems, municipal and industrial landfills, accidental chemical spills, careless use of solvents, illegal "midnight" dumping, and widespread use of agricultural chemicals. Groundwater contamination became the environmental issue of the 1980s.

In response to growing pressure to do something about hazardous wastes, Congress passed the *Comprehensive Environmental Response, Compensation, and Liability Act* (CERCLA, but more commonly known as Superfund) in 1980 to deal with already contaminated sites. In 1980 and 1984, Congress strengthened the *Resource Conservation and Recovery Act* (RCRA), which controls the manufacturing, transportation, and disposal of newly produced hazardous substances. These two acts, one cleaning up problems of the past, and one helping avoid problems in the future, have had a major impact. The billions of dollars that have been spent attempting to clean up contaminated soil and groundwater, along with the very high costs of handling and disposing of new hazardous wastes, have stimulated interest in pollution prevention as the environmental theme of the future.

Once contaminated, groundwater is difficult, if not impossible, to restore. A recent study by the National Research Council (1994) estimates that there are between 300,000 and 400,000 sites in the United States that may have contaminated soil or groundwater requiring some form of remediation. The bulk of these sites are leaking underground storage tanks. These sites have been costing around $100,000 each to clean up; more complex sites are costing an average of $27 million each (U.S. EPA, 1993). To clean up all of these sites, the estimated cost would be between $480 billion and $1 trillion (NRC, 1994). These enormous costs have called into question the goal of cleaning up such sites to drinking water quality, especially if the groundwater is unlikely to ever be needed for such purposes.

9 Aquifers

Rainfall and snow melt can flow into rivers and streams, return to the atmosphere by evaporation or transpiration, or seep into the ground to become part of the subsurface, or underground, water. As water percolates down through cracks and pores of soil and rock, it passes through a region called the *unsaturated zone,* which is characterized by the presence of both air and water in the spaces between soil particles. Water in the unsaturated zone, called *vadose water,* is essentially unavailable for human use. That is, it cannot be pumped, although plants certainly use soil water that lies near the surface. In the *saturated zone,* all spaces between soil particles are filled with water. Water in the saturated zone is called *groundwater,* and the upper boundary of the saturated zone is called the *water table.* There is a transition region between these two zones called the *capillary fringe,* where water rises into small cracks as a result of the attraction between water and rock surfaces. Figure 26 illustrates these regions of the subsurface.

An *aquifer* is a saturated geologic layer that is permeable enough to allow water to flow fairly easily through it. An aquifer sits on top of a *confining bed* or, as it is sometimes called, an *aquitard* or an *aquiclude,* which is a relatively impermeable layer that greatly restricts the movement of groundwater. The two terms, aquifer and confining bed, are not precisely defined and are often used in a relative sense.

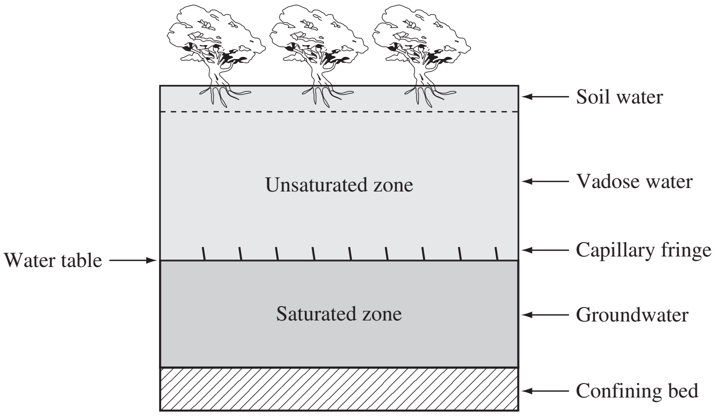


FIGURE 26 Identification of subsurface regions.

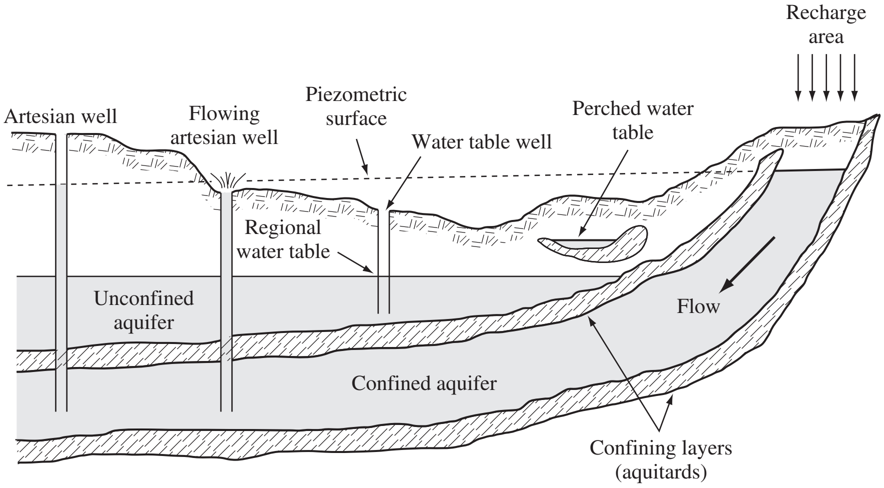


FIGURE 27 Confined and unconfined aquifers, a perched water table, water table well, and an artesian well.

The illustration in Figure 26 shows an *unconfined* aquifer. A *water table well* drilled into the saturated zone of such an aquifer will have water at atmospheric pressure at the level of the water table. Groundwater also occurs in *confined aquifers,* which are aquifers sandwiched between two confining layers, as shown in Figure 27. Water in a confined aquifer can be under pressure so that a well drilled into it will have water naturally rising above the upper surface of the confined aquifer, in which case the well is called an *artesian well.* A line drawn at the level to which water would rise in an artesian well defines a surface called the *piezometric surface* or the *potentiometic surface.* In some circumstances, enough pressure may exist in a confined aquifer to cause water in a well to rise above the ground level and flow without pumping. Such a well is called a *flowing artesian well.* Also shown in Figure 27 is a local impermeable layer in the midst of an unsaturated zone, above

TABLE 11

Representative Values of Porosity and Specific Yield

Material	Porosity (%)	Specific Yield (%)
Clay	45	3
Sand	34	25
Gravel	25	22
Gravel and sand	20	16
Sandstone	15	8
Limestone, shale	5	2
Quartzite, granite	1	0.5

Source: Linsley et al., 1992.

the main body of groundwater. Downward percolating water is trapped above this layer, creating a *perched water table.*

The amount of water that can be stored in a saturated aquifer depends on the *porosity* of the soil or rock that makes up the aquifer. Porosity (η) is defined to be the ratio of the volume of voids (openings) to the total volume of material:

$$\text{Porosity } (\eta) = \frac{\text{Volume of voids}}{\text{Total volume of solids and voids}} \tag{39}$$

While porosity describes the water-bearing capacity of a geologic formation, it is not a good indicator of the total amount of water that can be removed from that formation. Some water will always be retained as a film on rock surfaces and in very small cracks and openings. The volume of water that can actually be drained from an *unconfined* aquifer per unit of area per unit decline in the water table is called the *specific yield,* or the *effective porosity.* Representative values of porosity and specific yield for selected materials are given in Table 11. For confined aquifers, the estimated yield is somewhat affected by the pressure released as water is removed, so a different term, the *storage coefficient,* is used to describe the yield.

EXAMPLE 10 **Specific yield**

For an aquifer of sand, having characteristics given in Table 11, what volume of water would be stored in a saturated column with a cross-sectional area equal to 1.0 m^2 and a depth of 2.0 m? How much water could be extracted from that volume?

Solution The volume of material is $1.0\ m^2 \times 2.0\ m = 2.0\ m^3$, so the volume of water stored would be

$$\begin{aligned}\text{Volume of water} &= \text{Porosity} \times \text{Volume of material}\\ &= 0.34 \times 2.0\ m^3 = 0.68\ m^3\end{aligned}$$

The amount that could actually be removed would be

$$\text{Yield} = \text{Specific yield} \times \text{Volume of material} = 0.25 \times 2.0\ m^3 = 0.5\ m^3$$

10 Hydraulic Gradient

In an unconfined aquifer, the slope of the water table, measured in the direction of the steepest rate of change, is called the *hydraulic gradient*. It is important because groundwater flow is in the direction of the gradient and at a rate proportional to the gradient. To define it more carefully, we need to introduce the notion of the *hydraulic head*. As shown in Figure 28, the hydraulic head is the vertical distance from some reference datum plane (usually taken to be sea level) to the water table. It has dimensions of length, such as "meters of water" or "feet of water." If we imagine two wells directly in line with the groundwater flow, the gradient would be simply the difference in head divided by the horizontal distance between them. In an analogous way, the gradient for a confined aquifer is the slope of the piezometric surface along the direction of maximum rate of change.

If the flow in Figure 28 is from left to right in the plane of the page, the gradient would be

$$\text{Hydraulic gradient} = \frac{\text{Change in head}}{\text{Horizontal distance}} = \frac{h_2 - h_1}{L} \tag{40}$$

Notice that the gradient is dimensionless as long as numerator and denominator both use the same units. In a microscopic sense, gradient can be expressed as

$$\text{Hydraulic gradient} = \frac{dh}{dL} \tag{41}$$

If we imagine looking down onto the groundwater flow of Figure 28, we could imagine *streamlines* showing the direction of flow, and *equipotential lines*, perpendicular to the streamlines, which represent locations with equal head. The combination of streamlines and equipotential lines creates a two-dimensional *flow net*, as shown

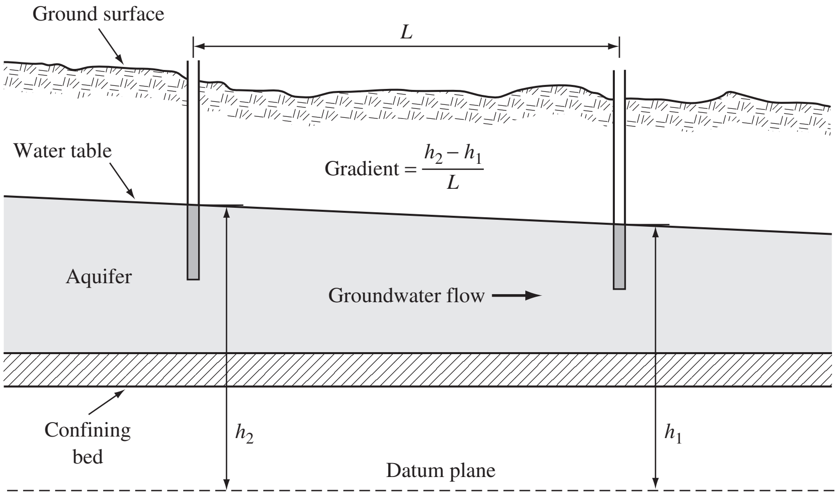


FIGURE 28 Hydraulic head and gradient in an unconfined aquifer. The groundwater flow is from left to right in the plane of the page, and the gradient is $\Delta h/L$.

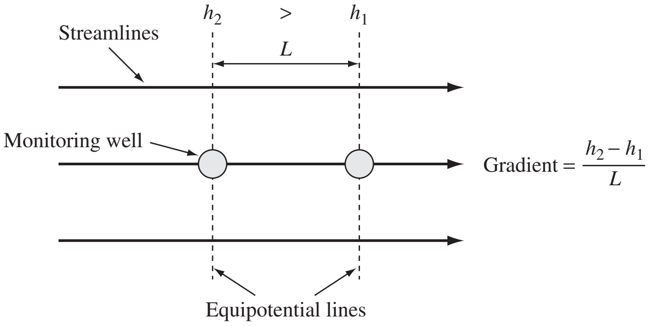


FIGURE 29 A two-dimensional flow net consisting of streamlines and equipotential lines. If two wells happen to fall on a streamline, the gradient is just $\Delta h/L$.

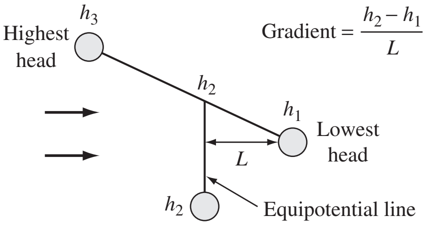


FIGURE 30 Using three wells to determine the gradient: Drawing a line between the wells with highest and lowest head and finding the spot along that line corresponding to the third well establishes an equipotential line between the spot and the third well.

in Figure 29. If we happened to have two wells aligned along a streamline as shown in Figure 29, the gradient would be simply the difference in head divided by the distance between equipotentials.

Figure 29 represents the special case of two wells along the same streamline. In the more general case, it is possible to estimate the gradient using measurements made at three wells using the following simple graphical procedure (Heath, 1983): Begin by finding the locations and hydraulic heads for three nearby wells, as shown in Figure 30. Then, follow these steps:

1. Draw a line between the two wells with the highest and lowest head and divide that line into equal intervals. Identify the location on the line where the head is equal to the head of the third (intermediate head) well.
2. Draw a line between the intermediate-head well and the spot on the line drawn in step 1 that corresponds to the head of the intermediate well. This is an *equipotential line,* meaning that the head anywhere along the line should be approximately constant. Groundwater flow will be in a direction perpendicular to this line.
3. Draw a line perpendicular to the equipotential line through the well with the lowest (or the highest) head. This is a *flow line,* which means groundwater flow is in a direction parallel to this line.
4. Determine the gradient as the difference in head between the head on the equipotential and the head at the lowest (or highest) well, divided by the distance from the equipotential line to that well.

EXAMPLE 11 Estimating the Hydraulic Gradient from Three Wells

Two wells are drilled 200 m apart along an east-west axis. The west well has a total head of 30.2 meters, and the east well has a 30.0 m head. A third well located 100 m due south of the east well has a total head of 30.1 m. Find the magnitude and direction of the hydraulic gradient.

Solution The locations of the wells are shown in Figure 31*a*. In Figure 31*b*, a line has been drawn between the wells with the highest (west, 30.2 m) and lowest (east, 30.0 m) head, and the location along that line with head equal to the intermediate well (south, 30.1 m) has been indicated. The line corresponding to a 30.1-m equipotential has been drawn.

In Figure 31*c*, streamlines perpendicular to the equipotential line have been drawn toward the east well. The direction of the gradient is thus at a 45° angle from the southwest. The distance between the equipotential and the east well is easily determined by geometry to be $L = 100/\sqrt{2}$. So the gradient is

$$\text{Hydraulic gradient} = \frac{(30.1 - 30.0)\text{ m}}{(100/\sqrt{2})\text{ m}} = 0.00141$$

Notice that the gradient is dimensionless.

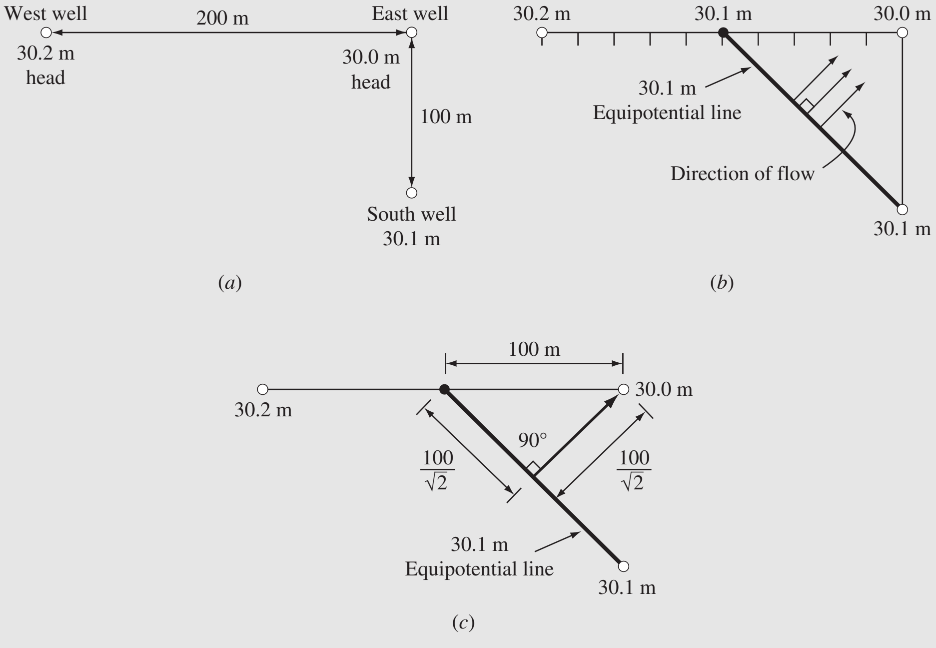


FIGURE 31 Finding the gradient for the well field in Example 11.

11 Darcy's Law

The basic equation governing groundwater flow was first formulated by the French hydraulic engineer Henri Darcy in 1856. Based on laboratory experiments in which he studied the flow of water through sand filters, Darcy concluded that flow rate Q is proportional to the cross sectional area A times the hydraulic gradient (dh/dL):

$$Q = KA\left(\frac{dh}{dL}\right) \tag{42}$$

where

Q = flow rate (m^3/day)
K = hydraulic conductivity, or coefficient of permeability (m/day)
A = cross-sectional area (m^2)
$\left(\frac{dh}{dL}\right)$ = the hydraulic gradient

Equation (42) is known as *Darcy's law* for flow through porous media. Darcy's law assumes linearity between flow rate and hydraulic gradient, which is a valid assumption in most, but not all, circumstances. It breaks down when flow is turbulent, which may occur in the immediate vicinity of a pumped well. It is also invalid when water flows through extremely fine-grained materials, such as colloidal clays, and it should only be used when the medium is fully saturated with water. It also depends on temperature. Some approximate values of the constant of proportionality K are given Table 12. It should be appreciated, however, that these values for the hydraulic conductivity are very rough. Conductivities can easily vary over several orders of magnitude for any given category of material, depending on differences in particle orientation and shape as well as relative amounts of silt and clay that might be present.

Aquifers that have the same hydraulic conductivity throughout are said to be *homogeneous,* whereas those in which hydraulic conductivity differs from place to place are *heterogeneous*. Not only may hydraulic conductivity vary from place to place within the aquifer, but it may also depend on the direction of flow. It is common, for example, to have higher hydraulic conductivities in the horizontal direction than in the vertical. Aquifers that have the same hydraulic conductivity in any

TABLE 12

Approximate Values of Hydraulic Conductivity

	Conductivity	
Material	(gpd/ft^2)	(m/day)
Clay	0.01	0.0004
Sand	1000	41
Gravel	100,000	4100
Gravel and sand	10,000	410
Sandstone	100	4.1
Limestone, shale	1	0.041
Quartzite, granite	0.01	0.0004

Source: Linsley et al., 1992.

flow direction are said to be *isotropic,* whereas those in which conductivity depends on direction are *anisotropic*. Although it is mathematically convenient to assume that aquifers are both homogeneous and isotropic, they rarely, if ever, are.

EXAMPLE 12 Flow Through an Aquifer

A confined aquifer 20.0 m thick has two monitoring wells spaced 500 m apart along the direction of groundwater flow. The difference in water level in the wells is 2.0 m (the difference in piezometric head). The hydraulic conductivity is 50 m/day. Estimate the rate of flow per meter of distance perpendicular to the flow.

Solution Figure 32 summarizes the data. The gradient is

$$\left(\frac{dh}{dL}\right) = \frac{2.0 \text{ m}}{500 \text{ m}} = 0.004$$

Using Darcy's law, with an arbitrary aquifer width of 1 m, yields

$$\begin{aligned} Q &= KA\left(\frac{dh}{dL}\right) \\ &= 50 \text{ m/d} \times 1.0 \text{ m} \times 20.0 \text{ m} \times 0.004 = 4.0 \text{ m}^3\text{/day per meter of width} \end{aligned}$$

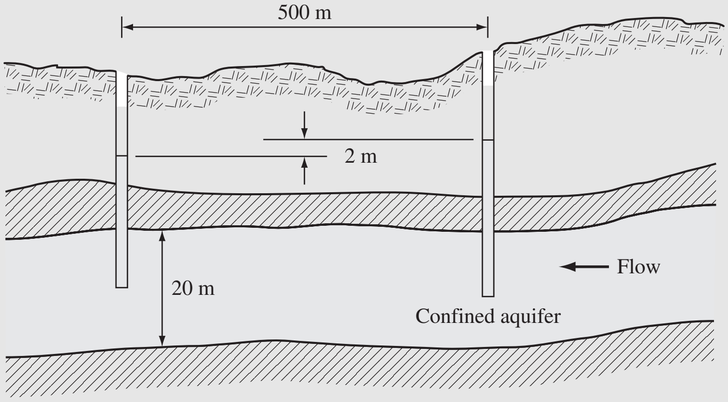


FIGURE 32 Example 12, flow through a confined aquifer.

Flow Velocity

It is often important to estimate the rate at which groundwater is moving through an aquifer, especially when a toxic plume exists upgradient from a water supply well. If we combine the usual relationship between flow rate, velocity, and cross-sectional area,

$$Q = Av \tag{43}$$

with Darcy's law, we can solve for velocity:

$$\text{Darcy velocity } v = \frac{Q}{A} = \frac{KA(dh/dL)}{A} = K\frac{dh}{dL} \tag{44}$$

The velocity given in (44) is known as the *Darcy velocity*. It is not a "real" velocity in that, in essence, it assumes that the full cross-sectional area *A* is available

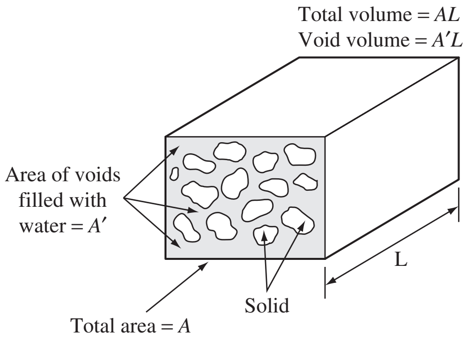


FIGURE 33 The cross-sectional area available for flow, A', is less than the overall cross-sectional area of the aquifer, A.

for water to flow through. Since much of the cross-sectional area is made up of solids, the actual area through which all of the flow takes place is much smaller, and as a result, the *real groundwater velocity is considerably faster than the Darcy velocity.*

As suggested in Figure 33, consider the cross-section of an aquifer to be made up of voids and solids, with A representing the total cross-sectional area and A' being the area of voids filled with water. Letting v' be the actual *average linear velocity* (sometimes called the *seepage velocity*), we can rewrite (44) as

$$Q = Av = A'v' \tag{45}$$

Solving for v' and introducing an arbitrary length of aquifer L gives

$$v' = \frac{Av}{A'} = \frac{ALv}{A'L} = \frac{\text{Total volume} \times v}{\text{Void volume}} \tag{46}$$

But recall that the ratio of void volume to total volume is just the porosity η introduced in (39). Therefore, the actual average linear velocity through the aquifer is the Darcy velocity divided by porosity;

$$v' = \frac{\text{Darcy velocity}}{\text{Porosity}} = \frac{v}{\eta} \tag{47}$$

or, using (44),

$$\text{Average linear velocity} \quad v' = \frac{K}{\eta}\left(\frac{dh}{dL}\right) \tag{48}$$

EXAMPLE 13 A Groundwater Plume

Suppose the aquifer in Example 12 has become contaminated upgradient of the two wells. Consider the upgradient well as a *monitoring* well whose purpose is to provide early detection of the approaching plume to help protect the second, drinking water well. How long after the monitoring well is contaminated would you expect the drinking water well to be contaminated? Make the following three assumptions (each of which will be challenged later):

1. Ignore dispersion or diffusion of the plume (that is, it does not spread out).
2. Assume the plume moves at the same speed as the groundwater.
3. Ignore the “pulling” effect of the drinking water well.

The aquifer has a porosity of 35 percent.

Solution The Darcy velocity is given by (44):

$$\text{Darcy velocity } v = K\frac{dh}{dL} = 50 \text{ m/d} \times 0.004 = 0.20 \text{ m/d}$$

The average linear velocity is the Darcy velocity divided by porosity:

$$\text{Average linear velocity } v' = \frac{0.20 \text{ m/d}}{0.35} = 0.57 \text{ m/d}$$

so the time to travel the 500-m distance would be

$$t = \frac{500 \text{ m}}{0.57 \text{ m/d}} = 877 \text{ days} = 2.4 \text{ yr}$$

As this example illustrates, groundwater moves very slowly.

12 Contaminant Transport

A few comments on the assumptions made in Example 13 are in order. The first assumption was that there would be no dispersion or diffusion, so the contamination would move forward with a sharp front, the so-called *plug flow* case. The second was that the plume moves at the same speed as the groundwater flow. The third was that the drinking water well did not pull the plume toward it, which would make the plume speed up as it approached the well. In reality, all three of these assumptions need to be examined.

Dispersion and Diffusion

When there is a difference in concentration of a solute in groundwater, molecular *diffusion* will tend to cause movement from regions of high concentration to regions where the concentration is lower. That is, even in the absence of groundwater movement, a blob of contaminant will tend to diffuse in all directions, blurring the boundary between it and the surrounding groundwater. A second process that causes a contaminant plume to spread out is *dispersion*. Since a contaminant plume follows irregular pathways as it moves, some finding large pore spaces in which it can move quickly, while other portions of the plume have to force their way through more confining voids, there will be a difference in speed of an advancing plume that tends to cause the plume to spread out. Since diffusion and dispersion both tend to smear the edges of the plume, they are sometimes linked together and simply referred to as *hydrodynamic dispersion*.

Since contamination spreads out as it moves, it does not arrive all at once at a given location downgradient. This effect is easily demonstrated in the laboratory by establishing a steady state flow regime in a column packed with a homogeneous granular material and then introducing a continuous stream of a nonreactive tracer, as shown in Figure 34*a*. If the tracer had no dispersion, a plot of concentration versus the time that it takes to leave the column would show a sharp jump. Instead, the front arrives smeared out, as shown in Figure 34*b*.

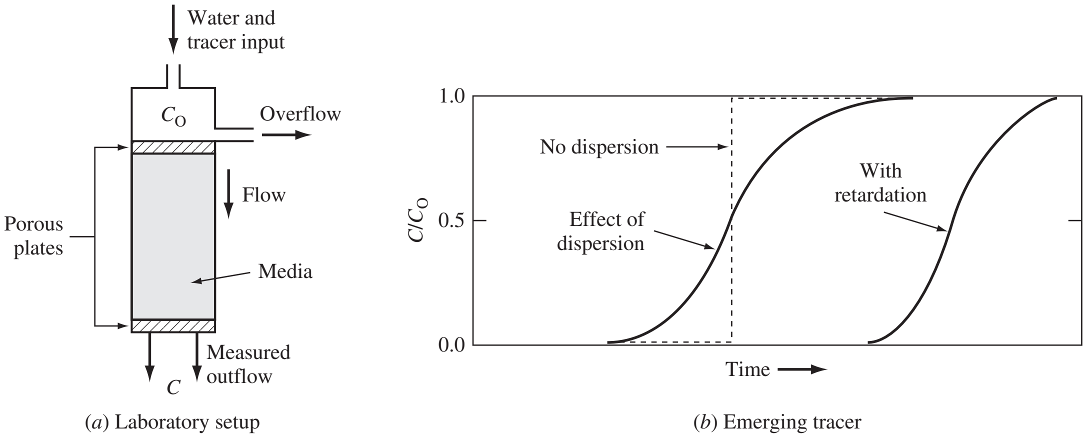


FIGURE 34 Dispersion and retardation as a continuous feed of tracer passes through a column. With no dispersion, the tracer emerges all at once. With retardation and dispersion, the tracer smears out and emerges with some delay.

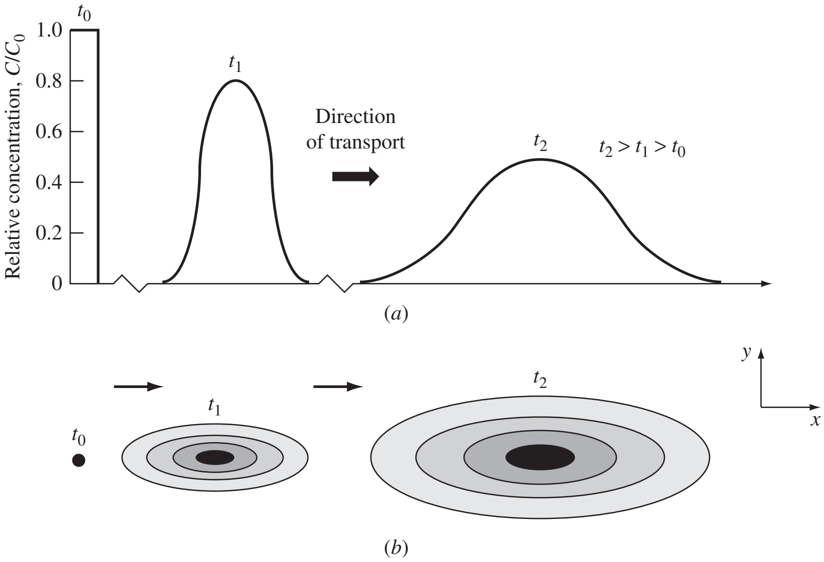


FIGURE 35 An instantaneous (pulse) source in a flowfield creates a plume that spreads as it moves downgradient: (*a*) in one dimension, (*b*) in two dimensions (darker colors mean higher concentrations).
(*Source: Ground Water Contamination* by Bedient/Rifai/Newell, © 1994. Reprinted by permission of Prentice-Hall, Inc., Upper Saddle River, NJ.)

While the column experiment described in Figure 34 illustrates longitudinal dispersion, there is also some dispersion normal to the main flow. Figure 35 shows what might be expected if an instantaneous (pulse) source of contaminant is injected into a flowfield, such as might occur with an accidental spill that contaminates groundwater. As the plume moves downgradient, dispersion causes the plume to spread in the longitudinal as well as orthogonal directions.

Retardation

A second assumption used in Example 13 was that the contaminants would move at the same speed as the groundwater, which may or may not be the case in reality. As contaminants move through an aquifer, some are *absorbed* by solids along the way, and some are *adsorbed* (that is, adhered to the surface of particles). The general term *sorption* applies to both processes. The ratio of total contaminant in a unit volume of aquifer to the contaminant dissolved in groundwater is called the *retardation factor.* For example, the retardation factor for chloride ions is 1, which means all of it is dissolved in groundwater. A retardation factor of 5 for some other contaminant means 20 percent is dissolved in groundwater, and 80 percent is sorbed to the aquifer solids.

When the amount sorbed reaches equilibrium in the aquifer, the retardation factor takes on a more intuitive meaning. It is the ratio of the average velocity of groundwater v' to the velocity of the sorbed material, v_s:

$$\text{Retardation factor} = R = \frac{\text{Average groundwater velocity, } v'}{\text{Velocity of the sorbed material, } v_s} \geq 1 \tag{49}$$

So, for example, a retardation factor of 5 means the groundwater travels five times as far in a given period of time as does the contaminant.

Figure 36 illustrates the impact of sorption on the time required for a pulse of contaminant to make its way to a monitoring well. A nonsorbing compound shows modest dispersion and emerges rather quickly. When a substance sorbs at a rate proportional to its concentration, it is said to exhibit linear sorption, and such compounds show retardation, but as shown in the figure, the pulse maintains its

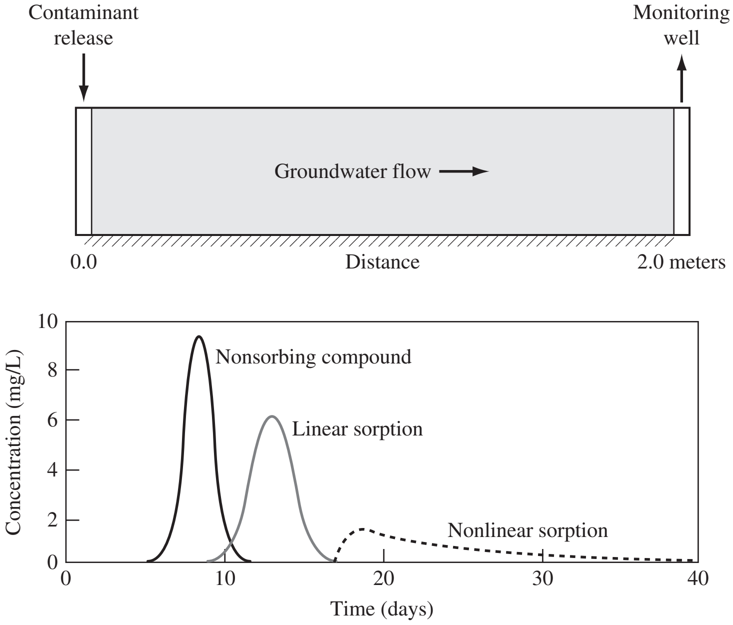


FIGURE 36 Influence of sorption on contaminant transport. A pulse of contaminant at $t = 0$ shows up in the monitoring well at different times depending on the extent of sorption. (*Source:* National Research Council, 1994.)

symmetrical shape. For nonlinear sorption, a long tail is shown, suggesting that much larger volumes of water will be required to flush the system.

It is relatively straightforward to perform retardation experiments in the laboratory, producing results such as are shown in Figure 36. In an actual aquifer, however, retardation experiments are much more difficult. In one such in situ experiment (Roberts, Goltz, and Mackay, 1986), a number of organic solutes, including carbon tetrachloride (CTET) and tetrachloroethylene (PCE), were injected into the groundwater, along with a chloride tracer, which is assumed to move at the same rate as the groundwater itself. The position of the three plumes roughly 21 months later, as indicated by two-dimensional contours of depth-averaged concentrations, is shown in Figure 37. As can be seen, the center of the chloride plume has moved roughly 60 m away from the point of injection ($x = 0, y = 0$), the CTET has moved about 25 m, and the PCE has moved only a bit over 10 m.

Using actual plume measurements such as are shown in Figure 37, Roberts et al. (1986) determined retardation factors for CTET and PCE, as well as for bromoform (BROM), dichlorobenzene (DCB), and hexachlorethane (HCE). As shown in

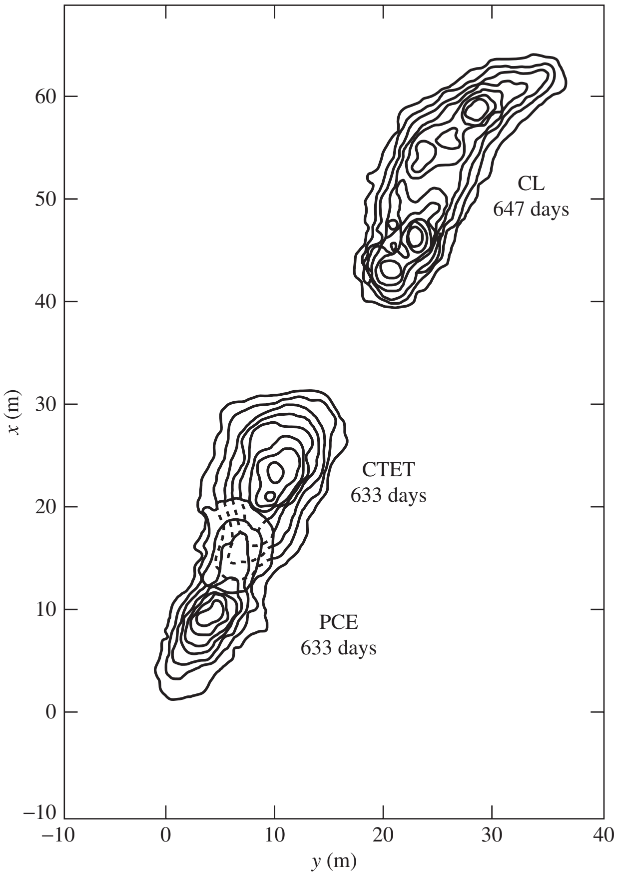


FIGURE 37 Plume separation for chloride (CL), carbon tetrachloride (CTET), and tetrachloroethylene (PCE) 21 months after injection.
(*Source:* Roberts et al., 1986.)

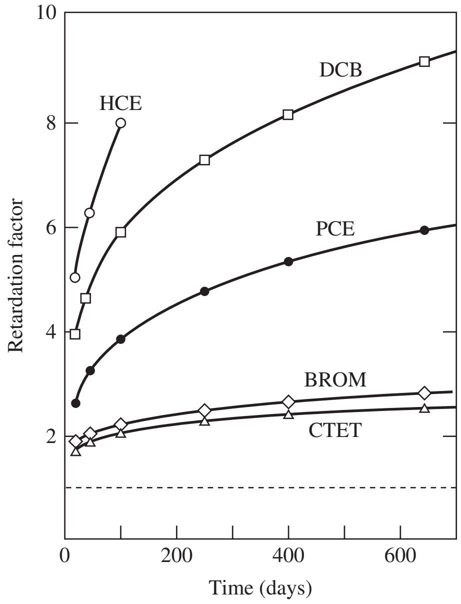


FIGURE 38 Retardation factors for carbon tetrachloride (CTET), tetrachloroethylene (PCE), bromoform (BROM), dichlorobenzene (DCB), and hexachloroethane (HCE).
(*Source:* Roberts et al., 1986.)

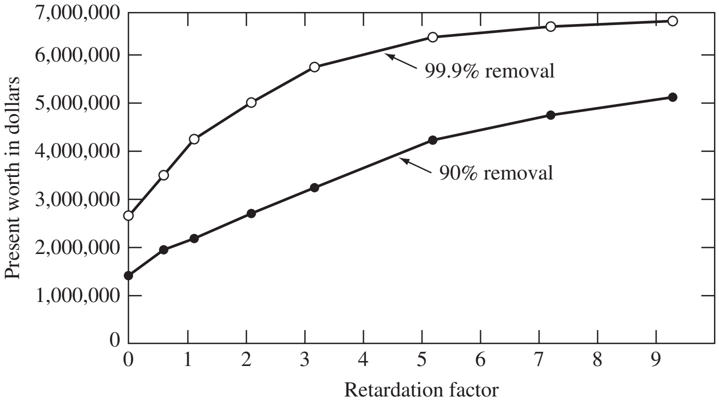


FIGURE 39 Retardation increases the operating cost of a pump-and-treat system.
(Based on $650,000 initial capital cost, $180,000 annual O&M, 3.5 percent discount factor, 25-year equipment life, and 1 pore-volume pumping per year.)
(*Source:* National Research Council, 1994.)

Figure 38, retardation factors are not constants but instead appear to increase over time and eventually reach a steady state value.

In a conventional pump-and-treat system for groundwater cleanup, retardation can greatly increase the cost of the system, since pumping must continue for a longer period of time. Figure 39 shows an estimate of the operating cost of such a system as a function of retardation.

13 Cone of Depression

Another assumption made in Example 13 was that pumping water from the aquifer would not affect the hydraulic gradient. That is, in fact, not the case, since there must be a gradient toward the well to provide flow to the well. Moreover, the faster the well is pumped, the steeper the gradient will be in the vicinity of the well. When a well is pumped, the water table in an unconfined aquifer, or the piezometric surface for a confined aquifer, forms a *cone of depression* in the vicinity of the well such as is shown in Figure 40.

If we make enough simplifying assumptions, we can use Darcy's law to derive an expression for the shape of the cone of depression, as follows. We will assume that pumping has been steady for a long enough time that the shape of the cone is no longer changing; that is, we assume equilibrium conditions. In addition, we will assume that the original water table is horizontal. If we also assume that the drawdown is small relative to the depth of the aquifer and that the well draws from the entire depth of the aquifer, then the flow to the well is horizontal and radial. Under these conditions, and using Figures 40 and 41, we can write Darcy's law for flow

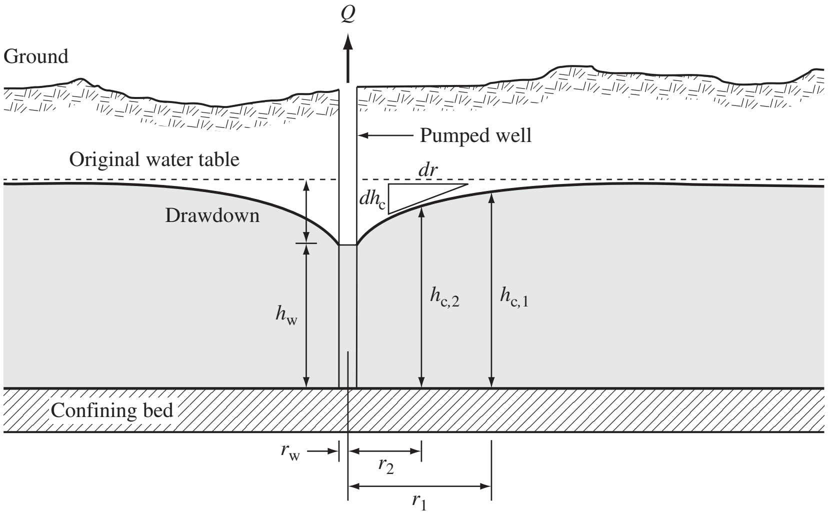


FIGURE 40 Cone of depression in an unconfined aquifer.

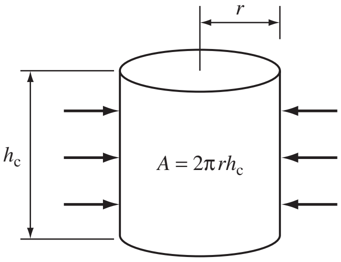


FIGURE 41 Flow to the well passes through a cylinder of area $A = 2\pi r h_c$.

passing through a cylinder of radius r, depth h_c, and hydraulic gradient (dh_c/dr) as

$$Q = KA\frac{dh_c}{dr} = K2\pi r h_c \frac{dh_c}{dr} \tag{50}$$

where the cross-sectional area of an imaginary cylinder around the well is $2\pi r h_c$, K is the hydraulic conductivity, and (dh_c/dr) is the slope of the water table at radius r. The flow through the cylinder toward the well equals the rate at which water is being pumped from the well, Q. It's important to recognize the subtle distinction between h_c that appears in (50) and the hydraulic head, h, that first appears in (40) and Figure 28. The hydraulic head, h, is measured from any suitable horizontal datum such as sea level, whereas h_c is the height of the water table above the confining layer at the bottom of the aquifer. (In other words, h_c is the depth of an unconfined aquifer at a distance r from the well.)

Rearranging (50) into an integral form,

$$\int_{r_2}^{r_1} Q\frac{dr}{r} = K2\pi \int_{h_{c,2}}^{h_{c,1}} h\, dh_c \tag{51}$$

Notice how the limits have been set up. The radial term is integrated between two arbitrary values, r and r_1, corresponding to heads $h_{c,2}$ and $h_{c,1}$. Integrating (51) gives

$$Q \ln\left(\frac{r_1}{r_2}\right) = \pi K(h_{c,1}^2 - h_{c,2}^2) \tag{52}$$

so

$$\text{Unconfined aquifer:} \qquad Q = \frac{\pi K(h_{c,1}^2 - h_{c,2}^2)}{\ln(r_1/r_2)} \tag{53}$$

Equation (53), for an unconfined aquifer, can be used in several ways. It can be used to estimate the shape of the cone of depression for a given pumping rate, or it can be rearranged and used to determine the aquifer hydraulic conductivity K. To estimate K, two observation wells capable of measuring heads $h_{c,2}$ and $h_{c,1}$ are set up at distances r_2 and r_1 from a well pumping at a known rate Q. The data obtained can then be used in (53), as the following example illustrates. Actually, since equilibrium conditions take fairly long to be established, and since more information about the characteristics of the aquifer can be obtained with a transient study, this method of obtaining K is not commonly used. A similar, but more complex, model based on a transient analysis of the cone of depression is used (see, for example, Freeze and Cherry, 1979).

EXAMPLE 14 Determining *K* from the Cone of Depression

Suppose a well 0.30 m in diameter has been pumped at a rate of 6,000 m^3/day for a long enough time that steady-state conditions apply. An observation well located 30 m from the pumped well has been drawn down by 1.0 m, and another well at 100 m is drawn down by 0.50 m. The well extends completely through an unconfined aquifer 30.0 m thick.

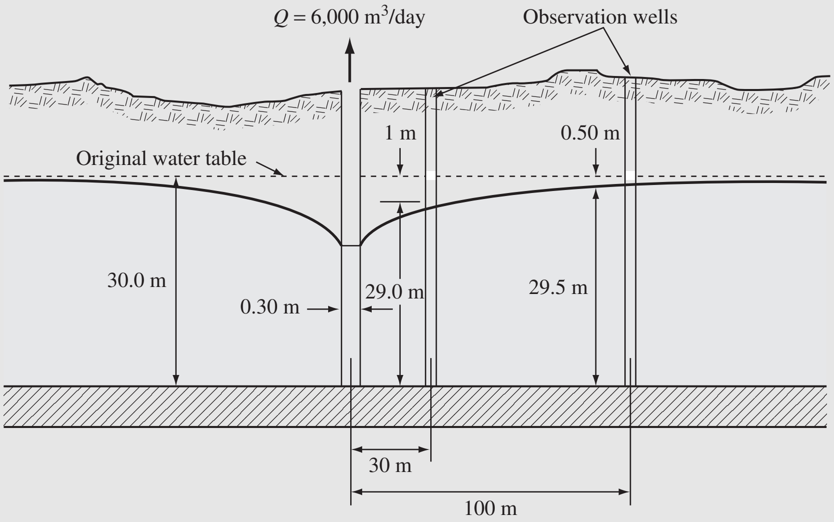


FIGURE 42 Data from Example 14 used to determine hydraulic conductivity.

a. Determine the hydraulic conductivity K.
b. Estimate the drawdown at the well.

Solution It helps to put the data onto a drawing, as has been done in Figure 42.

a. Rearranging (53) for K and then inserting the quantities from the figure gives

$$K = \frac{Q \ \ln(r_1/r_2)}{\pi(h_{c,1}^2 - h_{c,2}^2)}$$

$$= \frac{(6{,}000 \text{ m}^3/\text{d}) \ln(100/30.0)}{\pi[(29.5)^2 - (29.0)^2]\text{ m}^2} = 78.6 \text{ m/d}$$

b. To estimate the drawdown, let $r = r_w = 0.30/2 = 0.15$ m, and let us use the first observation well for $r_1 = 30$ m and $h_{c,1} = 29$ m. Using (55) to find the head at the outer edge of the well h_w gives

$$Q = \frac{\pi K(h_{c,1}^2 - h_w^2)}{\ln(r_1/r_w)} = \frac{\pi(78.6 \text{ m/d})(29.0^2 - h_w^2)\text{ m}^2}{\ln(30.0/0.15)} = 6{,}000 \text{ m}^3/\text{d}$$

Solving for h_w yields

$$h_w = 26.7 \text{ m}$$

so the drawdown would be $30.0 - 26.7 = 3.3$ m

Equation (53) was derived for an unconfined aquifer. The derivation for a confined aquifer is similar, but now the height of the cylinder at radius r, through which water flows to reach the well, is a constant equal to the thickness of the

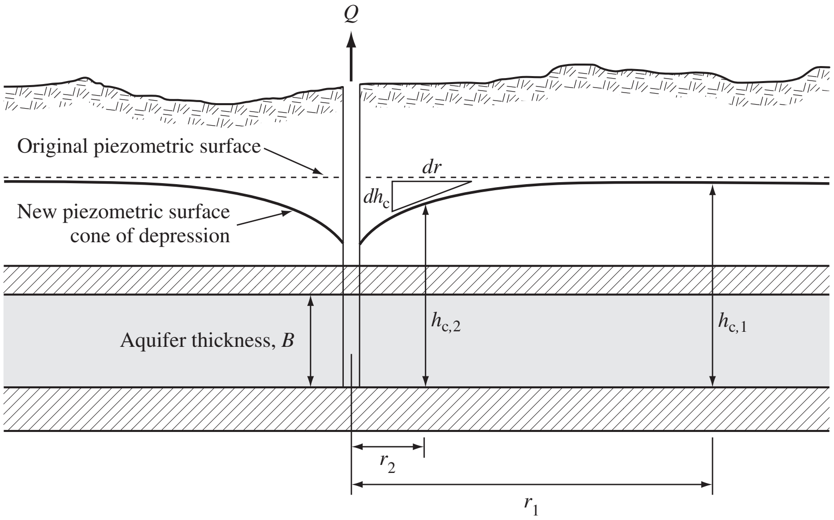


FIGURE 43 Cone of depression for a confined aquifer.

aquifer, B. Also, as shown in Figure 43, the cone of depression now appears in the piezometric surface.

With the same assumptions as were used for the unconfined aquifer, we can write Darcy's law as

$$Q = K2\pi rB\left(\frac{dh_c}{dr}\right) \tag{54}$$

which integrates to

$$\text{Confined aquifer:} \qquad Q = \frac{2\pi KB(h_{c,1} - h_{c,2})}{\ln (r_1/r_2)} \tag{55}$$

14 Capture-Zone Curves

The most common way to begin the cleanup of contaminated groundwater is to install extraction wells. As an extraction well pumps water from the contaminated aquifer, the water table (or piezometric surface, if it is a confined aquifer) in the vicinity of the well is lowered, creating a hydraulic gradient that draws the plume toward the well. With properly located extraction wells, the polluted groundwater can be pumped out of the aquifer, cleaned in an above-ground treatment facility, and used or returned to the aquifer. The process is referred to as *pump-and-treat* technology.

Figure 44 shows the effect of an extraction well located in a region with a uniform and steady regional groundwater flow that is parallel to and in the direction of the negative x-axis. As water is extracted, the natural streamlines are bent toward the well as shown in the figure. The outer envelope of the streamlines that converge on the well is called the *capture-zone* curve. Groundwater inside the capture zone is extracted; water outside the capture zone is not. Flow lines outside of the capture

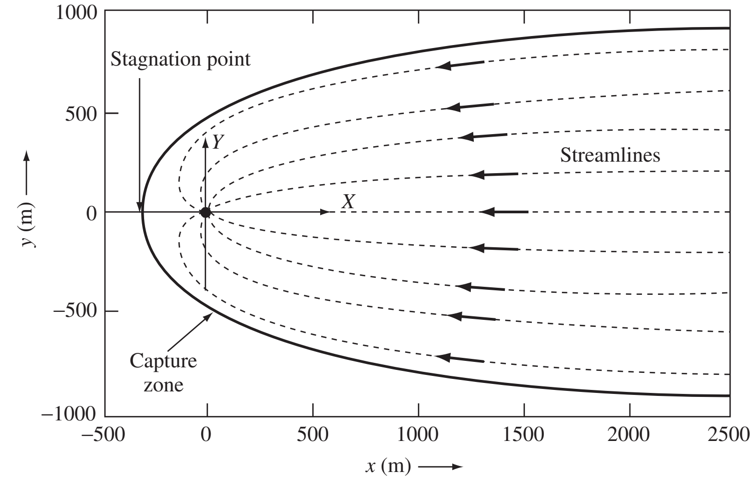


FIGURE 44 A single extraction well located at $x = 0, y = 0$, in an aquifer with regional flow along the x-axis. The capture zone is the region in which all flow lines converge on the extraction well. Drawn for $(Q/B\nu) = 2{,}000$.
(*Source:* Javendal and Tsang, 1986.)

zone may curve toward the well, but the regional flow associated with the natural hydraulic gradient is strong enough to carry that groundwater past the well.

Javandel and Tsang (1986) developed the use of capture-zone type curves as an aid to the design of extraction well fields for aquifer cleanup. Their analysis is based on an assumed ideal aquifer (that is, one that is homogeneous, isotropic, uniform in cross-section, and infinite in width). Also, it is either confined or unconfined with an insignificant drawdown relative to the total thickness of the aquifer. They assume extraction wells that extend downward through the entire thickness of the aquifer and are screened to extract uniformly from every level. These are very restrictive assumptions that are unlikely ever to be satisfied in any real situation; nonetheless, the resulting analysis does give considerable insight into the main factors that affect more realistic, but complex, models.

For a single extraction well located at the origin of the coordinate system shown in Figure 44, Javandel and Tsang derive the following relationship between the x and y coordinates of the envelope surrounding the capture zone:

$$y = \pm\frac{Q}{2B\nu} - \frac{Q}{2\pi B\nu}\tan^{-1}\frac{y}{x} \tag{56}$$

where

B = aquifer thickness (m)
ν = Darcy velocity, conductivity × gradient (m/day)
Q = pumping rate from well (m^3/day)

Equation (56) can be rewritten in terms of an angle ϕ (radians) drawn from the origin to the x,y coordinate of interest on the capture-zone curve, as shown in Figure 45. That is,

$$\tan\phi = \frac{y}{x} \tag{57}$$

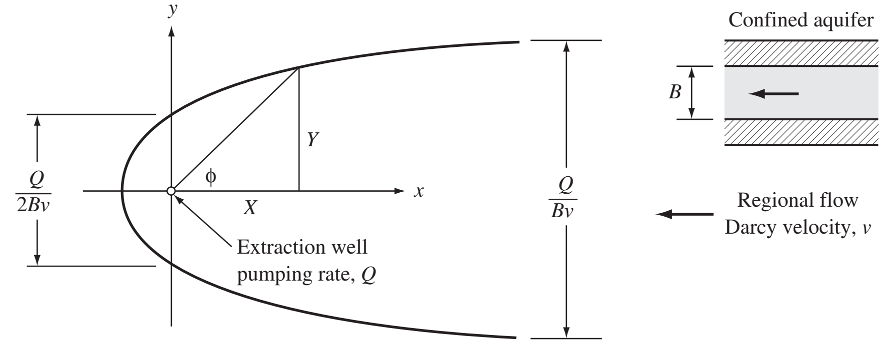


FIGURE 45 Capture-zone curve for a single extraction well located at the origin in an aquifer with regional flow velocity v, thickness B, and pumping rate Q.

so that, for $0 \leq \phi \leq 2\pi$,

$$y = \frac{Q}{2Bv}\left(1 - \frac{\phi}{\pi}\right) \tag{58}$$

Equation (58) makes it easy to predict some important measures of the capture zone. For example, as x approaches infinity, $\phi = 0$, and $y = Q/(2Bv)$, which sets the maximum total width of the capture zone as $2[Q/(2Bv)] = Q/Bv$. For $\phi = \pi/2$, $x = 0$ and y becomes equal to $Q/(4Bv)$. Thus, the width of the capture zone along the y-axis is $Q/(2Bv)$, which is only half as broad as it is far from the well. These relationships are illustrated in Figure 45.

The width of the capture zone is directly proportional to the pumping rate Q and inversely proportional to the product of the regional (without the effect of the well) Darcy flow velocity v and the aquifer thickness B. Higher regional flow velocities therefore require higher pumping rates to capture the same area of plume. Usually there will be some maximum pumping rate determined by the acceptable amount of drawdown at the well that restricts the size of the capture zone. Assuming that the aquifer characteristics have been determined and the plume boundaries defined, one way to use capture-zone type curves is first to draw the curve corresponding to the maximum acceptable pumping rate. Then by superimposing the plume onto the capture zone curve (drawn to the same scale), it can be determined whether or not a single well will be sufficient to extract the entire plume and, if it is, where the well can be located. Figure 46 suggests the approach, and the following example illustrates its use for a very idealized plume.

EXAMPLE 15 A Single Extraction Well

Consider a confined aquifer having thickness 20 m, hydraulic conductivity 1.0×10^{-3} m/s, and a regional hydraulic gradient equal to 0.002. The maximum pumping rate has been determined to be 0.004 m^3/s. The aquifer has been contaminated and, for simplicity, consider the plume to be rectangular, with width 80 m. Locate a single extraction well so that it can totally remove the plume.

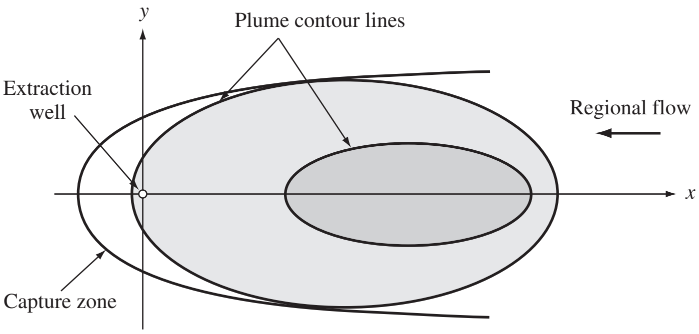


FIGURE 46 Superimposing the plume onto a capture-zone type curve for a single extraction well.

Solution *Let* us first determine the regional Darcy velocity:

$$\nu = K\frac{dh}{dx} = 1.0 \times 10^{-3}\ \text{m/s} \times 0.002 = 2.0 \times 10^{-6}\ \text{m/s}$$

Now find the critical dimensions of the capture zone. Along the y-axis, its width is

$$\frac{Q}{2B\nu} = \frac{0.004\ \text{m}^3/\text{s}}{2 \times 20\ \text{m} \times 2.0 \times 10^{-6}\ \text{m/s}} = 50\ \text{m}$$

and, at an infinite distance upgradient, the width of the capture zone is

$$\frac{Q}{B\nu} = 100\ \text{m}$$

So the 80-m-wide plume will fit within the capture zone if the well is located some distance downgradient from the front edge. Using Figure 47 as a guide, we can determine the distance x that must separate the plume from the well. From (60), with $y = 40$ m,

$$y = \frac{Q}{2B\nu}\left(1 - \frac{\phi}{\pi}\right) = 40 = 50\left(1 - \frac{\phi}{\pi}\right)$$

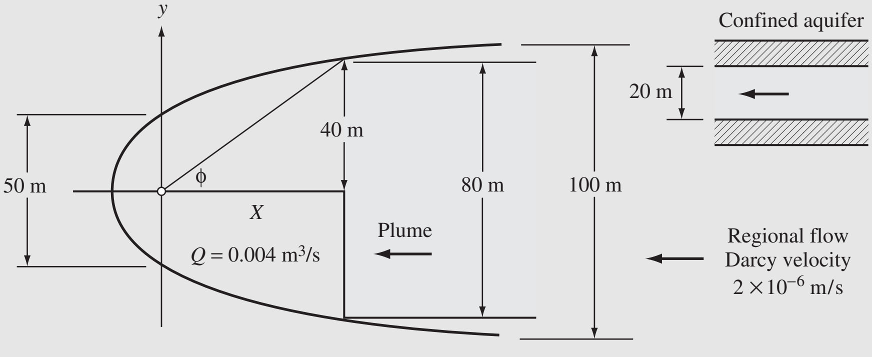


FIGURE 47 Example problem with single extraction well.

so the angle, in radians to the point where the plume just touches the capture zone, is

$$\phi = 0.2\pi \text{ rad}$$

and, from Figure 47,

$$x = \frac{y}{\tan\phi} = \frac{40}{\tan(0.2\pi)} = 55 \text{ m}$$

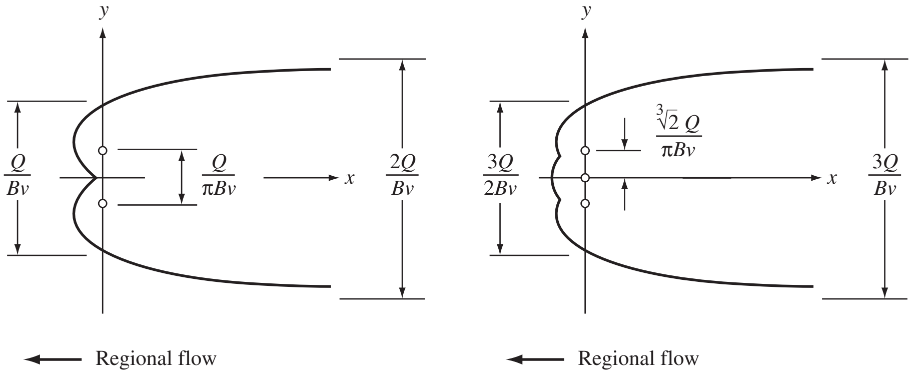
The extraction well should be placed 55 m ahead of the oncoming plume and directly in line with it.

The single-well solution found in Example 15 is not necessarily a good one. The extraction well is far downgradient from the plume, which means a large volume of clean groundwater must be pumped before any of the contaminated plume even reaches the well. That can add years of pumping time and raise total costs considerably before the aquifer is rehabilitated.

A better solution would involve more extraction wells placed closer to the head of the plume. Javandel and Tsang have derived capture-zone type curves for a series of n optimally placed wells, each pumping at the same rate Q, lined up along the y-axis. Optimality is defined to be the maximum spacing between wells that will still prevent any flow from passing between them. The separation distance for two wells has been determined to be $Q/(\pi B\nu)$. If the wells are any farther apart than this, some of the flow can pass between them and not be captured. With this optimal spacing, the two wells will capture a plume as wide as $Q/(B\nu)$ along the y-axis and as wide as $2Q/B\nu$ far upgradient from the wells, as shown in Figure 48*a*. Analogous parameters for the case of three optimally spaced wells are given in Figure 48*b*.

A general equation for the positive half of the capture-zone type curve for n optimally spaced wells arranged symmetrically along the y-axis is:

$$y = \frac{Q}{2B\nu}\left(n - \frac{1}{\pi}\sum_{i=1}^{n}\phi_i\right) \tag{59}$$



(*a*) (*b*)

FIGURE 48 Capture-zone type curves for optimally spaced wells along the y-axis, each pumping at the rate Q: (*a*) two wells; (*b*) three wells.

where ϕ_i is the angle between a horizontal line through the *i*th well and a spot on the capture-zone curve.

To demonstrate use of these curves, let us redo Example 15, but this time we will use two wells.

EXAMPLE 16 **Capture Zone for Two Wells**

Consider the same plume that was described in Example 15—that is, it is rectangular with width 80 m in a confined aquifer with thickness $B = 20$ m and Darcy velocity $v = 2.0 \times 10^{-6}$ m/s.

a. If two optimally located wells are aligned along the leading edge of the plume, what minimum pumping rate Q would assure complete plume capture? How far apart should the wells be?

b. If the plume is 1,000 m long, and the aquifer porosity is 0.30, how long would it take to pump an amount of water equal to the volume of water contained in the plume? Notice that it would take much, much longer to pump out the whole plume than this estimate would suggest since there will be some uncontaminated groundwater removed with the plume, and *we are ignoring retardation*. The effect of retardation on the amount of pumping needed to clean an aquifer will be considered later.

Solution

a. The plume width along the *y*-axis (also the leading edge of the plume) is 80 m, so from Figure 48,

$$\frac{Q}{Bv} = \frac{Q}{20 \text{ m} \times 2.0 \times 10^{-6} \text{ m/s}} = 80 \text{ m}$$

$$Q = 0.0032 \text{ m}^3\text{/s (each)}$$

From Figure 48, the optimal spacing between two wells is given as

$$\text{Optimal separation} = \frac{Q}{\pi B v}$$

$$= \frac{0.0032 \text{ m}^3\text{/s}}{\pi \times 20 \text{ m} \times 2.0 \times 10^{-6} \text{ m/s}} = 25.5 \text{ m}$$

These dimensions are shown in Figure 49.

b. The volume of contaminated water in the plume is the porosity times the plume volume:

$$V = 0.30 \times 80 \text{ m} \times 20 \text{ m} \times 1{,}000 \text{ m} = 480{,}000 \text{ m}^3$$

At a total pumping rate of $2 \times 0.0032 \text{ m}^3\text{/s} = 0.0064 \text{ m}^3\text{/s}$, it would take

$$t = \frac{480{,}000 \text{ m}^3}{0.0064 \text{ m}^3\text{/s} \times 3600 \text{ s/hr} \times 24 \text{ hr/day} \times 365 \text{ day/yr}} = 2.4 \text{ yr}$$

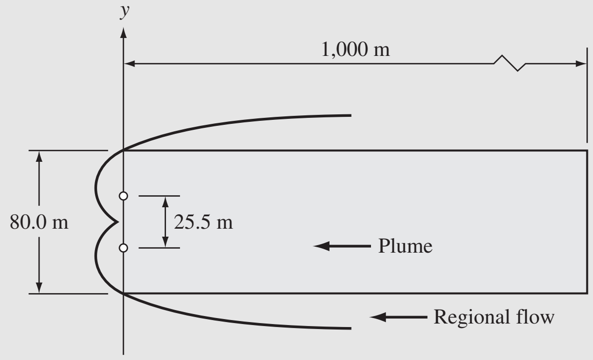


FIGURE 49 Example problem with two extraction wells.

to pump a volume of water equal to that contained in the plume. Again note, however, that the time required to pump the actual plume would be much greater than 2.4 years, depending on retardation and the fraction of the water pumped that is actually from the plume.

Since real contaminant plumes and aquifers are so much more complex than the ideal ones considered in this chapter, designing a well field is a much more difficult task than has been presented here. Interested readers are referred to more advanced texts such as those by Gupta (1989) or Bedient et al. (1994).

15 Control of Groundwater Plumes

We have seen how a column of extraction wells lined up in front of an approaching groundwater plume can be used to intercept the pollution, preventing it from contaminating the aquifer any further downgradient. Another way to protect downgradient uses, such as a drinking water well, is to use some combination of *extraction wells* and *injection wells*. Extraction wells are used to lower the water table (or piezometric surface), creating a hydraulic gradient that draws the plume to the wells. Injection wells raise the water table and push the plume away. Through careful design of the location and pumping rates of such wells, the hydraulic gradient can be manipulated in such a way that plumes can be kept away from drinking water wells and drawn toward extraction wells. Extracted, contaminated groundwater can then be treated and reinjected back into the aquifer, reused, or released into the local surface water system.

Figure 50 shows two strategies to control a contaminant plume that endangers a nearby well, here called a production well. In Figure 50*a* an injection well pushes the plume away from the production well, and an extraction well pulls it away from the production well. Additional wells would be placed in the plume itself to extract the contaminants. In Figure 50*b*, a different strategy is used. The injection well is used to push the plume into the vicinity of the extraction well so that the plume can be removed more quickly from the aquifer.

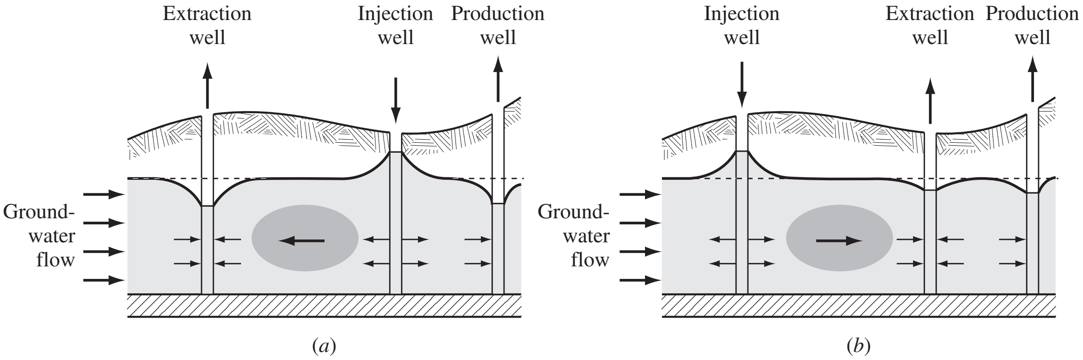


FIGURE 50 Manipulating the hydraulic gradient with multiple wells: (*a*) the injection and extraction wells push and pull the plume away from the production well; (*b*) the injection well pushes the plume into the extraction well.

Manipulating the hydraulic gradient to control and remove a groundwater plume is called *hydrodynamic control.* The well field used to create hydrodynamic control of a contaminant plume can be anything from a single extraction well, properly placed and pumping at the right rate, to a complex array of extraction wells and injection wells.

Hydrodynamic control of groundwater plumes is an effective way to protect production wells. Construction costs are relatively low, and if the original well field is not sufficient to control the plume adequately, it is always possible to add additional wells when necessary. In addition, the environmental disturbance on the surface is minimal. On the negative side, the operation and maintenance costs can be high since the wells must be pumped for many years.

An alternate approach to protecting production wells is literally to surround the plume with a wall of impermeable material that extends from the surface down to the aquitard. A number of such *physical containment* schemes involving different materials and construction techniques are possible. Perhaps the most common is a *slurry cutoff wall* in which a narrow trench, about 1 to 2 m wide, is dug around the plume, down to bedrock, and then backfilled with a relatively impermeable mixture of soil and bentonite. The slurry wall keeps the plume from migrating off-site while other remediation measures are applied to clean up the aquifer.

16 Contaminants in Groundwater

In the discussion of surface water contamination, the focus was on controlling the spread of infectious diseases, reducing the oxygen demand of the wastes, and removing nutrients to help reduce eutrophication. Surface waters are, of course, also subject to contamination by a range of toxic chemicals, including pesticides, metals, PCBs, and so forth, but those pollutants were mentioned only briefly in that context. In the study of groundwater contamination, however, toxic chemicals are the principal pollutants of concern.

The 25 most frequently detected contaminants found in groundwater at hazardous waste sites are listed in Table 13. Nine of these contaminants are

TABLE 13

The 25 Most Frequently Detected Groundwater Contaminants at Hazardous Waste Sites

Rank	Compound	Common Sources
1	Trichloroethylene	Dry cleaning; metal degreasing
2	Lead	Gasoline (prior to 1975); mining; construction material (pipes); manufacturing
3	Tetrachloroethylene	Dry cleaning; metal degreasing
4	Benzene	Gasoline; manufacturing
5	Toluene	Gasoline; manufacturing
6	Chromium	Metal plating
7	Methylene chloride	Degreasing; solvents; paint removal
8	Zinc	Manufacturing; mining
9	1,1,1-Trichloroethane	Metal and plastic cleaning
10	Arsenic	Mining; manufacturing
11	Chloroform	Solvents
12	1,1-Dichloroethane	Degreasing; solvents
13	1,2-Dichloroethene, *trans-*	Transformation product of 1,1,1-trichloroethane
14	Cadmium	Mining; plating
15	Manganese	Manufacturing; mining; occurs in nature as oxide
16	Copper	Manufacturing; mining
17	1,1-Dichloroethene	Manufacturing
18	Vinyl chloride	Plastic and record manufacturing
19	Barium	Manufacturing; energy production
20	1,2-Dichloroethane	Metal degreasing; paint removal
21	Ethylbenzene	Styrene and asphalt manufacturing; gasoline
22	Nickel	Manufacturing; mining
23	Di(2-ethylhexy)phthalate	Plastics manufacturing
24	Xylenes	Solvents; gasoline
25	Phenol	Wood treating; medicines

Source: National Research Council, 1994.

inorganic: lead (Pb), chromium (Cr), zinc (Zn), arsenic (As), cadmium (Cd), manganese (Mn), copper (Cu), barium (Ba), and nickel (Ni). Lead, arsenic, and cadmium are not only high on the list, but they are also among the most toxic chemicals found in groundwater. The rest of the top 25 are organic chemicals.

Nonaqueous-Phase Liquids

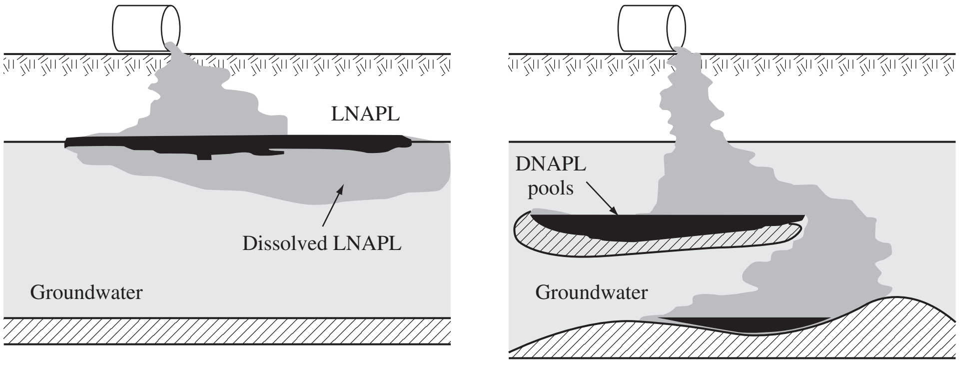
Many of the organic chemicals listed in Table 13 do not dissolve very well in water. They are called *nonaqueous-phase liquids* (NAPLs), and their presence makes the task of restoring an aquifer to drinking water quality very difficult or even impossible. NAPLs are generally divided into two categories: Those that are more dense than water are called dense NAPLs (DNAPLs), whereas those that are less dense than water are called light NAPLs (LNAPLs).

Examples of DNAPLs include chlorinated solvents, such as trichloroethylene and tetrachloroethylene, polychlorinated biphenyls (PCBs), pesticides such as chlordane, and polycyclic aromatic hydrocarbons (PAHs). Many of the LNAPLs are fuel

TABLE 14

Examples of Nonaqueous-Phase Liquids

Pollutant	Specific Gravity	Aqueous Solubility (mg/L)
DNAPLs		
Carbon tetrachloride	1.58	7.57×10^2
Trichloroethylene	1.47	1.10×10^3
Tetrachloroethylene	1.63	1.50×10^2
Phenol	1.24	9.3×10^4
LNAPLs		
Benzene	0.873	1.75×10^3
Toluene	0.862	5.35×10^2
p-Xylene	0.861	1.98×10^2



(*a*) LNAPLs (*b*) DNAPLs

FIGURE 51 Nonaqueous-phase liquids do not dissolve very well in groundwater: (*a*) LNAPLs float on top of groundwater; (*b*) DNAPLs form pools.

hydrocarbons, including crude oil, gasoline, and benzene. Table 14 lists some important NAPLs along with their solubility in water and their specific gravity.

As LNAPLs enter the unsaturated zone, some may dissolve into water in the pores, some may volatilize and become mixed with air in other pore spaces, and some may adsorb onto soil particles. When LNAPLs reach the water table, they do not dissolve well and instead spread out to form a layer of contaminant floating on top of the saturated zone, as suggested in Figure 51*a*. DNAPLs also sink, dissolve, adsorb, and volatilize in the unsaturated zone, but when they reach the water table, they keep on sinking until they reach a layer of relatively impermeable material. There they form pools that can overflow and sink to the next impermeable layer, as suggested in Figure 51*b*. Portions of DNAPLs tend to form small globules that become trapped in pore spaces, making them virtually impossible to remove by pumping. As they remain lodged in tiny crevices, and as they accumulate in pools on the bottom of aquifers, they slowly dissolve into passing groundwater. Their solubility is so low, however, that removal by dissolution can take centuries. Attempts to remove DNAPLs by extracting groundwater can be nearly impossible.

Table 14 lists aqueous solubilities for a selection of NAPLs, and those can be used to make a first-cut estimate of the length of time required to remove NAPLs by groundwater pumping. The solubilities found in Table 14 tend to be much higher than those actually found in contaminated aquifers, which tend to be about 10 percent of the theoretical value (NRCouncil, 1994). Consider the following example.

EXAMPLE 17 Estimated Time to Remove a TCE Spill[1]

Suppose 1 m^3 of aquifer is contaminated with 30 L of trichloroethylene (TCE). The aquifer has porosity of 0.3, groundwater moves through it with an actual speed of 0.03 m/day, and the TCE has a dissolved concentration equal to 10 percent of its aqueous solubility.

a. Find the mass of dissolved TCE and the mass of undissolved DNAPL.
b. Estimate the time for TCE to be removed.

Solution

a. From Table 14, the aqueous solubility of TCE is given as 1,100 mg/L, but the actual dissolved TCE is only 10 percent of that, or 110 mg/L. The porosity of the aquifer is 0.3, so the volume of fluid in 1 m^3 of aquifer is 0.3 m^3. The amount of dissolved TCE is therefore

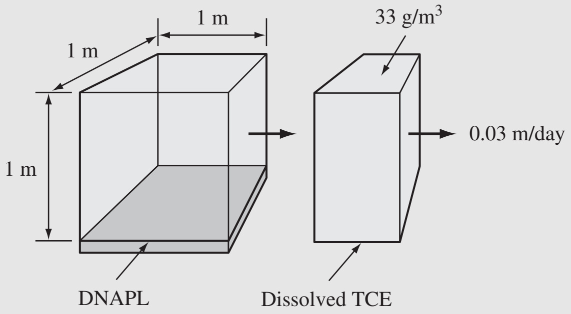
$$\text{Dissolved TCE} = 110\ \text{mg/L} \times 0.3\ \text{m}^3 \times 10^3\ \text{L/m}^3 = 33{,}000\ \text{mg} = 33\ \text{g}$$

Table 14 indicates that the specific gravity of TCE is 1.47. That is, it is 1.47 times the 1.0 kg/L density of water. The total mass of TCE in the aquifer is therefore

$$\text{Total TCE} = 30\ \text{L} \times 1.47 \times 1\ \text{kg/L} \times 10^3\ \text{g/kg} = 44{,}100\ \text{g}$$

Since 33 g are dissolved in groundwater, the remaining 44,097 g is NAPL mass. That is, 99.92 percent of the TCE has not dissolved.

b. If we picture the 1 m^3 of aquifer as a cube 1 m on a side, then the rate at which fluid leaves the cube is



$$\text{Fluid leaving} = 1\ \text{m}^2 \times 0.03\ \text{m/day} = 0.03\ \text{m}^3\text{/day}$$

---

[1] Based on an example given in National Research Council (1994).

taking away an amount of TCE equal to

$$\text{TCE flux through } 1 \text{ m}^2 = 1 \text{ m}^2 \times 0.03 \text{ m/day} \times 33 \text{ g/m}^3 = 0.99 \text{ g/day}$$

So the time needed to remove all 44,100 g of TCE would be

$$\text{Time to remove TCE} = \frac{44{,}100 \text{ g}}{0.99 \text{ g/day} \times 365 \text{ day/yr}} = 122 \text{ years}$$

While Example 17 yields an estimate of over a century to clean this aquifer completely, in actuality it would probably take even longer. As the globules of TCE dissolve, their surface area decreases, and the rate at which they dissolve drops, so the actual time required would be significantly longer than 122 years.

17 Groundwater Remediation Technologies

Groundwater cleanup may depend on treatment of the water after it has been extracted (ex situ treatment) or treatment of the water in place in the aquifer (in situ treatment). In this section, we first discuss some of the problems with using the traditional, ex situ pump-and-treat approach. This is followed by a description of another ex situ approach, but this time based on extracting contaminated gas rather than water. Finally, two newer in situ approaches are discussed, which, when successful, can greatly decrease the cost of groundwater remediation.

Conventional Pump-and-Treat Systems

Pump-and-treat technology for aquifer cleanup is based on extracting contaminated groundwater and then treating it above ground. Treated effluent can then be used for beneficial purposes or returned to the aquifer. There are an estimated 3,000 pump-and-treat sites in operation in the United States, which represents approximately three-quarters of all groundwater-remediation projects.

The cost of cleaning up an aquifer and the length of time required to do so escalate dramatically as the level of cleanup desired increases. A hypothetical illustration of the cost and time escalations for a conventional pump-and-treat system is presented in Figure 52. The example corresponds to a 190-million-liter plume with an average concentration of 1 mg/L of TCE, using EPA cost estimates, including a 4 percent discount factor. Achieving 80 percent contaminant removal is estimated to take 15 years and cost \$2.8 million (present value, including capital, operation, and maintenance costs). With a 99.99 percent removal goal, the system would cost \$6 million and take 84 years to complete (NR Council, 1994). Obviously, there are many assumptions and uncertainties in a calculation such as this one, but the implications are clear.

One assumption in Figure 52 is that the site can be cleaned up using pump-and-treat technology. A recent study by the Committee on Ground Water Cleanup Alternatives for the National Research Council evaluated 77 contaminated sites where pump-and-treat systems are operating (NR Council, 1994) and found only limited circumstances in which such systems would be likely to achieve cleanup

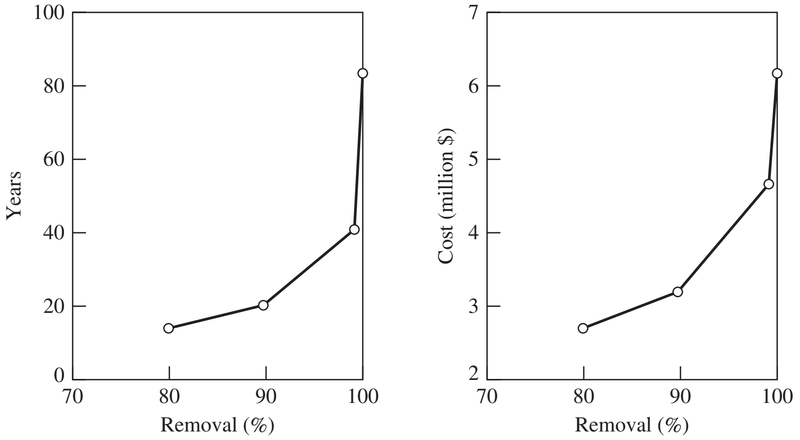


FIGURE 52 Impact of cleanup goal on (present value) cost and duration of a conventional pump-and-treat system. Upper data points correspond to 99 percent and 99.99 percent removal of TCE. Drawn for a plume volume of 190×10^6 L, a pumping rate of 380 L/min, a retardation factor of 4.8 (TCE), and a 4 percent discount factor. (*Source:* Based on data in NR Council, 1994.)

goals in a reasonable period of time at a reasonable cost. They established a site rating system based on contaminant chemistry and hydrogeology that attempts to rate sites by their relative ease of cleanup. Four ratings were developed, with Category 1 sites being the easiest to clean up and Category 4 sites the most difficult. Category 1 sites are characterized by a single-layer, homogeneous aquifer contaminated with pollutants that are mobile, dissolve easily, and are to some extent removed naturally by degradation or volatilization. Pump-and-treat technology should be able to restore groundwater to drinking water quality in these rare sites. At the other extreme, Category 4 sites have complex hydrogeology involving multiple layers of heterogeneous materials with complex contaminant pathways that are difficult to predict. The contaminants in Category 4 are NAPLs, with DNAPLs being the most difficult. Category 4 sites are unlikely to achieve drinking water-quality goals. Of the 77 sites studied by the National Research Council Committee, only 2 were placed in Category 1, while 42 were ranked Category 4. Table 15 shows the characteristics of each of the four categories.

Attempting to achieve groundwater cleanup to drinking water standards using conventional pump-and-treat technology appears to be a futile goal for most contaminated aquifers. This limited effectiveness has motivated the development of a number of other technologies that can be used to augment or replace the conventional pump-and-treat system. General descriptions follow of the most widely used alternatives.

Soil Vapor Extraction

Soil vapor extraction (SVE) systems are designed to remove organic vapors from the unsaturated zone. As shown in Figure 53, the suction side of a blower pulls soil vapors up extraction wells and sends them to a vapor treatment unit. The extraction wells consist of slotted, plastic pipe set into a permeable packing material. They may

TABLE 15

Relative Ease of Cleaning Up Contaminated Aquifers as a Function of Contaminant Chemistry and Hydrogeology

	Contaminant Chemistry					
Hydrogeology	Mobile Dissolved (degrades/ volatilizes)	Mobile, Dissolved	Strongly Sorbed, Dissolved (degrades/ volatilizes)	Strongly Sorbed, Dissolved	Separate Phase LNAPL	Separate Phase DNAPL
Homogeneous, single layer	1[a]	1–2	2	2–3	2–3	3
Homogeneous, multiple layers	1	1–2	2	2–3	2–3	3
Heterogeneous, single layer	2	2	3	3	3	4
Heterogeneous, multiple layers	2	2	3	3	3	4
Fractured	3	3	3	3	4	4

[a] Relative ease of cleanup, where 1 is easiest and 4 is most difficult.
Source: National Research Council, 1994.

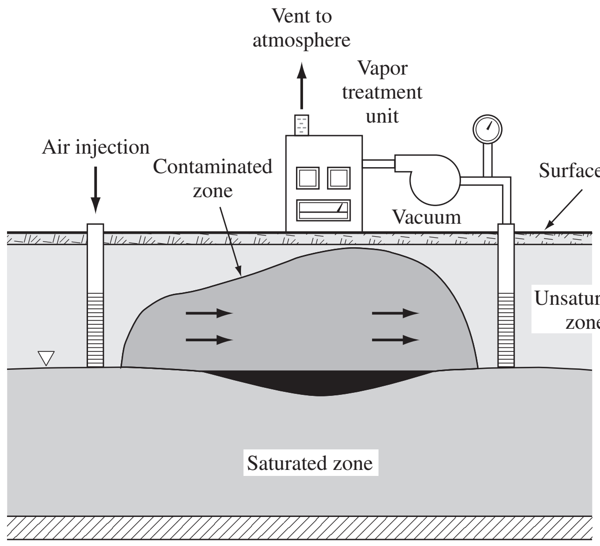


FIGURE 53 A soil vapor extraction system.

be placed vertically or horizontally depending on circumstances. Horizontal systems are often used when the contamination is near the surface. The system may include an air injection system that helps push vapors toward collection wells. The air injection system can be connected to the treatment system, creating a closed-loop system. A soil cap can help keep the vapors from venting directly to the atmosphere.

Soil vapor extraction systems are effective at removing volatile organic compounds that have leaked from underground storage tanks, such as leaky gasoline storage tanks at service stations. They may also help remove some NAPLs in the unsaturated zone. A combination of SVE and pump-and-treat can help remove LNAPLs that tend to sit on the surface of the saturated zone. When the pumping

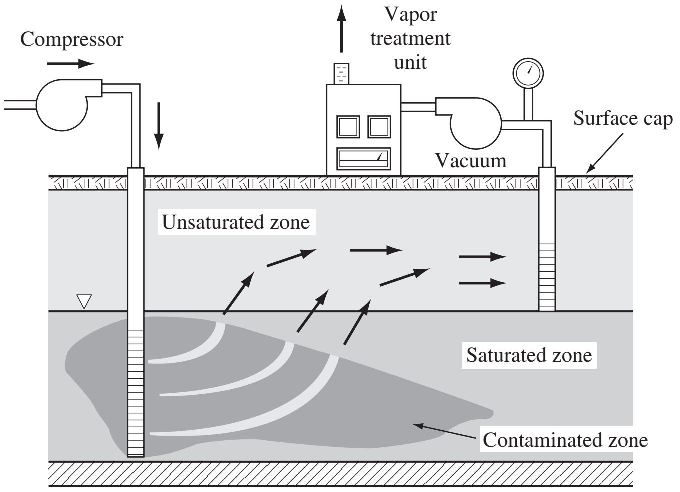


FIGURE 54 Combining air sparging with soil vapor extraction.

system lowers the water table, some of the LNAPLs are left behind. LNAPLs that remain sorbed onto particles that are now in an unsaturated zone can then be affected by the SVE system.

The performance of SVE systems is highly dependent on the characteristics of the subsurface and the contaminants. Low-permeability zones in the subsurface are not easily flushed, so contaminants located there will be difficult to remove, and contaminants that are strongly sorbed to subsurface particles will also resist removal. With highly volatile substances located in permeable soils, SVE systems perform well.

SVE systems can be augmented with *air sparging systems,* as shown in Figure 54. An air sparging system consists of a compressor that drives air through an injection well into the saturated zone. The injected air rises through the contaminant plume and captures volatile compounds as it moves. Adsorbed contaminants can desorb directly into the rising air stream, and dissolved volatile compounds can move from the liquid phase into the vapor phase to be vacuumed out by the SVE system. Some air sparging systems have been shown to be very effective at removing substantial quantities of volatile organic compounds, such as benzene and gasoline.

Air sparging can also be done using a horizontal injection distribution system that lies under the contamination with a horizontal suction system that is above the contamination, as shown in Figure 55.

In Situ Bioremediation

In situ systems are designed to degrade subsurface pollution in place without the need to capture and deliver contaminants to an above-ground treatment system. By treating contaminants in situ, the need for excavation, above-ground treatment, and transportation to a disposal site is eliminated, and the risk of human exposure to hazardous chemicals is greatly reduced. There are other advantages as well, including the potential to degrade compounds that are sorbed to subsurface materials.

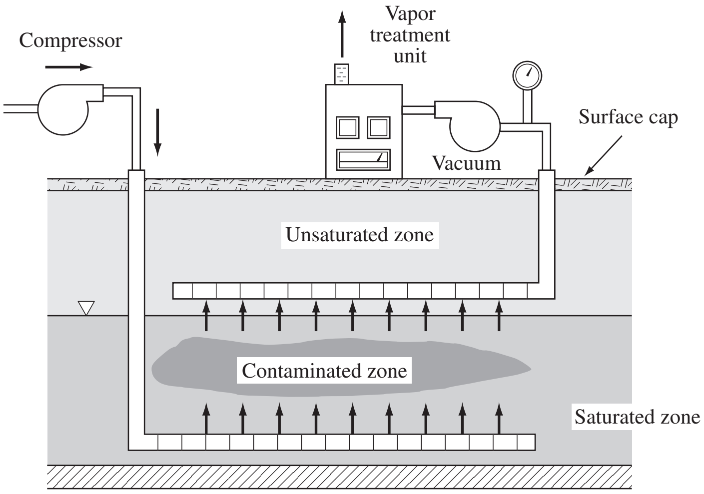


FIGURE 55 Air sparging with horizontal wells.

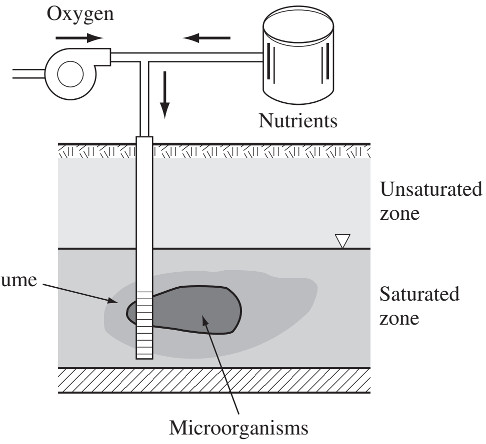


FIGURE 56 In situ bioremediation.

In situ bioremediation is based on stimulating the growth of microorganisms, primarily bacteria, that are indigeneous to the subsurface and that can biodegrade contaminants. If provided with oxygen (or other electron acceptors) and nutrients, microorganisms can degrade a number of common soil and groundwater contaminants, especially petroleum-based hydrocarbons, to carbon dioxide and water. If contamination is close to the land surface, the treatment system can be as simple as an infiltration gallery that allows nutrient- and oxygen-rich water to percolate down through the contamination zone. For deeper contamination, an injection system, such as is shown in Figure 56, can deliver the amended water to the desired locations. The oxygen needed for aerobic decomposition can be supplied by bubbling air or oxygen into the injected water or by adding hydrogen peroxide. The added nutrients are often nitrogen and phosphorus when the contaminants are hydrocarbons. Extraction wells may be added to the system to help control the flow of the injected amended water.

In situ bioremediation has also been applied to more difficult contaminants than hydrocarbons, including chlorinated solvents and metals. Biodegradation of chlorinated compounds can be accomplished either aerobically or anaerobically. The aerobic pathway involves methanotrophic bacteria, which derive their energy from methane. To stimulate their growth, methane is injected into the contaminated aquifer. In the process of consuming methane, methanotrophs produce an enzyme that has the ability to transform hazardous contaminants that are ordinarily resistant to biodegradation. Field studies have demonstrated high rates of transformation for some chlorinated compounds. Semprini et al. (1990) stimulated indigenous methanotrophic bacteria by injecting dissolved methane and oxygen into an aquifer under the Moffett Naval Air Station in California. Their experiment showed the following amounts of biodegradation in just a matter of days: trichloroethylene, 20 to 30 percent; *cis*-l,2-dichloroethylene, 45 to 55 percent; *trans*-l,2-dichloroethylene, 80 to 90 percent; and vinyl chloride, 90 to 95 percent.

Chlorinated solvents can also be degraded anaerobically. A different consortium of bacteria are involved, and the nutrients that need to be supplied include an electron donor such as methanol, glucose, or acetate. Some concern has been expressed for intermediate products of this biodegradation, such as vinyl chloride, that are themselves hazardous. Complete degradation to simple ethene and ethane is possible, but anaerobic processes proceed much more slowly than their aerobic counterparts, and such degradation may take years in the field.

Permeable Reactive Barriers

Since the mid-1990s, *permeable reactive barriers* (PRBs) have generated more attention and undergone more study than any other technique for remediating underground contamination (U.S. EPA, 1998). As shown in Figure 57, the PRB is a permeable wall of specially chosen material through which the contaminated groundwater passes. As the water passes, the pollutants are removed by one or more processes, including abiotic redox degradation, biotransformation, adsorption, and precipitation. Since the

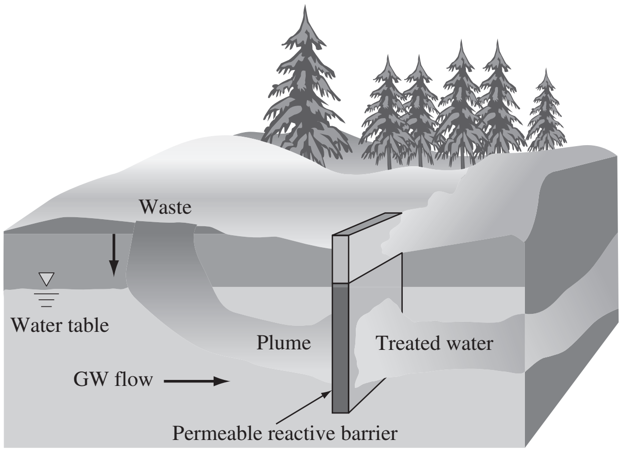


FIGURE 57 A permeable reactive barrier for treating a groundwater contaminant plume. (*Source:* US EPA, 1998.)

PRB is only effective if the plume moves through it, the hydraulic conductivity of the PRB must be greater than that of the aquifer around it. If this is not true, or if over time the PRB clogs due to filtration of particles, precipitation of solids, or buildup of solid reaction products, the water will divert around the wall.

The most common material used is granular zero valent iron (ZVI), which is typically placed in a 2 to 6 feet thick wall. Iron in the zero oxidation state (Fe(0)) is readily oxidized to ferrous iron (Fe(II)), and the electrons lost can destructively reduce many of the halogenated organic pollutants listed in Table 13. Using trichloroethylene as an example, complete reductive dechlorination may follow an overall reaction such as

$$3\ \mathrm{Fe}^0 \rightarrow 3\ \mathrm{Fe}^{2+} + 6\ \mathrm{e}^- \qquad \text{oxidation half reaction} \tag{60}$$

$$\underline{\mathrm{HC_2Cl_3} + 3\mathrm{H}^+ + 6\ \mathrm{e}^- \rightarrow \mathrm{H_4C_2} + 3\ \mathrm{Cl}^-} \qquad \text{reduction half reaction} \tag{61}$$

$$3\ \mathrm{Fe}^0 + \mathrm{HC_2Cl_3} + 3\mathrm{H}^+ \rightarrow 3\ \mathrm{Fe}^{2+} + \mathrm{H_4C_2} + 3\ \mathrm{Cl}^- \qquad \text{net redox reaction} \tag{62}$$

In a similar way to (62), ZVI may be used to treat certain redox-sensitive inorganic ions such as chromium from Cr(VI) to the less mobile Cr(III) state.

By using materials other than ZVI or in addition to it, removal of additional contaminants can be accomplished. If the nutrients needed for microbial growth are included in the reactive barrier, bioremediation and removal of some inorganic and organic pollutants can be achieved. Microbes stimulated by PRBs have not only been used to directly degrade a contaminant but also to produce reactants, such as sulfide, which in turn will either react with or cause the precipitation of the target contaminants. If lime is used in the barrier, the pH may be raised sufficiently that certain metals will precipitate out. In other applications, a sorbent, such as a hydrous ferric oxide, has been emplaced to adsorb contaminants such as arsenic and lead (U.S. EPA, 1998).

Reactive barriers are attractive because they are often relatively low cost, allow possible cleanup of contaminants that are hard to address otherwise, and require virtually no upkeep after they are in place. However, their long-term performance has not always been satisfactory. PRBs have failed or performed below expectations for many reasons, including loss of permeability, loss of iron or media reactivity, slower than expected reaction kinetics, media clogging or passivation by microbially generated biomass, and poor hydraulic connection between the aquifer and the permeable reactive zone. A good place to learn more about PRBs is the ITRC report (2005) and the additional references in it.

Other Remediation Technologies

None of the groundwater remediation technologies discussed are applicable to all contaminants in all conditions, and none are without drawbacks. Consequently, a very large array of other technologies have been employed with varying degrees of success and ranges of applicability and cost. Enhanced *solubilization/mobilization, thermal desorption,* and *steam extraction* techniques aim to make subsurface contaminants more mobile (reduce retardation and mobilize NAPLs) or more bioavailable, so they can be removed or degraded by one or more of the earlier discussed technologies. *Electrokinetic* processes depend on the creation of an electrical field in the subsurface that can enhance contaminant mobilization or, in a very few cases, degrade the target pollutant. Isolation techniques, such as *impermeable barriers, vitrification, encapsulation,* and *soil solidification*, attempt to permanently immobilize the contaminant. Abiotic chemical degradation uses injected compounds such as potassium permanganate or hydrogen peroxide to oxidize subsurface pollutants. Many variations of bioremediation and combination biotic/abiotic processes have been reported, including *phytoremediation, biosparging, biofiltration,* and *sequential anaerobic-aerobic* techniques. Further discussion of these and other alternative approaches is beyond the scope of this text, but details can be found in references such as by Bedient (1999) or from the resources listed by EPA (2001).

PROBLEMS

1 In a standard five-day BOD test,

(a) Why is the BOD bottle stoppered?

(b) Why is the test run in the dark (or in a black bottle)?

(c) Why is it usually necessary to dilute the sample?

(d) Why is it sometimes necessary to seed the sample?

(e) Why isn't ultimate BOD measured?

2 Incoming wastewater, with BOD_5 equal to about 200 mg/L, is treated in a well-run secondary treatment plant that removes 90 percent of the BOD. You are to run a five-day BOD test with a standard 300-mL bottle, using a mixture of treated sewage and dilution water (no seed). Assume the initial DO is 9.2 mg/L.

(a) Roughly what maximum volume of treated wastewater should you put in the bottle if you want to have at least 2.0 mg/L of DO at the end of the test (filling the rest of the bottle with water)?

(b) If you make the mixture half water and half treated wastewater, what DO would you expect after five days?

3 A standard five-day BOD test is run using a mix consisting of four parts distilled water and one part wastewater (no seed). The initial DO of the mix is 9.0 mg/L, and the DO after five days is determined to be 1.0 mg/L. What is BOD_5?

4 A BOD test is to be run on a sample of wastewater that has a five-day BOD of 230 mg/L. If the initial DO of a mix of distilled water and wastewater is 8.0 mg/L, and the test requires a decrease in DO of at least 2.0 mg/L with at least 2.0 mg/L of DO remaining at the end of the five days, what range of dilution factors (P) would produce acceptable results? In 300-mL bottles, what range of wastewater volumes could be used?

5 The following data have been obtained in a BOD test that is made to determine how well a wastewater treatment plant is operating:

	Initial DO (mg/L)	Final DO (mg/L)	Volume of Wastewater (mL)	Volume of Dilution Water (mL)
Untreated sewage	6.0	2.0	5	295
Treated sewage	9.0	4.0	15	285

What percentage of the BOD is being removed by this treatment plant? If this is a secondary treatment plant that is supposed to remove 85 percent of the BOD, would you say it is operating properly?

6 Figure P6 shows a plot of BOD remaining versus time for a sample of effluent taken from a wastewater treatment plant.

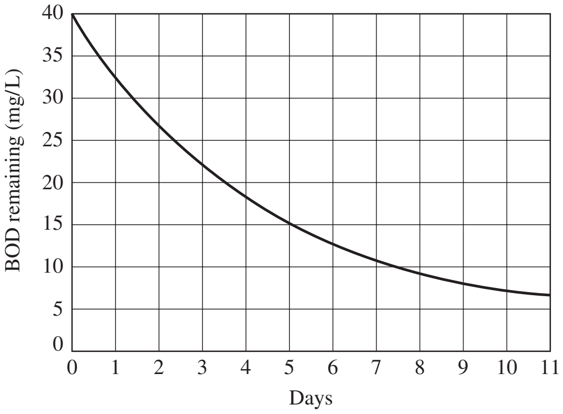


FIGURE P6

(a) What is the ultimate BOD (L_0)?

(b) What is the five-day BOD?

(c) What is L_5?

7 If the BOD_5 for some wastewater is 200 mg/L, and the ultimate BOD is 300 mg/L, find the reaction rate constants k (base e) and K (base 10).

8 A BOD test is run using 100 mL of treated wastewater mixed with 200 mL of pure water. The initial DO of the mix is 9.0 mg/L. After 5 days, the DO is 4.0 mg/L. After a long period of time, the DO is 2.0 mg/L, and it no longer seems to be dropping. Assume nitrification has been inhibited so the only BOD being measured is carbonaceous.

(a) What is the five-day BOD of the wastewater?

(b) Assuming no nitrification effects, estimate the ultimate carbonaceous BOD.

(c) What would be the remaining BOD after five days have elapsed?

(d) Estimate the reaction rate constant k (day^{-1}).

9 Suppose you are to measure the BOD removal rate for a primary wastewater treatment plant. You take two samples of raw sewage on its way into the plant and two samples of the effluent leaving the plant. Standard five-day BOD tests are run on the four samples, with no seeding, producing the following data:

Sample	Source	Dilution	DO_i(mg/L)	DO_f(mg/L)
1	Raw	1:30	9.2	2.2
2	Raw	1:15	9.2	?
3	Treated	1:20	9.0	2.0
4	Treated	?	9.0	>0

(a) Find BOD_5 for the raw and treated sewage, and the percent removal of BOD in the treatment plant.

(b) Find the DO that would be expected in Sample 2 at the end of the test.

(c) What would be the maximum volume of treated sewage for Sample 4 that could be put into the 300-mL BOD bottle and still have the DO after five days remain above 2 mg/L?

10 A standard BOD test is run using seeded dilution water. In one bottle, the waste sample is mixed with seeded dilution water giving a dilution of 1:30. Another bottle, the blank, contains just seeded dilution water. Both bottles begin the test with DO at the saturation value of 9.2 mg/L. After five days, the bottle containing waste has DO equal to 2.0 mg/L, while that containing just seeded dilution water has DO equal to 8.0 mg/L. Find the five-day BOD of the waste.

11 A mixture consisting of 30 mL of waste and 270 mL of seeded dilution water has an initial DO of 8.55 mg/L; after five days, it has a final DO of 2.40 mg/L. Another bottle containing just the seeded dilution water has an initial DO of 8.75 mg/L and a final DO of 8.53 mg/L. Find the five-day BOD of the waste.

12 Some wastewater has a BOD_5 of 150 mg/L at 20°C. The reaction rate k at that temperature has been determined to be 0.23/day.

(a) Find the ultimate carbonaceous BOD.

(b) Find the reaction rate coefficient at 15°C.

(c) Find BOD_5 at 15°C.

13 Some waste has a five-day BOD at 20°C equal to 210 mg/L and an ultimate BOD of 350 mg/L. Find the five-day BOD at 25°C.

14 A clever approach for finding L_0 involves daily measurements of BOD (that is, BOD_1, BOD_2, BOD_3). A straight line is fitted to a plot of BOD_{t+1} vs. BOD_t, and the intersection point of that line with a line drawn through the origin with slope = 1 is then found. That intersection point occurs where $\text{BOD}_{t+1} = \text{BOD}_t$; that is, it is the point where BOD is no longer changing; hence it is L_0, as shown in Figure P14.

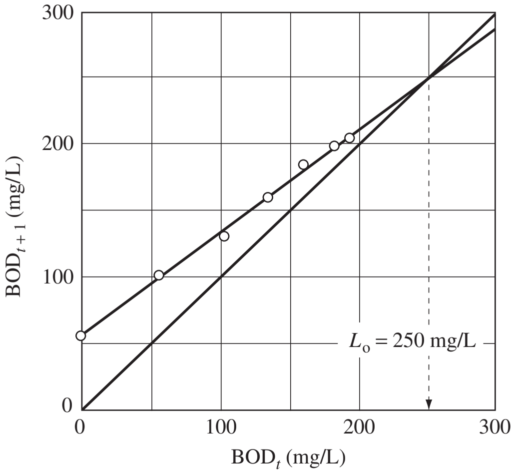


FIGURE P14

The following are BOD data for the sample waste graphed in P.14, along with three additional wastes. Determine L_0 for each of the additional wastes using this graphical procedure:

t	Example		t	Waste 1	Waste 2	Waste 3
(day)	BOD_t	BOD_{t+1}	(day)	BOD_t	BOD_t	BOD_t
0	0	57	0	0	0	0
1	57	102	1	62	38	41
2	102	134	2	104	72	79
3	134	160	3	142	104	101
4	160	184	4	179	123	121
5	184	199	5	200	142	140
6	199	207	6	222	151	152
7	207		7	230	167	159

15 Show that the procedure in Problem 14 works by verifying that a plot of BOD_{t+1} vs.BOD_t is linear; that is,

$$\text{BOD}_{t+1} = a\,\text{BOD}_t + b$$

at the intersection point where $\text{BOD}_{t+1} = \text{BOD}_t$, BOD utilized is no longer increasing so it is equal to L_0.

16 Suppose some wastewater has a BOD_5 equal to 180 mg/L and a reaction rate k equal to 0.22/day. It also has a total Kjeldahl nitrogen content (TKN) of 30 mg/L.

(a) Find the ultimate carbonaceous oxygen demand (CBOD).

(b) Find the ultimate nitrogenous oxygen demand (NBOD).

(c) Find the remaining BOD (nitrogenous plus carbonaceous) after five days have elapsed.

17 Suppose a pond's water contains 10.0 mg/L of some algae, which can be represented by the chemical formula $C_6H_{15}O_6N$. Using the following reactions:

$$C_6H_{15}O_6N + 6\,O_2 \longrightarrow 6\,CO_2 + 6\,H_2O + NH_3$$

$$NH_3 + 2\,O_2 \longrightarrow NO_3^- + H^+ + H_2O$$

(a) Find the theoretical carbonaceous oxygen demand (see Example 2.2).

(b) Find the total theoretical (carbonaceous plus nitrogenous) oxygen demand.

18 For a solution containing 200 mg/L of glycine [$CH_2(NH_2)COOH$] whose oxidation can be represented as

$$2\,CH_2(NH_2)COOH + 3\,O_2 \longrightarrow 4\,CO_2 + 2\,H_2O + 2\,NH_3$$

$$NH_3 + 2\,O_2 \longrightarrow NO_3^- + H^+ + H_2O$$

(a) Find the theoretical CBOD.

(b) Find the ultimate NBOD.

(c) Find the total theoretical BOD.

19 A sample contains 200 mg/L of casein ($C_8H_{12}O_3N_2$). Calculate the theoretical CBOD, NBOD, and total BOD. If none of the NBOD is exerted in the first five days, and $k = 0.25$/day, estimate the five-day BOD.

20 An approximate empirical formula for bacterial cells is $C_5H_7O_2N$. What would be the total carbonaceous and nitrogenous oxygen demand for 1 g of such cells?

21 A wastewater treatment plant discharges 1.0 m^3/s of effluent having an ultimate BOD of 40.0 mg/L into a stream flowing at 10.0 m^3/s. Just upstream from the discharge point, the stream has an ultimate BOD of 3.0 mg/L. The deoxygenation constant k_d is estimated at 0.22/day.

(a) Assuming complete and instantaneous mixing, find the ultimate BOD of the mixture of waste and river just downstream from the outfall.

(b) Assuming a constant cross-sectional area for the stream equal to 55 m^2, what ultimate BOD would you expect to find at a point 10,000 m downstream?

22 The wastewater in Problem 21 has DO equal to 4.0 mg/L when it is discharged. The river has its own DO, just upstream from the outfall, equal to 8.0 mg/L. Find the initial oxygen deficit of the mixture just downstream from the discharge point. The temperatures of sewage and river are both 15°C.

23 A single source of BOD causes an oxygen sag curve with a minimum downstream DO equal to 6.0 mg/L. If the BOD of the waste is doubled (without increasing the waste flow rate), what would be the new minimum downstream DO? In both cases, assume that the initial oxygen deficit just below the source is zero, and the saturated value of DO is 10.0 mg/L. (Note that when the initial deficit is zero, the deficit at any point is proportional to the initial BOD.)

24 The oxygen sag caused by a cannery reaches a minimum DO equal to 3.0 mg/L. Upstream from the cannery, the river DO is saturated at 10.0 mg/L, and it has no BOD of its own. Just downstream from the discharge point, the DO is still essentially saturated (i.e., consider the initial oxygen deficit to be zero so the downstream deficit is proportional to initial BOD). By what percentage should the BOD of the cannery waste be reduced to assure a healthy stream with at least 5.0 mg/L DO everywhere?

25 Two point-sources of BOD along a river (A and B) cause the oxygen sag curve shown in Figure P25.

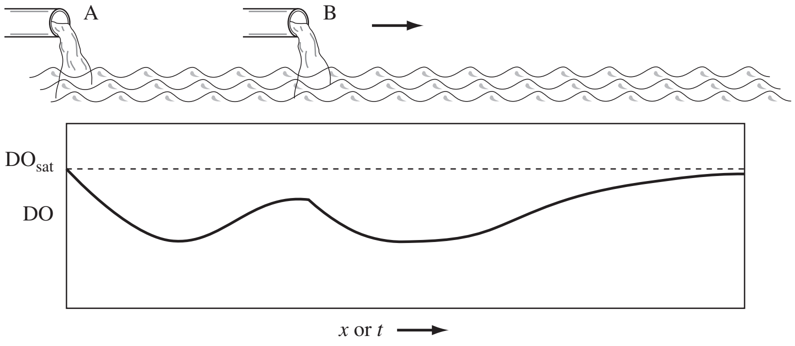


FIGURE P25

(a) Sketch the rate of reaeration vs. distance downriver.

(b) Sketch L_t (that is, the BOD remaining) as a function of distance downriver.

26 Untreated sewage with a BOD of 240 mg/L is sent to a wastewater treatment plant where 50 percent of the BOD is removed. The river receiving the effluent has the oxygen sag

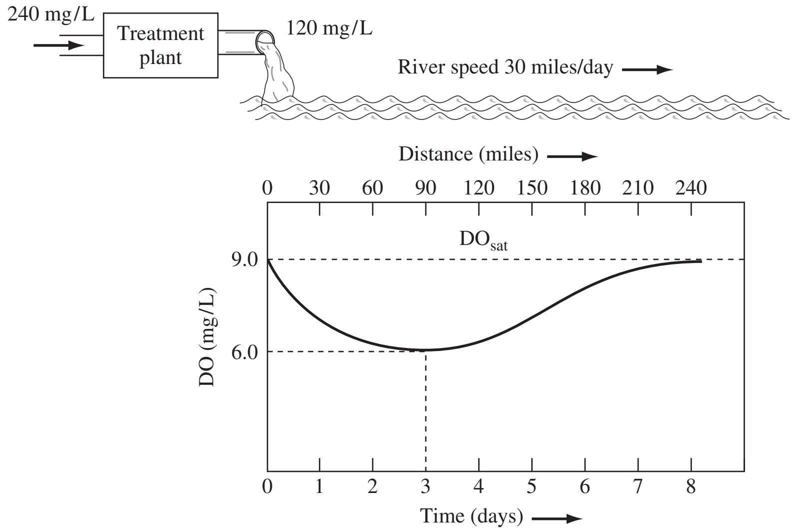


FIGURE P26

curve shown in Figure P26 (the river has no other sources of BOD). Notice that downstream is expressed both in miles and days required to reach a given spot.

(a) Suppose the treatment plant breaks down, and it no longer removes any BOD. Sketch the new oxygen sag curve a long time after the breakdown. Label the coordinates of the critical distance downriver.

(b) Sketch the oxygen sag curve as it would have been only four days after the breakdown of the treatment plant.

27 Suppose the only source of BOD in a river is untreated wastes that are being discharged from a food processing plant. The resulting oxygen sag curve has a minimum value of DO, somewhere downstream, equal to 3.0 mg/L (see Figure P27). Just below the discharge point, the DO of the stream is equal to the saturation value of 10.0 mg/L.

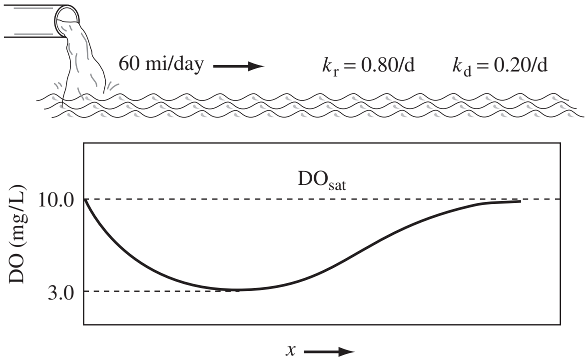


FIGURE P27

(a) By what percent should the BOD of the wastes be reduced to assure a healthy stream with at least 5.0 mg/L of DO everywhere? Would a primary treatment plant be sufficient to achieve this reduction?

(b) If the stream flows 60 miles per day, has a reaeration coefficient k_r equal to 0.80/day, and has a deoxygenation coefficient k_d of 0.20/day, how far downstream (miles) would the lowest DO occur?

(c) What ultimate BOD (L_0 mg/L) of the mixture of river and wastes just downstream from the discharge point would cause the minimum DO to be 5.0 mg/L?

(d) Sketch the oxygen sag curve before and after treatment recommended in (a), labeling critical points (DO_{min} location and value).

28 The ultimate BOD of a river just below a sewage outfall is 50.0 mg/L, and the DO is at the saturation value of 10.0 mg/L. The deoxygenation rate coefficient k_d is 0.30/day, and the reaeration rate coefficient k_r is 0.90/day. The river is flowing at the speed of 48.0 miles per day. The only source of BOD on this river is this single outfall.

(a) Find the critical distance downstream at which DO is a minimum.

(b) Find the minimum DO.

(c) If a wastewater treatment plant is to be built, what fraction of the BOD would have to be removed from the sewage to assure a minimum of 5.0 mg/L everywhere downstream?

29 If the river in Problem 28 has an initial oxygen deficit just below the outfall of 2.0 mg/L, find the critical distance downstream at which DO is a minimum and find that minimum DO.

30 A city of 200,000 people deposits 37 cubic feet per second (cfs) of sewage having a BOD of 28.0 mg/L and 1.8 mg/L of DO into a river that has a flow rate of 250 cfs and a flow speed of 1.2 ft/s. Just upstream of the release point, the river has a BOD of 3.6 mg/L and a DO of 7.6 mg/L. The saturation value of DO is 8.5 mg/L. The deoxygenation coefficient k_d is 0.61/day, and the reaeration coefficient k_r is 0.76/day. Assume complete and instantaneous mixing of the sewage and river.

(a) Find the initial oxygen deficit and ultimate BOD just downstream of the outfall.

(b) Find the time and distance to reach the minimum DO.

(c) Find the minimum DO.

(d) Find the DO that could be expected 10 miles downstream.

31 For the following waste and river characteristics just upstream from the outfall, find the minimum downstream DO that could be expected:

Parameter	Wastewater	River
Flow (m^3/s)	0.3	0.9
Ultimate BOD (mg/L)	6.4	7.0
DO (mg/L)	1.0	6.0
$k_d(\text{day}^{-1})$	—	0.2
$k_r(\text{day}^{-1})$	—	0.37
Speed (m/s)	—	0.65
DO_{sat} (mg/L)	8.0	8.0

32 Redo Example 8 at a temperature of 30°C.

33 Just downstream of the outfall from a pollution point-source, the DO of a river is 6 mg/L, and the mix of river and wastes has a BOD of 20 mg/L. The saturation value of DO is 9 mg/L. The deoxygenation constant is $k_d = 0.20$/day.

(a) Estimate the reaeration coefficient using the O'Connor and Dobbins relationship (24), assuming that the river speed is 0.25 m/s, and the average stream depth is 3 m.

(b) Find the critical time downstream at which minimum DO occurs.

(c) Find the minimum DO downstream.

(d) If the outfall is the only source of BOD, what percent removal of BOD would be needed to assure a minimum DO of 5 mg/L?

34 A lake contains the following concentrations of certain chemical species.

Species	Concentration
Selenium (Se)	0.015 mg/L
Phosphorus (P)	0.03 mg/L
Nitrogen (N)	0.12 mg/L
Silicon (Si)	0.27 mg/L
Carbon (C)	0.40 mg/L
Arsenic (As)	0.008 mg/L
Radon (Rn)	0.007 mg/L

What is the limiting nutrient for algal growth in the lake?

35 An often-used chemical representation of algae is $C_{106}H_{263}O_{110}N_{16}P$.

(a) Determine the mass (mg) of each element in 1 g of algae.

(b) Suppose there are 0.10 mg of N and 0.04 mg of P available for algal production per liter of water. Assuming adequate amounts of the other nutrients, which is the limiting nutrient?

(c) What mass of algae could be produced (milligrams algae per liter of water)?

(d) If the nitrogen source could be cut by 50 percent, how much algae (mg/L) could be produced?

(e) If the phosphorus source could be cut by 50 percent, how much algae could be produced?

36 Suppose the N and P content of some algae is as shown in the following table. The third column shows milligrams of nutrient available per liter of water.

Nutrient	Milligrams Per Gram of Algae	Milligrams Available Per Liter
Nitrogen	60	0.12
Phosphorus	10	0.03

(a) What percent reduction in nitrogen is needed to control algal production to 1.0 mg/L?

(b) What percent reduction in phosphorus is needed to control algal production to 1.0 mg/L?

37 Consider a lake with 100×10^6 m^2 of surface area for which the only source of phosphorus is the effluent from a wastewater treatment plant. The effluent flow rate is 0.4 m^3/s, and its phosphorus concentration is 10.0 mg/L (10.0 g/m^3). The lake is also fed by a stream having 20 m^3/s of flow with no phosphorus. If the phosphorus settling rate is estimated to be 10 m/yr, estimate the average phosphorus concentration in the lake. What level of phosphorus removal at the treatment plant would be required to keep the average lake concentration below 0.010 mg/L?

38 A 50 cm^3 sample of dry soil from an aquifer weighs 100 g. When it is poured into a graduated cylinder, it displaces 35 cm^3 of water.

(a) What is the porosity of the soil?

(b) What is the average density of the actual solids contained in the soil?

39 Using data from Table 11, what volume of water would be removed from an unconfined gravel aquifer 10,000 m^2 in area if the water table drops by 1 m? What fraction of the water contained in that portion of the aquifer would have been removed?

40 Consider three monitoring wells, each located at the vertex of an equilateral triangle. The distance between any pair of wells is 300 m. The head at each well, referenced to some common datum, is as follows: well 1, 100 m; well 2, 100.3 m; well 3, 100.3 m. Sketch the well field and find the magnitude and direction of the hydraulic gradient.

41 Three wells are located in an x, y plane at the following coordinates: well 1, (0,0); well 2, (100 m, 0); and well 3, (100 m, 100 m). The ground surface is level, and the distance from the surface to the water table for each well is as follows: well 1, 10 m; well 2, 10.2 m; well 3, 10.1 m. Sketch the well field and find the hydraulic gradient.

42 The aquifer described in Problem 40 has a gradient of 0.00115, a hydraulic conductivity of 1,000 m/d, and a porosity of 0.23.

(a) What is the Darcy velocity?

(b) What is the average linear velocity of the groundwater?

(c) If the front edge of a plume is perfectly straight, how long would it take to travel to well 1 after first arriving (simultaneously) at wells 2 and 3, assuming a retardation factor of 2?

43 A 750-m section of river runs parallel to a channel 1,000 m away (see Figure P43). An aquifer connecting the two has hydraulic conductivity equal to 7.0 m/day and an average thickness of 10 m. The surface of the river is 5.0 m higher than the surface of the channel. Estimate the rate of seepage from the river to the channel.

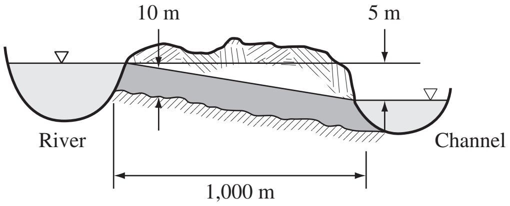


FIGURE P43

44 Based on observations of the rate of travel between two observation wells of a tracer element having retardation factor equal to 1.0, it is determined that the average linear flow velocity in an aquifer is 1.0 m/day when the hydraulic gradient is 0.0005. A sample of the aquifer is tested and found to have a porosity of 0.20. Estimate the hydraulic conductivity.

45 Derive the expression for the cone of depression in a confined aquifer as given in this chapter in (55).

46 Drawdown at a 0.20-m-diameter, fully penetrating well, which has been pumping at the rate of 1,000 m^3/day for a long enough time that steady-state conditions have been reached, is determined to be 0.70 m. The aquifer is unconfined and 10.0 m thick. An

observation well 10 m away has been drawn down by 0.20 m. Determine the hydraulic conductivity of the aquifer.

47 A confined aquifer 30.0 m thick has been pumped from a fully penetrating well at a steady rate of 5,000 m^3/day for a long time. Drawdown at an observation well 15 m from the pumped well is 3.0 m, and drawdown at a second observation well 150 m away is 0.30 m. Find the hydraulic conductivity of the aquifer.

48 Derive the following expression for the length of time required for groundwater to flow from an observation well to a pumped well in a confined aquifer:

$$t = \frac{\pi B \eta}{Q}(R^2 - r_w^2)$$

where

B = thickness of the confined aquifer
η = aquifer porosity
R = radial distance from the observation well to the pumped well
r_w = radius of the pumped well
Q = pumping rate

Hint: Combine (54), $Q = 2\pi r B K(dh_c/dr)$, with the average linear velocity from Eq. (48):

$$v' = -\frac{K}{\eta}\left(\frac{dh_c}{dr}\right) = \frac{dr}{dt}$$

49 For the aquifer described in Problem 47, and using the equation given in Problem 48, determine the time required for groundwater to travel from the observation well 15 m away to the pumped well with diameter 0.40 m. The porosity is 0.30.

50 A *stagnation point* in a capture-zone type curve is a spot where groundwater would have no movement. For the case of a single extraction well, the stagnation point is located where the capture-zone curve crosses the x-axis. Use the fact that for small angles $\tan\theta \approx \theta$ to show that the x-axis intercept of the capture-zone curve for a single well is $x = -Q/(2Bv\pi)$.

51 Suppose a spill of 0.10 m^3 of trichloroethylene (TCE) distributes itself evenly throughout an aquifer 10.0 m thick, forming a rectangular plume 2,000 m long and 250 m wide (see Figure P51). The aquifer has porosity 0.40, hydraulic gradient 0.001, and hydraulic conductivity 0.001 m/s.

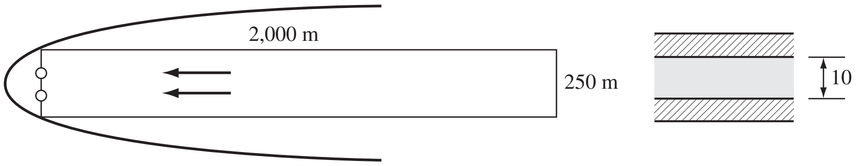


FIGURE P51

(a) Given the solubility of TCE, could this much TCE be totally dissolved in the aquifer? What would be the concentration of TCE (mg/L) in this idealized groundwater plume?

(b) Using capture-zone type curves, design an extraction field to pump out the plume under the assumption that the wells are all lined up along the leading edge of the plume, with each well to be pumped at the same rate, not to exceed 0.003 m^3/s per well. What is the smallest number of wells that could be used to capture the whole plume? What minimum pumping rate would be required for each well?

(c) What would the optimal spacing be between the wells (at that minimum pumping rate)?

52 Starting with (59), show that the width of the capture zone along the y-axis for n optimally spaced wells is equal to $nQ/(2Bv)$.

53 A single well is to be used to remove a symmetrical oblong plume of contaminated groundwater in an aquifer 20.0 m thick with porosity 0.30, hydraulic conductivity 1.0×10^{-4} m/s, and hydraulic gradient 0.0015. With the plume and capture-zone curve superimposed as shown in Figure P53, the angle from the well to the point where the two just touch is 45°, and the width of the plume is 100.0 m. What pumping rate would create these conditions?

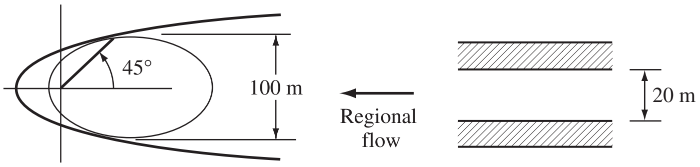


FIGURE P53

54 An underground storage tank leaks tetrachloroethylene (PCE) at a rate of 100 mL/d (about the volume of a small lemon per day). This continues for one year before the leak is discovered and stopped. The tank lies at the water table, so all of the PCE released enters the sand aquifer (but is not necessarily immediately dissolved in the groundwater). The aquifer is 4.0 m deep and composed of sand. The actual groundwater velocity is 0.09 m/d, and the hydraulic gradient is 0.0005. The retardation factor for PCE is 6.

(a) What is the hydraulic conductivity of the aquifer?

(b) How far would you expect the PCE to have traveled down gradient during the one-year period?

(c) Assume the dissolved PCE distributes uniformly in the aquifer and forms a plume 2.0 m wide, 4.0 m deep, and extending as far as you calculated. What is the average PCE concentration in the plume?

(d) If one well is placed on the centerline of the leading edge of the plume, what will its pumping rate need to be to capture the plume?

(e) What is the minimum duration the extraction well will need to be operated to remove all of the PCE from the aquifer? (Assume the concentration of PCE in the groundwater stays at its initial concentration until all of the PCE has been removed.)

55 A cubic meter of a gravel-and-sand aquifer has been contaminated with 20 L of tetrachloroethylene. The amount of tetrachloroethylene dissolved in aquifer water is 20 percent of its aqueous solubility.

(a) How much tetrachloroethylene is dissolved?

(b) How much remains as undissolved DNAPL mass?

(c) If the aquifer has a gradient of 0.001, use porosity and hydraulic conductivity data for gravel-and-sand aquifers to estimate the average linear velocity of the groundwater.

(d) How long would it take to remove the tetrachloroethylene?

REFERENCES

Bedient, P. B., H. S. Rifai, and C. J. Newell, 1994, *Ground Water Contamination, Transport and Remediation,* Prentice Hall, Englewood Cliffs, NJ.

Berk, S. G., J. H. Gunderson, A. L. Newsome, A. L. Farone, B. J. Hayes, K. S. Redding, N. Uddin, E. L. Williams, R. A. Johnson, M. Farsian, A. Reid, J. Skimmyhorn, and M. B. Farone, 2006, Occurrence of infected amoebae in cooling towers compared with natural aquatic environments: Implications for emerging pathogens, *Environmental Science and Technology,* 40(23).

Braghetta, A., 2006, Drawing the connection between malnutrition and lack of safe drinking water in Guatemala, *Journal of American Water Works Association,* 98(5).

Brunner, T. J., P. Wick, P. Manser, P. Spohn, R. N. Grass, L. K. Limbach, A. Bruinink, and W. J. Stark, 2006, In vitro cytotoxicity of oxide nanoparticles: Comparison to asbestos, silica, and the effect of particle solubility, *Environmental Science and Technology,* 40(14): 4374–4381.

Carpenter, S., N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley, and V. H. Smith, 1998, Nonpoint pollution of surface waters with phosphorus and nitrogen, *Issues in Ecology,* 3:1–12.

CAS (Chemical Abstracts Service), 2006, CHEMLIST (Regulated Chemicals Listing), American Chemical Society, Columbus, OH, retrieved from http://www.cas.org/CASFILES/chemlist.html.

Colborn, T., D. Dumanoski, and J. P. Myers, 1996, *Our Stolen Future,* Dutton, New York.

Coupe, R. H., and J. D. Blomquist, 2004, Water-soluble pesticides in finished water of community water supplies, *Journal of American Water Works Association,* 96(10).

Davis, M. L., and D. A. Cornwell, 1991, *Introduction to Environmental Engineering,* 2nd ed., McGraw-Hill, New York.

Frederick, K. D., and J. C. Hanson, 1982, *Water for Western Agriculture,* Resources for the Future, Washington, DC.

Freeze, R. A., and J. A. Cherry, 1979, *Groundwater,* Prentice-Hall, Englewood Cliffs, NJ.

Galloway, J. N., and E. B. Cowling, 1978, The effects of precipitation on aquatic and terrestrial ecosystems: A proposed precipitation chemistry network, *Journal of the Air Pollution Control Association,* 28(3).

Gleick, P. H., 1993, *Water in Crisis, A Guide to the World's Fresh Water Resources,* Oxford University Press, New York.

Gupta, R. S., 1989, *Hydrology and Hydraulic Systems,* Prentice-Hall, Englewood Cliffs, NJ.

Hammer, M. J., and M. J. Hammer, Jr., 1996, *Water and Wastewater Technology,* 3rd Edition, Prentice Hall, Englewood Cliffs, NJ.

Heath, R. C., 1983, *Basic Ground-Water Hydrology,* U.S. Geological Survey Water-Supply Paper 2220, Washington, DC.

Henriksen, A., 1980, *Proceedings of the International Conference on the Ecological Impact of Acid Precipitation,* D. Drablos and A. Tollan (eds.), p. 68.

Hutson, S. S., N. L. Barber, J. F. Kenny, K. S. Linsey, D. S. Lumia, and M. A. Maupin, 2004, *Estimated Use of Water in the United States in 2000,* U.S. Geological Survey Circular 1268, Reston, VA.

ITRC (Interstate Technology and Research Council), 2005, *Permeable Reactive Barriers: Lessons Learned/New Directions,* ITRC Permeable Reactive Team, PRB-4, The Interstate Technology & Regulatory Council, Washington, DC.

Javendel, I., and C. Tsang, 1986, Capture-zone type curves: A tool for aquifer cleanup, *Ground Water,* 24(5):616–625.

Jiang, S. C., 2006, Human adenoviruses in water: Occurrence and health implications: A critical review, *Environmental Science and Technology,* 40(23).

Kavlock, R. J., and G. P. Daston, 1996, Research needs for the risk assessment of health and environmental effects of endocrine disruptors: A report of the U.S. EPA-sponsored workshop, *Environmental Health Perspectives,* 104 (Supplement 4):715–740.

Linsley, R. K., M. A. Kohler, and J. L. H. Paulhus, 1992, *Hydrology for Engineers,* 3rd ed., McGraw-Hill, New York.

Monastersky, R., 1994, Cameroon's killer lakes: A rising threat, *Science News,* 145:215.

Muir, D. C. G., and P. H. Howard, 2006, Are there other persistent organic pollutants? A challenge for environmental chemists, *Environmental Science and Technology,* 40(23).

Nagler, J. J., J. Bouma, G. H. Thorgaard, and D. D. Dauble, 2001, High incidence of a male-specific genetic marker in phenotypic female Chinook salmon from the Columbia River, *Environmental Health Perspectives,* 109:67–69.

National Research Council (NRC), 1994, *Alternatives for Ground Water Cleanup,* National Academy Press, Washington, DC.

O'Connor, D. J., and W. E. Dobbins, 1958, Mechanism of reaeration in natural streams, *Transactions of the American Society of Civil Engineers,* 153:641.

OECD, 2004, *The 2004 OECD List of High Production Volume Chemicals,* Organization of Economic Cooperation and Development, Paris, France.

Postel, S., 1996, Forging a sustainable water strategy, in *State of the World 1996,* L. Brown and W. W. Norton (eds.), New York.

Pruden, A., R. Pei, H. Storteboom, and K. H. Carlson, 2006, Antibiotic resistance genes as emerging contaminants: Studies in Northern Colorado, *Environmental Science and Technology,* 40(23).

Reisner, M., 1986, *Cadillac Desert, The American West and Its Disappearing Water,* Viking, New York.

Roberts, P. V., M. N. Goltz, and D. M. Mackay, 1986, A natural gradient experiment on solute transport in a sand aquifer, 3, retardation estimates and mass balances for organic solutes, *Water Resources Research,* 22(13):2047–2058.

Sawyer, C. N., 1947, Fertilization of lakes by agricultural and urban drainage, *Journal of the New England Water Works Association,* 41(2).

Sawyer, C. N., P. L. McCarty, and G. F. Parkin, 1994, *Chemistry for Environmental Engineering,* 4th ed., McGraw-Hill, New York.

Scott, B. F., C. Spencer, S. A. Mabury, and D. C. G. Muir, 2006, Poly and perfluorinated carboxylates in North American precipitation, *Environmental Science and Technology,* 40(23):7167–7174.

Semprini, L., P. V. Roberts, G. D. Hopkins, and P. L. McCarty, 1990, A field evaluation of in situ biodegradation of chlorinated ethanes: Part 2—results of biostimulation and biotransformation experiments, *Ground Water,* 28(5):715–727.

Shepard, M., 1986, Restoring life to acidified lakes, *Electric Power Research Institute Journal,* April/May.

Shiklomanov, I. A., 1993, World fresh water resources, in *Water in Crisis, A Guide to the World's Fresh Water Resources,* P. H. Gleick (ed.), Oxford University Press, New York.

Sinclair, C. J., A. B. A. Boxall, S. A. Parsons, and M. R. Thomas, 2006, Prioritization of pesticide environmental transformation products in drinking water supplies, *Environmental Science and Technology,* 40(23).

Skogerboe, G. V., and J. P. Law, 1971, *Research Needs for Irrigation Return Flow Quality Control,* Project No. 13030, U.S. Environmental Protection Agency, Washington, DC.

Snyder, E. M., R. C. Pleus, and S. A. Snyder, 2005, Pharmaceuticals and EDCs in the US Water Industry—An Update, *Journal of American Water Works Association,* 97(11).

Streeter, N. W., and E. B. Phelps, 1925, U.S. Public Health Service Bulletin No. 146.

Streets, S. S., S. A. Henderson, A. D. Stoner, D. L. Carlson, M. F. Simcik, and D. L. Swackhamer, 2006, Partitioning and bioaccumulation of PBDEs and PCBs in Lake Michigan, *Environmental Science and Technology,* 40(23):7263–7269.

Stumm, W., and J. J. Morgan, 1981, *Aquatic Chemistry,* 2nd ed., Wiley, New York.

Tchobanoglous, G., and E. D. Schroeder, 1985, *Water Quality,* Addison-Wesley, Reading, MA.

Thomann, R. V., and J. A. Mueller, 1987, *Principles of Surface Water Quality Modeling and Control,* Harper & Row, New York.

UNEP, 1993, *Environmental Data Report, 1993–94,* United Nations Environment Programme, Blackwell, Oxford.

U.S. EPA, 1993, *Cleaning Up the Nation's Waste Sites: Markets and Technology Trends,* EPA 542-R-92-012, Office of Solid Waste and Emergency Response, Environmental Protection Agency, Washington, DC.

U.S. EPA, 1994, *The Quality of Our Nation's Water: 1992,* Environmental Protection Agency, EPA 841-S-94-002, Washington, DC.

U.S. EPA, 1996, *Strategic Plan for the Office of Research and Development,* Environmental Protection Agency, EPA/600/R-96/059, Washington, DC.

U.S. EPA, 1998, *Permeable Reactive Barrier Technologies for Contaminant Remediation,* Office of Research and Development, EPA600/R-98/125, Washington, DC.

U.S. EPA, 2001, *Innovative Remediation and Site Characterization Technology Resources,* Office of Solid Waste and Emergency Response, EPA 542-F-01-026B, Washington, DC.

U.S. EPA, 2002, *National Water Quality Inventory: 2000 Report,* Environmental Protection Agency, Office of Water, EPA-841-R-02-001, Washington, DC.

U.S. EPA, 2006, *Overview of Current Maximum Daily Load—TMDL—Program and Regulations,* updated March 14, 2006, retrieved from http://www.epa.gov/owow/tmdl/overviewfs.html.

U.S. GS, 1984, *Estimated Use of Water in the United States, 1980,* U.S. Geological Survey, Department of the Interior, Washington DC.

U.S. GS, 1985, *Study and Interpretation of the Chemical Characteristics of Natural Water,* Water-Supply Paper 2254, U.S. Geological Survey, Department of the Interior, Washington, DC.

U.S. Water Resources Council, 1978, *The Nation's Water Resources, 1975–2000,* Second National Water Assessment, Washington, DC, December.

Vollenweider, R. A., 1975, Input-output models with special reference to the phosphorus loading comcept in limnology, *Schweiz. Z. Hydrol.,* 37:53–83.

WHO (World Health Organization), 2004, *Water, Sanitation and Hygiene Links to Health—Facts and Figures,* updated November, 2004, retrieved from http://www.who.int/water_sanitation_health/publications/facts2004.

WHO/UNICEF (World Health Organization/United Nations Children's Fund), 2005, *Water for Life: Making It Happen,* Geneva, WHO Press.

Wright, R. F., 1984, Norwegian models for surface water chemistry: An overview, in *Modeling of Total Acid Precipitation Impacts,* J. L. Schnoor (ed.), Butterworth, Boston.

Water Quality Control

> Our planet is shrouded in water, and yet 8 million children under the age of five will die this year from lack of safe water.
>
> —*United Nations Environment Programme*

1 Introduction

One of the first things that world travelers worry about is whether it is safe to drink the water and whether uncooked foods washed in local water are safe to eat. The unfortunate answer in most places is "no." For more than two billion people in the developing countries of the world, access to safe drinking water is simply not possible. The rest of us usually assume (correctly, in most circumstances) that water coming out the tap is clean and safe. That important luxury is the result of the coordinated efforts of scientists, engineers, water plant operators, and regulatory officials.

Our traditional confidence in the quality of drinking water in the United States, however, has been shaken of late. In 1993, 400,000 people in Milwaukee

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

became ill, and more than 100 died, from an intestinal parasite, *Cryptosporidium,* in their drinking water. In the same year, residents of Manhattan and others in the Washington, DC, area were told to boil their water when surprising numbers of *E. coli* bacteria began to show up in their drinking water, despite heavy doses of chlorination. Compounding the chlorination problem has been the realization that byproducts of the disinfection process may be causing on the order of 10,000 cancer cases per year in the United States. The seriousness of these problems has led the Environmental Protection Agency's Office of Research and Development to identify drinking water disinfection as one of its 6 highest-priority research topics in its 1996 Strategic Plan (U.S. EPA, 1996).

Two complementary approaches are necessary to protect the quality of our water. Legislative bodies provide the laws that regulatory agencies use to define acceptable discharges and establish standards that govern the minimum quality of water for its many beneficial uses. The scientific and engineering community provides the technical guidance needed by legislators and regulators, as well as the technology that is used to achieve those standards. This chapter will explore both.

2 Municipal Water and Wastewater Systems

Two critical systems combine to break the carrier-feces-water-victim sequence responsible for the spread of waterborne diseases. The first is the water collection, treatment, and distribution system that provides safe drinking water. The principal legislation in the United States that regulates drinking water quality is the *Safe Drinking Water Act* (SDWA). The second is the wastewater collection and treatment system that removes contaminants before they are released back into the local stream, lake, estuary, or coastal waters. The *Clean Water Act* (CWA) governs the discharge of wastewater effluents and the water quality in the receiving body of water. With increased population and water shortages, the discharges from wastewater systems are becoming more closely linked to the source waters for drinking water systems. The time between when treated and, as is too often the case worldwide, untreated wastewater is discharged to the environment and when it is taken up again for human use is becoming increasingly short. Even now in most large U.S. municipalities, treated wastewater is used for nonpotable purposes such as turf irrigation and, within the near future, will be required for higher quality, even potable, uses. The primary responsibility of these two systems is to kill pathogens before and after water is used, although removal of other toxins such as industrial solvents, heavy metals, and even pharmaceutical products is gaining increased attention. Wastewater treatment systems also reduce biochemical oxygen demand (BOD) and nutrient loading to protect the ecological health of the receiving water.

As shown in Figure 1, municipal systems may get their water from a local stream, reservoir, or groundwater system. Larger cities tend to rely heavily on surface water supplies, whereas small community water systems more often take advantage of groundwater. In the United States, about half of the drinking water comes from surface water supplies and the other half from groundwater.

After water is used in households and businesses, it is collected in a sanitary sewer system and sent to the local wastewater treatment facility. Industrial wastewater

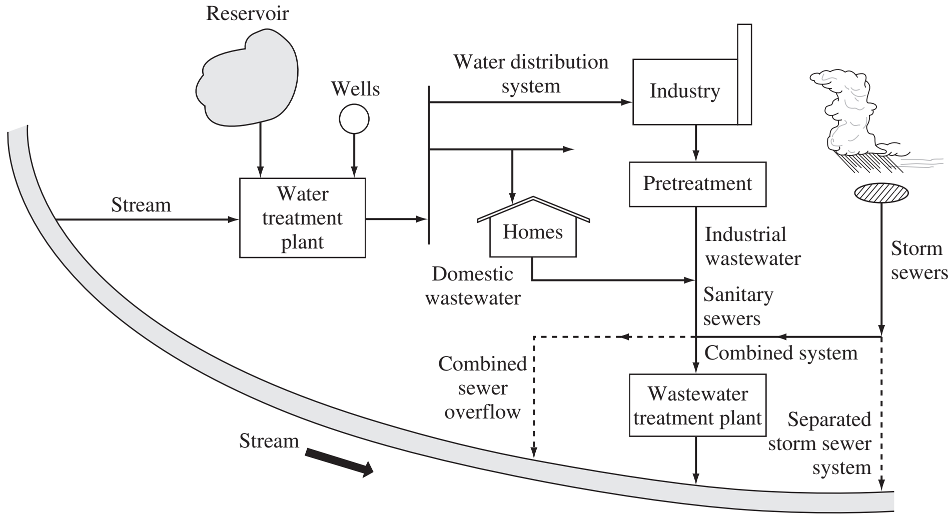


FIGURE 1 Water and wastewater systems. In older parts of cities, storm and sanitary sewers are combined, which can lead to untreated wastewater discharges during rainy conditions.

may be treated and released directly into the receiving body of water, or it may go to the municipal sanitary sewer system. The industrial effluent often must receive some pretreatment before it can be released. Wastewater dischargers must obtain permits issued under the *National Pollutant Discharge Elimination System* (NPDES). An NPDES permit requires the discharger to meet certain technology-based effluent limits and perform effluent monitoring.

Also shown in Figure 1 is a storm sewer system that collects run-off from urban streets. In older cities, the stormwater sewer lines join the sanitary system, and the combination of wastewaters flows to the municipal wastewater treatment plant. About 1,000 communities in the United States, primarily in the Northeast and Great Lakes region, use combined sewer systems (U.S. EPA, 1999). These *combined systems* are unsatisfactory when it rains because they often end up carrying more wastewater than the local treatment system can handle. When that happens, a portion of the flow, which includes raw sewage, must be diverted around the treatment plant and released directly into the receiving water. The result is contaminated shorelines that must be posted with warnings after almost every storm. Separating these combined systems is immensely expensive. The preferred approach has been to create massive reservoirs, often underground, that store the combined flow until the storm passes, after which time the reservoir is slowly drained back into the sanitary sewer system.

In newer cities, and in newer sections of old cities, the stormwater collection system is completely separated from the sanitary sewer system to avoid the problem of overloading. Separated systems are not without their own problems, however. Run-off from streets and impervious areas during both wet and dry periods may be passed untreated into the local receiving water and often carries a heavy load of organics, metals, and sediment.

3 The Safe Drinking Water Act (SDWA)

Legislation to protect drinking water quality in the United States began with the Public Health Service Act of 1912. In this Act, the nation's first water quality standards were created. These standards slowly evolved over the years, but it wasn't until the passage of the Safe Drinking Water Act (SDWA) of 1974 that federal responsibility was extended beyond interstate carriers to include all community water systems regularly serving 15 or more outlets, or 25 or more customers. This extended federal regulation to cover the drinking water of about 90 percent of the people in the United States. Most community water systems are large; in fact, about 7 percent of the community water systems serve about 81 percent of the people. This original SDWA had two basic thrusts: (1) it required the EPA to establish national standards for drinking water quality, and (2) it required the operators of some 160,000 public water systems in the country to monitor the quality of water being delivered to customers and to treat that water, if necessary, to assure compliance with the standards.

Since enactment of the 1974 SDWA, there have been more than 10 major revisions and additions to the act with the most significant being the 1986 and 1996 amendments. In 1986, EPA was directed to quicken the pace of standard setting, including requirement of the equivalent of filtration for all surface water supplies and disinfection for all water systems. The regulations also recognized the potential for lead contamination in water distribution systems, including the plumbing systems in homes and nonresidential facilities. To help prevent lead poisoning, the 1986 amendments require the use of lead-free pipe, solder, and flux for all such systems. (In the past, solder normally contained about 50 percent lead; now it may contain no more than 0.2 percent lead.) In 1996, Congress added a requirement for all large municipal water systems to report annually to their customers data on the quality of their drinking water. Other provisions require additional controls on harmful pollutants, such as *Cryptosporidium,* and that EPA determine if the benefits justify the costs of any new regulation. In addition, in 2002, Congress added amendments to address water system security in response to heightened concern that contaminants could be intentionally put into public drinking water systems.

Drinking water standards fall into two categories: primary standards and secondary standards. *Primary standards* are enforceable and based on health-related criteria. *Secondary standards* are unenforceable guidelines based on both aesthetics, such as taste, odor, and color of drinking water, as well as nonaesthetic characteristics, such as corrosivity and hardness. In setting primary standards, EPA is required to balance the public health benefits of the standard against what is technologically and economically feasible. In this way, they are quite different from National Ambient Air Quality Standards, which must be set at levels that protect public health regardless of cost or feasibility. Of the 92 primary standards enacted by 2005, 83 specified *Maximum Contaminant Levels* (MCLs), such as the 0.010 mg/L concentration set for arsenic. In 9 cases, a *treatment technique* (TT) is required rather than an MCL because the pollutant may be difficult or too time-consuming to detect and measure. For example, because detection of pathogens such as viruses, *Giardia,* and *Cryptosporidium* is difficult, the standard simply specifies water filtration as the requirement.

As a prerequisite to establishing a primary standard, EPA sets an unenforceable *Maximum Contaminant Level Goal* (MCLG) for each contaminant. These goals are set at levels that present no known or anticipated health effects, including a margin of safety, regardless of technological feasibility or cost. Because exposure to any amount of a cancer-causing chemical is deemed to pose some risk to the individual, the MCLGs for carcinogens are generally set to zero. The 1996 SDWA requires EPA to review every primary standard at least every six years to determine whether they can be brought closer to the desired MCLGs. In addition, under the impetus of the SDWA, the list of primary standards is continuously being updated and expanded to include previously unregulated contaminants, which are shown to pose unacceptable health risks. A list of primary drinking water standards is given in Table 1.

TABLE 1

Maximum Contaminant Levels (mg/L) for Drinking Water Organized by Category

Inorganic Contaminants			
Antimony	0.006	Cyanide (as free cyanide)	0.2
Arsenic	0.010	Fluoride	4.0
Asbestos (fiber length $> 10\ \mu m$)	7 MFL[a]	Lead	TT[b]
Barium	2	Mercury (inorganic only)	0.002
Beryllium	0.004	Nitrate (as N)	10
Cadmium	0.005	Nitrite (as N)	1
Chromium (total)	0.1	Selenium	0.05
Copper	TT[b]	Thallium	0.002
Organic Contaminants			
Acrylamide	TT[b]	Epichlorohydrin	TT[b]
Alachlor	0.002	Ethylbenzene	0.7
Atrazine	0.003	Ethylene dibromide (EDB)	0.00005
Benzene	0.005	Glyphosate	0.7
Benzo(a)pyrene	0.0002	Heptachlor	0.0004
Carbofuran	0.04	Heptachlor epoxide	0.0002
Carbon tetrachloride	0.005	Hexachlorobenzene	0.001
Chlordane	0.002	Hexachlorocyclopentadiene (HEX)	0.05
Chlorobenzene	0.1	Lindane	0.0002
2,4-D	0.07	Methoxychlor	0.04
Dalapon	0.2	Oxamyl (Vydate)	0.2
1,2-Dibromo-3-chloropropane (DBCP)	0.0002	Pentachlorophenol	0.001
o-Dichlorobenzene	0.6	Picloram	0.5
p-Dichlorobenzene	0.075	Polychlorinated biphenyls (PCBs)	0.0005
1,2-Dichloroethane	0.005	Simazine	0.004
1,1-Dichloroethylene	0.007	Styrene	0.1
cis-1,2-Dichloroethylene	0.07	Tetrachloroethylene (PCE)	0.005
trans-1,2-Dichloroethylene	0.1	Toluene	1
Dichloromethane	0.005	Toxaphene	0.003
1,2-Dichloropropane	0.005	2,4,5-TP (Silvex)	0.05
Di(ethylhexyl) adipate	0.4	1,2,4-Trichlorobenzene	0.07
Di(ethylhexyl) phthalate (DEHP)	0.006	1,1,1-Trichloroethane	0.2
Dinoseb	0.007	1,1,2-Trichloroethane	0.005
Dioxin (2,3,7,8-TCDD)	0.00000003	Trichloroethylene (TCE)	0.005
Diquat	0.02	Vinyl chloride	0.002
Endothall	0.1	Xylenes (total)	10
Endrin	0.002		

(continued)

TABLE 1 (continued)

Maximum Contaminant Levels (mg/L) for Drinking Water Organized by Category			
Radionuclides			
Combined Radium-226 and Radium-228	5 pCi/L	Beta particles and photon emitters	4 mrem/yr
Gross alpha emitters[c]	15 pCi/L	Uranium	0.030
Microbiological Contaminants			
Cryptosporidium	TT[b]	Total coliforms (fecal coliform & *E. coli*)	5.%[d]
Giardia lamblia	TT[b]	Turbidity	TT[b]
Heterotrophic plate count (HPC)	TT[b]	Enteric viruses	TT[b]
Legionella	TT[b]		
Disinfectants/Disinfectant By-Products			
Bromate	0.010	Chlorite	1.0
Chloramines (as Cl_2)	4.0[e]	Haloacetic acids (HAA5)	0.060
Chlorine (as Cl_2)	4.0[e]	Total trihalomethanes (TTHMs)	0.080
Chlorine dioxide (as ClO_2)	0.8[e]		

[a] 7 million fibers/liter.

[b] Treatment technique (TT) requirement rather than an MCL.

[c] Excludes radon and uranium contributions.

[d] No more than 5% of samples test positive for coliforms in a month.

[e] Maximum residual disinfectant level—highest concentration of a disinfectant allowed in drinking water.

Chemical Standards

Inorganic chemicals include highly toxic metals, such as arsenic, cadmium, lead, and mercury; nitrite (NO_2) and nitrate (NO_3), which can cause methemoglobinemia (blue-baby syndrome); fluoride, which is purposely added to water to help prevent dental caries but which can cause mottling of teeth, if the exposure is excessive; and asbestos fibers, which are especially dangerous when inhaled, but their danger is less certain for ingestion (the MCL is written in terms of millions of fibers per liter with lengths greater than 10 microns).

Organic chemical contaminants for which MCLs have been promulgated are classified using the following three groupings:

1. *Synthetic Organic Chemicals* (SOCs) are compounds used in the manufacture of a wide variety of agricultural and industrial products. This includes primarily insecticides and herbicides.
2. *Volatile Organic Chemicals* (VOCs) are synthetic chemicals that readily vaporize at room temperature. These include degreasing agents, paint thinners, glues, dyes, and some pesticides. Representative chemicals include benzene; carbon tetrachloride; 1,1,1-trichloroethane (TCA); trichlorethylene (TCE); and vinyl chloride.
3. *Disinfectant byproducts* (DBPs) are the byproducts formed when a disinfectant reacts with chemicals in the water to form a toxic product. There are both organic and inorganic DBPs. The two types of regulated organic DBPs, trihalomethanes (THMs) and haloacetic acids (HAAs), are both formed when chlorine is used as a disinfectant and reacts with certain types of organic compounds in the water. Both THMs and HAAs are associated with an increased

risk of cancer. This creates somewhat of a dilemna in drinking water treatment because the addition of disinfectants such as chlorine is clearly beneficial for inactivation of microbial pathogens that cause most waterborne diseases, yet byproducts of their addition may cause cancer. This is an example of why the 1996 SDWA amendments specify that the overall health risk from contaminants must be minimized by balancing between competing risks. For DBPs, the most common solution is to remove the DBP-forming compounds (called precursors) from the water before the disinfectant is added.

Radionuclides

Radioactivity in public drinking water supplies is another category of contaminants regulated by the SDWA. Some radioactive compounds, or *radionuclides,* are naturally occurring substances such as radon and radium-226, which are often found in groundwater. Others, such as strontium-90 and tritium, are surface water contaminants resulting from atmospheric nuclear weapons testing fallout. The MCLs for most radionuclides are expressed as picocuries per liter (pCi/L), where 1 pCi/L corresponds to 2.2 radioactive decays per minute (which is the decay rate of 1 gram of radium). For example, the gross alpha particle activity MCL (including radium-226 but excluding radon and uranium) is 15 pCi/L. The MCL for beta particles and photon radioactivity is an annual dose, either to the whole body or to any particular organ, of 4 millirem/year.

Although, as of 2005, an MCL had not yet been finally established for radon, it is perhaps the most important radionuclide associated with drinking water. It is a colorless, odorless, and tasteless gas that occurs naturally in some groundwater. It is an unusual contaminant because the danger arises not from drinking radon-contaminated water but from breathing the gas after it has been released into the air. When radon-laden water is heated or agitated, such as occurs in showers or washing machines, the dissolved radon gas is released. Inhaled radon gas is thought to be a major cause of lung cancer. The proposed radon drinking water standard encourages states to adopt what is called a Multimedia Mitigation (MMM) program by allowing 4,000 pCi/L of radon in drinking water, if an authorized program is implemented to minimize radon exposure from nonwater-related indoor air sources. If an MMM program is not implemented, then the proposed MCL for radon is 300 pCi/L.

Microbiological Standards

Another category of primary MCLs is *microbiological contaminants*. Although it would be desirable to evaluate the safety of a given water supply by individually testing for specific pathogenic microorganisms, such tests are too difficult to perform on a routine basis or to be used as a standard. Instead, a much simpler technique is used, based on testing water for evidence of any fecal contamination. In this test, coliform bacteria (typically *Escherichia coli*) are used as *indicator organisms* whose presence suggests that the water is contaminated. Because the number of coliform bacteria excreted in feces is on the order of 50 million per gram, and the concentration of coliforms in untreated domestic wastewater is usually several million per 100 mL, it would be highly unlikely that water contaminated with human wastes would have no coliforms. That conclusion is the basis for the drinking water standard for microbiological contaminants, which specifies that for large water systems (serving more than 10,000 people), no more than 5 percent of the test samples can show any coliforms; for smaller systems testing fewer than 40 samples per month, no more than 1 sample can test positive.

The assumption that the absence of coliforms implies an absence of pathogens is based primarily on the following two observations. First, in our society, it is the excreta from relatively few individuals that adds pathogens to a wastestream, whereas the entire population contributes coliforms. Thus the number of coliforms should far exceed the number of pathogens. Secondly, for many of the waterborne diseases that have plagued humankind, the survival rate of pathogens outside the host is much lower than the survival rate of coliforms. The combination of these factors suggests that, statistically speaking, the ratio of pathogens to coliforms should be sufficiently small that we can conclude that it is extremely unlikely that a sample of water would contain a pathogen without also containing numerous coliforms.

This approach to testing for microbiological purity has been effective, but it is not an absolutely certain measure. For example, some nonbacterial pathogens, notably viruses and cysts from protozoa (such as *Giardia* and *Cryptosporidium*), survive considerably longer outside of their hosts than coliform bacteria, which increases the probability of encountering pathogens without accompanying coliforms. An additional problem exists in relying totally on coliforms as an indicator organism for microbiological contamination. For many cases, coliforms are more easily inactivated than other pathogens, such as viruses and protozoa, by chemical disinfectants. Thus, post-disinfectant sampling may indicate a lack of viable coliforms, while other pathogens are still active. To overcome this limitation in solely relying on coliforms as indicator organisms, a number of somewhat complex rules have been made or proposed in the past decade that specify credits for various treatment techniques and require that the sum of the credits in a particular water treatment process ensure that there is 99.99 percent inactivation or removal of enteric viruses, 99.9 percent of *Giarda,* and 99 percent of *Cryptosporidium.* These requirements are termed 4-log, 3-log, and 2-log removal of viruses, *Giardia,* and *Cryptosporidium,* respectively.

The coliform test is also used to assess the safety of water-contact recreational activities, with many states recommending a limit of 1,000 coliforms per 100 mL. However, proper interpretation of a coliform test made on surface water is complicated by the fact that fecal coliforms are discharged by animals as well as humans. Thus a high fecal coliform count is not necessarily an indication of human contamination. When it is important to distinguish between human and animal contamination, more sophisticated testing can be performed. Such testing is based on the fact that the ratio of fecal coliform to fecal streptococci is different in human and animal discharges.

Secondary Standards

EPA's secondary standards are nonenforceable, maximum contaminant levels intended to protect public welfare. Public welfare criteria include factors such as taste, color, corrosivity, and odor, rather than health effects. The limits suggested in Table 2 are in large part based on taste, and although not required by EPA, have been adopted by some states as enforceable requirements. Excessive sulfate is undesirable because

TABLE 2

Secondary Standards for Drinking Water

Contaminant	Level	Contaminant Effects
Aluminum	0.05–0.2 mg/L	Water discoloration
Chloride	250 mg/L	Taste, pipe corrosion
Color	15 color units	Aesthetic
Copper	1.0 mg/L	Taste, porcelain staining
Corrosivity	Noncorrosive	Pipe leaching of lead
Fluoride	2.0 mg/L	Dental fluorosis
Foaming agents	0.5 mg/L	Aesthetic
Iron	0.3 mg/L	Taste, laundry staining
Manganese	0.05 mg/L	Taste, laundry staining
Odor	3 threshold odor number	Aesthetic
pH	6.5–8.5	Corrosive
Silver	0.10 mg/L	Skin discoloration
Sulfate	250 mg/L	Taste, laxative effects
Total dissolved solids	500 mg/L	Taste, corrosivity, detergents
Zinc	5 mg/L	Taste

of its laxative effect; iron and manganese are objectionable because of taste and their ability to stain laundry and fixtures; foaming and color are visually upsetting; excessive fluoride causes a brownish discoloration of teeth; and odor from various dissolved gases may make water unacceptable to the drinker.

4 Water Treatment Systems

The purpose of water treatment systems is to bring raw water up to drinking water quality. The particular type of treatment equipment required to meet these standards depends largely on the source of water. About half of the drinking water in the United States comes from groundwater, and half comes from surface water. Most large cities rely more heavily on surface water, whereas most small towns or communities depend more on groundwater. Typically surface water treatment focuses on particle removal, and groundwater treatment focuses on removal of dissolved inorganic contaminants such as calcium and iron. Producing a water free of microbial pathogens is critical for any water source, but surface water has a much greater chance of microbial contamination, so filtration is now a requirement for surface water.

As suggested in Figure 2, a typical treatment plant for surface water might include the following sequence of steps:

1. *Screening* and *grit removal* take out relatively large floating and suspended debris and the sand and grit that settles very rapidly which may damage equipment.
2. *Primary sedimentation* (also called *settling* or *clarification*) removes the particles that will settle out by gravity alone within a few hours.

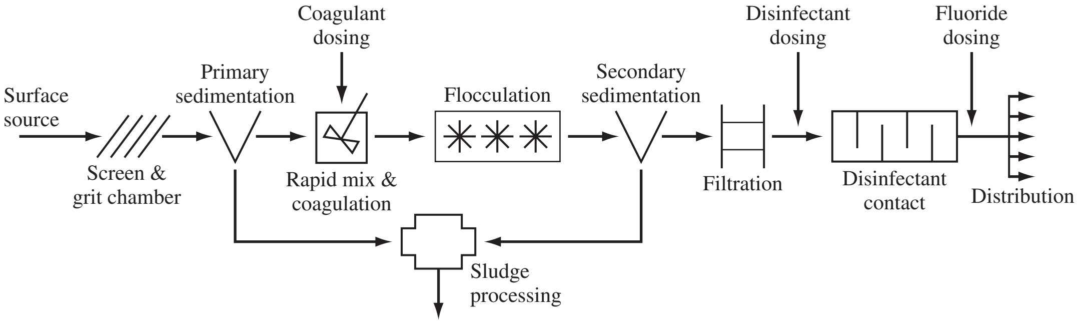


FIGURE 2 Schematic of a typical surface water treatment plant.

3. *Rapid mixing* and *coagulation* use chemicals and agitation to encourage suspended particles to collide and adhere into larger particles.
4. *Flocculation,* which is the process of gently mixing the water, encourages the formation of large particles of floc that will more easily settle.
5. *Secondary settling* slows the flow enough so that gravity will cause the floc to settle.
6. *Filtration* removes particles and floc that are too small or light to settle by gravity.
7. *Sludge processing* refers to the dewatering and disposing of solids and liquids collected from the settling tanks.
8. *Disinfection contact* provides sufficient time for the added disinfectant to inactivate any remaining pathogens before the water is distributed.

Groundwater is much freer of particles and pathogens than surface water, and in many places, it is delivered after disinfection alone. However, because groundwater often moves through the soils and minerals of the subsurface for long periods before withdrawal, it may contain high levels of dissolved minerals or objectionable gases. The most common dissolved mineral contaminants are calcium and magnesium, which are termed *hardness*. The calcium and magnesium can be removed by precipitating them as particles, so some softening steps are similar to the particle removal steps for a surface water. Figure 3 shows a typical groundwater treatment plant with the following unit operations:

1. *Aeration* removes excess and objectionable gases.
2. *Flocculation* (and *precipitation*) follows chemical addition, which forces the calcium and magnesium above their solubility limits.
3. *Sedimentation* removes the hardness particles that will now settle by gravity.
4. *Recarbonation* readjusts the water pH and alkalinity and may cause additional precipitation of hardness-causing ions.
5. *Filtration, disinfection,* and *solids processing* serve the same purposes as for surface water treatment.

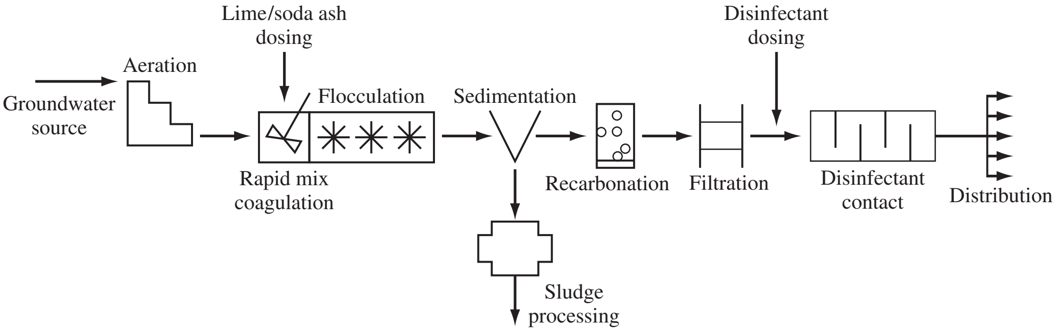


FIGURE 3 Schematic of a typical water treatment plant for groundwater, including softening for calcium and magnesium removal.

Sedimentation

Sedimentation or gravitational settling of particles from water is one of the oldest and simplest forms of water treatment. Simply allowing water to sit quietly in anything from a jar to a reservoir, decanting the water, and then using the undisturbed surface water often considerably improves the water's quality. A sedimentation basin or clarifier is a large circular or rectangular tank designed to hold the water for a long enough time to allow most of the suspended solids to settle out. The longer the detention time, the bigger and more expensive the tank must be, but correspondingly, the better will be the tank's performance. Clarifiers are usually equipped with a bottom scraper that removes collected sludge.

Sedimentation can remove particles that are contaminants themselves or may harbor other contaminants, such as pathogens or adsorbed metals. Although particles have very irregular shapes, their size may be described by an equivalent diameter that is determined by comparing them with spheres having the same settling velocity. The equivalent diameter is the *hydrodynamic diameter* when we speak of particles settling in water, and *aerodynamic diameter* for particles settling in air.

We can use a fairly simple analysis to calculate the settling velocity of a spherical particle. When such a particle reaches its terminal velocity, the gravitational force pulling it down is balanced by a drag force and a buoyancy force

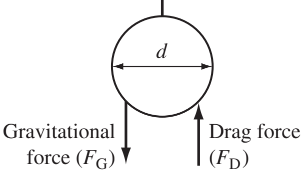


FIGURE 4 Forces acting on a particle settling in water.

(Figure 4). The drag force is a function of the particle's Reynolds number, which is expressed as

$$\mathrm{Re} = \frac{\rho v_s d_p}{\mu} \tag{1}$$

where

ρ = density of water
v_s = particle settling velocity
d_p = particle hydrodynamic diameter
μ = absolute viscosity of water

Although particles in water may have Reynolds numbers in the laminar ($\mathrm{Re} < 1$), transition ($1 < \mathrm{Re} < 10{,}000$), or turbulent ($\mathrm{Re} > 10^4$) regions, we only need to consider the laminar region case to understand the basics of sedimentation basin operation. In the laminar region, the drag force is

$$F_D = 3\pi\mu v_s d_p \tag{2}$$

Respectively, the gravitational and buoyancy forces are

$$F_G = mg = V_p \rho_p g \quad \text{and} \quad F_B = m_w g = V_p \rho g \tag{3}$$

where

g = gravitational acceleration (9.807 m/s^2)
m = the particle mass
m_w = the mass of the water displaced by the particle
V_p = particle volume = $\pi d_p^3/6$
ρ_p = particle density

From Figure 4, it can be seen that the balance of forces on a particle when it is at its terminal settling velocity is

$$F_G = F_D + F_B$$

Substituting (2) and (3) into this force balance gives

$$\frac{\pi d_p^3 \rho_p g}{6} = \frac{\pi d_p^3 \rho g}{6} + 3\pi\mu v_s d \tag{4}$$

which simplifies to Stokes' law

$$v_s = \frac{g(\rho_p - \rho) d_p^2}{18\mu} \tag{5}$$

Figure 5 shows the trajectory of a particle that is being carried by the horizontal flow of water from one end to the other of a rectangular settling basin. It settles a distance, h_p, as it traverses the basin such that

$$h_p = v_s \theta = \frac{v_s V_b}{Q} \tag{6}$$

where θ is the hydraulic detention time of the basin, V_b is the basin's volume, and Q is the volumetric flow rate of water through the basin. Settling basins work on the

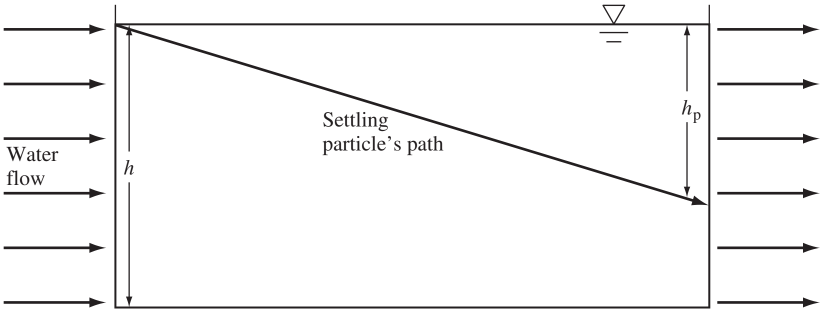


FIGURE 5 Trajectory of a particle settling as it is carried from left to right by the flow of water through a sedimentation basin.

principle that if a particle reaches the floor of the basin, then it is removed from the water. For the particle to reach the basin floor, its settling velocity (ν_s) must be equal to or greater than a *critical settling velocity,* ν_o, of

$$\nu_o = \frac{h}{\theta} = \frac{hQ}{V_b} = \frac{hQ}{hA_b} = \frac{Q}{A_b} \tag{7}$$

where A_b is the surface area of the rectangular basin.

The critical settling velocity (also known as the *surface loading rate* or the *overflow rate*) is the minimum settling velocity that a particle can have and still be guaranteed to be removed in the settling basin. In other words, 100 percent of the particles whose settling velocity is equal to or greater than the critical settling velocity, ν_o, will be removed by the basin. Therefore, based on (7) and (5), if we want to design a clarifier to remove all particles of a size, d_p, from a water stream with a flow rate, Q, the surface area, A_b, of the rectangular basin must be

$$A_b = \frac{18Q\mu}{g(\rho_p - \rho)d_p^2} \tag{8}$$

Although many clarifiers in both water and wastewater treatment plants are circular rather than rectangular, (7) and (8) still apply; however, for the circular clarifier, the basin surface area is π times the basin's radius squared, rather than its width times its length as with a rectangular basin. In most circular settling tanks, the influent water enters at the center, and the treated effluent leaves the perimeter of the tank as it overflows across a weir. If the tank is too small and wastewater flows too quickly toward the overflow weir, the solids will be moving too fast to settle. Overflow rates typically are in the range of 1 to 2.5 $\text{m}^3 \cdot \text{m}^{-2} \cdot \text{h}^{-1}$. Another parameter that influences clarifier performance is the detention time of the tank. The hydraulic detention time in any tank is its volume divided by the influent flow rate. This is numerically equal to the time that would be required to fill an empty tank with water entering at the daily average flow rate. Although the overflow rate does not depend on the depth of the sedimentation basin, the detention time does. Typical detention times range from 1 to 4 hours.

EXAMPLE 1 Silt Removal in a Clarifier

A drinking water treatment plant uses a circular sedimentation basin to treat 3.0 MGD of river water. (*MGD* stands for million gallons per day and is a common U.S. measure of flow rate used for water and wastewater treatment. 1.0 MGD = 0.0438 m^3/s). After storms occur upstream, the river often carries 0.010 mm silt particles with an average density of 2.2 g/cm^3, and the silt must be removed before the water can be used. The plant's clarifier is 3.5 m deep and 21 m in diameter. The water is 15°C.

a. What is the hydraulic detention time of the clarifier?

b. Will the clarifier remove all of the silt particles from the river water?

Solution

a. The hydraulic detention time, θ (theta), of a flow-through vessel is

$$\theta = V/Q \tag{9}$$

where V is the vessel's volume, and Q is the flow rate through the vessel. The volume of a circular basin is

$$V = \frac{\pi d_b^2 h}{4}$$

where d_b is the basin's diameter, and h is its depth. Thus, the clarifier detention time is

$$\theta = \frac{\pi d_b^2 h}{4Q} = \frac{\pi(21\text{ m})^2(3.5\text{ m})}{4(3.0\text{ MGD})(0.0438\text{ m}^3/\text{s/MGD})} = 9{,}226\text{ s} = 2.6\text{ hr}$$

b. To find if the silt particles will be removed, we need to calculate their settling velocity and compare it to the critical velocity for the sedimentation basin. We first need the water density ρ and the viscosity μ. "Density and Viscosity of Water and Air," the water density and viscosity at 15°C are found to be 999.1 $kg \cdot m^{-3}$ and 0.00114 $kg \cdot m^{-1} \cdot s^{-1}$, respectively. The gravitational acceleration constant g is found, "Useful Constants." Then using (5), the particle settling velocity is

$$v_s = \frac{(9.807\text{ m}\cdot\text{s}^{-2})(2{,}200 - 999.1\text{ kg}\cdot\text{m}^{-3})(10^{-5}\text{ m})^2}{18(0.00114\text{ kg}\cdot\text{m}^{-1}\cdot\text{s}^{-1})} = 5.74 \times 10^{-5}\text{ m/s}$$

The critical velocity for the sedimentation basin is found using (7) as

$$v_o = \frac{h}{\theta} = \frac{3.5\text{ m}}{9{,}226\text{ s}} = 38 \times 10^{-5}\text{ m/s}$$

The silt's settling velocity is much less than the basin's critical velocity, so all of the silt will not be removed by the treatment plant's clarifier. Note that the critical velocity could also have been calculated by dividing the flow rate by the clarifier's surface area.

Coagulation and Flocculation

Raw water may contain suspended particles that are too small to settle by gravity in a reasonable time period and cannot be removed by simple filtration. Many of these particles are colloids (particles in the size range of about 0.001 to 1 μm). The goal of coagulation is to alter the particle surfaces in such a way as to permit them to adhere to each other. Thus, they can grow to a size that will allow removal by sedimentation or filtration. Coagulation is considered to be a *chemical* treatment process that destabilizes particles (makes them "sticky"), as opposed to a *physical* treatment operation such as flocculation, sedimentation, or filtration.

Most colloids and nonsettleable particles of interest in water treatment remain suspended in solution because they have a net negative surface charge that causes the particles to repel each other. The coagulant's purpose is to neutralize the surface charge, thus allowing the particles to come together to form larger particles that can be more easily removed. The usual coagulant is alum, $Al_2(SO_4)_3 \cdot 18H_2O$, although $FeCl_3$, $FeSO_4$, and other coagulants, such as polyelectrolytes, can be used. Because the intention here is simply to introduce the concepts of water treatment, leaving the complexities for more specialized books, let us just look at the reactions involving alum. Alum ionizes in water, producing Al^{3+} ions, some of which neutralize the negative charges on the colloids. Most of the aluminum ions, however, react with alkalinity in the water (bicarbonate) to form insoluble aluminum hydroxide, $Al(OH)_3$. The overall reaction is

$$\underset{\text{Alum}}{Al_2(SO_4)_3 \cdot 18H_2O} + 6HCO_3^- \rightleftharpoons \underset{\substack{\text{Aluminum}\\ \text{hydroxide}}}{2Al(OH)_3\downarrow} + 6CO_2 + 18H_2O + \underset{\text{Sulfates}}{3SO_4^{2-}} \quad (10)$$

If insufficient bicarbonate is available for this reaction to occur, the pH must be raised, usually by adding lime $Ca(OH)_2$ or sodium carbonate (soda ash), Na_2CO_3. The aluminum hydroxide precipitate forms a light, fluffy *floc* that adsorbs destabilized particles on its surface as it settles. The destabilized particles may also aggregate and grow by colliding with each other. The degree of particle destabilization is quantified by the *collision efficiency factor*, α, which is defined as the fraction of the number of collisions between particles that result in aggregation. If the particles are completely destabilized, then every collision results in aggregation, and $\alpha = 1.0$, whereas, for instance, if $\alpha = 0.25$, only one in four collisions between particles results in the colliding particles sticking together.

Coagulants are added to the raw water in a *rapid mix/coagulation* tank that has quickly rotating impellers to mix the chemicals. Detention times in the tank are typically less than one-half minute. *Flocculation* follows in a tank that provides gentle agitation for approximately one-half hour. During this time, the precipitating aluminum hydroxide forms a plainly visible floc. The mixing in the flocculation tank must be done very carefully. It must be sufficient to encourage particles to make contact with each other and enable the floc to grow in size, but it cannot be so vigorous that the fragile floc particles will break apart. Mixing also helps keep the floc from settling in this tank, rather than in the sedimentation tank that follows. Figure 6 shows a cross-section of a mixing tank followed by a sedimentation tank.

The level of agitation in the flocculation tank is characterized by a parameter, G, the root mean square *velocity gradient* or more simply the *mixing intensity*. G is

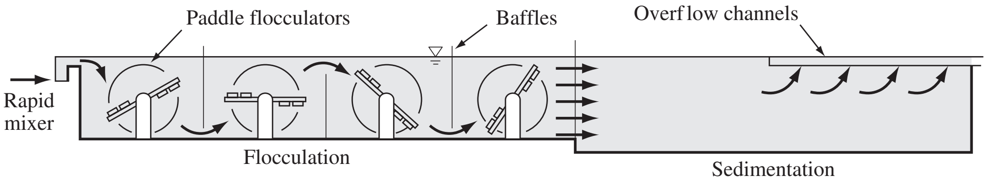


FIGURE 6 Cross-section of flocculation and sedimentation tanks (Hammer and Hammer, 1996).

the primary parameter the engineer uses to maximize the rate of collisions between particles, without mixing so intensely that the flocs break up. The mixing intensity G depends on the power input to the paddles as given by

$$G = \sqrt{\frac{P}{\mu V_b}} \tag{11}$$

where P is the power input to the paddles (or other mixing method), V_b is the volume of the vessel, and μ is the viscosity of the water.

The rate of aggregation of particles can be seen as the rate of change of the total particle number concentration. The particle number concentration, N, is the total number of particles (where each particle counts as a single particle, whether it's an aggregate or a singlet) divided by the volume of water. Each time two particles collide to form an aggregate, the number concentration is decreased by one. Thus, the rate of flocculation can be mathematically expressed as dN/dt. This is analogous to how the rate of change of the contaminant concentration with time was represented. Because mixing is inherent in flocculation basins (Figure 6), each chamber acts as a continuously stirred tank reactor (CSTR) with the effluent number concentration from one chamber being the influent number concentration for the next chamber. Detailed coagulation/flocculation theory is complex and beyond this text's scope. However, a useful understanding of the processes can be gained by considering a simple, yet still practically relevant set of conditions. The following analysis assumes the simple case in which all the particles are initially nearly the same size (called a *monodisperse distribution*), the water mixing is relatively mild (termed *laminar mixing*), and an aggregate's volume is the sum of the volumes of the individual particles comprising the aggregate (termed *coalescing aggregation*). For this case, the flocculation rate can be expressed as a first-order process.

$$\mathrm{r}(N) = -kN \tag{12}$$

where

$$k = \frac{\alpha 4 \Omega G}{\pi} \tag{13}$$

α is the collision efficiency factor discussed earlier, and Ω is the floc volume, which is defined as

$$\Omega = \frac{\pi d_p^3 N_0}{6} \tag{14}$$

where N_0 is the initial number concentration of monodisperse particles of diameter, d_p. You can think of the floc volume Ω as the total volume of solids per volume of water. In other words, it is the particle volume concentration, as opposed to the particle number concentration. Because we assume coalescing aggregation, the floc volume concentration is constant, even though the particle number concentration changes as aggregation occurs.

EXAMPLE 2 Silt Flocculation Before Sedimentation

To improve their settling, the 0.010 mm silt particles in Example 1 are completely destabilized by adding alum and are passed through one of two side-by-side, well-mixed flocculation chambers. The chambers are cubic with each dimension being 3.5 m. They are mixed with paddle mixers that input 2.50 kW of power into the water in each chamber. The water entering the flocculation chamber contains 10^5 particles per mL. What is the average diameter of the aggregates leaving the flocculation chambers?

Solution Each chamber handles half of the plant's 3.0 MGD, so we will analyze one chamber receiving 1.5 MGD. First, we find the mixing intensity, G, using (11).

$$G = \sqrt{\frac{(2.50\ \text{kW})(1{,}000\ \text{W/kW})}{(0.001145\ \text{kg}\cdot\text{m}^{-1}\cdot\text{s}^{-1})(3.5\ \text{m})^3}} = 226\ \text{s}^{-1}$$

The flocculation chamber acts as a steady state CSTR in which a first-order decay in particle number concentration is occurring (Figure 7).

The mass balance for a steady state CSTR with first-order decay is input rate = output rate + $V_b kC$. For the flocculation chamber, we write this equation in terms of number concentration N, rather than mass concentration C, and use the variables defined in Figure 7 to give

$$QN_0 = QN + V_b kN$$

where N_0 is the influent number concentration, and N is effluent concentration. By using (13) and dividing through by N, this becomes

$$\frac{N_0}{N} = 1 + k\frac{V_b}{Q} = 1 + \frac{\alpha 4\Omega G V_b}{\pi Q}$$

To find the floc volume Ω in (14), we need the particle diameter d_p and the initial particle density N_0. The particle diameter was given in Example 1 as 1×10^{-5} m, and the initial particle density is given previously as 10^5 particles/mL.

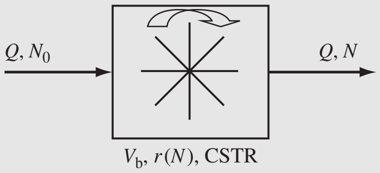


FIGURE 7 Schematic of single flocculation basin with variables defined.

The floc volume is then

$$\Omega = \frac{\pi(1 \times 10^{-5}\ \text{m})^3 \left(\frac{10^5}{\text{mL}}\right)\left(\frac{10^6\ \text{mL}}{\text{m}^3}\right)}{6} = 5.24 \times 10^{-5}$$

and $\alpha = 1.0$, since the silt particles are completely destabilized, so

$$\frac{N_0}{N} = 1 + \frac{4(5.24 \times 10^{-5})(226\ \text{s}^{-1})}{\pi}\left(\frac{(3.5\ \text{m})^3}{(1.5\ \text{MGD})\ (0.0438\ \text{m}^3 \cdot \text{s}^{-1})/\text{MGD}}\right) = 10.8$$

This means that the average aggregate contains 10.8 of the original 0.010 mm diameter silt particles. Since this approach assumes the particles coalesce upon aggregation the average aggregate volume, V_a, is 10.8 times the singlet volume, V_p, or $V_a = 10.8V_p$. To find the average aggregate's diameter, d_a, we use this relationship to write

$$V_a = \frac{\pi}{6}d_a^3 = 10.8\frac{\pi}{6}d_p^3$$

from which

$$d_a = \left(10.8d_p^3\right)^{1/3} = (10.8\ (0.010\ \text{mm})^3)^{1/3} = 0.0221\ \text{mm}$$

Filtration

Filtration is one of the most widely used and effective means of removing small particles from water. This also includes pathogens, which are essentially small particles. For drinking water filtration, the most common technique is called *rapid depth filtration*. The rapid depth filter consists of a layer or layers of carefully sieved filter media, such as sand, anthracite coal, or diatomaceous earth, on top of a bed of graded gravels. The pore openings between the media grains are often greater than the size of the particles that are to be removed, so much of the filtration is accomplished by means other than simple straining. Adsorption, continued flocculation, and sedimentation in the pore spaces are important removal mechanisms. When the filter becomes clogged with particles, the filter is shut down for a short period of time and cleaned by forcing water backwards through the media. After backwashing, the media settles back in place and operation resumes.

Rapid depth filters typically operate with *filtration rates* (also called *loading rates* or *superficial velocities*) between 2 and 10 gpm/ft^2 (5–25 m^{3}m^{-2}hr^{-1}). The filtration rate, ν_a, is the volumetric flow rate divided by the filter's cross-sectional area, A_f:

$$\nu_a = \frac{Q}{A_f} \tag{15}$$

A filter's surface area is normally less than 100 m^2, so banks of multiple filters are operated to treat large volumes of water. Having multiple filters allows continuous production of water even though at any given time one or more filters are being backwashed. Because some of the clean water produced in every filtration cycle must be used to backwash (and rinse the backwash water from the filter), less than 100 percent of the water that is filtered is actually available for distribution. The fraction of the water that is usable is the *filter efficiency* (or *production efficiency*), η_f, and the net rate of water production is the *effective filtration rate*, r_{ef}. These measures of a filter's performance are defined as

$$\eta_f = \frac{V_f - V_b - V_r}{V_f} \tag{16}$$

and

$$r_{ef} = \frac{\left(\frac{V_f}{A_f} - \frac{V_b}{A_f} - \frac{V_r}{A_f}\right)}{t_f + t_b + t_r} = \frac{(V_f - V_b - V_r)}{A_f t_c} \tag{17}$$

where

V_f = volume of water filtered per filter cycle (m^3)
V_b = volume of water used to backwash the filter (m^3)
V_r = volume of water used to rinse the filter after backwash (m^3)
t_f = duration of time that clean water is produced (hr)
t_b = duration of time that the filter is backwashing (hr)
t_r = duration of time that the filter is rinsing (hr)
t_c = duration of time for a complete filter cycle (hr)

Comparing the effective filtration rate (r_{ef}) to the filtration rate (ν_a) indicates how much the overall throughput rate of the filter is decreased by backwashing and rinsing. Rapid depth filters are designed with filter efficiencies of greater than 95 percent and, therefore, effective filtration rates approaching their loading rates. Filtration rates, ν_a, typically range from 2–8 gpm/ft^2.

Disinfection

The final, primary unit operation in most water treatment trains is disinfection. Disinfection has to meet two objectives: *primary disinfection,* to kill any pathogens in the water, and *secondary* (or *residual*) *disinfection* to prevent pathogen regrowth in the water during the period before it is used. Although traditionally a single disinfectant was added to the water to serve both purposes, it is becoming more common to use one chemical as the primary disinfectant and another as the residual disinfectant.

The most commonly used method of disinfection in the U.S. is *free chlorine* disinfection because it is cheap, reliable, and easy to use. Free chlorine in water is developed by dosing with either chlorine gas ($Cl_{2(g)}$), sodium hypochlorite (NaOCl) or calcium hypochlorite ($Ca(OCl)_2$). The dosed chemical reacts in the water to produce dissolved chlorine gas ($Cl_{2(aq)}$), hypochlorous acid (HOCl), and hypochlorite (OCl^-), which all contribute to the free chlorine concentration. The relative proportion of each is a function of pH, as shown in the following two reactions:

$$Cl_{2(aq)} + H_2O = HOCl + H^+ + Cl^- \tag{18}$$

$$HOCl = OCl^- + H^+$$

Hypochlorous acid is a much stronger disinfectant than hypochlorite, so free chlorine disinfection is usually conducted in slightly acidic water (see Example 2.8). Dissolved chlorine gas does not contribute significantly to the free chlorine concentration unless the water pH is less than 3, which is too acidic (corrosive) for practical use. Although free chlorine is very effective against bacteria, its effectiveness is less with protozoan cysts, most notably those of *Giardia lamblia* and *Cryptosporidium,* and with viruses. In almost all circumstances, complete treatment, including coagulation, sedimentation, filtration, as well as disinfection, is used to meet the microbiological standards of the Safe Drinking Water Act. (More information on pathogens and the requirements for their removal from drinking water was provided earlier in this chapter in Section 3.)

A principal advantage of chlorination over other forms of disinfection is that a chlorine residual is created that provides secondary disinfection of the treated water after leaving the treatment plant. This guards against possible contamination that might occur in the water distribution system. To increase the lifetime of this residual, some systems add ammonia to the treated water. Ammonia reacts with free chlorine to form *chloramines* (NH_2Cl, $NHCl_2$, and NCl_3), which are termed *combined chlorine*. Chloramines are less effective as oxidants than HOCl and hence are seldom used as primary disinfectants. However, they are more persistent than free chlorine and maintain secondary disinfection in distribution systems even with long residence times.

A disadvantage of free chlorine is the formation of halogenated *disinfectant byproducts* (DBPs). DBPs include *trihalomethanes* (*THMs*), such as the carcinogen, chloroform ($CHCl_3$), and *haloacetic acids* (*HAAs*). THMs and HAAs are created when free chlorine combines with natural organic substances, such as decaying vegetation, that may be present in the water. One approach to reducing THMs is to remove more of the organics before chlorination takes place. It was common practice in the past to chlorinate the incoming raw water before coagulation and filtration, but it has since been realized that this contributes to the formation of DBPs. By moving the free chlorine primary disinfection point to follow coagulation/flocculation, sedimentation, and filtration, many of the organic DBP precursors are removed, and some degree of THM control is achieved.

The problem of free chlorine DBPs has spurred interest in alternative primary disinfectants, such as chlorine dioxide (ClO_2) and ozone (O_3). Each has the advantage of not creating THMs or HAAs, but as more research is done, it has been found that they create other DBPs. For instance, ClO_2 may break down to form chlorite, which has an MCL of 1.0 mg/L, and ozone creates bromate, which has an MCL of 0.010 mg/L, when bromide is present in the water. The relative risk posed by these DBPs and their likelihood of formation relative to free chlorine DBPs is the subject of considerable current research. Chlorine dioxide is a potent bactericide and viricide. However, it is a very costly method of disinfection and because it is an explosive gas, it must be generated on site. Ozone is the most powerful disinfectant and is more effective against cysts and viruses than free chlorine or ClO_2. It has the added advantage of leaving no taste or odor problems. Although ozonation is widely used in European water treatment facilities, it has the disadvantages of not forming a protective residual in the treated water and being more expensive than chlorination.

The rate of inactivation of pathogens during disinfection is often modeled using the *Chick-Watson model* or variations of it. The model predicts the change in the pathogen concentration, N (number of pathogens/unit volume), with time as

$$\mathrm{r}(N) = \frac{dN}{dt} = -kC^nN \tag{19}$$

where C is the concentration of disinfectant, n is an empirical constant called the *coefficient of dilution,* and k is an empirical rate constant called the *coefficient of specific lethality*. For many cases, the model can be simplified and still works well if the coefficient of dilution is taken as 1.0. The coefficient of specific lethality is a function of the particular pathogen target, particular disinfectant being used, temperature, and pH. Sometimes (19) is expressed as a first-order reduction of pathogen concentration by assuming C^n is constant and lumping k and C^n into a single rate constant, k^*, as

$$\mathrm{r}(N) = \frac{dN}{dt} = -k^*N \tag{20}$$

EXAMPLE 3 Giardia Inactivation by Ozone

Because of the purity of its groundwater source, a water utility only has to disinfect the water before delivery. The utility uses ozone as its primary disinfectant and treats 10 MGD (million gallons per day) of water. The ozone is bubbled into the water as it passes sequentially through the four chambers of a disinfectant contact tank (see Figure 8). Each chamber of the tank is well mixed, holds 50,000 gallons of water, and receives the same ozone dose. As discussed earlier in Section 3, EPA requires that the *Giardia* concentration be reduced to 1/1,000 of its initial concentration (referred to as 3-log reduction). What is the minimum ozone concentration that must be used to achieve the required removal of *Giardia*? The coefficient of specific lethality and coefficient of dilution for ozone inactivation of *Giardia* under these conditions are 1.8 $\mathrm{L \cdot mg^{-1} \cdot min^{-1}}$ and 1.0, respectively.

Solution We need to do a mass balance on the tank in terms of the *Giardia* concentration, but the tank as a whole is not well modeled as a continuous stirred

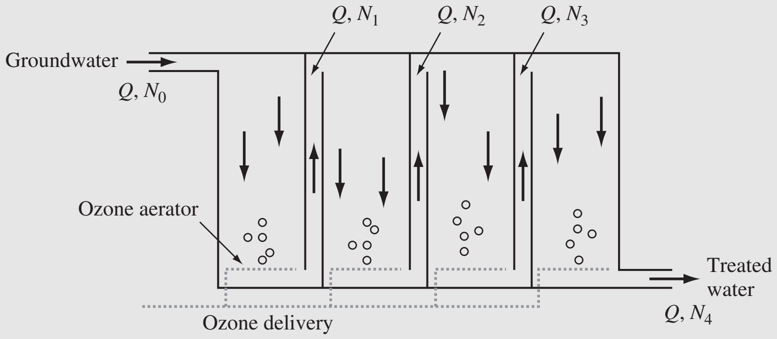


FIGURE 8 Four-chamber ozone contact tank for drinking water disinfection. Heavy arrows denote the water flow path through the tank.

tank reactor (CSTR) or a plug flow reactor (PFR). However, because each chamber is well mixed and the effluent of one chamber is the influent of the next, we can model the system as a series of CSTRs. First we develop an equation for a series of steady state CSTRs in which the volume V_c, the residence time θ_c, the flow rate Q, and the ozone concentration C, are the same in each CSTR. We make this derivation general to any series of CSTRs in which first-order decay is occurring by using (20) as the reaction rate expression and later substituting the product of k (the specific lethality coefficient) and C^n in place of k^*. Using (1.23) and (20), the steady state mass balance around the first chamber is

$$0 = QN_0 - QN_1 - Vk^*N_1 \tag{21}$$

Solving for N_1 gives

$$N_1 = \frac{QN_0}{Q + Vk^*}$$

Remembering that $\theta_c = V_c/Q$, and rearranging yields

$$\frac{N_1}{N_0} = \frac{1}{1 + \theta_c k^*} \tag{22}$$

Similarly, a steady state CSTR, mass balance around the second chamber yields

$$\frac{N_2}{N_1} = \frac{1}{1 + \theta_c k^*} \tag{23}$$

Because

$$\frac{N_2}{N_0} = \frac{N_2}{N_1} \times \frac{N_1}{N_0}$$

then the product of (22) and (23) is

$$\frac{N_2}{N_0} = \frac{1}{(1 + \theta_c k^*)^2} \tag{24}$$

Extending this approach to m number of CSTRs in series gives a general equation for first-order decay in a series of steady state CSTRs as

$$\frac{N_m}{N_0} = \frac{1}{(1 + \theta_c k^*)^m} \tag{25}$$

For the particular case of the ozone contact tank's four chambers ($m = 4$), the coefficient of dilution $n = 1$, and $k^* = kC^n$, (25) can be rearranged to give the ozone concentration in each chamber as

$$C = \frac{\left(\frac{N_0}{N_4}\right)^{1/4} - 1}{\theta_c k} \tag{26}$$

Three-log removal means the effluent concentration after disinfection is 1/1,000th of the influent concentration. In other words, $N_0/N_4 = 1{,}000$. Remembering from (9) that $\theta = V/Q$, for an influent flow rate of 10 MGD and a contactor compartment volume of 50,000 gallons, the minimum ozone concentration required in each chamber is

$$C = \frac{(1{,}000)^{1/4} - 1}{\dfrac{(50{,}000\text{ gal})(1{,}440\text{ min/day})}{(10^7\text{ gal/day})}(1.8\text{ L}\cdot\text{mg}^{-1}\cdot\text{ min}^{-1})} = 0.36\text{ mg/L}$$

Hardness and Alkalinity

The presence of multivalent cations, most notably calcium and magnesium ions, is referred to as water *hardness*. Groundwater is especially prone to excessive hardness. Hardness causes two distinct problems. First, the reaction between hardness and soap produces a sticky, gummy deposit called "soap curd" (the ring around the bathtub). Essentially all home cleaning activities, from bathing and grooming to dishwashing and laundering, are made more difficult with hard water.

Although the introduction of synthetic detergents has decreased but not eliminated the impact of hardness on cleaning, the second problem, that of scaling, remains significant. When hard water is heated, calcium carbonate ($CaCO_3$) and magnesium hydroxide ($Mg(OH)_2$) readily precipitate out of solution, forming a rock-like scale that clogs hot water pipes and reduces the efficiency of water heaters, boilers, and heat exchangers. Pipes filled with scale must ultimately be replaced, usually at great expense. Heating equipment that has scaled up not only transmits heat less readily, thus increasing fuel costs, but also is prone to failure at a much earlier time. For both of these reasons, if hardness is not controlled at the water treatment plant itself, many individuals and industrial facilities find it worth the expense to provide their own water softening.

Hardness is defined as the concentration of all multivalent metallic cations in solution, including iron (Fe^{2+}), manganese (Mn^{2+}), strontium (Sr^{2+}), and aluminum (Al^{3+}). However, the principal ions causing hardness in natural water are calcium (Ca^{2+}) and magnesium (Mg^{2+}), so typically hardness can be operationally defined as the sum of only Ca^{2+} and Mg^{2+}.

It is conventional practice to calculate hardness (as well as alkalinity) using the concepts of *equivalents* and *equivalent weights*. Chemists in the eighteenth century noted, for example, that 8 g of oxygen combine with 1 g of hydrogen (to form H_2O), while 3 g of carbon combine with 1 g of hydrogen (to form CH_4). They inferred from this that 8 g of oxygen ought to combine with 3 g of carbon (which it does in CO_2). In this example, the equivalent weight (EW) of oxygen would be 8 g, and that of carbon would be 3 g, and 1 EW of oxygen would combine with 1 EW of carbon (to form CO_2). This idea of equivalents, when first proposed, was not taken very seriously because it is complicated by the fact that individual elements can have more than one value for their equivalent weight, depending on the reaction in question (e.g., carbon forms CO as well as CO_2). However, with the understanding of the structure of the atom that came later, it regained favor as a convenient method of handling certain chemical computations.

The equivalent weight of a substance is its atomic or molecular weight divided by a number, n, that relates to its valence or ionic charge. For the reactions of interest in hardness and alkalinity calculations, n is simply the ionic charge. For neutral compounds, it is the number of hydrogen ions that would be required to replace the cation:

$$\text{Equivalent weight (EW)} = \frac{\text{Atomic or molecular weight}}{n} \tag{27}$$

Thus, for example, to find the EW of $CaCO_3$, we note it would take two hydrogen ions to replace the cation (Ca^{2+}). Its EW is therefore

$$EW_{CaCO_3} = \frac{(40 + 12 + 3 \times 16)}{2} = \frac{100}{2} = 50\text{g/eq} = 50 \text{ mg/meq} \tag{28}$$

where (mg/meq) are the units of milligrams per milliequivalent. For the calcium ion itself (Ca^{2+}), which has an atomic weight of 40.1 and a charge of 2, the equivalent weight is

$$EW_{Ca^{2+}} = \frac{40.1}{2} \approx 20.0 \text{ mg/meq}$$

Table 3 lists the equivalent weights for a number of common chemicals important in water hardness calculations.

In measuring hardness, the concentrations of the multivalent cations are expressed as either meq/L using the following expression

$$\text{meq/L of X} = \frac{\text{concentration of X(mg/L)}}{\text{EW of X(mg/meq)}} \tag{29}$$

TABLE 3

Equivalent Weights for Selected Ions, and Compounds

Name	Symbol or Formula	Atomic or Molecular Weight	Charge *n*	Equivalent Weight (mg/meq)
Cations				
Aluminum	Al^{3+}	27.0	3	9.0
Calcium	Ca^{2+}	40.1	2	20.0
Hydrogen	H^+	1.0	1	1.0
Magnesium	Mg^{2+}	24.3	2	12.2
Potassium	K^+	39.1	1	39.1
Sodium	Na^+	23.0	1	23.0
Anions				
Bicarbonate	HCO_3^-	61.0	1	61.0
Carbonate	CO_3^{2-}	60.0	2	30.0
Chloride	Cl^-	35.5	1	35.5
Hydroxide	OH^-	17.0	1	17.0
Nitrite	NO_2^-	46.0	1	46.0
Nitrate	NO_3^-	62.0	1	62.0
Orthophosphate	PO_4^{3-}	95.0	3	31.7
Sulfate	SO_4^{2-}	96.0	2	48.0
Neutrals				
Calcium carbonate	$CaCO_3$	100.0	2	50.0
Carbon dioxide	$CO_{2(aq)}$	44.0	2	22.0
Magnesium hydroxide	$Mg(OH)_2$	58.3	2	29.2

or alternatively, as *mg/L as* $CaCO_3$ by the equation

$$\text{mg/L of X as CaCO}_3 = \frac{\text{concentration of X (mg/L)} \times 50.0 \text{ mg CaCO}_3\text{/meq}}{\text{EW of X (mg/meq)}} \quad (30)$$

The following example shows how to work with these units.

EXAMPLE 4 Expressing the Concentration of Hardness

A sample of groundwater has 100 mg/L of Ca^{2+} and 10 mg/L of Mg^{2+}. Express its hardness in units of meq/L and mg/L as $CaCO_3$.

Solution Using (29) and Table 3, the contribution of calcium in meq/L is

$$\frac{100 \text{ mg/L}}{20.0 \text{ mg/meq}} = 5.0 \text{ meq/L}$$

Using (30), the calcium concentration in mg/L as $CaCO_3$ is

$$\frac{(100 \text{ mg/L})(50.0 \text{ mg CaCO}_3\text{/meq})}{(20.0 \text{ mg/meq})} = 250 \text{ mg/L as CaCO}_3$$

Similarly the Mg^{2+} hardness is calculated as 0.82 meq/L and 41.0 mg/L as $CaCO_3$. The water's hardness is

$$5.0 \text{ meq/L} + 0.82 \text{ meq/L} = 5.8 \text{ meq/L}$$

or,

$$(250.0 + 41.0) \text{ mg/L as CaCO}_3 = 291 \text{ mg/L as CaCO}_3$$

Although there are no absolute distinctions, and public acceptance of hardness is dependent on the past experiences of individual consumers, the qualitative classification of hardness given in Table 4 is often used. Hardness above about 150 mg/L as $CaCO_3$ is noticed by most people.

Another important characteristic of water is its *alkalinity,* which is a measure of the water's capability to absorb hydrogen ions without significant pH change. That is, alkalinity is a measure of the acid buffering capacity of water. In most natural water, the total amount of H^+ that can be neutralized is dominated by the

TABLE 4

Hardness Classifications of Water

	Hardness	
Description	meq/L	mg/L as $CaCO_3$
Soft	<1	<50
Moderately Hard	1–3	50–150
Hard	3–6	150–300
Very Hard	>6	>300

Source: From Tchobanoglous and Schroeder, 1985.

alkalinity provided by the carbonate buffering system. Thus, alkalinity is operationally defined as

$$\text{Alkalinity (mol/L)} = [\text{HCO}_3^-] + 2[\text{CO}_3^{2-}] + [\text{OH}^-] - [\text{H}^+] \tag{31}$$

where concentrations in [ ] denote units of mol/L. Notice that the concentration of carbonate $[\text{CO}_3^{2-}]$ in (31) is multiplied by 2, because each carbonate ion can neutralize two H^+ ions. This assumes the concentrations are being measured in molarity units (mol/L). Often, the concentrations are measured in terms of equivalents, or in mg/L as CaCO_3, in which case the 2 is already accounted for in the conversions, so the concentrations are added directly and

$$\text{Alkalinity (meq/L)} = (\text{HCO}_3^-) + (\text{CO}_3^{2-}) + (\text{OH}^-) - (\text{H}^+) \tag{32}$$

where the quantities in ( ) are concentrations in meq/L or mg/L as CaCO_3.

The following example demonstrates these alkalinity calculations.

EXAMPLE 5 **Calculating Alkalinity**

A sample of water at pH 10 has 32.0 mg/L of CO_3^{2-}. Find the alkalinity as CaCO_3.

Solution The EW of carbonate is given in Table 3 as 30 mg/meq. Converting 32.0 mg/L of carbonate into mg/L as CaCO_3 gives

$$\begin{aligned}(\text{CO}_3^{2-}) &= 32.0\text{ mg/L} \times \frac{1}{30.0\text{ mg/meq}} \times 50.0\text{ mg CaCO}_3/\text{meq}\\ &= 53.3\text{ mg/L as CaCO}_3\end{aligned}$$

Equilibrium between carbonate and bicarbonate is reached very rapidly, so the concentration of bicarbonate can be calculated knowing that concentrations in equilibrium expressions must be expressed in mol/L:

$$[\text{HCO}_3^-] = \frac{[\text{CO}_3^{2-}]10^{-\text{pH}}}{K_2} = \frac{\left(\dfrac{32\text{ mg/L}}{60\text{ mg/mmol}}\right)10^{-10}}{4.68 \times 10^{-11}} = 1.140\text{ mmol/L} \tag{33}$$

Because the ionic charge of bicarbonate is 1, 1.14 mmol/L HCO_3^- equals 1.14 meq/L HCO_3^-. Therefore, the concentration of bicarbonate in mg/L as CaCO_3 is

$$(\text{HCO}_3^-) = \left(\frac{1.140\text{ meq}}{\text{L}}\right)\left(\frac{50\text{ mg CaCO}_3}{\text{meq}}\right) = 57.0\text{ mg/L as CaCO}_3$$

The pH is 10, so $[\text{H}^+] = 1 \times 10^{-10}$ mol/L, and its EW is 1 mg/meq, so it is

$$\begin{aligned}(\text{H}^+) &= 1 \times 10^{-10}\text{ mol/L} \times 1\text{ g/mol} \times \frac{10^3\text{ mg/g}}{1\text{ mg/meq}} \times 50.0\text{ mg CaCO}_3/\text{meq}\\ &= 5.0 \times 10^{-6}\text{ mg/L as CaCO}_3\end{aligned}$$

Because $[H^+][OH^-] = 1 \times 10^{-14}$, then $[OH^-] = 1 \times 10^{-4}$ mol/L. The hydroxyl concentration is therefore

$$(OH^-) = 1 \times 10^{-4}\ \text{mol/L} \times 17.0\ \text{g/mol} \times \frac{10^3\ \text{mg/g}}{17.0\ \text{mg/meq}} \times 50.0\ \text{mg } CaCO_3/\text{meq}$$

$$= 5.0\ \text{mg/L as } CaCO_3$$

Total alkalinity, then, is just $(CO_3^{2-}) + (HCO_3^-) + (OH^-) - (H^+)$

$$\text{Alkalinity} = 53.3 + 57.0 + 5.0 - 5.0 \times 10^{-6} = 115.3\ \text{mg/L as } CaCO_3$$

For nearly neutral water (pH between 6 and 8.5), the concentrations of (H^+) and (OH^-) are very small. In this same pH range, the bicarbonate/carbonate equilibrium strongly favors bicarbonate

$$\text{Alkalinity (meq/L)} = (HCO_3^-) \quad \text{for } 6 \le \text{pH} \le 8.5 \tag{34}$$

Understanding the nature of both hardness and alkalinity is important to be able to understand water softening. In fact, how the components of hardness are defined is a combination of the concepts of hardness and alkalinity. As noted earlier, the *total hardness* (TH) is the sum of individual hardness components

$$\text{Total Hardness} = Ca^{2+} + Mg^{2+} \tag{35}$$

where it has been assumed in (35) that calcium and magnesium are the only two multivalent cations with appreciable concentrations. To understand water softening, it also is useful to separate total hardness, into two components: *carbonate hardness* (CH), associated with the anions HCO_3^- and CO_3^{2-}, and *noncarbonate hardness* (NCH) associated with other anions. If the concentration of HCO_3^- plus CO_3^{2-} equals or exceeds the total hardness (in essence, alkalinity exceeds TH), then CH = TH and NCH = 0. On the other hand, if alkalinity is less than TH, then CH = alkalinity and there is also some NCH which equals TH − CH. CH is especially important since it leads to scaling, as the following reaction suggests:

$$Ca^{2+} + \underset{\text{bicarbonate}}{2HCO_3^-} \leftrightarrow \underset{\substack{\text{calcium}\\\text{carbonate}}}{CaCO_3} + CO_2 + H_2O \tag{36}$$

CH is sometimes referred to as "temporary hardness" because it can be removed by simply heating the water. The carbonate and noncarbonate hardness components can be further subdivided into the calcium and magnesium components of each. These subdivisions of hardness are illustrated in the following example and are important in understanding water softening in the next section.

EXAMPLE 6 Chemical Analysis of a Sample of Water

The analysis of a sample of water with pH 7.5 has produced the following concentrations (mg/L):

Cations		Anions	
Ca^{2+}	80	Cl^-	100
Mg^{2+}	30	SO_4^{2-}	201
Na^+	72	HCO_3^-	165
K^+	6		

Find the total hardness (TH), the carbonate hardness (CH), the noncarbonate hardness (NCH), and the alkalinity, all expressed as $CaCO_3$. Find the total dissolved solids (TDS) in mg/L.

Solution It is helpful to set up a table in which each of the concentrations can be expressed in terms of $CaCO_3$ (or alternatively, meq). For Ca^{2+}, we get

$$\begin{aligned}(Ca^{2+}) &= 80 \text{ mg/L} \times \frac{1}{20.0 \text{ mg/meq}} \times 50.0 \text{ mg } CaCO_3\text{/meq} \\ &= 200 \text{ mg/L as } CaCO_3\end{aligned}$$

Other rows of the table are found in a similar way. Equivalent weights are from Table 3.

Ion	mg/L	mg/meq	mg/L as $CaCO_3$
Ca^{2+}	80	20.0	200.0
Mg^{2+}	30	12.2	123.0
Na^+	72	23.0	156.5
K^+	6	39.1	7.7
Cl^-	100	35.5	140.8
SO_4^{2-}	201	48.0	209.4
HCO_3^-	165	61.0	135.2

As a first check on the accuracy of the chemical analysis, we can compare the sum of the concentrations of cations and anions as $CaCO_3$ (or as meq/L) to see if they are nearly equal.

$$\Sigma \text{ cations} = 200.0 + 123.0 + 156.5 + 7.7 = 487.2 \text{ mg/L as } CaCO_3$$

$$\Sigma \text{ anions} = 140.8 + 209.4 + 135.2 = 485.4 \text{ mg/L as } CaCO_3$$

This is quite close. The difference would probably be associated with small concentrations of other ions as well as measurement error. Figure 9 depicts the ionic composition of the water.

a. The total hardness is the sum of the multivalent cations, (Ca^{2+}) and (Mg^{2+}):

$$TH = 200.0 + 123.0 = 323.0 \text{ mg/L as } CaCO_3$$

So this is very hard water (Table 4).

b. The carbonate hardness is that portion of total hardness associated with carbonates, which in this case is just bicarbonate HCO_3^-:

$$CH = 135.2 \text{ mg/L as } CaCO_3$$

In turn, the calcium carbonate hardness (Ca-CH), is the portion of the calcium associated with CH. In this case, Ca-CH is 135.2 mg/L as $CaCO_3$. This may be most easily seen in the bar graph of the water's composition and components shown in Figure 9.

c. The NCH is the difference between the TH and the CH:

$$NCH = TH - CH = 323.0 - 135.2 = 187.8 \text{ mg/L as } CaCO_3$$

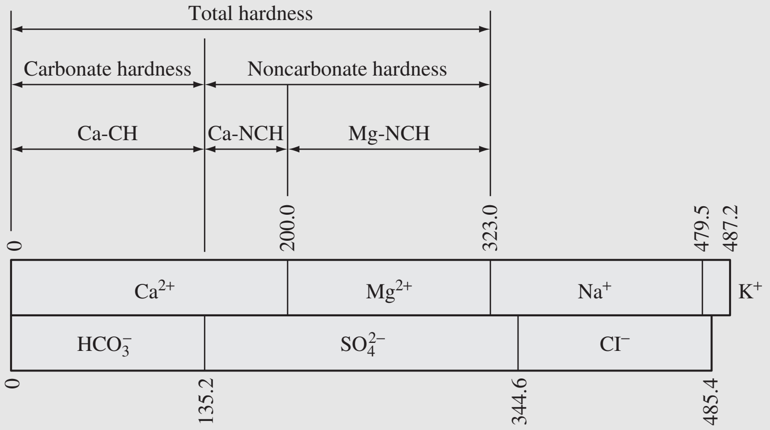


FIGURE 9 Bar graph of ionic constituents of the water in Example 6. This water contains calcium carbonate hardness (Ca-CH), calcium noncarbonate hardness (Ca-NCH), and magnesium noncarbonate hardness (Mg-NCH), but no magnesium carbonate hardness (Mg-CH). Concentrations are given as mg/L as $CaCO_3$.

The calcium noncarbonate hardness (Ca-NCH) is the portion of the calcium not associated with CH:

$$\text{Ca-NCH} = \text{Ca} - \text{Ca-CH} = 200.0 - 135.2 = 64.8 \text{ mg/L as } CaCO_3$$

And likewise, the magnesium noncarbonate hardness (Mg-NCH) is

$$\begin{aligned}\text{Mg-NCH} &= \text{TH} - \text{CH} - \text{Ca-NCH} = 323 - 135.2 - 64.8\\ &= 123.0 \text{ mg/L as } CaCO_3\end{aligned}$$

d. Because the pH is nearly neutral, the concentrations of (H^+) and (OH^-) are negligible. Thus, the alkalinity is given by (34) as just the bicarbonate:

$$\text{Alkalinity} = (HCO_3^-) = 135.2 \text{ mg/L as } CaCO_3$$

e. The total dissolved solids (TDS) is simply the sum of the cation and anion concentrations expressed in mg/L:

$$\text{TDS} = 80 + 30 + 72 + 6 + 100 + 201 + 165 = 654 \text{ mg/L}$$

It is usually very helpful to display the ionic constituents of water using a bar graph such as that shown in Figure 9.

Softening

Hard water causes scaling of pipes and makes laundering more difficult, so many water treatment plants provide water softening. Surface waters seldom have hardness levels above 200 mg/L as $CaCO_3$, so softening is not usually part of the treatment process. For groundwater, however, where hardness levels are sometimes over 1,000 mg/L, it is quite common. There are two popular approaches to softening water: the *lime-soda ash process* and the *ion-exchange process*. Either may be used in a central treatment plant prior to distribution, but individual home units only use the ion-exchange process.

In the lime-soda ash process, *hydrated (slaked) lime* ($Ca(OH)_2$) is added to the water, raising the pH to about 10.3. At this pH, calcium carbonate ($CaCO_3$) and magnesium hydroxide ($Mg(OH)_2$) are relatively insoluble, and most of the calcium and magnesium ions precipitate out as these solids. Because of the strong dependence of $CaCO_3$ and $Mg(OH)_2$ solubility on pH, lime-soda ash softening is able to achieve the seemingly odd feat of reducing the calcium ion concentration in water by adding calcium (as lime) to the water. The key is increasing the pH to above 10, so there is a high enough concentration of hydroxide ions (OH^-) to promote $Mg(OH)_2$ precipitation and to provide sufficient carbonate, from a combination of the water's alkalinity and added soda ash (Na_2CO_3), to promote $CaCO_3$ precipitation. Five reactions describe how lime and soda ash addition softens water. The first reaction raises the pH by overcoming the primary base buffering component, dissolved carbon dioxide ($CO_{2(aq)}$). $CO_{2(aq)}$ can equally well be denoted as carbonic acid. The other four reactions describe how the various components of hardness— Ca-CH, Mg-CH, Ca-NCH, and Mg-NCH—are removed.

The five softening reactions are the following:

1. Removal of $CO_{2(aq)}$:

$$CO_{2(aq)} + Ca(OH)_2 \leftrightarrow CaCO_{3(s)} + H_2O \tag{37}$$

2. Removal of calcium carbonate hardness (Ca-CH):

$$Ca^{2+} + 2HCO_3^- + Ca(OH)_2 \leftrightarrow 2CaCO_{3(s)} + 2H_2O \tag{38}$$

3. Removal of magnesium carbonate hardness (Mg-CH):

$$Mg^{2+} + 2HCO_3^- + 2Ca(OH)_2 \leftrightarrow 2CaCO_{3(s)} + Mg(OH)_{2(s)} + 2H_2O \tag{39}$$

4. Removal of calcium noncarbonate hardness (Ca-NCH):

$$Ca^{2+} + Na_2CO_3 \leftrightarrow CaCO_{3(s)} + 2Na^+ \tag{40}$$

5. Removal of magnesium noncarbonate hardness (Mg-NCH):

$$Mg^{2+} + Ca(OH)_2 + Na_2CO_3 \leftrightarrow CaCO_{3(s)} + Mg(OH)_{2(s)} + 2Na^+ \tag{41}$$

It has been found in practice that a modest excess of lime has to be added beyond that accounted for in equations (37–41) to achieve the needed pH and good softening. This excess varies between 20 and 40 mg/L as $CaCO_3$ (0.4– 0.8 meq/L) depending on the magnesium removal required, but for understanding the basics of softening, we can assume the excess lime requirement to always be 20 mg/L as $CaCO_3$ (0.4 meq/L). Practically, precipitative softening can only reduce the hardness of water down to about 40 mg/L as $CaCO_3$ (0.8 meq/L).

EXAMPLE 7 Lime-Soda Ash Softening

A utility treats 12 MGD of the water in Example 6 to its maximum softness before distribution.

a. How much lime and soda ash (in mg/L as $CaCO_3$) must the utility add?
b. What mass of softening sludge will the plant generate daily?

Solution

a. To solve this problem, the first step is to calculate the amount of $CO_{2(aq)}$ that must be neutralized. The water pH is 7.5, and its bicarbonate concentration was given in Example 6 as 165 mg/L (or 0.002705 M). Remembering to use molar units, we can calculate

$$[CO_{2(aq)}] = \frac{[HCO_3^-][H^+]}{K_1} = \frac{(0.002705\ M)(10^{-7.5}\ M)}{4.47 \times 10^{-7}\ M} = 1.91 \times 10^{-4}\ M$$

which converts to a $CO_{2(aq)}$ concentration of 19.14 mg/L as $CaCO_3$.

The second step is to draw a bar graph for the system like the one in Figure 9, but abbreviated to only show the hardness components of interest. Figure 10 shows the graph.

Based on the values in Figure 10 (and also calculated in Example 6) and the carbonic acid concentration calculated previously, we set up Table 5 for the components removed in the five water softening equations (37–41) and fill in the first two columns of the table.

The rest of the table can now be completed. This is done row-by-row by referring to the stoichiometry of the equation for that row. For instance in the first row, (37) indicates 1 mg/L $CaCO_3$ of lime is required for each mg/L $CaCO_3$ of $CO_{2(aq)}$ removed and, in turn, this produces 1 mg/L of $CaCO_{3(s)}$. You can stoichiometrically equate concentrations of different compounds such as $CO_{2(aq)}$, lime, and calcium carbonate solid, because they have all been converted to the common basis of mg/L as $CaCO_{3(s)}$. For (37), no soda ash is needed or $Mg(OH)_{2(s)}$ produced. Likewise, to complete the Ca-CH row, (38) shows that to remove the 135 mg/L $CaCO_3$ of Ca-CH, an equal concentration of lime must be added, but the reaction will produce

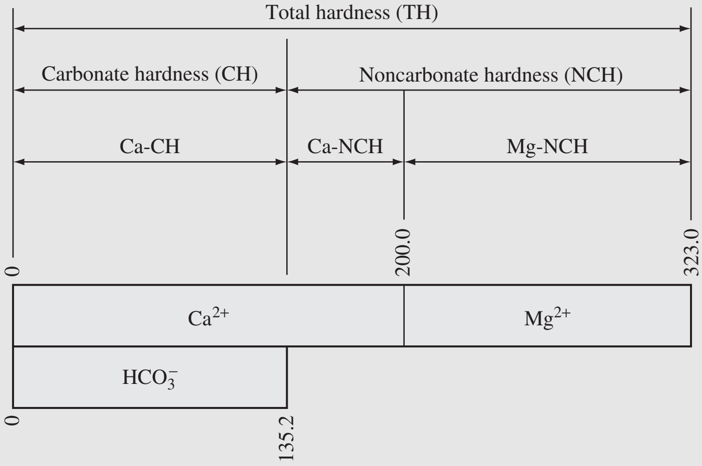


FIGURE 10 Bar graph of the hardness components in the water of Example 6. This is simply an abbreviated form of Figure 9 but is all that is needed for solving lime-soda ash softening problems.

TABLE 5

Components, Lime and Soda Ash Dosage, and Solids Generated in Softening the Water in Example 6

Component	Concentration (mg/L $CaCO_3$)	Lime (mg/L $CaCO_3$)	Soda Ash (mg/L $CaCO_3$)	$CaCO_{3(s)}$ (mg/L $CaCO_3$)	$Mg(OH)_{2(s)}$ (mg/L $CaCO_3$)
$CO_{2(aq)}$ (37)	19.1	19.1	0	19.1	0
Ca-CH (38)	135	135	0	270	0
Mg-CH (39)	0	0	0	0	0
Ca-NCH (40)	64.8	0	64.8	64.8	0
Mg-NCH (41)	123	123	123	123	123
Excess		20			
Totals		297	188	477	123

double that concentration of $CaCO_{3(s)}$. After completing the rows for equations (37) through (41), the 20 mg/L $CaCO_3$ of excess lime that is required is added to the lime column, and each column is summed to quantify the total concentrations required or generated. This shows that the treatment plant must add 297 mg/L $CaCO_3$ of slaked lime and 188 mg/L $CaCO_3$ of soda ash to soften the water.

b. The calcium carbonate and magnesium hydroxide solids that are precipitated during softening are called *softening sludges* because, in fact, they are very watery masses rather than solids in the normal sense of the word. A softening sludge will typically have only about 5 percent solids and about 95 percent water. For this example, 477 mg/L $CaCO_3$ of calcium carbonate and 123 mg/L $CaCO_3$ of magnesium hydroxide sludge were produced. Thus, the mass concentration (rather than concentration as $CaCO_3$) of solids would be

$$\text{mass conc.} = \left(\frac{477\text{ mg CaCO}_3}{\text{L}}\right) + \left(\frac{123\text{ mg as CaCO}_3}{\text{L}}\right)\left(\frac{58.3\text{ mg Mg(OH)}_2}{100\text{ mg CaCO}_3}\right)$$
$$= 549\text{ mg/L}$$

and the daily mass of solids produced would be

$$(549\text{ mg/L})(12\text{ MGD})\left(\frac{(0.0438\text{ m}^3\cdot\text{s}^{-1})}{\text{MGD}}\right)\left(\frac{86{,}400\text{ s}}{\text{day}}\right)\left(\frac{10^3\text{ L}}{\text{m}^3}\right)\left(\frac{\text{kg}}{10^6\text{ mg}}\right)$$
$$= 24{,}900\text{ kg}$$

This 24,900 kg (~25 metric tons) is only the solids mass and, based on 5 percent solids in sludge, it will be accompanied by about 500 metric tons of water. Almost always the sludge will be further processed and thickened to decrease its water content before disposal. However, even after thickening, the sludge will be from 50 to 85 percent water. Obviously, sludge handling and disposal is a major concern in lime-soda ash softening.

After softening, the water is at a high pH and too corrosive to distribute without adjusting the pH back to near-neutral. Usually, this is done by *recarbonation,* where $CO_{2(g)}$ is bubbled into the water. This not only decreases the pH but also adds carbonate to the water, which may cause some residual calcium hardness to precipitate as calcium carbonate. Precipitative softening is actually a sequence of several of the unit operations discussed earlier. The lime and soda ash are added and mixed in a rapid mix tank, the calcium carbonate and magnesium hydroxide solids form in what is basically a slowly agitated flocculation basin, the solids are removed in a sedimentation basin, and, after recarbonation, the water is filtered to remove residual turbidity and particles that form during recarbonation.

The alternative to softening by precipitation uses an ion-exchange process. The exchange process involves replacement of the unwanted ions in solution with ions of a different species that are attached to an insoluble resin. The ion-exchange process can be used to remove nitrate ions, metal ions, and other contaminant ions as well as the ions responsible for hardness. Laboratory quality deionized water is obtained in this way, but the processing required to obtain such high-quality water is too expensive to use in municipal treatment plants.

In the ion-exchange process, hard water flows through a column containing solid resin beads. The resins were originally made of naturally occurring clays called *zeolites* but are now nearly always synthetically produced. In an ion-exchange unit, the resin removes Ca^{2+} and Mg^{2+} ions from the water and replaces them with sodium ions, which form soluble salts. Using $Ca(HCO_3)_2$ as an example, the reaction can be represented as

$$Ca(HCO_3)_2 + Na_2R \leftrightarrow CaR + 2NaHCO_3 \tag{42}$$

where R represents the solid ion-exchange resin. The alkalinity (HCO_3^-) remains unchanged. The sodium salts that are formed do not cause hardness, but the dissolved sodium ions remain in the treated water and may be harmful to individuals with heart problems.

The hardness removal is essentially 100 percent effective as long as the ion-exchange resin has sodium remaining. When the sodium is depleted, the ion-exchange bed must be regenerated by removing it from service and backwashing it with a solution of concentrated NaCl, forming new Na_2R. The wastewater produced during regeneration must be properly disposed of because it contains a high concentration of chloride, sodium, and other cations. The regeneration reaction involving Ca^{2+} can be represented as

$$CaR + 2NaCl \leftrightarrow Na_2R + CaCl_2 \tag{43}$$

The ion-exchange process can be used in wastewater treatment as well as water treatment. In such cases, the process can enable recovery of valuable chemicals for reuse or harmful ones for disposal. It is, for example, often used to recover chromic acid from metal finishing waste for reuse in chrome-plating baths. It is even used for the removal of radioactivity.

Membrane Processes

Although membranes have only been employed in water and wastewater treatment for the past 50 years and only to any large extent in the past 25 years, they are now

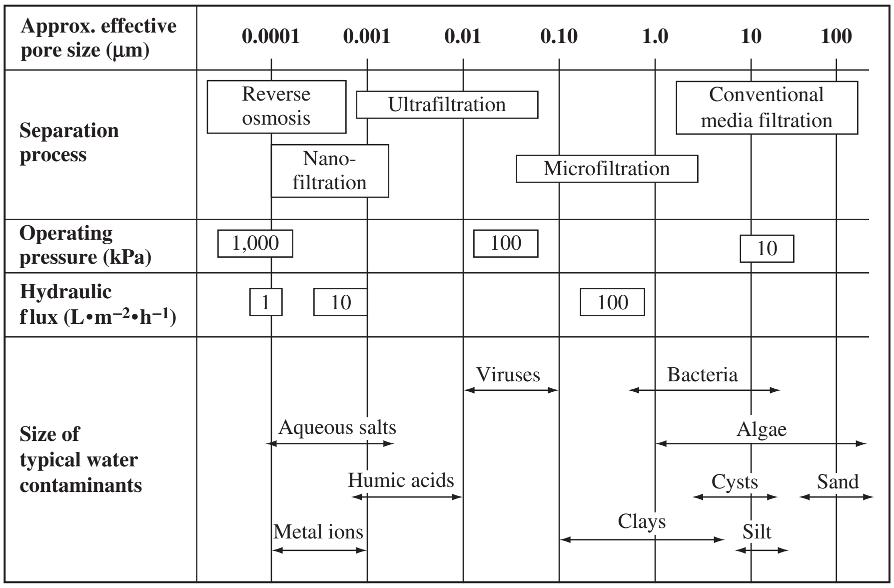


FIGURE 11 Size ranges and types of contaminants removed by membrane processes. (*Source:* Adapted from Taylor and Wiesner, 1999.)

being employed in one form or another to address nearly all contaminants in water, including pathogens, hardness and dissolved solids, organics and disinfectant byproduct precursors, metals, and suspended particles. Generically, all membranes can be considered *semipermeable* physical barriers because they allow the passage of water and, to a greater or lesser degree, some water constituents, while stopping or severely restricting the permeation of other water constituents. The purified water is called the *permeate,* and the influent water for treatment is called the *feed*. All membrane processes generate an aqueous wastestream, called the *concentrate* (or *retentate*), which contains the rejected contaminants. Traditionally, membranes have been classified as *microfiltration, ultrafiltration, nanofiltration,* and *reverse osmosis,* based roughly on the minimum size of the contaminant that is excluded from passing through the particular membrane (Figure 11). As the size of the contaminant excluded becomes smaller, the membrane is said to become "tighter."

As the variety of membranes has increased, the traditional classifications have become blurred, and now, for instance, one manufacturer may list a loose ultrafiltration membrane for the same removal of viruses as another manufacturer's tight microfiltration membrane. A more intuitive classification is based on whether the membrane is designed to reject particles, including pathogens, or dissolved substances, including salts and hardness. Particulate separation processes include microfiltration and ultrafiltration, whereas solute (dissolved substance) separation processes include nanofiltration and reverse osmosis. Most particulate separation membranes are constructed as hollow fibers with thousands of fibers in a single pressure vessel. The fibers are typically between 0.5 and 2.0 mm outside diameter. The water may be forced through the fiber wall from the feed on the inside to the permeate space on the outside (an "inside-out" flow) or from feed water on the outside of the fiber to a permeate channel in the center of the fiber (an "outside-in" flow). The particles that accumulate on the feed side of the membrane must be periodically cleaned away by backwashing the module. In contrast, most solute separation membranes are configured as layered sheets of membranes, which are wrapped in a spiral wound configuration. The channels between the sheets alternate between high pressure feed, and low pressure permeate channels, and the water is purified as it is forced through the membrane from the feed to the permeate channel.

Both particle and solute separation membrane performance is quantified by two primary criteria: contaminant *rejection* and water *recovery*. Rejection, R, quantifies the fraction of the target contaminant that is retained in the concentrate as

$$R = 1 - \frac{C_P}{C_F} \tag{44}$$

where C_P is the concentration of the contaminant in the permeate stream, and C_F is its concentration in the feed stream. Because particle separation membranes may often reduce the contaminant concentration by many orders of magnitude, the rejection may also be quantified as a log value,

$$R_{\log} = -\log(1 - R) = -\log\left(\frac{C_P}{C_F}\right) = \log\left(\frac{C_F}{C_P}\right) \tag{45}$$

It is not uncommon for particle separation membranes used for disinfection to achieve over 5-log rejection of viruses and cysts, whereas solute separation membranes typically achieve no better than 99 percent rejection of salts and hardness.

Because all membrane processes generate an aqueous wastestream, the recovery, or fraction of the feed flow that becomes permeate flow, is often of critical importance. The recovery, r, is defined as

$$r = \frac{Q_P}{Q_F} \tag{46}$$

where Q_P is the permeate flow rate, and Q_F is the feed flow rate.

Recovery for loose, particle separation membranes is typically greater than 95 percent and often as high as 98 percent, whereas solute separation membranes seldom achieve better than 90 percent recovery. The small, under-the-sink reverse osmosis units that are now used in many U.S. homes often have recoveries of 50 percent or less!

In the 1960s, desalination of seawater and low-quality, brackish water was optimistically viewed as an answer to growing shortages of potable fresh water, and advances in reverse osmosis were considered a key to achieving this goal. The alternative to membrane desalination is thermal distillation, which is based on the fact that salts do not evaporate with water. Thus, if water can be boiled and the vapor condensed, the vapor will be pure water. About 40 percent of the desalination capacity in the world is by thermal distillation, although most of the production is concentrated in oil-rich Middle Eastern nations because thermal processes are relatively energy intensive (Cooley et al., 2006). In 2005, in the United States, over 90 percent of the desalination capacity was based on membrane processes (Cooley et al., 2006), and it is expected that membranes will account for an increasing share of the world desalination capacity as energy costs rise faster than membrane production costs. However, by 2005, desalination still produced much less than 1 percent of both the world's and the U.S. total freshwater consumption. This is not to say that desalination, and membrane desalination in particular, will not continue to see widespread growth and play an important role in providing safe potable water in many water short areas; however, it is unlikely that it will provide a large fraction of the world's total freshwater demand in the near future. Cooley et al. (2006) summarize this situation well when they write, "The potential benefits of ocean desalination are great, but the economic, cultural, and environmental costs of wide commercialization remain high."

5 Wastewater Treatment

Municipal wastewater is typically over 99.9 percent water. The characteristics of the remaining portion vary somewhat from city to city, with the variation depending primarily on inputs from industrial facilities that mix with the somewhat predictable composition of residential flows. Given the almost limitless combinations of chemicals found in wastewater, it is too difficult to list them individually. Instead, they are often described by a few general categories, as has been done in Table 6.

In Table 6, a distinction is made between *total dissolved solids* (TDS) and *suspended solids* (SS). The sum of the two is *total solids* (TS). The suspended solids portion is, by definition, the portion of total solids that can be removed by a membrane filter (having a pore size of about 1.7 μm). The remainder (TDS) that cannot be filtered includes dissolved solids, colloidal solids, and very small suspended particles.

Wastewater treatment plants are usually designated as providing *primary, secondary,* or *advanced* treatment, depending on the degree of purification. Primary treatment plants utilize physical processes, such as screening, skimming, and sedimentation, to remove pollutants that are settleable, floatable, or too large to pass through simple screening devices. Primary treatment typically removes about 35 percent of the BOD and 60 percent of the suspended solids. In the early 1970s, the sewage of about 50 million people in the United States was receiving no better treatment than this. While the most visibly objectionable substances are removed in primary treatment, and some degree of safety is normally provided by disinfection after primary treatment, the effluent still has enough BOD to cause oxygen depletion problems and enough nutrients, such as nitrogen and phosphorus, to accelerate eutrophication.

The Clean Water Act (CWA), in essence, requires at least secondary treatment for all publicly owned treatment works (POTWs) by stipulating that such facilities provide at least 85 percent BOD removal (with possible case-by-case variances that allow lower percentages for marine discharges). This translates into an effluent

TABLE 6

Composition of Untreated Domestic Wastewater

Constituent	Abbreviation	Concentration (mg/L)
5-day biochemical oxygen demand	BOD_5	100–350
Chemical oxygen demand	COD	250–1,000
Total dissolved solids	TDS	250–1,000
Suspended solids	SS	100–400
Total Kjeldahl nitrogen	TKN	20–80
Total phosphorus (as P)	TP	5–20

requirement of 30 mg/L for both 5-day BOD and suspended solids (monthly average). In secondary treatment plants, the physical processes that make up primary treatment are augmented with processes that involve the microbial oxidation of wastes. Such biological treatment mimics nature by utilizing microorganisms to oxidize the organics, with the advantage being that the oxidation can be done under controlled conditions in the treatment plant itself, rather than in the receiving body of water. When properly designed and operated, secondary treatment plants remove about 90 percent of the BOD and 90 percent of the suspended solids.

Although the main purpose of primary treatment (in addition to disinfecting the wastes) is to remove objectionable solids, and the principal goal of secondary treatment is to remove most of the BOD, neither is effective at removing nutrients, dissolved material (metals, salts), or biologically resistant (refractory) substances. For example, typically no more than half of the nitrogen and one-third of the phosphorus are removed during secondary treatment. This means the effluent can still be a major contributor to eutrophication problems. In circumstances where either the raw sewage has particular pollutants of concern or the receiving body of water is especially sensitive, so-called advanced treatment (previously called *tertiary* treatment) may be required. In many parts of the United States, advanced nutrient removal is now required in POTWs. Advanced treatment processes are varied and specialized, depending on the nature of the pollutants that must be removed. In most circumstances, advanced treatment follows primary and secondary treatment, although in some cases, especially in the treatment of industrial waste, it may completely replace those conventional processes.

An example flow diagram for a wastewater treatment plant that provides primary and secondary treatment is illustrated in Figure 12.

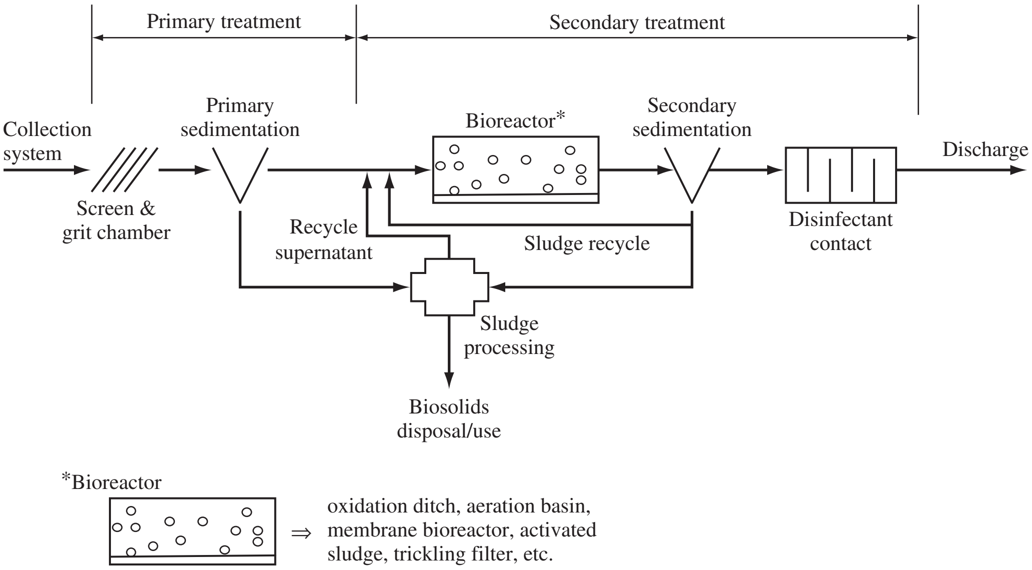


FIGURE 12 Schematic of a typical wastewater treatment facility providing primary and secondary treatment. The bioreactor may involve any number of different techniques for BOD oxidation, and the sludge processing may include several different types of biomass reduction and water removal (thickening) operations.

Primary Treatment

As suggested in Figure 12, primary treatment begins with simple screening. Screening removes large floating objects such as rags, sticks, old shoes, and whatever else is present that might otherwise damage the pumps or clog small pipes. Screens vary but typically consist of parallel steel bars spaced anywhere from 2 cm to 7 cm apart, perhaps followed by a wire mesh screen with smaller openings. One way to avoid the problem of disposal of materials collected on screens is to use a device called a comminuter, which grinds those coarse materials into small enough pieces that they can be left right in the wastewater flow.

After screening, the wastewater passes into a grit chamber where it is held for up to a few minutes. The *detention time* (the tank volume divided by the flow rate) is chosen to be long enough to allow sand, grit, and other heavy material to settle out, but is too short to allow lighter, organic materials to settle. Detention times are typically 20 to 30 s (seconds). By collecting only these heavier materials, the disposal problem is simplified because those materials are usually nonoffensive and can be easily disposed of in a landfill.

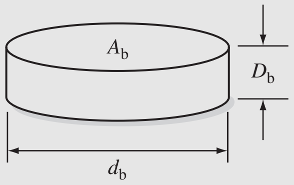
From the grit chamber, the sewage passes to a *primary settling tank* (also known as a *sedimentation basin* or a *primary clarifier*) where the flow's speed is reduced sufficiently to allow most of the suspended solids to settle out by gravity. Detention times of approximately 1.5 to 3 hours are typical, resulting in the removal of from 50 to 65 percent of the suspended solids and 25 to 40 percent of the BOD. Primary settling tanks are either round or rectangular, and their behavior is similar to that of the clarifiers described earlier in this chapter for water treatment facilities. The solids that settle, called *primary sludge* or raw sludge, are removed for further processing, as is the grease and scum that floats to the top of the tank. If this is just a primary treatment plant, the effluent at this point is disinfected (typically by chlorine or UV light) to destroy bacteria and help control odors. Then it is released.

As described for water treatment, sizing of primary settling tanks (clarifiers) is based primarily on the overflow rate and the detention time. These parameters have been found to be the key determinants of successful performance. High-rate clarifiers have been used in Europe for several decades and are now starting to be used in U.S. POTWs, as the cost and availability of land becomes more restrictive (Tchobanoglous et al., 2003). By addition of various coagulants and ballast (additives such as sand that make the settling particles more dense) and specialized settling equipment within the clarifier, wastewater particles can be made to settle more rapidly, so the same particle removal efficiency can be achieved at a higher overflow rate. The range of overflow rates, ν_o, for high-rate clarification is $1{,}000–3{,}000\ \text{m}^3 \cdot \text{m}^{-2} \cdot \text{d}^{-1}$, compared to $30–100\ \text{m}^3 \cdot \text{m}^{-2} \cdot \text{d}^{-1}$ for conventional clarification. Example 8 illustrates how significantly high-rate clarification can decrease the required clarifier size.

EXAMPLE 8 Sizing a Primary Clarifier

A town of 30,000 sends 0.5 m^3 of wastewater per person per day to the wastewater treatment plant. A conventional circular primary clarifier would be designed to have an average detention time of 2.0 hours and an average overflow rate of 40 m/d, whereas a high-rate clarifier would have an average overflow rate of 1,500 m/d.

a. What would be the dimensions of the conventional clarifier?
b. How much area would be saved by using a high-rate clarifier?



Solution

a. At 0.5 m^3/person/d, the daily flow for 30,000 people would be 15,000 m^3/d. The overflow rate, ν_o, is defined by (7), so the clarifier surface area is

$$A_b = \frac{Q}{\nu_o} = \frac{15{,}000\ \text{m}^3 \cdot \text{d}^{-1}}{40\ \text{m}^3 \cdot \text{m}^{-2} \cdot \text{d}^{-1}} = 375\ \text{m}^2$$

The tank diameter is therefore

$$d_b = \sqrt{\frac{4\ A_b}{\pi}} = \sqrt{\frac{4 \times 375\ \text{m}^2}{\pi}} = 22\ \text{m}$$

The detention time (9), θ, is the ratio of volume to flow rate, and the volume is area times depth, so the clarifier depth is

$$D_b = \frac{Q\theta}{A_b} = \frac{15{,}000\ \text{m}^3/\text{d} \times 2.0\ \text{h}}{375\ \text{m}^2 \times 24\ \text{h/d}} = 3.3\ \text{m}$$

As it turns out, this depth is about the minimum that would be considered good design practice.

b. Repeating (7) for the high-rate clarifier case,

$$A_b = \frac{Q}{\nu_o} = \frac{15{,}000\ \text{m}^3 \cdot \text{d}^{-1}}{1{,}500\ \text{m}^3 \cdot \text{m}^{-2} \cdot \text{d}^{-1}} = 10\ \text{m}^2$$

Thus, the high-rate clarifier would use about 3% of the land area as the conventional clarifier. In reality, the area savings would not be as large as this, because some additional area would be needed for the coagulant/ballast dosing equipment and for a possible flocculation tank.

Maximum flow rates over weirs per lineal meter of weir, minimum depths of tanks, constraints related to hour-to-hour flow variations, as well as seasonal variations, and so forth, complicate the design of an actual clarifier, so Example 8 is merely a first cut at specifying the tankage needed.

Secondary (Biological) Treatment

The main purpose of secondary treatment is to provide additional BOD and suspended solids removal, beyond what is achievable by simple sedimentation. There are two commonly used approaches, both of which take advantage of the ability of microorganisms to convert organic wastes into stabilized, low-energy compounds. In the first approach, *suspended growth treatment,* the microorganisms are suspended in and move with the water. In contrast, the microorganisms in *attached growth treatment* processes are fixed on a stationary surface, and the water flows past the microorganisms, much like the water passes by the media in a drinking water treatment filter. There are many different configurations and variations of each approach, and the most common type of each will be discussed later. However, to understand the functioning of secondary treatment processes, the basics of microbial kinetics must first be understood.

Microbial Kinetics

The rate of consumption of BOD is the primary rate of interest in secondary treatment. The organic matter that microorganisms consume is generically called *substrate,* and it is normally measured as mg/L of BOD. Kinetics are also important in terms of the mass of organisms, which may increase or decrease with time, depending on the growth conditions and availability of substrate. Rather than try to quantify the number concentration of the very diverse community of microbes in a bioreactor, they are normally estimated using a surrogate measure such as mg/L of *volatile suspended solids* (VSS) or mg/L *volatile solids* (VS). The term *volatile* arises from the analytic technique used, which measures the difference in mass of solids in a known volume of water between that weighed after drying at 105°C and after burning off at 500°C. Although VSS and VS contain more than just the viable mass of microbes, they are simple to measure and have been found to provide a reasonable first approximation of the microorganism concentration. Thus, the kinetic analysis of a bioreactor involves, at a minimum, three mass balances, which quantify the flow and changes in water, substrate, and microbe mass. However, these mass balances are interrelated. The rate of substrate entering and leaving the reactor is affected by the water rate entering and leaving, and the rate of microbial growth is affected by changes in the mass of substrate available.

During the exponential phase of microbial growth, the increase in the *microbial mass growth rate,* r_g, can be modeled by the first-order rate expression

$$r_g = \frac{dX}{dt} = \mu X \tag{47}$$

where X is the concentration of microorganisms (typically mg VSS/L), and μ is the specific biomass growth rate constant (time^{-1}). As you would expect, the growth rate constant is dependent on the availability of substrate. Monod (1949) reasoned that when the substrate concentration is zero, then the growth rate constant should be zero, and when there is substrate in excess, the microbial growth rate should reach a maximum that is determined by the maximum speed at which microbes can reproduce. Figure 13 shows this dependence of the growth rate constant on the substrate concentration, S. The Monod equation expresses it mathematically as

$$\mu = \frac{\mu_m S}{(K_s + S)} \tag{48}$$

where μ_m is the maximum specific growth rate (time^{-1}), S is the substrate concentration (typically, mg BOD$_5$/L), and K_s is the half velocity constant. K_s is defined as the substrate concentration when the growth rate constant is half of its maximum value (μ_m). Based on (47) and (48), the microbial mass growth rate can be expressed:

$$r_g = \frac{\mu_m X S}{(K_s + S)} \tag{49}$$

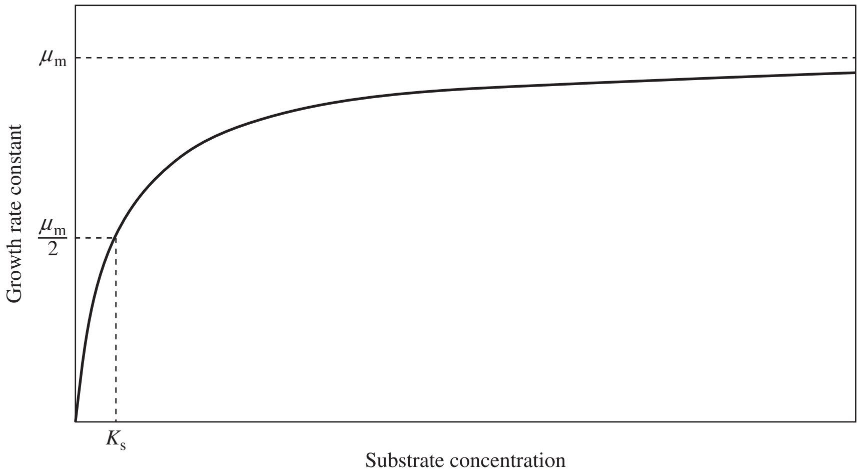


FIGURE 13 The Monod equation empirically describes the dependence of the biomass growth rate constant on the substrate concentration.

Not only does the rate of microbial growth depend on the substrate concentration, but also the amount of microbial mass generated should be proportional to the amount of substrate consumed. In other words, for a particular substrate and set of growth conditions, a certain proportion of the substrate mass consumed should be converted into a predictable amount of new microbial cell mass. The proportionality constant is the *yield coefficient,* Y, and has the units of mg VSS/mg BOD_5. Microbes derive energy for new cell synthesis and normal metabolism by catalyzing oxidation-reduction reactions. The fraction of the substrate that becomes microbial mass depends on the particular biochemical transformation that takes place and is the subject of the study of bioenergetics (McCarty, 1971). For instance, if glucose (a simple sugar substrate) is microbially oxidized by reducing oxygen (O_2) rather than by reducing carbon dioxide (CO_2), about ten times more cell mass is synthesized per gram of glucose consumed based on values from McCarty (1971, 1975). The yield coefficient relates the rate of substrate consumption (r_{su}) under particular conditions to the rate of microbe growth:

$$r_{su} = \frac{dS}{dt} = \frac{-r_g}{Y} \tag{50}$$

The maximum specific growth rate, μ_m, is related by the yield coefficient to the maximum specific substrate utilization constant, k:

$$k = \frac{\mu_m}{Y} \tag{51}$$

Combining the previous three equations, the substrate utilization rate is expressed as

$$r_{su} = \frac{dS}{dt} = \frac{-\mu_m XS}{Y(K_s + S)} = \frac{-kXS}{(K_s + S)} \tag{52}$$

We have made one simplification that cannot be overlooked. In (47) and (49), we assumed that all of the microbes in the population would continue to reproduce.

TABLE 7

Typical Microbial Kinetic Parameters (All are given for 20°C)

Parameter	Typical Units	Range	Default Value
k	mg BOD_5/(mg VSS · d)	2–10	5
μ_m	d^{-1}	1–10	3
K_s	mg BOD_5/L	25–100	60
Y	mg VSS/mg BOD_5	0.2–0.8	0.6
k_d	d^{-1}	0.025–0.075	0.06

Source: Adapted from Metcalf and Eddy, 1991.

This is akin to assuming there is a birth rate but no death rate. The death rate for microbes, r_d, should also be a first-order process but now with a negative reaction term.

$$r_d = \frac{dX}{dt} = -k_d X \tag{53}$$

where k_d is the endogenous decay (death) rate constant (time^{-1}). The net or observed rate of change of microbe concentration, r'_g, is

$$r'_g = r_g + r_d = \frac{\mu_m XS}{(K_s + S)} - k_d X \tag{54}$$

Using (54) and (50) gives a very useful expression:

$$r'_g = -Yr_{su} - k_d X \tag{55}$$

(54) tells us the reaction rate in a mass balance equation on microbe mass is the difference between the rate they are generated (49) and the rate they die (53). Similarly, (55) states that the reaction rate for microbial growth is proportional to the rate that substrate is consumed (r_{su}) minus the rate microbes die (r_d). Table 7 provides ranges and typical values for the microbial kinetic parameters that have been measured in wastewater treatment bioreactors.

EXAMPLE 9 BOD Consumption in Pond

The shallow pond depicted in Figure 14 stays well mixed due to the wind and the steady flow through of a small creek.

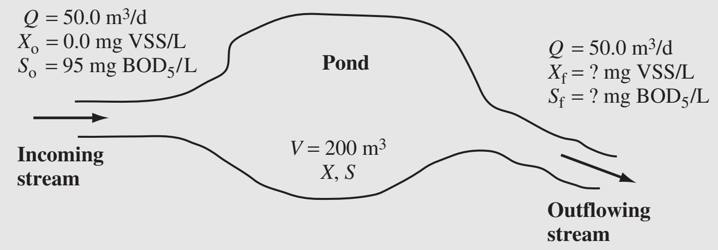


FIGURE 14 A pond in which microbial activity consumes organic matter.

If the microbes in the pond consume the inflowing biodegradable organic matter according to typical kinetics, determine:

a. The BOD_5 leaving the pond
b. The biodegradable organic matter removal efficiency of the pond
c. The concentration of volatile suspended solids leaving the pond

Solution The pond can be modeled as a steady state CSTR because it is well-mixed and, we'll assume, the stream flow has been at a steady flow rate and composition for a long period.

a. First, set up a mass balance on the microbial mass using the pond as the control volume. From the variables in Figure 14, we can write

$$0 = QX_o - QX_f + Vr'_g$$

Substituting in (54) and recognizing that $X_o = 0$, $X_f = X$, and $S_f = S$ for a CSTR leads to

$$0 = -QX + V\frac{\mu_m XS}{(K_s + S)} - Vk_d X \tag{56}$$

Because the hydraulic retention time of the pond is $\theta = V/Q$, (56) simplifies to

$$1 = \theta\frac{\mu_m S}{(K_s + S)} - \theta k_d$$

This can be solved algebraically for the substrate concentration to give

$$S = \frac{K_s(1 + \theta k_d)}{(\theta\mu_m - 1 - \theta k_d)} \tag{57}$$

It is important to note that the effluent substrate concentration, (57), is not a function of the influent substrate concentration and, interestingly, it was derived from a microbial mass balance, rather than a substrate mass balance. Using the default values in Table 7 and $\theta = 200\ \mathrm{m}^3/50\ \mathrm{m}^3/\mathrm{d} = 4.0\ \mathrm{d}$, the outflowing creek's BOD_5 concentration is

$$S = \frac{60\ \dfrac{\mathrm{mg\ BOD_5}}{\mathrm{L}}(1 + 4\ \mathrm{d} \times 0.06\ \mathrm{d}^{-1})}{(4\ \mathrm{d} \times 3\ \mathrm{d}^{-1} - 1 - 4\ \mathrm{d} \times 0.06\ \mathrm{d}^{-1})} = 6.9\ \frac{\mathrm{mg\ BOD_5}}{\mathrm{L}}$$

b. The efficiency of a bioreactor such as the pond is given by

$$\mathrm{Eff} = \frac{(S_o - S_f)}{S_o} \times 100\% = \frac{\left(95 - 6.9\ \dfrac{\mathrm{mg\ BOD_5}}{\mathrm{L}}\right)}{95\ \dfrac{\mathrm{mg\ BOD_5}}{\mathrm{L}}} \times 100\% = 93\% \tag{58}$$

c. To calculate the microbe concentration (as VSS) in the outflowing stream, we construct a steady state, substrate mass balance for the pond as

$$0 = QS_o - QS_f + Vr_{su} \tag{59}$$

Substituting in (52) and solving for the microbe mass concentration, X, yields

$$X = \frac{(K_s + S)(S_o - S)}{\theta k S} \tag{60}$$

Again, using Table 7 default values,

$$X = \frac{\left(60 + 6.9\,\frac{\text{mg BOD}_5}{\text{L}}\right)\left(95 - 6.9\,\frac{\text{mg BOD}_5}{\text{L}}\right)}{\left(4\text{ d} \times 5\,\frac{\text{mg BOD}_5}{\text{mg VSS}\cdot\text{d}} \times 6.9\,\frac{\text{mg BOD}_5}{\text{L}}\right)} = 43\,\frac{\text{mg VSS}}{\text{L}}$$

It is worth remembering that the BOD_5 in the outflowing creek is underestimated substantially because much of the microbial mass generated could actually be used as substrate by other predatory microbes and, in fact, cause a significant oxygen demand. The BOD_5 removal efficiency calculated is based only on the decrease in the original quantity of biodegradable organic matter entering the pond and doesn't account for the organic matter synthesized in the pond.

Suspended Growth Treatment

The mathematical expressions just described allow us to quantify the rate that microbes grow and consume BOD. These can now be applied to understanding and designing the various biological reactors that are employed in secondary wastewater treatment. Most modern plants employ some form of suspended growth bioreactor in which the microbes are freely suspended in the water as it is treated. The following section describes the two most widely used suspended growth processes, activated sludge and aerated lagoons, and one recently introduced suspended growth process, membrane bioreactors, which is rapidly gaining acceptance, particularly in densely populated urban areas.

Activated Sludge. The most commonly employed type of suspended growth process is *activated sludge*. There are many variations of the activated sludge process, but all have two things in common: they depend on injection of extra oxygen into the reactor, and they artificially increase the microbial mass in the reactor by separating out and returning back to the activated sludge tank a large portion of the microbial cells in the reactor effluent. Oxygen is injected into the bioreactor by diffusers, somewhat like the bubblers used in home aquariums. The oxygen maintains aerobic (oxygenated) conditions, which facilitates rapid microbial decomposition of organic matter and agitates the mixture of wastewater and microbial cells. This promotes thorough mixing of the substrate with the microbes and maintains the entire mass in suspension. The mixture of aerated water and microbial mass is called the *mixed liquor*.

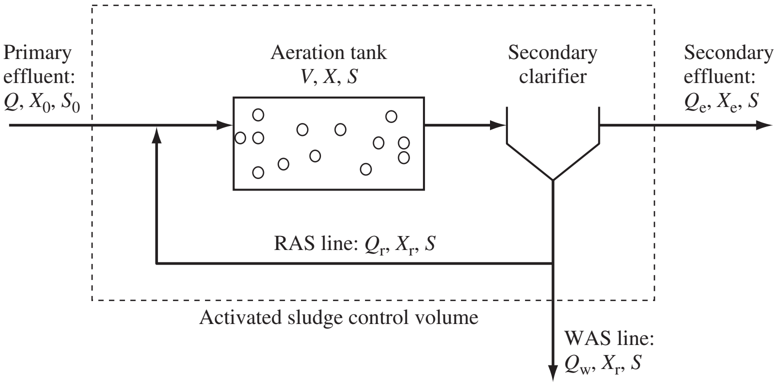


FIGURE 15 Control volume and parameter definitions used for analysis of an activated sludge wastewater treatment process.

As shown in Figures 12 and 15, the activated sludge process is, at least, two unit operations: the activated sludge bioreactor and the secondary clarifier. Because the microbial cells generated in the activated sludge reactor are particles, they can be separated from the water by sedimentation. This is done in the secondary clarifier. Much of the cell mass that is settled out in the clarifier is returned to the bioreactor via the *return activated sludge* (RAS) line (Figure 15). However, to maintain a steady state concentration in the bioreactor, a portion must also be wasted through a *waste activated sludge* (WAS) line to compensate for the new cell mass being synthesized in the activated sludge tank. Because more cell mass than water is recycled, the average cell retention time is longer than the average hydraulic retention time in an activated sludge process. The assumption is normally made that all of the BOD consumption occurs in the bioreactor, so the hydraulic retention time is defined as $\theta = V/Q$ (Figure 15). The average *cell retention time* (which is also known as *solids retention time,* SRT, or *sludge age*), θ_c, is

$$\theta_c = \frac{VX}{Q_e X_e + Q_w X_w} \approx \frac{VX}{Q_w X_w} \tag{61}$$

The simplifying approximation shown in (61) is often made based on the assumption that the secondary clarifier is very effective at settling out VSS, so $X_e \approx 0$. In addition to the sludge age, another important, related design and operational parameter is the food to microbe ratio, *F/M:*

$$F/M = \frac{QS_0}{VX} \tag{62}$$

The cell retention time, θ_c, and the F/M are related and depend largely on the activated sludge wastage rate from the WAS line (Figure 15). If the F/M is too high (meaning also that θ_c is too low) because cells are being wasted too rapidly, then the BOD removal efficiency will be low. In contrast, if cells are wasted too slowly (low F/M, long θ_c), then the microbes will be starved, and the process will be inefficient in terms of energy needed to aerate the mixed liquor and recycle microbes from the secondary clarifier.

Table 8 gives typical values for sludge age and F/M ratio for POTWs treating municipal wastewater.

TABLE 8

Typical Mean Cell Retention Times and Food to Microbe (*F*/*M*) Ratios for Activated Sludge Treatment of Municipal Wastewater

Cell Retention Time (days)	*F*/*M* Ratio (g BOD/g VSS · d)
5	0.3
7	0.5
20	0.1
30	0.05

Source: Adapted from Tchobanoglous et al., 2003.

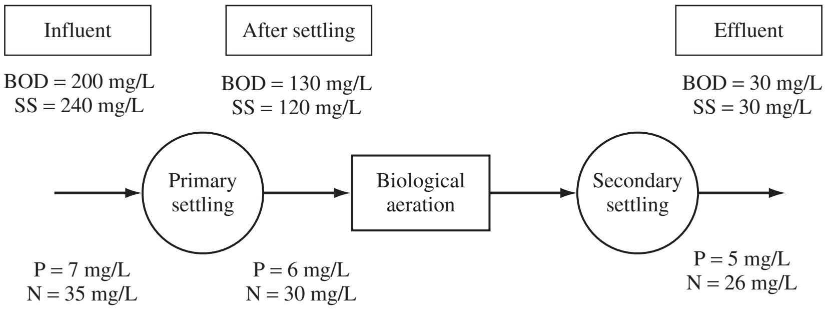


FIGURE 16 Approximate concentrations of BOD_5, suspended solids, total nitrogen, and total phosphorus as wastewater passes through a secondary wastewater treatment plant. (*Source:* Based on Hammer and Hammer, 1996.)

By allowing greater contact between microorganisms and wastewater in a given volume of space, activated sludge tanks can take up considerably less land area for equivalent performance compared to the attached growth processes described later. They are also less expensive to construct than trickling filters, have fewer problems with flies and odors, and can achieve higher rates of BOD removal. They do, however, require more expertise to operate and more energy for pumps and blowers, and hence have higher operating costs. Figure 16 shows representative values of BOD, suspended solids, total nitrogen, and total phosphorus as wastewater passes through primary sedimentation, biological aeration (e.g., activated sludge treatment), and secondary settling.

Membrane Bioreactors. Since the mid-1990s, microfiltration membranes have been employed in suspended growth bioreactors to decrease the size of secondary treatment tankage and improve the suspended solids separation efficiency. These *membrane biological reactors* (MBRs) draw water from the mixed liquor into hollow fiber membranes submerged in the activated sludge aeration tank, thus avoiding the need for a secondary clarifier in the conventional activated sludge process. The microfiltration fibers have a pore size of about 0.2 μm, so they are very effective in producing low TSS effluent. (See the section, "Membrane Processes," earlier in this chapter for more discussion of membrane separation processes.) Because MBRs exclude nearly all microbial cells from the secondary effluent, some of the activated sludge mixed liquor must be continuously bled off of the aeration tank by a waste activated sludge line to maintain the desired mixed liquor concentration. The very efficient removal of bacteria and cysts (but not so much viruses) makes MBRs attractive in systems where wastewater reuse and reclamation is desired. As you might expect, avoiding fouling of the membrane fibers is one of the biggest engineering challenges with MBRs. Although different MBR vendors employ different, often proprietary, processes, commonly fouling control involves a combination of bursts of bubbles injected immediately below the hollow fibers, frequent backwashing by inside-out flushes of clean water, and periodic chemical cleaning of the fibers. In addition to fouling problems, MBRs are more expensive both to build and to operate than most other treatment options, and because they are a new process innovation, there is little data on how long they will last or how frequently the costly membranes may need replacement.

Aerated Lagoons and Oxidation Ponds. *Oxidation ponds* are large, shallow ponds, typically 1 to 2 m deep, where raw or primary treated sewage is decomposed by microorganisms. The conditions are similar to those that prevail in a eutrophic lake. Oxidation ponds can be designed to maintain aerobic conditions throughout, but more often, the decomposition taking place near the surface is aerobic, whereas that near the bottom is anaerobic. Such ponds, having a mix of aerobic and anaerobic conditions, are called *facultative ponds*. In ponds, the oxygen required for aerobic decomposition is derived from surface aeration and algal photosynthesis; deeper ponds, called *lagoons,* are mechanically aerated. A schematic diagram of the reactions taking place in a facultative pond is given in Figure 17.

Oxidation ponds can be designed to provide complete treatment of raw sewage, but they require a good deal of space. These ponds have been used extensively in small communities where land constraints are not critical. The amount of pond surface area required is considerable, with 1 hectare per 240 people (1 acre per 100 people) often being recommended, although in areas with warm climates and

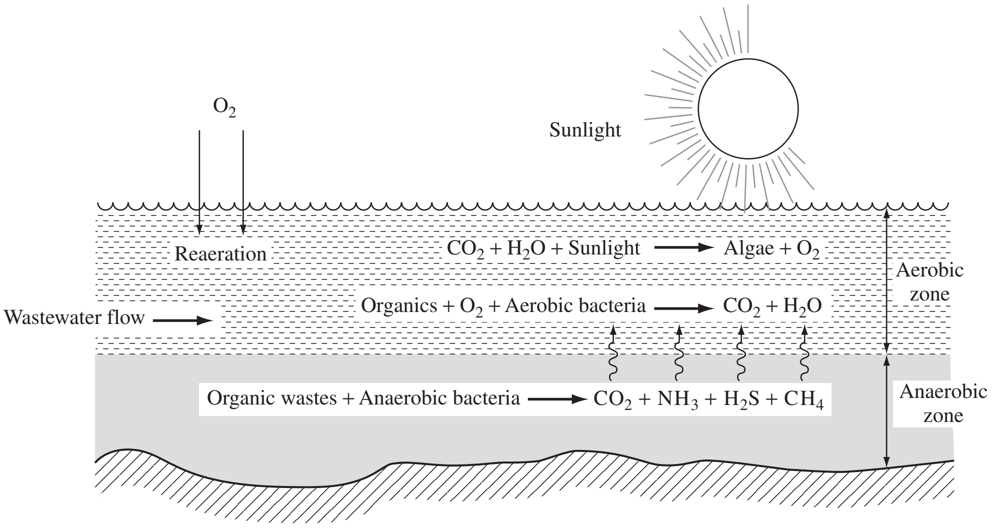


FIGURE 17 Schematic of an oxidation pond.

mild winters, such as in the southwestern United States, about half that area is often used (Viessman and Hammer, 1985).

Ponds are easy to build and manage, they accommodate large fluctuations in flow, and they can provide treatment that approaches that of more sophisticated biological systems but at a much lower cost. The effluent, however, may contain undesirable concentrations of algae, and especially in the winter when less oxygen is liberated by photosynthesis, they may produce unpleasant odors. Used alone, they also have the disadvantage that the effluent may not meet the EPA secondary treatment requirement of 30 mg/L of both BOD_5 and suspended solids. Their simplicity and effectiveness in destroying pathogenic organisms, however, make these ponds especially useful in developing countries.

Oxidation ponds are also used to augment secondary treatment, in which case they are often called *polishing ponds*.

Attached Growth Treatment

Although not as common in modern wastewater treatment plants as suspended growth processes, attached growth processes are still often used both as the sole secondary treatment process and as a pretreatment step before an activated sludge process. Suspended growth processes are also attractive for application in rural and underdeveloped areas where simple-to-construct, low-maintenance, and low-energy secondary treatment is required.

Trickling Filters. *Trickling filters* were first used in 1893, and they have been used successfully ever since, but they do cost more to build, are more temperature sensitive, and remove less BOD than the activated sludge plants that were later developed. A trickling filter consists of a rotating distribution arm that sprays the feed wastewater over a circular bed of plastic packing or other coarse materials (Figure 18). "Fist size" rocks were originally used as packing (and are still used for simple, low cost operations), but plastic media is now preferred as it is easier to work with, allows taller structures, and can be engineered to have much greater surface area for microbial growth per volume of filter. Tall trickling filters filled with plastic media are often called *biotowers*. The spaces between the packing allow air

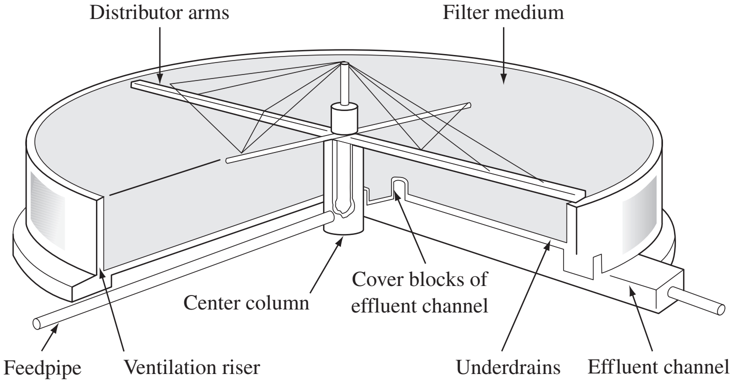


FIGURE 18 Cross-section of a trickling filter.
(*Source:* Hammer and Hammer, 1996.)

to circulate easily so that aerobic conditions can be maintained. Of course, the size of the openings is such that there is no actual filtering taking place, so the name trickling *filter* is somewhat of a misnomer. Instead, the media in the bed is covered by a layer of biological slime that adsorbs and consumes the wastes trickling through the bed. This slime consists mainly of bacteria, but it may also include fungi, algae, protozoa, worms, insect larvae, and snails. The accumulating slime periodically sloughs off the packing surface and is passed on with the water to the secondary settling tank, where it is removed. Not shown in Figure 18 is a provision for returning some of the effluent from the filter back into the incoming flow. Such recycling not only enables more effluent organic removal, but it also provides a way to keep the biological slimes from drying out and then dying during low flow conditions.

Rotating Biological Contactors. Trickling filters (and biotowers) are examples of devices that rely on microorganisms that grow on the surface of rocks, plastic, or other media. A variation on this attached growth idea is provided by the *rotating biological contactor* (RBC). An RBC consists of a series of closely spaced, circular, plastic disks, typically 3.6 m in diameter, that are attached to a rotating horizontal shaft. The bottom 40 percent of each disk is submersed in a tank containing the wastewater to be treated. The biomass film that grows on the surface of the disks moves into and out of the wastewater as the RBC rotates. While the microorganisms are submerged in the wastewater, they adsorb organics; while they are rotated out of the wastewater, they are supplied with needed oxygen. By placing modular RBC units in series, greater treatment efficiencies than with trickling filters can be achieved (Figure 19). These devices have been used in the United States only since 1969, and although early units suffered from assorted mechanical problems, they are now generally accepted. They are easier to operate under varying load conditions than trickling filters because it is easier to keep the solid medium wet at all times.

Hybrid Suspended/Attached Growth Systems

A large number and variety of secondary wastewater treatment systems, combining attributes of both suspended and fixed growth biological reactors, have been developed and employed since the mid-1980s. The simplest employ a trickling filter between the primary clarifier and the activated sludge aeration basin. This can decrease the cost of and time required for removal of the most easily degraded organic matter in the primary effluent. In these systems, the trickling filter effluent is

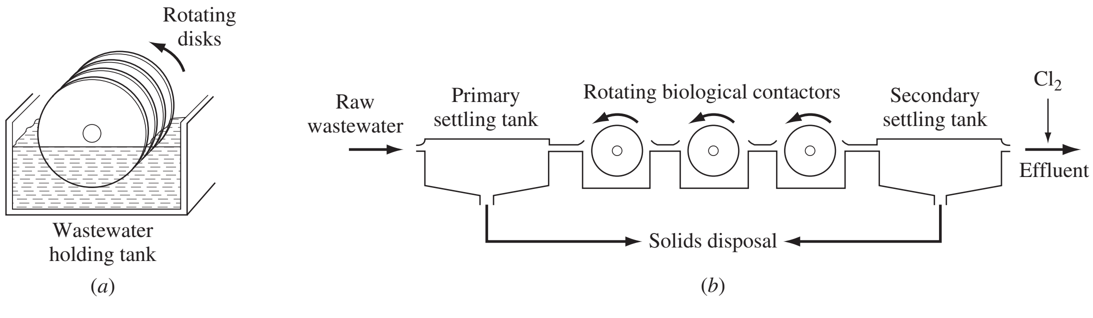


FIGURE 19 Rotating biological contactor cross-section and treatment system: (*a*) RBC cross-section; (*b*) RBC series included in a secondary wastewater treatment system.

not clarified before it is sent to the activated sludge tank. More innovative and recent hybrid systems have employed either fixed or suspended media submerged in the activated sludge aeration tank. Microbial biofilms, which grow on the media, consume the organic matter in the surrounding aerated and agitated fluid. These systems improve effluent clarification because it is simple to retain the attached cells in the aeration basin, while providing very high mixed liquor microbial populations (low F/M ratios). The interested reader can find details of many of these systems, including moving-bed biofilm reactors, fluidized bed biological reactors, submerged fixed-film activated sludge reactors, and submerged rotating biological reactors in Tchobanoglous et al. (2003).

Sludge Treatment

The processes described thus far have the purpose of removing solids and BOD from the wastewater before the liquid effluent is released to the environment or reclaimed for immediate human use such as landscape irrigation. What remains to be disposed of is a mixture of solids and water, called sludge, which is withdrawn from the bottom of the primary and secondary clarifiers. The collection, processing, and disposal of sludge can be the most costly and complex aspect of wastewater treatment.

The quantity of sludge produced may be as high as 2 percent of the original volume of wastewater, depending somewhat on the treatment process being used. Because sludge can be as much as 97 percent water, and the cost of disposal will be related to the volume of sludge being processed, one of the primary goals of sludge treatment is to separate as much of the water from the solids as possible. The other goal is to stabilize the solids so that they are no longer objectionable or environmentally damaging.

The traditional method of sludge processing utilizes anaerobic digestion. That is, it depends on bacteria that thrive in the absence of oxygen. Anaerobic digestion is slower than aerobic digestion but has the advantage that only a small percentage of the wastes are converted into new bacterial cells. A typical yield coefficient (Y) for aerobic degradation is 0.6 mg VSS/mg BOD_5, whereas the yield for anaerobic degradation of the same substrate would be about 0.1 mg VSS/mg BOD_5. Instead, most of the organics are converted into carbon dioxide and methane gas. The digestion process is complex, but can be summarized by the two steps shown in Figure 20. In the first phase, complex organics, such as fats, proteins, and carbohydrates, are biologically converted into simpler organic materials, mostly organic fatty acids. The bacteria that perform this conversion are commonly referred to as *acid formers*. They are relatively tolerant to changes in temperature and pH, and they grow much faster than the *methane formers* (*methanogens*) that carry out the second stage of digestion.

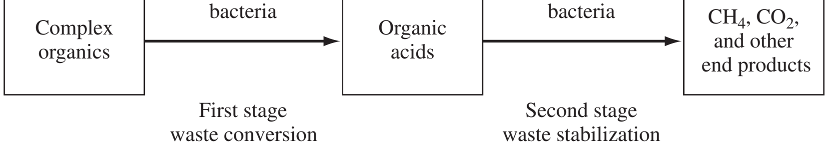


FIGURE 20 The two phases of anaerobic digestion.

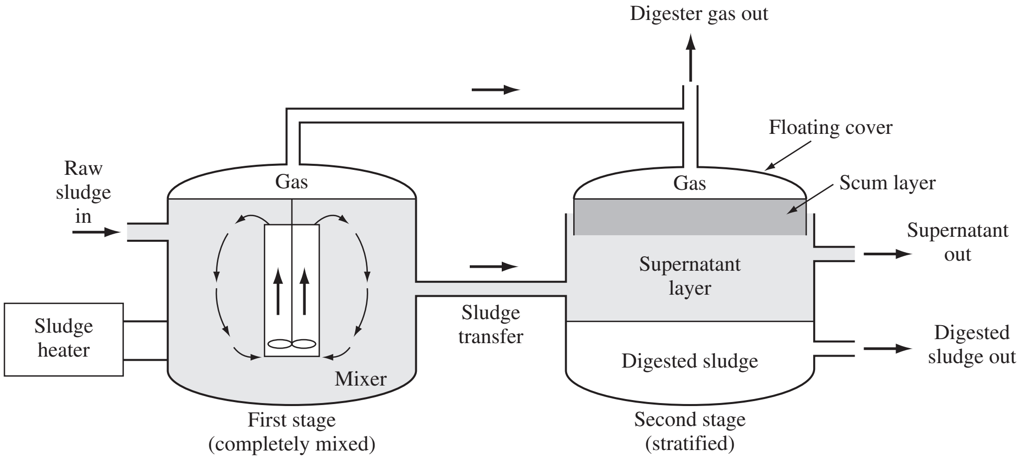


FIGURE 21 Schematic of a two-stage anaerobic digester.

Methane-forming bacteria slowly convert organic acids into CO_2, CH_4, and other stable end products. These bacteria are very sensitive to temperature, pH, toxins, and oxygen. If their environmental conditions are not just right, the rate at which they convert organic acids to methane slows, and organic acids begin to accumulate, dropping the pH. A positive feedback loop can be established where the acid formers continue to produce acid while the methane formers, experiencing lower and lower pH, become more and more inhibited. When this occurs, the digester is said to have gone sour, and massive doses of lime may be required to bring it back to operational status.

Many treatment plants that incorporate anaerobic digestion for sludge stabilization use a two-stage digester such as in Figure 21. Sludge in the first stage is thoroughly mixed and heated to increase the rate of digestion. Typical detention times are between 10 and 15 days. The second stage tank is neither heated nor mixed and is likely to have a floating cover to accommodate the varying amount of gas being stored. Stratification occurs in the second stage, which allows a certain amount of separation of liquids (called supernatant) and solids, as well as the accumulation of gas. The supernatant is returned to the main treatment plant for further BOD removal, and the settled sludge is removed, dewatered, and disposed of. The gas produced in the digester is about 60 percent methane, which is a valuable fuel with many potential uses within the treatment plant. The methane may be used to heat the first stage of the digester, and it can fuel an engine/generator set to power pumps, compressors, and other electrical equipment.

Digested sludge removed from the second stage of the anaerobic digester is still mostly liquid. The solids have been well digested, so there is little odor. Traditionally, dewatering was accomplished by pumping the sludge onto large sludge drying beds where evaporation and seepage removed the water. However, vacuum filters, filter presses, centrifuges, and incinerators are more commonly employed in modern installations as they require less land and time than drying beds. The digested and dewatered sludge is potentially useful as a soil conditioner, but most often it is simply trucked away and disposed of in a landfill.

Nutrient Removal

It used to be that anything that followed conventional primary and biological treatment was considered to be advanced treatment. As such, extra stages designed to remove nutrients, especially nitrogen and phosphorus, were considered part of advanced treatment. Nutrient removal is now so common, however, that the term "advanced" is used to describe additional steps taken to remove various toxic substances, such as metals and other hazardous wastes. Those treatment technologies will be described later in this chapter.

Nitrogen Removal. As bacteria decompose waste, nitrogen that was bound up in complex organic molecules is released as ammonia nitrogen. Subsequent oxidation of ammonia requires oxygen, which, if it occurs in the receiving body of water, contributes to oxygen depletion problems. In addition, nitrogen is an important nutrient for algal growth and, as Figure 16 indicates, only about 30 percent is normally removed in a conventional secondary treatment facility. To avoid these oxygen-demand and eutrophication problems, treatment plants need to be augmented to achieve higher rates of nitrogen removal.

Increased nitrogen control utilizes aerobic bacteria to convert ammonia (NH_4^+) to nitrate (NO_3^-), which is *nitrification,* followed by an anaerobic stage in which different bacteria convert nitrate to nitrogen gas (N_2), which is *denitrification*. The overall process then is referred to as *nitrification/denitrification.*

The nitrification step actually occurs in two bacterially driven stages. Ammonia is converted to nitrite (NO_2^-) by *Nitrosomonas,* while *Nitrobacter* oxidize nitrite to nitrate. The combination of steps can be summarized by

$$NH_4^+ + 2\,O_2 + \xrightarrow{\text{bacteria}} NO_3^- + 2H^+ + H_2O \qquad (63)$$

Nitrification does not begin to be important until domestic wastewater is at least 5 to 8 days old. Thus, if this method of nitrogen control is to be used, the wastewater must be kept in the treatment plant for a much longer time than would normally be the case. Detention times of 15 days or more are typically required. If reaction (63) takes place in the treatment plant rather than in the receiving body of water, at least the oxygen demand for nitrification is satisfied. The nitrogen, however, remains in the effluent, and if the process were to stop here, that nitrogen could go on to contribute to unwanted algal growth. To avoid this, a denitrification step is required.

The second phase of the nitrification/denitrification process is anaerobic denitrification

$$2NO_3^- + \text{organic matter} \xrightarrow{\text{bacteria}} N_2\uparrow + CO_2 + H_2O \qquad (64)$$

which releases harmless, elemental nitrogen gas. The energy to drive this reaction comes from the organic matter indicated in (64). Because this denitrification process occurs after waste treatment, there may not be enough organic material left in the wastestream to supply the necessary energy. Thus, an additional source of biodegradable organics, usually methanol (CH_3OH), must be provided.

Phosphorus Removal. Only about 30 percent of the phosphorus in municipal wastewater is removed during conventional primary and biological treatment

(Figure 16). Because phosphorus is very often the limiting nutrient, its removal from the wastestream is especially important when eutrophication is a problem.

Phosphorus in wastewater exists in many forms, but all of it ends up as orthophosphate ($H_2PO_4^-$, HPO_4^{2-}, and PO_4^{3-}). Removing phosphate is most often accomplished by adding a coagulant, usually alum [$Al_2(SO_4)_3$] or lime [$Ca(OH)_2$]. The pertinent reaction involving alum is

$$Al_2(SO_4)_3 + 2PO_4^{3-} \rightarrow 2AlPO_4\downarrow + 3SO_4^{2-} \tag{65}$$

Alum is sometimes added to the aeration tank when the activated sludge process is being used, thus eliminating the need for an additional clarifier.

6 Hazardous Wastes

As infectious diseases such as cholera and typhoid have been brought under control in the industrialized countries, there has been a shift in attention toward noninfectious human health problems such as cancer and birth defects that are induced by toxic substances. Dramatic incidents in the late 1970s such as occurred at Love Canal, New York, where hazardous substances from an abandoned dump site oozed into backyards and basements, and Times Beach, Missouri, where tens of thousands of gallons of dioxin-laced oil were carelessly sprayed onto the dusty streets, helped motivate the following decades of hazardous waste controls.

In response to growing public pressure to do something about hazardous wastes, Congress passed the *Comprehensive Environmental Response, Compensation, and Liability Act* (CERCLA) in 1980 to deal with *already contaminated* sites, and in 1980 and 1984, it strengthened the *Resource Conservation and Recovery Act* (RCRA) that controls *new sources* of hazardous waste. In essence, CERCLA deals with problems of the past, while RCRA attempts to prevent future problems. These two laws have provided the driving force behind almost all efforts at hazardous waste control.

What is a hazardous waste? Unfortunately, the answer is largely a matter of definition with various pieces of environmental legislation defining it somewhat differently. In general, the federal government defines hazardous waste as "anything which, because of its quantity, concentration, or physical, chemical, or infectious characteristics may cause, or significantly contribute to, an increase in mortality; or cause an increase in serious irreversible, or incapacitating reversible, illness; or pose a substantial present or potential hazard to human health and the environment when improperly treated, stored, transported, or disposed of, or otherwise managed." At times, a distinction needs to be made between a hazardous substance and a hazardous waste. A hazardous substance, or material, has some commercial value, whereas a hazardous waste is a material that has been used, spilled, or is no longer needed.

More specifically, EPA uses two procedures to define wastes as hazardous. One is based on a listing of substances and industrial process wastes that have been designated as hazardous. This list is published in the Code of Federal Regulations (CFR). The second defines a substance as hazardous if it possesses any of the following four characteristic attributes: *ignitability, corrosivity, reactivity,* or *toxicity.* These attributes are described briefly here:

- *Ignitable* substances are easily ignited and burn vigorously and persistently. Examples include volatile liquids, such as solvents, whose vapors ignite at relatively low temperatures (defined as 60°C or less).
- *Corrosive* substances include liquids with pH less than 2 or greater than 12.5 and those that are capable of corroding metal containers.
- *Reactive* substances are unstable under normal conditions. They can cause explosions and/or liberate toxic fumes, gases, and vapors when mixed with water.
- *Toxic* substances are harmful or fatal when ingested or absorbed. Toxicity of a waste is determined using an extraction test that has been designed to simulate the expected leaching action that might occur, for example, in a sanitary landfill. A substance is designated as being toxic if the extract from the test

TABLE 9

Examples of Hazardous Wastes Generated by Business and Industries

Waste Generators	Waste Types
Chemical manufacturers	Strong acids and bases Spent solvents Reactive wastes
Vehicle maintenance shops	Heavy metal paint wastes Ignitable wastes Used lead acid batteries Spent solvents
Printing industry	Heavy metal solutions Waste inks Spent solvents Spent electroplating wastes Ink sludges containing heavy metals
Leather products manufacturing	Waste toluene and benzene
Paper industry	Paint wastes containing heavy metals Ignitable solvents Strong acids and bases
Construction industry	Ignitable paint wastes Spent solvents Strong acids and bases
Cleaning agents and cosmetics manufacturing	Heavy metal dusts Ignitable wastes Flammable solvents Strong acids and bases
Furniture and wood manufacturing and refinishing	Ignitable wastes Spent solvents
Metal manufacturing	Paint wastes containing heavy metals Strong acids and bases Cyanide wastes Sludges containing heavy metals

contains any of a long list of chemicals in concentrations greater than allowed. EPA changed the extraction test procedure in 1990 from one called the EP Toxicity Test to the Toxicity Characteristic Leaching Procedure (TCLP) test.

Note that the term *hazardous substance* is more all-encompassing than *toxic substance,* although the two are often used interchangeably.

While these four characteristic attributes provide guidelines for considering a substance to be hazardous, EPA maintains an extensive list of specific hazardous wastes. EPA *listed wastes* are organized into three categories: *source-specific wastes, generic wastes,* and *commercial chemical products.* Source-specific wastes include sludges and wastewaters from treatment and production processes in *specific industries* such as petroleum refining and wood preserving. The list of generic wastes includes wastes from common manufacturing and industrial processes such as solvents used in degreasing operations. The third list contains specific chemical products such as benzene, creosote, mercury, and various pesticides. All listed wastes are presumed to be hazardous regardless of their concentrations. Well over 1,000 chemicals are listed by EPA as hazardous, some of which are shown in Table 9.

7 Hazardous Materials Legislation

Many federal environmental laws regulate hazardous materials in the United States. Table 10 gives a very brief description of the most important of these, along with the agency in charge of enforcement. In this section, we will focus on the three acts that are most crucial to the current management programs for hazardous materials. The first is the *Toxic Substances Control Act* (TSCA), which authorizes EPA to regulate new chemicals as they are being developed as well as existing chemicals when there are new concerns for their toxicity. The second is the Resource Conservation and Recovery Act (RCRA), which provides guidelines for prudent management of new and future hazardous substances. The third is the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), which deals primarily with mistakes of the past—inactive and abandoned hazardous waste sites.

Toxic Substances Control Act

The Toxic Substances Control Act (TSCA) gives EPA the authority to gather information on the toxicity of particular chemicals, to assess whether those chemicals cause unreasonable risks to humans and the environment, and to institute appropriate controls when deemed necessary. TSCA covers the full lifecycle of specific chemicals, from premanufacturing, to manufacturing, importation, processing, distribution, use, and disposal.

TSCA gives EPA the authority to compile an inventory of existing chemical substances that are being manufactured or imported for commercial uses. The inventory contains information on well over 60,000 chemical substances. Entities that are about to manufacture or import new chemicals that are not in the inventory must give EPA a 90-day advance notification of their intent to do so. Included in the notification must be the identity and molecular structure of the chemical; an estimate of the amounts to be manufactured and the proposed uses; the byproducts resulting from manufacture, processing, use, and disposal of the chemical; and test

TABLE 10

Environmental Laws Controlling Hazardous Substances

Law	Description
Atomic Energy Act (Nuclear Regulatory Commission)	Regulates nuclear energy production and nuclear waste disposal
Clean Air Act (EPA)	Regulates the emission of hazardous air pollutants
Clean Water Act (EPA)	Regulates the discharge of hazardous pollutants into the nation's surface water
Comprehensive Environmental Response, Compensation, and Liability Act (Superfund) (EPA)	Provides for the cleanup of inactive and abandoned hazardous waste sites
Emergency Planning and Community Right-to-Know Act (EPA)	Requires written emergency response plans for chemical releases and establishes the Toxic Release Inventory
Federal Insecticide, Fungicide, and Rodenticide Act (EPA)	Regulates the manufacture, distribution, and use of pesticides and the conduct of research into their health and environmental effects
Hazardous Materials Transportation Act (Department of Transportation)	Regulates the transportation of hazardous materials
Marine Protection, Research, and Sanctuaries Act (EPA)	Regulates waste disposal at sea
Occupational Safety and Health Act (Occupational Safety and Health Administration)	Regulates hazards in the workplace, including worker exposure to hazardous substances
Pollution Prevention Act (EPA)	Establishes priority of prevention, recycling, treatment, and waste minimization, before disposal in landfills
Resource Conservation and Recovery Act (EPA)	Regulates hazardous waste generation, storage, transportation, treatment, and disposal
Safe Drinking Water Act (EPA)	Regulates contaminant levels in drinking water and the disposal of wastes into injection wells
Surface Mining Control and Reclamation Act (Department of the Interior)	Regulates the environmental aspects of mining (particularly coal) and reclamation
Toxic Substances Control Act (EPA)	Regulates the manufacture, use, and disposal of specific chemicals

data relating to the health and environmental effects of the chemical. Similarly, significant new uses of chemicals that have already been approved for other purposes must also be approved by EPA. Based on the data submitted, EPA may prohibit the manufacture or importation until adequate data are developed to show that the chemical will not present an unreasonable risk of injury to health or the environment. A manufacturer may also be required to recall a substance that is later found to present an unreasonable risk.

TSCA also contains a provision that requires an exporter to notify EPA when certain chemicals are going to be exported. EPA in turn notifies the importing country's government of the export and provides available information on the chemicals, but, in general, it does not restrict such exports. However, chemicals that are manufactured for export only can be covered by TSCA if EPA finds that international use poses an unreasonable risk within the United States.

Both TSCA and the RCRA (discussed next) regulate the disposal of hazardous substances. The key distinction between the two is that TSCA deals with individual, specific chemicals (e.g., PCBs), whereas RCRA generally deals with combinations of chemicals contained in wastestreams.

Resource Conservation and Recovery Act

The Resource Conservation and Recovery Act (RCRA) regulates the generation, storage, transportation, treatment, and disposal of hazardous substances. It is our single most important law dealing with the management of hazardous waste, and it is perhaps the most comprehensive piece of legislation that EPA has ever promulgated. Its origins are in the 1965 Solid Waste Disposal Act, which was the first federal law to address the enormous problem of how to safely dispose of household, municipal, commercial, and industrial refuse. Congress amended that law in 1970 with the passage of the Resource Recovery Act, and finally, RCRA was passed in 1976. Revisions to RCRA were made in 1980 and again in 1984. The 1984 amendments, referred to as the *Hazardous and Solid Waste Amendments* (HSWA), significantly expanded the scope of RCRA, particularly in the area of land disposal.

The portions of RCRA that regulate hazardous wastes are contained in a section designated as Subtitle C. A major section of RCRA, Subtitle D, controls solid waste disposal, "Solid Waste Management and Resource Recovery."

Transportation, Storage, and Disposal. The key concept in RCRA is that hazardous substances must be properly managed from the moment they are generated until their ultimate disposal. This step-by-step management is often referred to as the *cradle-to-grave* approach, and it has three key elements:

1. A tracking system, in which a *manifest* document accompanies any waste that is transported from one location to another
2. A *permitting* system that helps assure safe operation of facilities that treat, store, or dispose of hazardous wastes
3. A system of controls and restrictions governing the *disposal* of hazardous wastes onto, or into, the land

About 96 percent of U.S. hazardous wastes are treated or disposed of at the site where they were originally generated (U.S. EPA, 1987a). The remaining 4 percent still represents a substantial volume of material that is transported from the source to *treatment, storage, or disposal* (TSD) facilities. To help eliminate improper handling of such transported wastes, EPA requires generators to prepare a *hazardous waste manifest* that must accompany the waste. The manifest identifies the type and quantity of waste, the generator, the transporter, and the TSD facility to which the waste is being shipped. One copy of the manifest is sent to EPA when the waste leaves the generator, and another when the waste arrives at a TSD facility. The generator, who is ultimately responsible for the waste tracking system, also receives a copy of the manifest after the waste arrives at the TSD facility. Figure 22 illustrates the manifest system.

Under RCRA, the treatment, storage, and disposal facilities that accept hazardous waste must first obtain permits from EPA. The permitting process has been

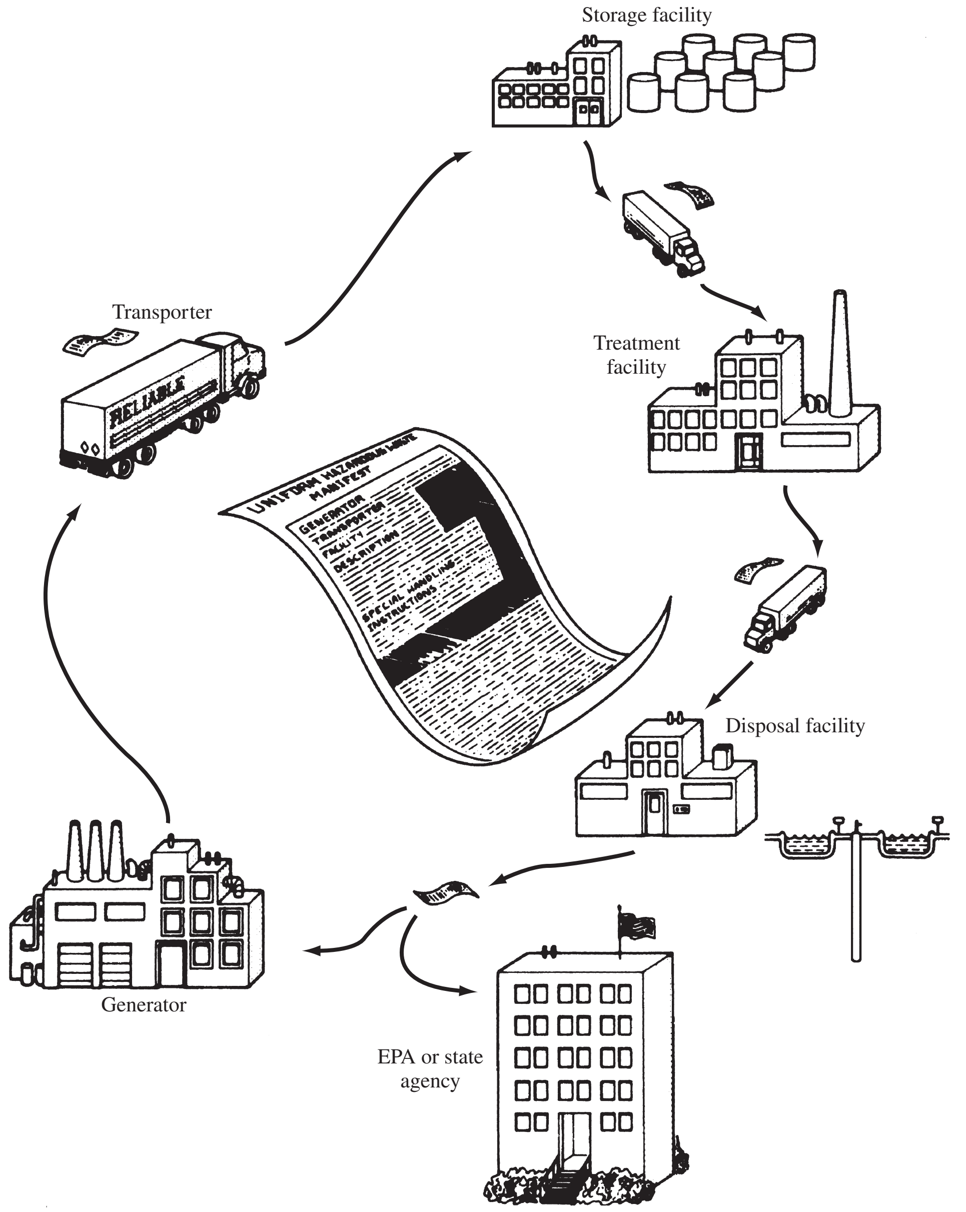


FIGURE 22 A one-page manifest must accompany every waste shipment. The resulting paper trail documents the waste's progress through treatment, storage, and disposal, providing a mechanism for alerting the generator and EPA if irregularities occur. (*Source:* U.S. EPA, 1986.)

established to give EPA power to enforce a number of standards and requirements, including the power to inspect facilities and to bring civil actions against violators of any of RCRA's provisions. To help assure proper management of hazardous wastes in the longer term, RCRA regulations require that TSD facilities acquire sufficient financial assurance mechanisms to ensure that the facility can be properly maintained after it ultimately closes, including a provision for 30 years of facility maintenance, security measures, and groundwater monitoring.

The Hazardous and Solid Waste Amendments of 1984. The 1984 Hazardous and Solid Waste Amendments (HSWA) to RCRA represented a turning point in the relationship between Congress and EPA. Before HSWA, Congress allowed considerable discretion to EPA with regard to the implementation of environmental laws. During the early 1980s, however, it was perceived that EPA had become politicized and was not meeting congressional expectations, especially with regard to implementation of CERCLA. When EPA decided to allow the disposal of liquid hazardous waste in land disposal facilities, Congress decided EPA had gone too far, and it was time to return policy control back to Congress, which it did with HSWA.

HSWA has four main provisions. First, it imposed severe restrictions on land disposal of hazardous waste. Second, it significantly reduced the level below which the rate of generation of hazardous waste was exempt from RCRA provisions. Third, it began the regulation of hundreds of thousands of leaking underground storage tanks (LUST). And fourth, it closed a loophole that had allowed some hazardous waste sites to be exempt from either RCRA or CERCLA regulations.

In the past, about 80 percent of U.S. hazardous wastes were disposed of on land either by pumping them underground using deep injection wells, burying them in landfills, or containing them in surface impoundments. The 1984 HSWA to RCRA significantly restricted disposal of wastes on land. In fact, land disposal is banned unless EPA determines that for a particular site and type of waste, there will be no migration of hazardous constituents for as long as the wastes remain hazardous. HSWA has been written in such a way that if EPA does not follow through with such a determination within a specified period of time, then the land ban is automatically imposed. This "hammer" provision was specifically intended to keep EPA from dragging its feet on implementation and is a direct result of the conflict between Congress and EPA during the Reagan administration.

To the extent that land disposal can continue to be used, new landfills and surface impoundments are required to have double liners, leachate collection systems, and groundwater monitoring facilities to help assure long-term containment. Given the ongoing problem of groundwater contamination from landfills, all liquid hazardous wastes are banned. Further details on land disposal technologies are described later in this chapter.

Waste Reduction. Clearly, rising economic costs of hazardous waste management are dictating a fresh look at the very origins of the problem itself. Is it really necessary for us to generate more than one ton of hazardous waste per person per year to provide the goods and services that we have come to expect? It is almost always cheaper and simpler to control the amount of pollution generated in the first place than to attempt to find engineering fixes that deal with it after it has been created. The 1984 amendments to RCRA recognize this important role of waste reduction by stating:

> *The Congress hereby declares it to be the national policy of the United States that, wherever feasible, the generation of hazardous waste is to be reduced or eliminated as expeditiously as possible.*

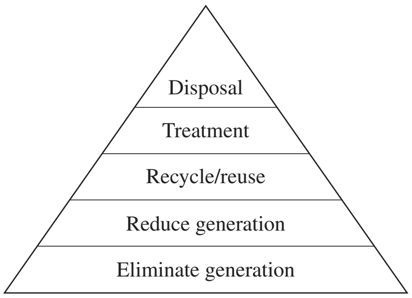


FIGURE 23 Hierarchy of priorities in hazardous waste management.

Figure 23 suggests a priority system for ranking various approaches to hazardous waste reduction. The first priority is to find ways to *eliminate* uses of hazardous substances altogether. Elimination might be achieved by changing manufacturing processes or by substituting products that can satisfy the same need without creating hazardous wastes (an example would be the substitution of concrete posts for toxic, creosote-treated wooden posts). The next priority is to *reduce* the amounts generated. Again, manufacturing process changes can be important. The third strategy is to *recycle* hazardous substances, such as solvents and acids, to maximize their use before treatment and disposal becomes necessary. Finally, hazardous substances can be *treated* to reduce their volume and toxicity. Only after all of those options have been exhausted should disposal be considered.

Eliminating, reducing, and recycling hazardous substances reduces treatment and disposal costs, and often generators can find it in their own economic best interest to actively seek out these source-reduction opportunities as a way to reduce the overall cost of doing business. RCRA, in fact, requires generators to certify on their hazardous waste manifests that they have taken steps to reduce the volume and toxicity of the hazardous waste that they are creating.

The Comprehensive Environmental Response, Compensation, and Liabilities Act

The Comprehensive Environmental Response, Compensation, and Liabilities Act (CERCLA) was enacted in 1980 to deal with abandoned hazardous waste sites. It is more often referred to as the Superfund law as a result of one of its key provisions in which a \$1.6 billion trust fund was created from special crude oil and chemical company taxes. In 1986, CERCLA was reauthorized, and the trust fund was increased to \$8.5 billion with half of the money coming from petroleum and chemical companies, and most of the rest coming from a special corporate environmental tax augmented by general revenues. The revisions to CERCLA enacted in 1986 are designated as the *Superfund Amendments and Reauthorization Act* (SARA).

CERCLA is an unusual environmental law in that it does not involve the usual regulations and permitting requirements. Instead, it is focused on identifying hazardous waste sites, preparing cleanup plans, and then forcing those deemed to be responsible parties (RPs) to pay for the remediation. Under CERCLA, EPA can deal with both short-term, emergency situations triggered by the actual or potential

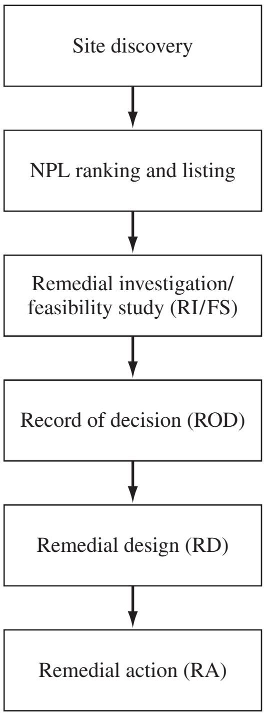


FIGURE 24 The Superfund remedial process.

release of hazardous substances into the environment, as well as with long-term problems involving abandoned or uncontrolled hazardous waste sites for which more permanent solutions are required.

Short-term *removal actions* can be taken at any site where there is an imminent threat to human health or the environment, such as might occur during a spill or fire, or when wastes that have been illegally disposed of are discovered ("midnight dumping"). In addition to removing and disposing of hazardous substances and securing the endangered area, EPA may take more extensive actions as necessary, such as providing alternate drinking water supplies to local residents if their drinking water has been contaminated, or even temporarily relocating residents.

More complex and extensive problems that are not immediately life-threatening are handled under EPA's *remedial* programs, which are outlined in Figure 24. Central to the Superfund process is the creation of a National Priorities List (NPL) of sites that are eligible for federally financed remedial activities. The NPL identifies the worst sites in the nation based on such factors as the quantities and toxicity of wastes involved, the exposure pathways, the number of people potentially exposed, and the importance and vulnerability of the underlying groundwater.

After a site is listed, EPA begins a *remedial investigation/feasibility study* (RI/FS) to determine appropriate remedial actions for the site. The remedial investigation phase involves gathering information needed to characterize the contamination problem at the site, including environmental and public health risks. The feasibility study that follows uses the remedial investigation information to identify, evaluate, and select cleanup alternatives. These alternative cleanup approaches are analyzed based on their relative effectiveness and cost. The RI/FS process culminates with the signing of a *record of decision* (ROD), in which the remedial action that has been selected is set forth. After the ROD has been signed, the detailed *remedial design* of the selected alternative takes place. For complex sites, the entire process from listing to the beginning of the actual cleanup itself has often taken five years or more.

The final *remedial action* may consist of a number of short-term steps, such as installation of surface water run-off controls, excavating soil, building below-grade containment walls, and capping the site. Most remedial actions have involved pump-and-treat technologies, which have serious limitations and often require decades of pumping. However, from 1986 to 1999, the percentage of groundwater Superfund RODs which selected only pump-and-treat remediation, dropped from 92 percent to 30 percent as other remedial strategies became more advanced and accepted (U.S. EPA, 2002a). The cleanup of a typical NPL site is a slow, complex, expensive process, which has led some to question whether it is really worth all that effort.

One of the most important, guiding policies of CERCLA is that those parties who are responsible for hazardous waste problems will be forced to pay the entire cost of cleanup. *Responsible parties* (RPs) may be the historic owners or operators of the site, any generators who disposed of wastes at the site (whether legally or not), or even the transporters who brought wastes to the site. The courts have recognized the concept of *retroactive, strict,* and *joint-and-several* liability for costs of cleanup. *Retroactive* covers problems created before Superfund was enacted; *strict liability* refers to the fact that liability does not depend on whether the RPs were negligent or diligent in their disposal practices; and *joint-and-several* means, in essence, that if damages cannot be individually apportioned by the RPs themselves, then each and every party is subject to liability for the entire cleanup cost. In other words, no matter when their contribution occurred, no matter how careful they were, and no matter how little an individual RP may have contributed to the overall problem, they can, theoretically, be liable for the entire cost of the cleanup. If RPs do not voluntarily perform the appropriate response actions, EPA can use Superfund monies to clean up the site. If they do so, then EPA is empowered to collect three times the cleanup cost from the RPs.

The retroactive, strict, joint-and-several provisions of Superfund have been heavily criticized. Banks and lending institutions worry that they could be liable for cleanup costs just because they loaned money to firms that were named as RPs. Insurance companies are fighting lawsuits when RPs claim they were covered for cleanup costs under their general liability insurance. Municipalities that have been named as RPs when their landfills have leaked don't feel they should be liable for enormous cleanup costs. To many, it seems unfair to be held responsible for disposal practices that were perfectly legal in the past, especially when their contributions to the problem were minimal. And finally, it seems particularly unfair to some that owners of property are at risk for cleanup costs even though they were not responsible in any way for the contamination.

CERCLA has also been criticized because it requires that the cleanup solution selected provides a permanent solution to the contamination problem. For some sites, it might be argued that temporary measures can cost very little while virtually eliminating health risks. For example, consider a site located some distance from residences with contaminated soil but not contaminated groundwater. It can be argued that capping the site, surrounding it with fences and signs, and monitoring the soil and groundwater would be a much better use of resources than a permanent site remediation project. However, it might not be allowed under CERCLA.

Brownfields. In an effort to address some of the liability concerns and fairness issues of CERCLA, as well as to help revitalize communities affected by contamination, EPA launched its *Brownfields Economic Redevelopment Initiative* in 1995. *Brownfields* are defined to be “abandoned, idled, or under-used industrial and commercial facilities where expansion or redevelopment is complicated by real or perceived environmental contamination.” The Initiative is based on the belief that public fear of contamination and lender concerns for liability are causing companies to leave brownfields in the urban core and head for greenfields outside the cities. It is hoped that implementation of EPA’s Brownfields Action Agenda will help reverse the spiral of unaddressed contamination, declining property values, and increased unemployment often found in inner-city industrial areas (U.S. EPA, 1996).

Key components of the Brownfields initiative include removing eligible sites from the Superfund site tracking system, known as the Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS). Initial actions removed 27,000 such sites out of the 40,000 that were being tracked. Removing sites from CERCLIS is intended to convey assurance that EPA will no longer pursue Superfund action at these sites. A second, very important component of the Initiative is a clarification of liability issues that frequently cause potential buyers and lenders to avoid contaminated properties. For example, EPA will not take action against owners when hazardous substances enter the property from aquifers contaminated elsewhere, as long as the landowner did not in any way cause the problem, and as long as the property owner did not have a contractual arrangement with the polluter.

The Brownfields issue is helping to focus attention on risk-based corrective actions in which the most feasible remediation technologies are coupled with best management practices to provide cost-effective cleanup without compromising the protection of public health, water quality, and the environment.

8 Hazardous Waste Treatment Technologies

Even with a much more vigorous hazardous waste reduction program, as RCRA requires, there will still be large quantities of hazardous wastes that will require treatment and disposal. In the past, there was little treatment, and disposal was most often on land. In both SARA (Superfund) and the 1984 Hazardous and Solid Waste Amendments of RCRA, emphasis is on the development and use of alternative and innovative treatment technologies that result in permanent destruction of wastes or a reduction in toxicity, mobility, and volume. Land disposal is greatly restricted under the 1984 RCRA amendments.

Treatment technologies are often categorized as being chemical, biological, physical, thermal, or fixation/stabilization. These categories are reasonably well defined, although there is room for confusion when technologies have overlapping characteristics:

- *Chemical, biological, and physical wastewater treatment processes* are currently the most commonly used methods of treating aqueous hazardous waste. Chemical treatment transforms waste into less hazardous substances using such techniques as pH neutralization, oxidation or reduction, and precipitation. Biological treatment uses microorganisms to degrade organic compounds in the wastestream. Physical treatment processes include gravity separation, phase change systems such as air and steam stripping of volatiles from liquid wastes, and various filtering operations, including carbon adsorption.
- *Thermal destruction processes* include incineration, which is increasingly becoming a preferred option for the treatment of hazardous wastes, and pyrolysis, which is the chemical decomposition of waste brought about by heating the material in the absence of oxygen.
- *Fixation/stabilization* techniques involve removal of excess water from a waste and solidifying the remainder either by mixing it with a stabilizing agent, such as Portland cement or polymeric substances, or vitrifying it to create a glassy substance. Solidification is most often used on inorganic sludges.

Choosing an appropriate technology to use in any given situation is obviously beyond the scope of this text. Not only are there many different kinds of hazardous wastes, in terms of their chemical makeup, but the treatability of the wastes depends on their form. A technology suitable for treating PCBs in sludges, for example, may not be appropriate for treating the same contaminant in dry soil. Table 11 gives a partial listing of available treatment technologies appropriate for a variety of types of hazardous wastestreams along with the applicable form of waste (liquid, gaseous, solids/sludges). For a more complete list, as well as detailed descriptions of each technology, see *Standard Handbook of Hazardous Waste Treatment and Disposal* (Freeman, 1989).

Physical Treatment

The use of the physical treatment processes described following is not limited to hazardous waste treatment. For instance, sedimentation was also discussed earlier in this chapter as a treatment method commonly used in both water and wastewater treatment plants. In fact, all of the physical processes except monitored natural attenuation are used to some degree in drinking water or wastewater plants.

Sedimentation. The simplest physical treatment systems that separate solids from liquids take advantage of gravity settling and natural flotation. Special sedimentation tanks and clarification tanks are designed to encourage solids to settle so they can be collected as a sludge from the bottom of the tank. Some solids will float naturally to the surface and they can be removed with a skimming device. It is also possible to encourage flotation by introducing finely divided bubbles into the wastestream. The bubbles collect particles as they rise, and the combination can be skimmed from the surface. Separated sludges can then be further concentrated by

TABLE 11

A Partial List of Treatment Technologies for Various Hazardous Wastestreams

	Hazardous wastestreams												Form of waste		
Treatment process	Corrosives	Cyanides	Halogenated solvents	Nonhalogenated organics	Chlorinated organics	Other organics	Oily wastes	PCBs	Aqueous with metals	Aqueous with organics	Reactives	Contaminated soils	Liquids	Solids/sludges	Gases
Separation/filtration		X	X	X	X	X			X	X			X		
Carbon adsorption									X	X	X		X		X
Air and steam stripping			X	X	X	X				X			X		
Electrolytic recovery									X				X		
Ion exchange	X								X	X			X		
Membranes									X	X			X		
Chemical precipitation	X								X				X		
Chemical oxidation/reduction		X								X			X		
Ozonation		X		X		X					X		X		X
Evaporation			X	X	X	X	X						X	X	
Solidification	X	X										X	X	X	
Liquid injection incineration			X	X	X	X	X						X		X
Rotary kilns			X	X	X	X	X	X				X	X	X	X
Fluidized bed incineration			X	X	X	X	X	X				X	X	X	X
Pyrolysis			X	X	X	X						X	X	X	
Molten glass			X	X	X	X	X			X			X	X	X

Source: Based on Freeman, 1989.

evaporation, filtration, or centrifugation. An example of a vacuum filter is shown in Figure 25.

Adsorption. Physical treatment can also be used to remove small concentrations of hazardous substances dissolved in water that would never settle out. One of the most commonly used techniques for removing organics involves the process of *adsorption*, which is the physical adhesion of chemicals onto the surface of a solid. The effectiveness of the adsorbent is directly related to the amount of surface area available to attract the molecules or particles of contaminant. The most commonly used adsorbent is a very porous matrix of *granular activated carbon* (GAC), which has an enormous surface area (on the order of 1,000 m^2/g). A single handful of GAC has an internal surface area of about 1 acre.

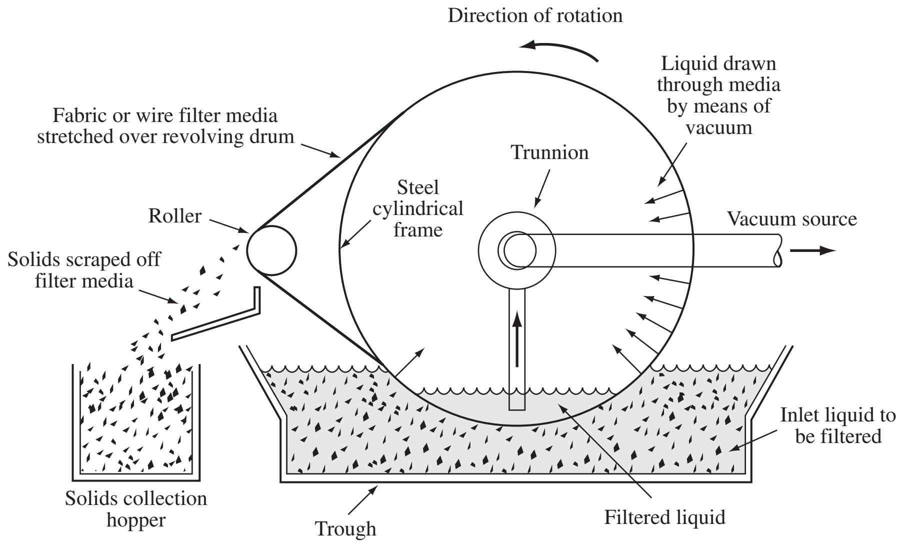


FIGURE 25 Schematic of a typical vacuum filter.
(*Source: Hazardous and Industrial Waste Treatment* by Haas and Vamos, © 1995. Reprinted by permission of Prentice-Hall, Inc., Upper Saddle River, NJ.)

Granular-activated carbon treatment systems usually consist of a series of large vessels partially filled with adsorbent. Contaminated water enters the top of each vessel, passes down through the GAC, and is released at the bottom. After a period of time, the carbon filter becomes saturated with adsorbed contaminants and must be either replaced or regenerated. Regeneration can be an expensive, energy-intensive process, usually done off site. During regeneration, the contaminants are usually burned from the surface of the carbon granules, although in some cases a solvent is used to remove them. GAC that cannot be regenerated due to its contaminant composition must be properly managed for disposal.

Aeration. For chemicals that are relatively volatile, another physical process, *aeration,* can be used to drive the contaminants out of solution. These stripping systems typically use air, although in some circumstances steam is used. In the most commonly used air stripper, contaminated water trickles downward over the surface of packing material in a tower, while air is blown upward carrying away the volatiles with it. Such a *packed tower* can easily remove over 95 percent of the volatile organic compounds (VOCs), including such frequently encountered ones as trichloroethylene, tetrachloroethylene, trichloroethane, benzene, toluene, and other common organics derived from solvents. Another type of stripper, called an *induced-draft stripper,* does not use a blower or packing material. In the induced-draft tower, a carefully engineered series of nozzles sprays contaminated water horizontally through the sides of a chamber. Air passing through the chamber draws off the volatiles. Induced-draft strippers are cheaper to build and operate, but their performance is much lower than a packed tower.

By combining air stripping with GAC, many volatile and nonvolatile organic compounds can be removed from water to nondetectable levels. By passing contaminated water first through the air stripper, most of the volatile organics are removed

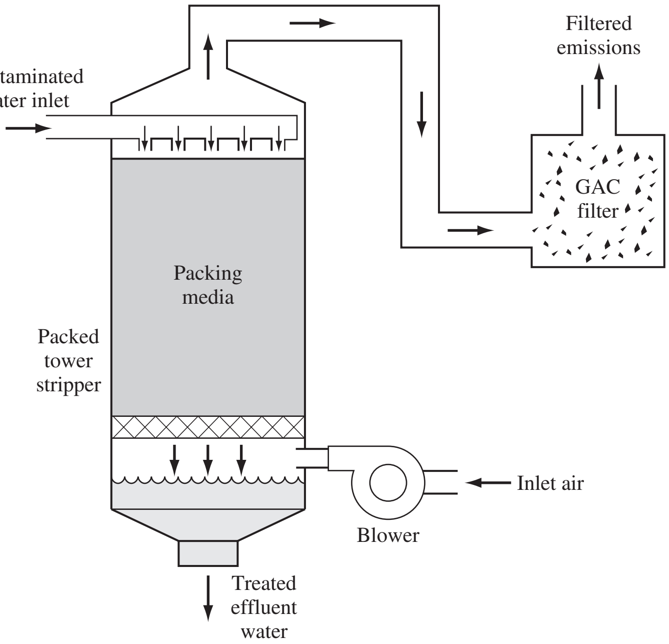


FIGURE 26 An air stripping tower followed by a gas-phase granular-activated carbon filter provides effective removal of VOCs. To remove nonvolatiles, the treated water coming out of the tower may be pumped through liquid-phase GAC adsorbers.

before reaching the GAC system, which extends the life of the carbon before regeneration or replacement is required.

The volatiles removed in an air stripper are, in some circumstances, released directly to the atmosphere. When discharge into the atmosphere is unacceptable, a GAC treatment system can be added to the exhaust air, as shown in Figure 26.

Other Physical Processes. Other physical processes that are sometimes used to treat hazardous wastes include reverse osmosis, ion exchange, and electrodialysis. *Reverse osmosis* devices use pressure to force contaminated water against a semipermeable membrane. The membrane acts as a filter, allowing the water to be pushed through its pores but restricting the passage of larger molecules that are to be removed. *Ion exchange* is a process wherein ions to be removed from the wastestream are exchanged with ions associated with a special exchange resin. Membrane filtration and ion exchange have been discussed earlier in the context of drinking water and wastewater treatment. In the context of hazardous wastes, ion exchange is often used to remove toxic metal ions from solution. *Electrodialysis* uses ion-selective membranes and an electric field to separate anions and cations in solution. In the past, electrodialysis was most often used for purifying brackish water, but it is now finding a role in hazardous waste treatment. Metal salts from plating rinses are sometimes removed in this way.

Although it is not strictly a physical remediation process, *monitored natural attenuation* (MNA) has been increasingly implemented at Superfund sites. MNA depends on the natural physical, chemical, and biological processes in the subsurface to degrade and immobilize the contaminants. It is implemented together with a carefully controlled monitoring program to ensure that natural processes are successfully remediating the site and stopping further environmental damage or human exposure from occurring. Natural bioremediation by indigenous microbes in the subsurface is often a key component of successful MNA. Between 1982 and 1999, Superfund site RODs with specified remedies listed MNA as the only groundwater remedy selected at 12 percent of the sites and MNA along with pump-and-treat or in situ bioremediation at an additional 9 percent of the sites (U.S. EPA, 2002a). MNA can be a cost-effective and environmentally responsible alternative to active cleanup when it is applied with the same regard to initial site characterization and long-term oversight as more active approaches.

Chemical Treatment

Chemically treating hazardous waste has the potential advantage of not only converting it to less hazardous forms, but it can also produce useful byproducts in some circumstances. By encouraging resource recovery, the treatment cost can sometimes be partially offset by the value of the end products produced.

Neutralization. Many chemical processes can be used to treat hazardous wastes, and the process decision depends primarily upon the characteristic of the waste. For example, recall that one of RCRA's categories of hazardous waste is anything corrosive, that is, having a pH of less than 2 or more than 12.5. Such wastes can be chemically *neutralized*. Acidic wastewaters are usually neutralized with slaked lime [$Ca(OH)_2$] in a continuously stirred tank reactor. The rate of addition of lime is controlled with a feedback control system that monitors pH and adjusts the feed rate accordingly.

Alkaline wastewaters may be neutralized by adding acid directly or by bubbling in gaseous CO_2, forming carbonic acid (H_2CO_3). The advantage of CO_2 is that it is often readily available in the exhaust gas from any combustion process at the treatment site. Simultaneous neutralization of acid and caustic waste can be accomplished in the same vessel as is suggested by Figure 27.

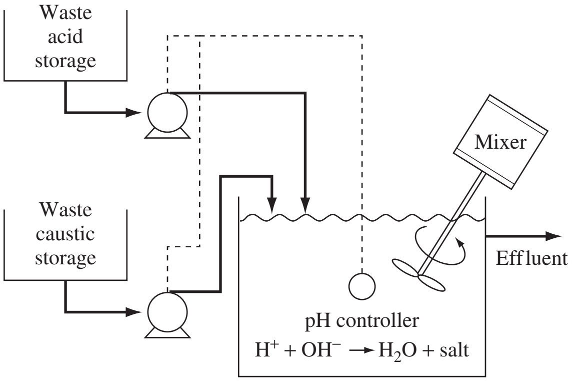


FIGURE 27 Simultaneous neutralization of acid and caustic waste.
(*Source:* U.S. EPA, 1987b.)

Chemical Precipitation. The ability to adjust pH is important not only for waste neutralization, but also because it facilitates other chemical processes that actually remove undesirable substances from the wastestream. For example, a common method for removing heavy metals from a liquid waste is via *chemical precipitation,* which is pH dependent. By properly adjusting pH, the solubility of toxic metals can be decreased, leading to formation of a precipitate that can be removed by settling and filtration.

Frequently, precipitation involves the use of lime, $Ca(OH)_2$, or caustic, NaOH, to form metal hydroxides. For example, the following reaction suggests the use of lime to form the hydroxide of a divalent metal (M^{2+}):

$$M^{2+} + Ca(OH)_2 \rightarrow M(OH)_2 + Ca^{2+} \tag{66}$$

Metal hydroxides are relatively insoluble in basic solutions, and, as shown in Figure 28, they are *amphoteric,* that is, they have some pH at which their solubility is a minimum. Because each metal has its own optimum pH, it is tricky to control precipitation of a mix of different metals in the same waste. For a waste containing several metals, it may be necessary to use more than one stage of precipitation to allow different values of pH to control the removal of different metals.

Although hydroxide precipitation using lime is the most common metal removal process, even lower concentrations of metals in the effluent can be obtained by precipitating the metals as sulfides. As can be seen in Figure 28, metal sulfides are considerably less soluble than metal hydroxides. A disadvantage of sulfide precipitation is the potential formation of odorous and toxic hydrogen sulfide gas.

Chemical Reduction-Oxidation. *Reduction-oxidation* (redox) reactions provide another important chemical treatment alternative for hazardous wastes. When electrons are removed from an ion, atom, or molecule, the substance is *oxidized;* when electrons are added, it is *reduced*. Both oxidation and reduction occur in the same reaction, hence the abbreviation redox. One of the most important redox treatment processes is the reduction of hexavalent chromium (Cr VI) to trivalent chromium (Cr III) in large electroplating operations. Sulfur dioxide is often used as the reducing agent, as shown in the following reactions:

$$SO_2 + H_2O \rightarrow H_2SO_3 \tag{67}$$

$$2\ CrO_3 + 3\ H_2SO_3 \rightarrow Cr_2\ (SO_4)_3 + 3\ H_2O \tag{68}$$

The trivalent chromium formed in reaction (68) is much less toxic and more easily precipitated than the original hexavalent chromium. Notice the chromium in reaction (68) is reduced from an oxidation state of +6 to +3, and the sulfur is oxidized from +4 to +6.

Another important redox treatment involves the oxidation of cyanide wastes, which are also common in the metal finishing industry. In the following reactions, cyanide is first converted to a less toxic cyanate using alkaline chlorination (pH above 10); further chlorination oxidizes the cyanate to simple carbon dioxide and nitrogen gas. Nearly complete destruction of cyanide results.

$$NaCN + Cl_2 + 2\ NaOH \rightarrow NaCNO + 2\ NaCl + H_2O \tag{69}$$

$$2\ NaCNO + 3\ Cl_2 + 4\ NaOH \rightarrow 2\ CO_2 + N_2 + 6\ NaCl + 2\ H_2O \tag{70}$$

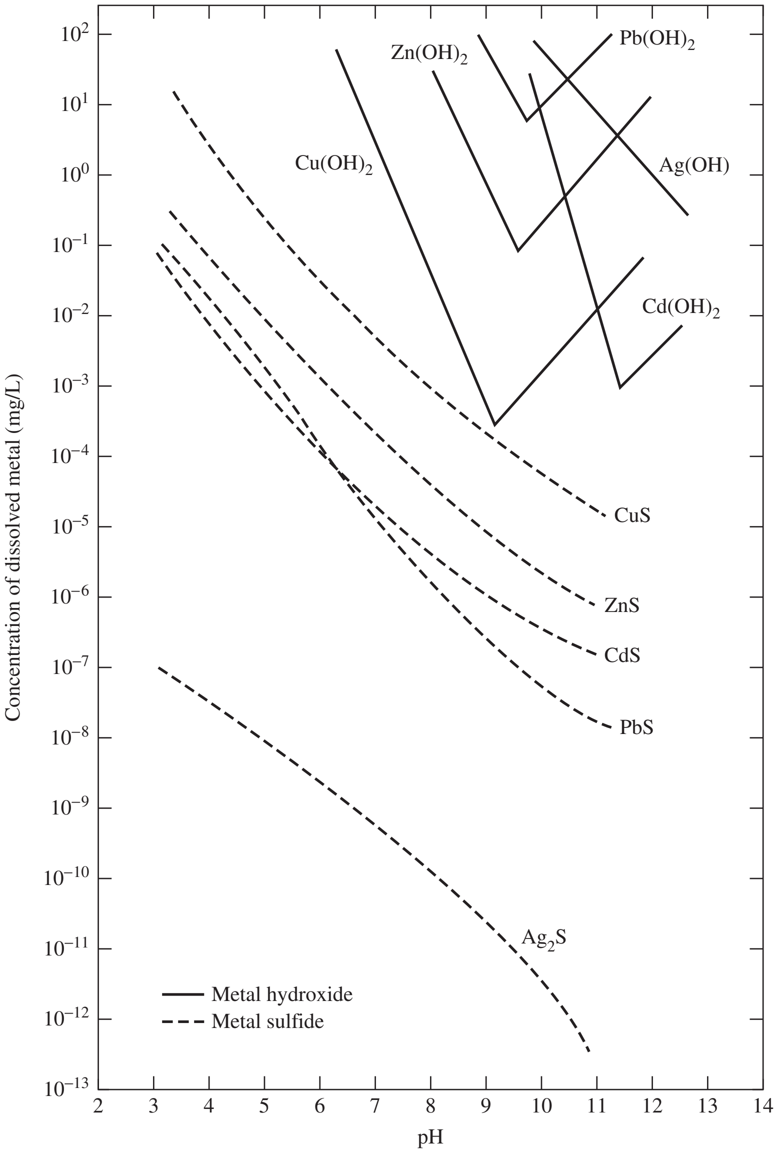


FIGURE 28 Chemical precipitation of metals can be controlled by pH. Metal sulfides are less soluble than metal hydroxides.
(*Source:* U.S. EPA, 1980.)

Wastes that can be treated via redox oxidation include benzene, phenols, most organics, cyanide, arsenic, iron, and manganese. Those that can be successfully treated using reduction treatment include chromium (VI), mercury, lead, silver, chlorinated organics like PCBs, and unsaturated hydrocarbons (U.S. EPA, 1988a).

Ultraviolet Radiation/Oxidation. A promising approach to destruction of dissolved organic compounds is oxidation, which converts organics into simple carbon dioxide and water. A number of oxidants are possible, including hydrogen peroxide, oxygen, ozone, and potassium permanganate. One of the most potent oxidizers is the hydroxyl free radical ($^{\bullet}OH$), which easily initiates the oxidation of hydrogen-containing organic molecules. The formation of hydroxyl radicals can be enhanced by exposing an oxidant, such as hydrogen peroxide (H_2O_2) or ozone (O_3), to ultraviolet (UV) light, as the following reactions suggest:

$$O_3 + h\nu + H_2O \rightarrow H_2O_2 + O_2 \tag{71}$$

$$H_2O_2 + h\nu \rightarrow 2\ OH^{\bullet} \tag{72}$$

where $h\nu$ represents a photon of light with frequency ν. Low-pressure mercury vapor lamps produce photons of about the right wavelength (254 nm) needed to efficiently convert the hydrogen peroxide in (72). The initial oxidization of an organic molecule using the hydroxyl radical can be represented as

$$^{\bullet}OH + HR \rightarrow H_2O + R^{\bullet} \rightarrow \ldots \rightarrow CO_2 \tag{73}$$

where $R^{\bullet}$ represents a carbon-based radical. The rate of destruction of various chlorinated organics using oxidation by ozone alone is compared with destruction using ozone and UV light together in Figure 29.

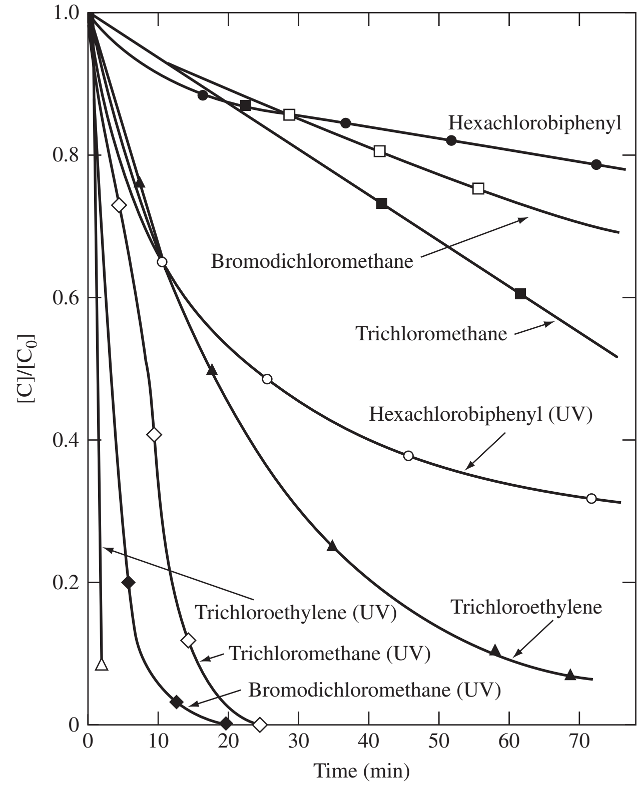


FIGURE 29 Comparison of destruction of chlorinated organics using ozone alone versus ozone plus UV light.

(*Source:* Glaze, 1987. Reprinted with permission from *Environmental Science & Technology*, American Chemical Society.)

Biological Treatment

Virtually all municipal wastewater treatment plants in the United States and a large number of industrial systems rely on biological treatment processes to decompose organic wastes. Biological treatment systems use microorganisms, mainly bacteria, to metabolize organic material, converting it to carbon dioxide, water, and new bacterial cells. Because biological systems rely on living organisms to transform wastes, considerable care must be exercised to assure conditions are conducive to life. Microbes need a source of carbon and energy, which they can get from the organics they consume, as well as nutrients such as nitrogen and phosphorus. They are sensitive to pH and temperature. Some need oxygen.

As living organisms, microbes are susceptible to toxic substances, which at first glance makes biological treatment of hazardous wastes seem an unlikely choice. Surprisingly, though, most hazardous organics are amenable to biological treatment (*bioremediation*) provided that the proper distribution of organisms can be established and maintained. For any given organic substance, some organisms may find that substance to be an acceptable food supply, whereas others may find it toxic. Moreover, organisms that flourish with the substance at one concentration may die when the concentration is increased beyond some critical level. Finally, even though a microbial population may have been established that can handle a particular kind of organic waste, it may be destroyed if the characteristics of the waste are changed too rapidly. If changes are made slowly enough, however, selection pressures may allow the microbial consortium to adjust to the new conditions and thereby remain effective.

Aqueous Waste Treatment. It is convenient to consider biological treatment of various sorts of wastewaters, including leachates from hazardous waste landfills, separately from in situ biological treatment of soils and groundwater. When liquid hazardous wastes can be conveyed to the treatment facility, it is possible to carefully control the characteristics of the waste that reach the biological portion of the facility, increasing the likelihood of a successful degradation process.

Biological treatment is often just one step in an overall treatment system. As suggested in Figure 30, an example system would include a chemical treatment stage

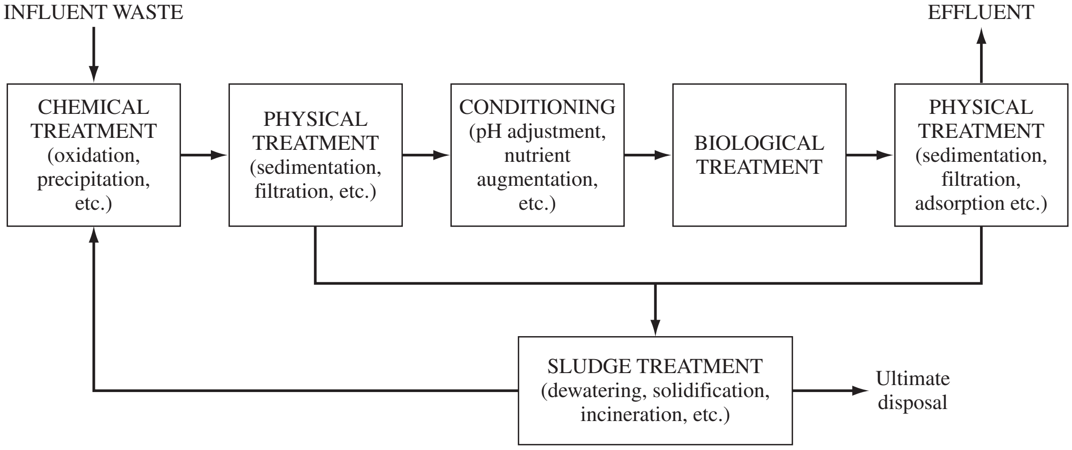


FIGURE 30 General flow diagram for treatment of liquid hazardous waste.

to oxidize and precipitate some of the toxics, followed by physical treatment to separate the resulting solids from the wastestream. The effluent from the physical treatment step may then be conditioned to give it the right pH and nutrient supply needed by the microorganisms in the biological treatment step.

The biological treatment stage utilizes processes already described for municipal wastewater treatment plants. After biological treatment, further sedimentation and clarification followed by carbon adsorption can be used to polish the effluent. Inorganic sludges produced during chemical processing and organic sludges from the biotreatment stage are separated from the liquid wastestream and treated. These sludges must be dewatered and disposed of in accordance with RCRA regulations because they are likely to be hazardous.

Waste Incineration

Waste incineration is being advocated by EPA as its technology of choice for many types of hazardous wastes. Incineration is particularly effective with organic wastes, not only in soils but in other solids, gases, liquids, and slurries (thin mixtures of liquids and solids) and sludges (thick mixtures) as well. Carcinogens, mutagens, and teratogens as well as pathological wastes can all be completely detoxified in a properly operated incinerator. Incinerators are not, however, capable of destroying inorganic compounds, although they can concentrate them in ash, making transportation and disposal more efficient. In addition, metals that volatilize at temperatures below 2,000°F pose a particular problem because, once vaporized, they are difficult to remove using conventional air pollution control equipment.

The principal measure of an incinerator's performance is known as the *destruction and removal efficiency* (DRE). A DRE of 99.99 percent, for example, (commonly called "four nines DRE") means that 1 molecule of an organic compound is released to the air for every 10,000 molecules entering the incinerator. RCRA requires a minimum DRE of 99.99 percent for most organic compounds, and a DRE of 99.9999 percent (six nines) for dioxins and dibenzofurans. The Toxic Substances Control Act regulations cover thermal destruction of PCBs, and although they are written somewhat differently from RCRA, they in essence require a 99.9999 percent DRE.

As is the case for all combustion processes, the most critical factors that determine combustion completeness are (1) the temperature in the combustion chamber, (2) the length of time that combustion takes place, (3) the amount of turbulence or degree of mixing, and (4) the amount of oxygen available for combustion. Controlling these factors, which is crucial to obtaining the high levels of performance required by law, is made especially difficult in hazardous waste incinerators because of the variability of the wastes being burned. In addition to combustion controls, stack gas cleaning systems are a necessary part of the system. Proper operation and maintenance of these complex incineration systems requires highly trained personnel, diligent and qualified supervisory staff, and an alert governmental agency to assure compliance with all regulations.

Although there are a number of types of hazardous waste incinerators, only two principal designs account for most of the existing units in operation: the *liquid*

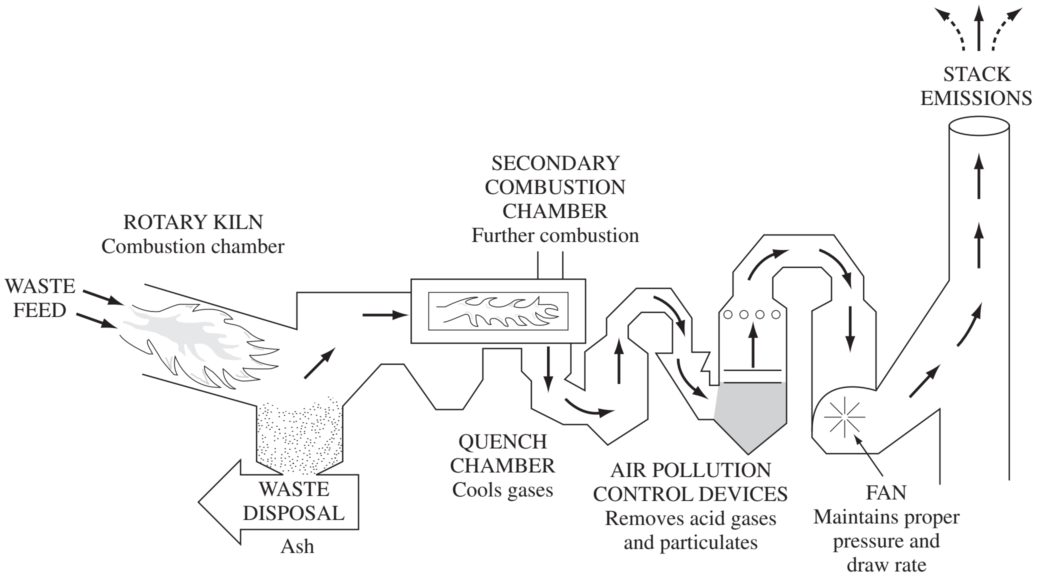


FIGURE 31 A rotary kiln hazardous waste incinerator.
(*Source:* U.S. EPA, 1988b.)

injection incinerator and the *rotary kiln incinerator*. Liquid injection incinerators are the most common even though they are usable only for gases, liquids, and slurries thin enough to be pumped through an atomizing nozzle. The nozzle emits tiny droplets of waste that are mixed with air and an auxiliary fuel such as natural gas or fuel oil. The resulting gaseous mixture is burned at a very high temperature. The atomizing nozzle used in a liquid injection incinerator must be designed to accommodate the particular characteristics of the expected wastestream, which limits the types of waste that any given incinerator can treat.

The rotary kiln incinerator is more versatile than the liquid injection type, being capable of handling gases, liquids, sludges, and solids of all sorts, including drummed wastes. Figure 31 shows a diagram of such an incinerator. The main unit consists of a slightly inclined, rotating cylinder perhaps 2 to 5 meters in diameter and 3 to 10 meters long. Wastes and auxiliary fuel are introduced into the high end of the kiln, and combustion takes place while the cylinder slowly rotates. The rotation helps increase turbulence, which improves combustion efficiency. Partially combusted waste gases are passed to a secondary combustion chamber for further oxidation. Rotary kiln incinerators are commercially available as mobile units and fixed installations.

In spite of numerous controls, hazardous waste incinerators have the potential to emit amounts of noxious gases that may be unacceptable to neighbors. Emissions may include unburned organic compounds from the original waste, various *products of incomplete combustion* (PICs) formed in the incinerator, odors, carbon monoxide, nitrogen and sulfur oxides, hydrogen chloride, and particulates. The unburned ash and sludge from the air pollution control devices are considered hazardous wastes and must be treated as such. If they are transported off site, then not only are there hazardous materials transported into the facility, but there are some leaving as well. The perception of potentially adverse impacts associated with incineration has made the siting of these facilities an extremely difficult task.

9 Land Disposal

The EPA regulates the land disposal of hazardous wastes under its Land Disposal Restrictions (LDR) program. About 23 million tons of hazardous waste is land disposed in the United States each year (U.S. EPA, 2001). Land disposal techniques include landfills, surface impoundments, underground injection wells, and waste piles. Historically, land disposal has been the most common method of getting rid of hazardous wastes in this country. Unfortunately, many of these disposal sites have been poorly engineered and monitored and the results have sometimes been tragic, as well as very expensive to rectify. They have been used extensively in the past because they were the most convenient and inexpensive method of disposal. However, remediation at older sites that have leaked toxics into the soil and groundwater has proven to be tremendously costly, and the originally perceived economic advantage of land disposal is now seen to often have been short-sighted.

As was mentioned earlier, the 1984 Hazardous and Solid Waste Amendments (HSWA) to RCRA ban unsafe, untreated wastes from land disposal (U.S. EPA, 1991). The HSWA provisions are LDRs, where land disposal includes placement of hazardous waste in the following:

- Landfills
- Surface impoundments
- Waste piles
- Injection wells
- Land treatment facilities
- Salt domes or salt bed formations
- Underground mines or caves
- Concrete vaults or bunkers, intended for disposal purposes

HWSA requires that EPA assess all hazardous wastes to determine whether land disposal can be used, and if so, under what conditions. Without an EPA determination on a particular waste, that waste is automatically banned from land disposal. RCRA goes on to provide new restrictions and standards for those land disposal facilities that will be allowed to accept hazardous substances, including the following:

- Banning liquids from landfills
- Banning underground injection of hazardous waste within a quarter-mile of a drinking water well
- Requiring more stringent structural and design conditions for landfills and surface impoundments, including two or more liners, leachate collection systems above and between the liners, and groundwater monitoring
- Requiring cleanup or corrective action if hazardous waste leaks from a facility
- Requiring information from disposal facilities on pathways of potential human exposure to hazardous substances

- Requiring location standards that are protective of human health and the environment, for example, allowing disposal facilities to be constructed only in suitable hydrogeologic settings

Landfills. In accordance with these new, more stringent RCRA requirements, the design and operation of hazardous waste landfills has become much more sophisticated. A hazardous waste landfill is now designed as a modular series of three-dimensional (3D) control cells. By incorporating separate cells, it becomes possible to segregate wastes so that only compatible wastes are disposed of together. Arriving wastes are placed in an appropriate cell and covered at the end of each working day with a layer of cover soil.

Beneath the hazardous wastes, there must be a double-liner system to stop the flow of liquids, called *leachate,* from entering the soil and groundwater beneath the site. The upper liner must be a *flexible-membrane lining* (FML) usually made of sheets of plastic or rubber. Commonly used plastics include polyvinyl chloride (PVC), high-density polyethylene (HDPE), and chlorinated polyethylene (CPE). Rubber FMLs include chlorosulfonated polyethylene (CSPE) and ethylene propylene diene monomer (EPDM). Depending on the material chosen for the FML, the thickness is typically anywhere from 0.25 mm (10 mils) to over 2.5 mm (100 mils). The lower liner is usually an FML, but recompacted clay at least 3 feet thick is also considered acceptable.

Leachate that accumulates above each liner is collected in a series of perforated drainage pipes and pumped to the surface for treatment. To help reduce the amount of leachate formed by precipitation seeping into the landfill, a low permeability cap is placed over completed cells. When the landfill is finally closed, a cap that may consist of an FML along with a layer of compacted clay is placed over the entire top with enough slope to assure drainage away from the wastes.

The landfill must also include monitoring facilities. The groundwater flowing beneath the site should be tested with monitoring wells placed up-gradient and down-gradient from the site. There may only need to be one up-gradient well to test the “natural” quality of the groundwater before it flows under the site, but there should be at least three or more monitoring wells placed down-gradient to assure detection of any leakage from the site. In addition, the soil under the site, above the water table, should be tested using devices called suction lysimeters.

A cross-section of a completed hazardous waste landfill is shown in Figure 32.

Surface Impoundments. *Surface impoundments* are excavated or diked areas used to store liquid hazardous wastes. Usually storage is temporary unless the impoundment has been designed to eventually be closed as a landfill. Impoundments have been popular because they have been cheap and because wastes remain accessible, allowing some treatment to take place during storage. Typical treatment technologies used in surface impoundments include neutralization, precipitation, settling, and biodegradation.

Historically, surface impoundments have often been poorly constructed and monitored. In a survey of 180,000 surface impoundments, EPA estimated that prior to 1980, only about one-fourth were lined, and fewer than 10 percent had monitoring programs (U.S. EPA, 1984). The same survey also found that surface impoundments were usually poorly sited. More than half were located over very thin or very

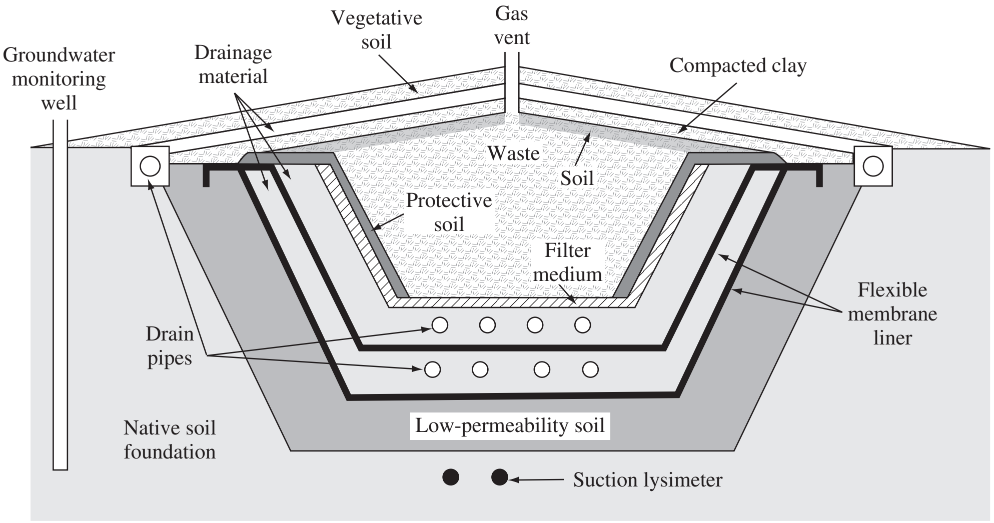


FIGURE 32 Schematic of a hazardous waste landfill.

permeable soils that would allow easy transport of leachate to groundwater. Over three-fourths of the impoundments were located over very thick and permeable aquifers that would allow relatively rapid dispersion of contaminants should they reach the water table. Moreover, about 98 percent of the surface impoundments were located less than one mile from sources of high-quality drinking water.

As a result of these poor siting, construction, and management practices, surface impoundments are the principal source of contamination in a large number of Superfund sites. Current EPA regulations require new surface impoundments, or expansions to existing impoundments, to have two or more liners, a leachate-collection system, and monitoring programs similar to those required for landfills. However, the legacy of past practices will undoubtedly take billions of dollars and decades of time to remediate.

Underground Injection. The most popular way to dispose of liquid hazardous wastes has been to force them underground through deep injection wells (Figure 33). About 89 percent of the land disposed hazardous waste and more than 700 million gallons of fluids annually are disposed in underground injection wells (U.S. EPA, 2002b). To help assure that underground drinking water supplies won't become contaminated, since 1984, injection wells used to dispose of hazardous industrial wastes are required to extend below the lowest formation containing underground sources of drinking water. Typical injection depths are more than 700 meters below the surface. Because the main concern with underground injection is the potential for contaminating underground drinking water supplies, the regulation of such systems has come under the Safe Drinking Water Act of 1974.

Unfortunately, a number of hazardous waste injection wells have had leakage problems, so such wells cannot be considered entirely safe. Regulations covering construction, operation, and monitoring of injection wells are becoming more stringent, and as is the case for all land disposal options, continued reliance on this technology is being discouraged.

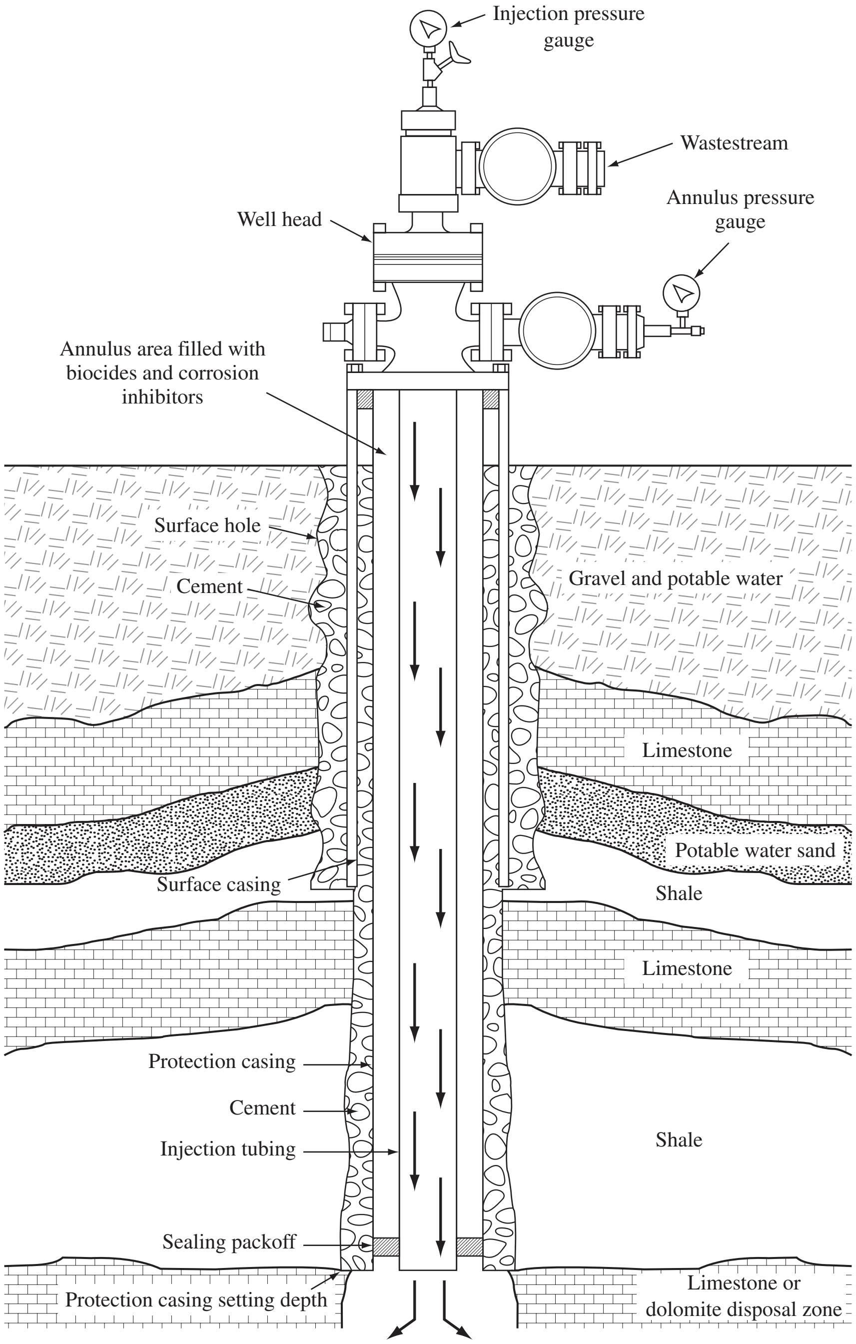


FIGURE 33 Cross-section of a hazardous waste disposal well.
(*Source:* Wentz, 1989.)

PROBLEMS

1 (a) What is the difference between a primary and a secondary drinking water standard?

(b) What is the difference between a Maximum Contaminant Level (MCL) and a Maximum Contaminant Level Goal (MCLG)? Give an example of a contaminant that has an MCLG but not a numeric MCL.

2 How does the Clean Water Act differ from the Safe Drinking Water Act?

3 A municipal wastewater treatment plant employs two circular primary clarifiers arranged in parallel, following the bar screen and grit removal chamber. The plant receives 5.0 MGD. Each clarifier is center-fed (water enters at the center and exits at the perimeter). The clarifier radius is 43.0 ft, and depth is 10.0 ft.

(a) What is the detention time in each clarifier?

(b) What is the hydraulic loading rate (critical velocity) in units of $ft^3 \cdot ft^{-2} \cdot hr^{-1}$?

(c) What is the weir loading rate in units of $ft^3 \cdot ft^{-1} \cdot hr^{-1}$?

4 (a) What is the terminal settling velocity of a particle with a specific gravity of 1.4 and a diameter of 0.010 mm in 20°C water?

(b) Would particles of the size in part (a) be completely removed in a settling basin with a width of 10.0 m, a depth of 3.00 m, a length of 30.0 m, and a flow rate of 7,500 m^3/d?

(c) What is the smallest diameter particle of specific gravity 1.4 that would be removed in the sedimentation basin described in part (b)?

5 You are trying to remove 10.0 μm diameter particles in a water treatment plant. The water is at 20°C, and the particle density is 1.2 g/mL. The plant treats 0.100 m^3/s of water. It is proposed to use a 3.5 m deep, rectangular sedimentation tank with a length to width ratio of 5:1. What is the minimum required width of the basin?

6 Horizontal shaft paddle wheels are used in a tapered flocculation basin comprised of three compartments. The power input to the paddles progressively decreases through the basin with the first compartment paddle driving with 186 W of power, the second compartment paddle with 30.0 W of power, and the final compartment paddle with 7.50 W of power. Each paddle wheel mixes a compartment 4.17 m deep, 3.75 m wide, and 4.17 m long. The normal water temperature is 15°C, and the average flow rate is 16,000 m^3/d.

(a) What is the mixing intensity in each of the three compartments?

(b) If the influent water contains a nearly monodisperse (single-sized) particle size distribution of 20.0 μm particles with a concentration of 1.8×10^5 particles/mL, what will be the average number of singlet particles in an aggregate leaving the first compartment?

7 Jar tests are used to study flocculation in laboratories. A jar test apparatus is a 4.0 liter batch reactor in which a water sample is placed, coagulant is initially added, and it is stirred at a constant rate. Samples are taken of the water over time, and the total number of particles (singlets plus aggregates) are counted.

(a) Draw a picture of the setup, and write a mass balance describing the total particle number concentration (N) as a function of time in the batch reactor.

(b) If the stirrer imparts 6.5×10^{-5} W($kg \cdot m^2 \cdot s^{-3}$) of power at a given stirring speed and 15°C, what is the mean velocity gradient (G) in the reactor?

(c) There are 1.0×10^7 particles initially in the 4.00 L of water put in the jar apparatus. All of the particles are initially 50 micrometers (5×10^{-5} m) diameter. Assuming the stirring speed in part (b), how many particles per liter will there be after 30 minutes of stirring, if the particles are completely destabilized?

(d) What is the average number of singlet particles in an aggregate after the 30 minutes of stirring?

(e) What would the particle concentration be after 30 minutes based on part (c), but for the case where the particles are partially destabilized and only one in every five collisions between particles results in the colliding particles adhering?

8 A rapid sand filter has a loading rate of 8.00 m/h, surface dimensions of 10 m × 8 m, an effective filtration rate of 7.70 m/h, and a production efficiency of 96 percent. A complete filter cycle duration is 52 h and the filter is rinsed for 20 minutes at the start of each cycle.

(a) What flow rate (m^3/s) does the filter handle during production?

(b) What volume of water is needed for backwashing plus rinsing the filter in each filter cycle?

9 A conventional water treatment plant treats $2.40\ m^3 \cdot s^{-1}$ of river water. The filters in the plant are 8.0×8.0 m surface area and operate on the following schedule. A filter is cleaned once in each 24-hour period, and the backwash rate is $10.0\ L \cdot m^{-2} \cdot s^{-1}$ for a period of 8 min. It takes an additional 8 min to drain the filter and break up the media with air injection. Of the total filter cycle time of 24 hours, 15 minutes is used for rinsing the filter after backwashing. During production, water is applied to the filter at a rate of $5.50\ L \cdot m^{-2} \cdot s^{-1}$.

(a) What is the filter efficiency?

(b) What is the minimum number of filters required by the plant?

10 Chlorine is used for primary disinfection at a water treatment plant that treats $0.20\ m^3/s$ of reservoir water. *Giardia* (a protozoan) is the primary pathogen of concern, and the *Giardia* inactivation rate constant for free chlorine disinfection is $0.53\ min^{-1}$. Note that the inactivation rate constant is k^* in the simplified Chick-Watson equation where:

$$r(N) = -k^* N \text{ and } k^* = kC^n$$

The plant achieves 3-log removal of *Giardia* at pH 6.0–9.0 in a plug flow contact chamber using free chlorine concentration of 2.0 mg/L.

(a) What is the volume of contact chamber?

(b) If a CSTR, rather than PFR, contact chamber were used, what would the required volume need to be?

(c) What volume of contact chamber would be needed if it were constructed as five separate CSTR chambers in series? The volume asked for is the combined volume of all five separate chambers.

(d) Assuming a coefficient of dilution of 1.0, what is the coefficient of specific lethality for *Giardia* under these conditions?

11 The PureNSafe Company is designing a solar-driven, portable water disinfection system to be used by the military. The system is simply an inlet tube, a pump, a well-mixed chamber, and an outlet tube. Water is pumped continuously into the chamber where it is irradiated with UV light before exiting through the outlet tube into a collection vessel. The disinfection rate constant for UV light is $7.80\ s^{-1}$, and the bacterial count in the water must be reduced by 99.9 percent. Because the army needs the unit to fit into a backpack,

they require that the unit's chamber be no larger than 2.00 liters. How much water will the system be able to produce during 10 hours of sunlight?

12 A sample of groundwater has 150 mg/L of Ca^{2+} and 60 mg/L of Mg^{2+}. Find the total hardness expressed in milliequivalents per liter (meq/L) and mg/L as $CaCO_3$. Using Table 4, how would this water be classified (e.g., soft, hard, etc.)?

13 For a solution with pH equal to 9.0, express the concentrations of H^+ and OH^- in meq/L and mg/L as $CaCO_3$.

14 A sample of water at pH 10.5 has 39.0 mg/L of CO_3^{2-} and 24.5 mg/L of HCO_3^-.

(a) Ignoring the contribution of $[H^+]$ and $[OH^-]$ to alkalinity, what is the alkalinity as $CaCO_3$?

(b) Including the contribution of $[H^+]$ and $[OH^-]$, find the alkalinity as $CaCO_3$.

15 A chemical analysis of a surface water yields the following data:

Ion	Concentration mg/L
Ca^{2+}	90
Mg^{2+}	30
Na^+	72
K^+	6
Cl^-	120
SO_4^{2-}	225
HCO_3^-	165
pH	7.5

(a) Determine the alkalinity expressed as $CaCO_3$.

(b) Determine the hardness as $CaCO_3$.

(c) Estimate the total dissolved solids.

16 The Cl^- concentration in the water analysis given in Problem 15 has been questioned. What concentration of Cl^-(mg/L) would make the cations and anions balance?

17 The water with the composition shown in the following table is to be softened.

Component	Concentration mg/L
$CO_{2(aq)}$	14.5
Ca^{2+}	110.0
Mg^{2+}	50.7
Na^+	75
HCO_3^-	350
SO_4^{2-}	85.5
Cl^-	16.2
pH	8.2

(a) What is the concentration (expressed in meq/L) of magnesium carbonate hardness in the water?

(b) What is the concentration (expressed in meq/L) of magnesium noncarbonate hardness in the water?

(c) What concentration (expressed in meq/L) of slaked lime $[Ca(OH)_2]$ must be added to remove the carbon dioxide?

18 A water is analyzed and found to have the composition shown in the following table.

Component	Concentration
CO_2	6.0 mg/L
Ca^{2+}	50.0 mg/L
Mg^{2+}	20.0 mg/L
Na^+	5.0 mg/L
Alkalinity	3.1 millimole/L
SO_4^{2-}	85 mg/L
pH	7.6

(a) What concentrations (expressed as mg/L) of slaked lime [$Ca(OH)_2$] and soda ash (Na_2CO_3) must be added to the water to remove the maximum practical hardness?

(b) If the water shown in the table is the source water for a 15 MGD water softening plant, estimate the mass of sludge (in kg/day) produced by the softening process.

19 For the water given in Problem 15, determine the following:

(a) The concentration (expressed as mg/L) of slaked lime [$Ca(OH)_2$] added to the water to achieve the maximum practical softness

(b) The concentration (expressed as mg/L) of soda ash (Na_2CO_3) added to achieve the maximum practical softness

20 A sample of water has the following concentration of ions (and the pH is near neutral):

Cations	mg/L	Anions	mg/L
Ca^{2+}	95.0	HCO_3^-	160.0
Mg^{2+}	26.0	SO_4^{2-}	135.0
Na^+	15.0	Cl^-	73.0

(a) What is the total hardness (TH)?

(b) What is the carbonate hardness (CH)?

(c) What is the noncarbonate hardness (NCH)?

(d) What is the alkalinity?

(e) What is the total dissolved solids concentration?

(f) Draw an ion concentration bar graph.

21 For the water given in Problem 20, determine the following:

(a) The concentration (expressed as mg/L) of slaked lime [$Ca(OH)_2$] added to the water to achieve the maximum practical softness

(b) The concentration (expressed as mg/L) of soda ash (Na_2CO_3) added to achieve the maximum practical softness

22 A sample of water has the following concentrations of ions:

Cations	mg/L	Anions	mg/L
Ca^{2+}	40.0	HCO_3^-	110.0
Mg^{2+}	10.0	SO_4^{2-}	67.2
Na^+	?	Cl^-	11.0
K^+	7.0		

(a) Assuming no other constituents are missing, use an anion-cation balance to estimate the concentration of Na^+.

(b) What is the total hardness (TH)?

(c) Draw an ion concentration bar graph.

23 For the water given in Problem 22, determine the following:

(a) The concentration (expressed as mg/L) of slaked lime [$Ca(OH)_2$] added to the water to achieve the maximum practical softness.

(b) The concentration (expressed as mg/L) of soda ash (Na_2CO_3) added to achieve the maximum practical softness.

(c) If the water is the source water for a 37 MGD water softening plant, estimate the mass of sludge (in kg/day) produced by the softening process.

24 A reverse osmosis plant desalinates 5×10^6 L/day of feedwater containing 1,500 mg/L of salts, yielding 3×10^6 L/day of product water with 75 mg/L of salt (see Figure P24). What would the salt concentration be in the brine?

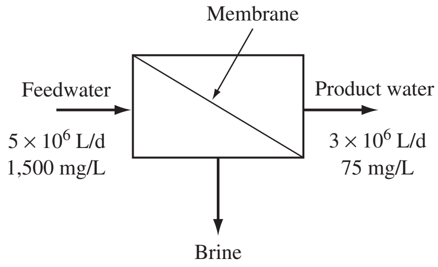


FIGURE P24

25 Determine the following for the reverse osmosis plant described in Problem 24:

(a) The water recovery for the plant

(b) The salt rejection for the process

(c) The log rejection of salt for the process

26 (a) What is the main constituent of concern in wastewater treatment?

(b) For each of the following unit operations in a wastewater treatment train briefly describe how it removes some of the constituent you identified in part a:

- grit chamber
- primary sedimentation basin
- biological reactor
- secondary clarifier
- digestor

27 Briefly describe how primary wastewater treatment differs from secondary wastewater treatment.

28 A rectangular primary clarifier for a domestic wastewater plant is to be designed to settle 2,000 m^3/day with an overflow rate of 32 m^3/m^2-day. The tank is to be 2.4 m deep and 4.0 m wide. How long should it be and what detention time would it have?

29 A final settling tank for a 2-million-gallon-per-day (2 MGD) activated-sludge treatment plant has an average overflow rate of 800 g/day-ft^2. The tank needs to have a minimum detention time of 2.0 hr, and to allow proper settling it must be at least 11 ft deep. If the tank is circular, what should its diameter and depth be?

30 A perfectly mixed aeration pond with no recycle (return line) serves as the biological reactor for a small community. The pond receives 30 m^3/d of influent with a BOD$_5$ of 350 mg/L that must be reduced to 20 mg/L before discharge. It has been found that the kinetic constants for the system are $K_s = 100$ mg/L BOD$_5$, $k_d = 0.10\ \text{d}^{-1}$, $\mu_m = 1.6\ \text{d}^{-1}$, and Y is 0.60 mg VSS/mg BOD$_5$.

(a) What must the hydraulic detention time be in the aeration pond?

(b) What mass of microbes will be produced in the pond each day?

31 A rectangular, well-mixed aeration lagoon is 60 m long, 5 m wide, and 2 m deep. It receives 400 m^3/d of wastewater with a BOD$_5$ of 336 mg BOD$_5$/L for treatment. The biodegradation rate constants for the wastewater are yield coefficient = 0.8 mg VSS/mg BOD$_5$, endogenous decay constant = 0.08 d^{-1}, maximum specific microbial growth rate = 1.10 d^{-1}, and the half velocity constant = 76.0 mg BOD$_5$/L.

(a) What is the efficiency of the lagoon?

(b) The pond's efficiency depends on oxygen being always available for the microbes to use in degrading the wastewater organic carbon. How much oxygen (in kg) must be supplied daily to the pond?

32 A wastewater treatment uses an activated sludge process for secondary treatment of 0.300 m^3/s of primary effluent. The mixed liquor has a concentration of 2,100 mg VSS/L, and the RAS concentration is 10,000 mg VSS/L. The substrate concentration in the primary effluent is 220 mg BOD$_5$/L. The F/M ratio for the activated sludge tank is 0.52 mg BOD$_5 \cdot$ mg VSS$^{-1} \cdot$ d^{-1}, and the cell residence time is 9.0 d.

(a) What is the volume of the activated sludge tank?

(b) What is the WAS flow rate?

(c) What is the flow rate of secondary treated effluent?

(d) What is the hydraulic residence time for the activated sludge tank?

33 A 20 MGD wastewater treatment plant's primary clarifier influent has 800 mg/L solids. The clarifier removes 18 percent of the solids. The sludge flow rate is 0.070 MGD.

(a) What is the solids concentration, C_s, of primary clarifier sludge?

(b) What mass of solids (in kg/yr) is removed annually by the primary clarifier?

REFERENCES

Cooley, H., P. H. Gleick, and G. Wolff, 2006, *Desalination with a Grain of Salt: A California Perspective,* Pacific Institute, Oakland, CA.

Freeman, H. M. (ed.), 1989, *Standard Handbook of Hazardous Waste Treatment and Disposal,* McGraw Hill, New York.

Glaze, W. H., 1987, Drinking water treatment with ozone, *Environmental Science and Technology,* 21 (3).

Gleick, 1993, *Water in Crisis, A Guide to the World's Fresh Water Resources,* Oxford University Press, New York.

Haas, C. N., and R. J. Vamos, 1995, *Hazardous and Industrial Waste Treatment,* Prentice Hall, Englewood Cliffs, NJ.

Hammer, M. J., and M. J. Hammer, Jr., 1996, *Water and Wastewater Technology,* 3rd ed., Prentice-Hall, Englewood Cliffs, NJ.

McCarty, P. L., 1971, Energetics and bacterial growth, in *Organic Compounds in Aquatic Environments,* S. D. Faust and J. V. Hunter (eds.), Marcel Dekker, New York.

McCarty, P. L., 1975, Stoichiometry of biological reactions, *Progress in Water Technology,* 7, 157–172.

Metcalf, and Eddy, 1991, *Wastewater Engineering Treatment and Reuse,* 3rd ed., McGraw-Hill, Inc., New York.

Monod, J., 1949, The growth of bacterial cultures, *Annual Review of Microbiology,* vol. III, 371–394.

Robeck, G. G., N. A. Clarke, and K. A. Dostal, 1962, Effectiveness of water treatment processes in virus removal, *Journal American Water Works Association,* 54, 1275–1290.

Taylor, J. S., and M. Wiesner, 1999, Membranes, in *Water Quality and Treatment: A Handbook of Community Water Supplies,* 5th ed., R. D. Letterman (ed.), American Water Works Association, McGraw-Hill, Inc, New York.

Tchobanoglous, G., and E. D. Schroeder, 1985, *Water Quality,* Addison-Wesley, Reading, MA.

Tchobanoglous, G., F. L. Burton, and H. D. Stensel, 2003, *Wastewater Engineering: Treatment and Reuse,* 4th ed., McGraw-Hill, Inc., New York.

U.S. EPA, 1980, *Summary Report, Control and Treatment Technology for Metal Finishing Industry,* Environmental Protection Agency, EPA 625/8-80-003, Washington, DC.

U.S. EPA, 1984, *Surface and Impoundment Assessment National Report,* Environmental Protection Agency, EPA-570/9-84-002, Washington, DC.

U.S. EPA, 1986, *Solving the Hazardous Waste Problem: EPA's RCRA Program,* Office of Solid Waste, Environmental Protection Agency, Washington, DC.

U.S. EPA, 1987a, *The Hazardous Waste System,* Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, DC.

U.S. EPA, 1987b, *A Compendium of Technologies Used in the Treatment of Hazardous Wastes,* Center For Environmental Research Information, EPA/625/8-87/014, Environmental Protection Agency, Washington, DC.

U.S. EPA, 1988a, *Technology Screening Guide for Treatment of CERCLA Soils and Sludges,* Environmental Protection Agency, Office of Solid Waste and Emergency Response, EPA/502/2-88/004, Washington, DC.

U.S. EPA, 1988b, *Hazardous Waste Incineration: Questions and Answers,* Environmental Protection Agency, Office of Solid Waste, EPA/530-SW-88-018, Washington, DC.

U.S. EPA, 1991, *Land Disposal Restrictions, Summary of Requirements,* Environmental Protection Agency, OSWER 9934.0-1A, Washington, DC.

U.S. EPA, 1996, *Brownfields Action Agenda,* EPA/500/F-95/001, Environmental Protection Agency, Washington, DC.

U.S. EPA, 1999, *Combined Sewer Overflow Management Fact Sheet: Sewer Separation,* Office of Water, Environmental Protection Agency, EPA-832-F-99-041, Washington, DC.

U.S. EPA, 2001, *Land Disposal Restrictions: Summary of Requirements,* Environmental Protection Agency, Office of Solid Waste and Emergency Response, EPA/530-R-01-007, Washington, DC.

U.S. EPA, 2002a, *Groundwater Remedies Selected at Superfund Sites,* Environmental Protection Agency, Office of Solid Waste and Emergency Response, EPA/542-R-01-022, Washington, DC.

U.S. EPA, 2002b, *Protecting Drinking Water Through Underground Injection Control: Drinking Water Pocket Guide* #2, Environmental Protection Agency, Office of Groundwater and Drinking Water, EPA/8162-K-02-001, Washington, DC.

Veissman, W., Jr., and M. J. Hammer, 1985, *Water Supply and Pollution Control,* 4th ed., Harper & Row, New York.

Wentz, C. A., 1989, *Hazardous Waste Management,* McGraw Hill, New York.

Air Pollution

. . . from a particulate exposure standpoint, a 2-percent decrease in environmental tobacco smoke (passive smoking) would be equivalent to eliminating all the coal-fired power plants in the country.

—*Kirk Smith, 1993*

1 Introduction

Air pollution is certainly not a new phenomenon. Early references date to the Middle Ages, when smoke from burning coal was already considered such a serious problem that in 1307, King Edward I banned its use in lime kilns in London. In more recent times, though still decades ago, several serious episodes focused attention on the need to control the quality of the air we breathe. The worst of these occurred in London, in 1952. A week of intense fog and smoke resulted in over 4,000 excess

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

deaths that were directly attributed to the pollution. In the United States, the most alarming episode occurred during a 4-day period in 1948 in Donora, Pennsylvania, when 20 deaths and almost 6,000 illnesses were linked to air pollution. At the time, Donora had a population of only 14,000, making this the highest per capita death rate ever recorded for an air pollution episode.

Those air pollution episodes were the results of exceptionally high concentrations of sulfur oxides and particulate matter, the primary constituents of *industrial smog* or *sulfurous smog*. Sulfurous smog is caused almost entirely by combustion of fossil fuels, especially coal, in stationary sources such as power plants and smelters. In contrast, the air pollution problem in many cities is caused by emissions of carbon monoxide, oxides of nitrogen, and various volatile organic compounds that swirl around in the atmosphere reacting with each other and with sunlight to form *photochemical smog*. Although stationary sources also contribute to photochemical smog, the problem is most closely associated with motor vehicles. A major effect of efforts in the United States to control both sulfurous smog and photochemical smog has been the elimination of those dramatic, peak concentrations of pollution that were responsible for the air pollution episodes just mentioned. In their place, however, is the more insidious problem of morbidity and mortality increases associated with long-term exposure to lower concentrations of pollution. The human toll is much more difficult to document, but estimates place the current excess deaths caused by air pollution (mostly small particles) at several tens of thousands per year in the United States alone.

Much of the work on air pollution in the past few decades has centered on a small set of six substances, called *criteria pollutants,* that have been identified as contributors to both sulfurous and photochemical smog problems. The sources, transport, effects, and methods of controlling these criteria pollutants will be a principal focus of this chapter.

More recently, attention has been shifting toward the characterization and control of a growing list of especially hazardous air pollutants, many of which we are exposed to in our homes and workplaces, where we spend roughly 90 percent of our time. As the preceding quote suggests, modest improvements in indoor air quality can improve public health as much as major reductions in the traditional outdoor sources that have been the focus of most of the scientific and political efforts of the past 50 years.

2 Overview of Emissions

There are many sources of the gases and particulate matter that pollute our atmosphere. Substances that are emitted directly into the atmosphere are called *primary* pollutants, whereas others that are created by various physical processes and chemical reactions that take place in the atmosphere are called *secondary* pollutants. For example, nitrogen oxides and hydrocarbons emitted when fuels are burned are primary pollutants, but the ozone that is created when those chemicals react with each other in the atmosphere is a secondary pollutant.

The sources of primary pollutant emissions can be conveniently categorized by the processes that create them. Most primary pollutants enter the atmosphere as a result of combustion, evaporation, or grinding and abrasion. Automobile exhaust emissions and power plant stack gases are created during combustion; volatile substances such as gasoline, paints, and cleaning fluids enter the atmosphere by evaporation; whereas dust kicked up when land is plowed and asbestos fibers that flake off of pipe insulation are examples of grinding and abrasion. Of these, combustion accounts for the great majority of emissions, and the gases and particulate matter released when fuels are burned have been the focus of most of the technical and legislative pollution control efforts.

To investigate the origins of primary pollutants resulting from combustion, let's begin with complete combustion of a pure hydrocarbon fuel such as methane (CH_4):

$$CH_4 + 2\ O_2 \rightarrow CO_2 + 2\ H_2O \tag{1}$$

The products of combustion are simple carbon dioxide and water, neither of which had been considered an air pollutant until we realized that the accumulation of CO_2 in the atmosphere was enhancing Earth's natural greenhouse effect.

If the temperature of combustion isn't high enough, or there isn't enough oxygen available, or if the fuel isn't given enough time to burn completely, then the fuel will not be completely oxidized, and some of the carbon will be released as carbon monoxide (CO) instead of CO_2. Also, some of the fuel will not be completely burned, so there will be emissions of various partially combusted hydrocarbons that we will represent by (HC). So we can write the following descriptive reaction to represent incomplete combustion of our pure hydrocarbon fuel, methane:

$$CH_4 + O_2 \rightarrow \text{mostly } (CO_2 + 2\ H_2O) + \text{traces of } [CO + (HC)] \tag{2}$$

Of course, most combustion takes place in air, not in a pure oxygen environment, and air is roughly 78 percent N_2 and 21 percent O_2. When the temperature of combustion is high enough, some of that nitrogen reacts with the oxygen in air to form various nitrogen oxides (NO_x). Since this NO_x is formed when combustion temperatures are high, it is referred to as *thermal* NO_x.

$$\text{air } (N_2 + O_2) + \text{Heat} \rightarrow \text{Thermal } NO_x \tag{3}$$

So far, we have assumed the fuel being burned was a pure hydrocarbon such as methane. In reality, of course, most fuels have a number of other elements in them such as nitrogen, sulfur, lead (in gasoline), mercury (in coal), and other unburnable materials called ash. Burning fuel with these "impurities" in them releases additional NO_x (called *fuel* NO_x), oxides of sulfur (SO_x), lead (Pb), mercury (H_g), more particulate matter, and ash.

Combining the effects of incomplete combustion, combustion in air, and combustion of fuels that are not pure hydrocarbons yields the following qualitative description of combustion:

$$\begin{aligned}&\text{Fuel } (H, C, S, N, Pb, Hg, \text{ash}) + \text{Air } (N_2 + O_2) \rightarrow \\ &\qquad \text{Emissions } (CO_2, H_2O, CO, NO_x, SO_x, Pb, Hg, \text{particulates}) + \text{ash}\end{aligned} \tag{4}$$

Now let's add a simple representation of the photochemical reactions that produce ozone (O_3) and other constituents of photochemical smog. Hydrocarbons and other organic compounds that readily vaporize are called *volatile organic compounds* (VOCs). VOCs react with NO_x in the presence of sunlight to produce photochemical smog:

$$\text{VOCs} + NO_x + \text{Sunlight} \rightarrow \text{Photochemical smog } (O_3 + \text{etc.}) \quad (5)$$

To distinguish between the ozone that is formed near the ground by (5) from the ozone that exists in the stratosphere (next chapter), the designations *ground-level ozone* and *stratospheric ozone* are sometimes used. As we shall see, ground-level ozone is harmful to our health, whereas stratospheric ozone protects our health by shielding us from ultraviolet radiation from the sun.

Reactions (1) to (5) are greatly simplified, of course, but they do introduce the principal players in air pollution: CO, NO_x, SO_x, O_3, VOCs, Hg, Pb, and particulate matter.

One way to approach emissions and controls of air pollutants is to categorize the sources as being *mobile* sources or *stationary* sources. Mobile sources include highway vehicles (automobiles and trucks), and other modes of transportation, including railroads, aircraft, farm vehicles, and boats and ships. Stationary sources are often categorized as stationary *fuel combustion,* which includes electric power plants and industrial energy systems; *industrial processes,* such as metals processing, petroleum refineries, and other chemical and allied product manufacturing; and *miscellaneous sources.* Very roughly speaking, mobile sources are responsible for most of the CO and almost half of the NO_x, whereas stationary sources are responsible for most of the SO_x, Hg, particulates, and VOCs, along with a bit more than half of the NO_x. As we shall see, regulatory approaches to controlling emissions, as well as technologies used to do so, are quite different for mobile and stationary sources.

3 The Clean Air Act

Initial efforts on the part of the U.S. Congress to address the nation's air pollution problem began with the passage of the Air Pollution Control Act of 1955. Although it provided funding only for research and not control, it was an important milestone because it opened the door to federal participation in efforts to deal with air pollution. Until that time, it had been thought to be a state and local problem. This was followed by a series of legislative actions by Congress that included the Clean Air Act Amendments of 1963, 1966, 1970, 1977, and 1990, all of which are sometimes lumped together and referred to as simply the *Clean Air Act* (CAA). More recently, Congress unsuccessfully attempted to modify the CAA with the controversial *Clear Skies Act of 2003,* which many analysts say would have weakened the CAA. And, in 2005, the Environmental Protection Agency (EPA) established a new "Clean Air Mercury Rule" to create a market-based cap-and-trade program to reduce mercury emissions from power plants. This Rule was challenged, but the courts reaffirmed it in 2006.

Much of the real structure to the Clean Air Act was established in the 1970 Amendments. In those amendments, the U.S. Environmental Protection Agency

(U.S. EPA) was required to establish *National Ambient Air Quality Standards* (NAAQS), and states were required to submit *State Implementation Plans* (SIPs) that would show how they would meet those standards. In addition, the Act required *New Source Performance Standards* (NSPS) to be established that would limit emissions from certain specific types of industrial plants and from motor vehicles.

Air Quality and Emission Standards

The Clean Air Act requires the EPA to establish both air quality standards (NAAQS) and emission standards (NSPS), and it is important to keep in mind the fundamental difference between the two. Ambient air quality standards are acceptable *concentrations* of pollution in the atmosphere, whereas emission standards are allowable *rates* at which pollutants can be released from a source.

National Ambient Air Quality Standards have been established by the EPA at two levels: *primary* and *secondary*. Primary standards are required to be set at levels that will protect public health and include an “adequate margin of safety,” regardless of whether the standards are economically or technologically achievable. Primary standards must protect even the most sensitive individuals, including the elderly and those already suffering from respiratory and cardiopulmonary disorders. NAAQS are, therefore, conceptually different from Maximum Contaminant Levels (MCLs) that have been set for drinking water. Recall that the Safe Drinking Water Act requires the EPA to balance public health benefits with technological and economic feasibility in establishing drinking water MCLs.

Secondary air quality standards are meant to be even more stringent than primary standards. Secondary standards are established to protect public welfare (e.g., structures, crops, animals, fabrics, etc.). Given the difficulty in achieving primary standards, secondary standards have played almost no role in air pollution control policy, and, in fact, they have usually been set at the same levels as primary standards.

National Ambient Air Quality Standards now exist for six *criteria* pollutants: carbon monoxide (CO), lead (Pb), nitrogen dioxide (NO_2), ground-level ozone (O_3), sulfur dioxide (SO_2), and two categories of particulate matter (Table 1). The Clean Air Act requires that the list of criteria pollutants be reviewed periodically and that standards be adjusted according to the latest scientific information. Past reviews have modified both the list of pollutants and their acceptable concentrations. For example, lead has been added to the list, and the broad category of hydrocarbons has been dropped. The older standard for a general category of “oxidants” has been replaced by a more specific ozone standard. And in 2005, the one-hour-averaging-time O_3 standard was dropped (in all but a very limited number of regions of the country), but the eight-hour standard remains. Also, very importantly, the original particulate standard did not refer to the size of particulates, but as of 1987, it was modified to include only particles with an aerodynamic diameter of less than, or equal to, 10 micrometers (PM_{10}). Ten years later, in 1997, an additional fine particles category for particulates smaller than 2.5 microns was added ($PM_{2.5}$). In 2006, the short-term $PM_{2.5}$ standard was cut from 65 to $35 \mu g/m^3$, and the long-term standard was revoked.

The rules for determining whether a region of the country is in compliance with NAAQS are rather complicated. For CO and SO_x, for example, the standards

TABLE 1

National Ambient Air Quality Standards (NAAQS) and California Standards

Pollutant	Averaging Times	NAAQS Primary	California Standard	Most Relevant Health Effects
Carbon monoxide (CO)	8-hour	9 ppm	9 ppm	Aggravation of angina pectoris; decreased exercise tolerance; risk to fetuses
	1-hour	35 ppm	20 ppm	
Lead	3-months	1.5 μg/m^3		Impaired blood formation; infant development effects
Nitrogen dioxide (NO_2)	Annual mean	0.053 ppm		Aggravation of respiratory disease;
	1-hr	—	0.25 ppm, 1-hr	atmospheric discoloration
Particulate matter (PM_{10})	Annual mean	—	20 μg/m^3	Aggravated asthma; coughing; painful breathing; chronic bronchitis; decreased lung function; premature death in heart and lung patients
	24-hour	150 μg/m^3	50 μg/m^3	
Particulate matter ($PM_{2.5}$)	Annual mean	15.0 μg/m^3	12 μg/m^3	
	24-hour	35 μg/m^3		
Ozone (O_3)	8-hour	0.08 ppm		Decreased pulmonary function; surrogate for eye irritation
	1-hour		0.09 ppm	
Sulfur dioxide (SO_2)	Annual mean	0.03 ppm		Wheezing, shortness of breath, chest tightness; premature deaths
	24-hour	0.14 ppm	0.04 ppm	
	1-hour	—	0.25 ppm	

cannot be exceeded more than once per calendar year. For most of the others, compliance is based on a running three-year average. It should be noted that the law allows states to establish standards that are more stringent than the NAAQS, which California has done.

For the gases in Table 1, the concentrations may be expressed in parts per million by volume (ppm) as well as in mass per unit volume (μg/m^3 or mg/m^3). The volumetric units (ppm) are preferred since those are independent of pressure and temperature. The mass-per-volume concentrations assume a temperature of 25°C and 1 atmosphere of pressure. The following example illustrates the procedure.

EXAMPLE 1 Air Quality Standards as Mass per Unit Volume

Table 1 shows that California's air quality one-hour standard for NO_2 is 0.25 ppm. Express this as a concentration by mass at 25°C and 1 atm pressure.

Solution 1 mol of an ideal gas at 1 atm and 25°C occupies a volume of 24.465 L (24.465×10^{-3} m^3). The molecular weight of NO_2 is

$$\text{MW} = 1 \times 14 + 2 \times 16 = 46 \text{ g/mol}$$

therefore

$$[\mathrm{NO_2}] = \frac{0.25 \times 10^{-6}\ \mathrm{m^3(NO_2)/m^3(air)} \times 46\ \mathrm{g/mol}}{24.465 \times 10^{-3}\ \mathrm{m^3/mol}}$$

$$= 470 \times 10^{-6}\ \mathrm{g/m^3} = 470\ \mu\mathrm{g/m^3}$$

Notice that parts per million by volume (ppm) is a dimensionless volume fraction, independent of temperature and pressure, so it is the preferred measure.

Besides establishing National Ambient Air Quality Standards, the Clean Air Act also requires the EPA to establish emission standards for mobile sources such as cars and trucks. The 1970 Amendments to the Clean Air Act gave the auto industry a five-year deadline to achieve a 90-percent reduction in emissions from new cars. At the time, it was not even known whether such reductions were technologically possible, let alone how they could be implemented in such a short period of time. This "technology-forcing" legislation predictably led to numerous clashes between Congress and the automobile industry, and the standards were modified and delayed for many years. Later in this chapter, the emission controls that eventually were implemented will be described in some detail.

The EPA is also required to establish emissions standards for certain large stationary sources such as fossil-fuel-fired power plants, incinerators, Portland cement plants, nitric acid plants, petroleum refineries, sewage treatment plants, and smelters of various sorts. The methods of achieving the emission standards for stationary sources will be explored later in this chapter.

The Clean Air Act Amendments of 1977

The goal of the 1970 Amendments was to attain clean air by 1975, as defined by the NAAQS, with allowable extensions in certain circumstances until 1977. For a number of reasons, only about one-third of the air quality control regions in the nation were meeting the standards by 1977. This forced Congress to re-address the problem through the Clean Air Act Amendments of 1977. Besides extending the deadlines, the 1977 Amendments had to deal with two important questions. First, what measures should be taken in *nonattainment areas* that were not meeting the standards? Second, should air quality in regions where the air is cleaner than the standards be allowed to degrade toward the standards, and if so, by how much?

For nonattainment areas, the 1970 Act appeared to prohibit any increase in emissions whatsoever, which would have eliminated industrial expansion and severely curtailed local economic growth. To counter this, the EPA adopted a policy of *emission offsets*. To receive a construction permit, a major new source of pollution in a nonattainment area must first find ways to reduce emissions from existing sources. The reductions, or offsets, must exceed the anticipated emissions from the new source. The net effect of this offset policy is that progress is made toward meeting air quality standards in spite of new emission sources being added to the airshed.

Offsets can be obtained in a number of ways. For example, emissions from existing sources in the area might be reduced by installing better emission controls on equipment that may, or may not, be owned by the permit seeker. In some cases, a permit seeker may simply buy out existing emission sources and shut them down. Emission offsets can be "banked" for future use, or they can be sold or traded to other companies for whatever the market will bear. In addition to offsets, new sources in nonattainment areas must use emission controls that yield the *Lowest Achievable Emission Rate* (LAER) for the particular process. LAER technology is based on the most stringent emission rate achieved in practice by similar sources, regardless of the economic cost or energy impacts.

The 1970 Amendments were not specific about regions that were cleaner than ambient standards required and in fact appeared to allow air quality to deteriorate to those standards. The 1977 Amendments settled the issue of whether or not this would be allowed by establishing the concept of *prevention of significant deterioration* (PSD) in attainment areas. Attainment areas are put into one of three classes and the amount of deterioration allowed is determined by the class. In PSD areas, *Best Available Control Technology* (BACT) is required on major new sources. BACT is less stringent than LAER, as it does allow consideration of economic, energy, and environmental impacts of the technology, but it can be more strict than allowed by NSPS. In all PSD areas, the allowable increments of air quality degradation are constrained by the NAAQS. That is, in no circumstance would air quality be allowed to deteriorate to the point where the area is no longer in compliance with ambient air quality standards.

The Clean Air Act Amendments of 1990

The Clean Air Act Amendments of 1990 significantly strengthened the EPA's efforts to assure healthful air in the United States and broadened its scope to include control of pollutants that affect a global problem—stratospheric ozone depletion. Principal changes in the Act included the following:

- Creation of a new *Acid Rain Program* (Title IV)
- New requirements for nonattainment areas (Title I)
- Tightened automobile emission standards and new fuel requirements (Title II)
- New toxic air pollution controls (Title I)
- Phase-out schedule for ozone-depleting substances (Title VI)

The Acid Rain Program. One of the most important shortcomings of the Clean Air Act before the 1990 Amendments was its inability to deal effectively with acid rain (or, more correctly, acid deposition). As will be described later, acid deposition results from emissions of sulfur dioxide (SO_2) that convert to sulfuric acid and nitrogen oxides (NO_x) that become nitric acid. The 1990 Amendments approached this problem in a new and creative way by creating a market-based "cap-and-trade" system to control emissions.

The cap-and-trade system is based on a more flexible *allowance* system, where one allowance authorizes the owner to emit 1 ton of SO_2. Large coal-fired power plants are not allowed to emit any more tons of SO_2 than the number of allowances they own. If insufficient allowances are owned to cover emissions, the owners are subject to an excess-emissions penalty. The intent is for these allowances to be bought and sold or banked in the same way that other commodities are traded. New sources that have no allowances have to purchase them from existing sources or from annual EPA auctions. The idea, of course, is that major sources will find the least-expensive ways to cut their emissions and then sell some of their allowances to others who can't reduce their emissions as cheaply. The goal is a least-cost emission limitation that allows sources the flexibility they need to make the most cost-effective choices. By ratcheting down the number of allowances that the EPA issues each year, the cap placed on total emission is intended to result in a 50 percent reduction in SO_2 emissions from 1980 levels by 2010.

The cap-and-trade system went into effect in 1995, so we now have more than a decade of results to use to evaluate how well it is working. A recent study by Chestnut and Mills (2005) concludes the program is accomplishing much more than expected, at far less cost than anticipated. Between 1993 and 2002, the EPA's Acid Rain Program resulted in an SO_2 emission reduction of over 30 percent, and it is well on its way to meeting its 2010 goal of cutting emissions in half compared to 1980 emissions.

Success with the SO_2 cap-and-trade program led to a similar effort to control NO_x emissions called the *NO_x Budget Trading Program* (NBP). The program became operational in 2003, and early results are promising. In 2005, out of the 2,570 NO_x emitters affected under NBP, all but 3 sources were in compliance, and emissions were 11 percent lower than 2004 and 57 percent lower than in 2000 before implementation of the NBP (U.S. EPA, 2006a).

By switching to lower-sulfur coal (which also contains less mercury) and installing scrubbers, not only have SO_x and NO_x been reduced, as Title IV intended, but also highly toxic mercury emissions have been cut as well as ambient levels of $PM_{2.5}$ and ozone. The combined annual health and welfare benefits in 2010 of the Acid Rain Program in the United States and southern Canada are estimated to be over \$120 billion (over 90 percent of which is attributable to projected reductions in $PM_{2.5}$), whereas the annual costs of SO_x and NO_x controls is estimated at only \$3 billion (Chestnut and Mills, 2005). Table 2 summarizes the emission reductions projected for 2010 attributable to Title IV.

National Emission Standards for Hazardous Air Pollutants (NESHAP). Earlier versions of the Clean Air Act established NESHAP, and through it, a small list of pollutants (asbestos, benzene, beryllium, coke oven emissions, inorganic arsenic, mercury, radionuclides, and vinyl chloride) were controlled. The 1990 Amendments extend that list to include 189 pollutants listed in the legislation. The list can be changed by the EPA, and if there are objections to any additions, the burden of proof is on the petitioner, who must show that the chemical may not

TABLE 2

Annual U.S. Power Generation Emissions Reductions with Title IV

Emissions from U.S. Power Generation	SO_2 (Million Tons)	NO_x (Million Tons)	Mercury (Tons)
2010 emissions without Title IV	17.3	7.6	52
2010 emissions with Title IV	9.3	5	42
The difference Title IV makes	8.0	2.6	10

Source: Chestnut and Mills, 2005.

reasonably be anticipated to cause any adverse human health or environmental effects. Emission standards for chemicals on the list are based on the *Maximum Achievable Control Technology* (MACT).

Ozone Depleting Substances. Finally, Title VI of the 1990 Amendments was written to protect the stratospheric ozone layer by phasing out ozone-depleting substances such as chlorofluorocarbons (CFCs). The phase-out was mandated to be at least as strict as that required by the international treaty known as the Montreal Protocol.

4 Progress in Controlling Emissions and Improving Air Quality

The Clean Air Act has definitely helped reduce emissions of the key criteria pollutants, and the resulting air quality improvements are significant. As Figure 1 indicates, emissions of CO, NO_x, SO_x, and VOCs have been cut in half compared with 1970 levels, and with the 1995 elimination of lead as an antiknock additive in gasoline, those emissions have dropped 98 percent. PM_{10} is down significantly, but NO_x remains a more difficult pollutant to control, with emissions down only 17 percent. Since ozone is a secondary pollutant—that is, it is not actually emitted but rather is formed by photochemical reactions that take place in the atmosphere—it is not shown in this emissions figure. Instead, VOCs are used as indicators of the potential for ozone formation.

Those emission reductions over the past 30 years are even more impressive when compared with the growth in key factors that would normally be tied to pollutant emissions. As shown in Figure 2, while an aggregate of these six key pollutants has dropped by half, U.S. population and energy consumption have grown by about 40 percent, while GDP and vehicle miles traveled have more than doubled.

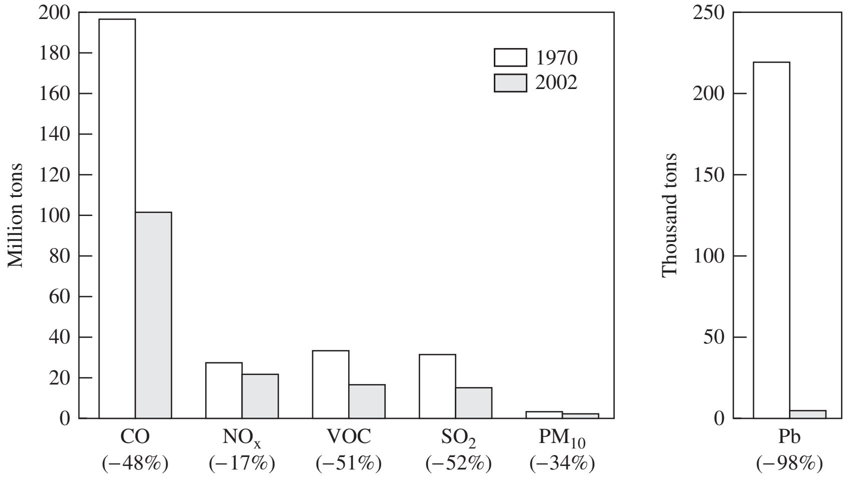


FIGURE 1 Comparison of U.S. emissions in 2002 with those of 1970.
(*Source:* U.S. EPA, *Air Trends Report,* 2003.)

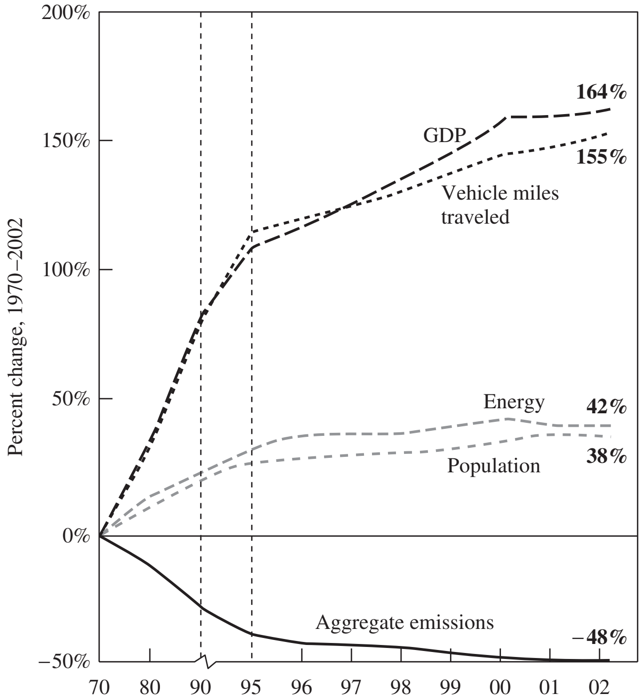


FIGURE 2 Emission reductions of an aggregate of six principal pollutants compared with growth in GDP, vehicle miles traveled, energy, and population.
(*Source:* U.S. EPA, *Air Trends Report,* 2003.)

The NAAQS form the basis for the EPA's *Air Quality Index* (AQI), which is used to report to the public an overall assessment of a given day's air quality. The AQI integrates air quality data for five of the criteria pollutants into a single number that represents the worst daily air quality in an urban area. Descriptive terms ranging from "good" to "hazardous" are assigned to ranges of the AQI, as shown in Table 3, and messages such as the following may be issued in local weather reports. An AQI value of 100 indicates that at least one pollutant has just reached its ambient air quality standard on that day. An index of 500 indicates an imminent and substantial endangerment to public health.

- **Good** (0–50): No health impacts are expected in this range.
- **Moderate** (51–100): Unusually sensitive people should consider reducing prolonged or heavy exertion.
- **Unhealthy for Sensitive Groups** (101–150): Active children and adults, and people with lung disease, such as asthma, should reduce prolonged or heavy exertion outdoors. The general public is not likely to be affected.
- **Unhealthy** (151–200): Everyone may begin to experience health effects. Members of sensitive groups may experience more serious health effects.
- **Very Unhealthy** (201–300): Air quality in this range triggers a health alert, meaning everyone may experience more serious health effects.
- **Hazardous** (301–500): Air quality in this range triggers health warnings of emergency conditions. The entire population is more likely to be affected.

TABLE 3

Air Quality Index (AQI) Categories and Ranges

Category	AQI	8-hr O_3 (ppm)	1-hr O_3 (ppm)	24-hr $PM_{2.5}$ ($\mu g/m^3$)	24-hr PM_{10} ($\mu g/m^3$)	8-hr CO (ppm)	24-hr SO_2 (ppm)	1-hr NO_2 (ppm)
Good	0–50	0.000–0.064	—	0.0–15.4	0–54	0.0–4.4	0.000–0.034	—
Moderate	51–100	0.065–0.084	—	15.5–40.4	55–154	4.5–9.4	0.035–0.144	—
Unhealthy for Sensitive Groups	101–150	0.085–0.104	0.125–0.164	40.5–65.4	155–254	9.5–12.4	0.145–0.224	—
Unhealthy	151–200	0.105–0.124	0.165–0.204	65.5–150.4	255–354	12.5–15.4	0.225–0.304	—
Very Unhealthy	201–300	0.125–0.374	0.205–0.404	150.5–250.4	355–424	15.5–30.4	0.305–0.604	0.65–1.24
Hazardous	301–400	use 1-hr	0.405–0.504	250.5–350.4	425–504	30.5–40.4	0.605–0.804	1.25–1.64
	401–500	use 1-hr	0.505–0.604	350.5–500.4	505–604	40.5–50.4	0.805–1.004	1.65–2.04

The number of days that the AQI is above 100 is often used to describe progress that is being made in air quality. Since almost all such days are triggered by high levels of ozone, special attention is paid to that criteria pollutant. A special network of *Photochemical Assessment Monitoring Stations* (PAMs) has been established in ozone nonattainment areas that report detailed data for ozone and its precursors. Figure 3 shows that PAMS sites, in general, have not shown much reduction in the number of days with AQI above 100 in the 1990s; however, during the same period of time, California saw a reduction of over half.

While the gains shown in Figures 1 to 3 are impressive, roughly half of the U.S. population lives in counties where at least one NAAQS continues to be exceeded. The pollutant most responsible for those excessive exposures is ground-level ozone (Figure 4). Equally disturbing is the fact that 20 percent live in counties where the more lethal $PM_{2.5}$ pollution doesn't meet its air quality standard.

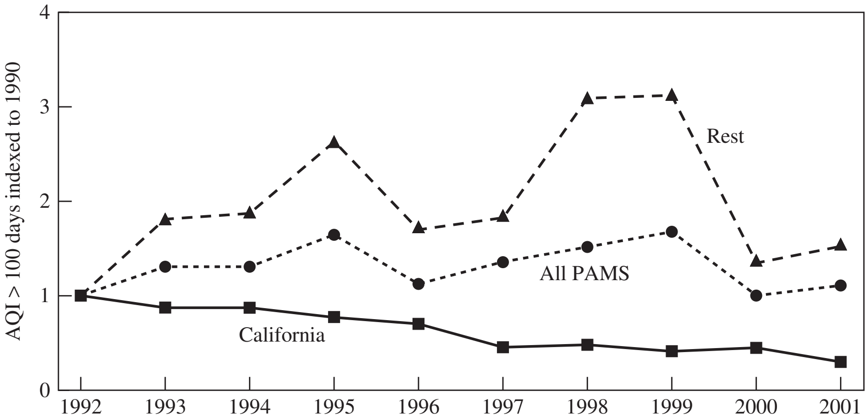


FIGURE 3 Annual number of days when AQI was greater than 100 for the most polluted American cities, indexed to 1990.
(*Source:* U.S. EPA, *Air Trends Report,* 2003.)

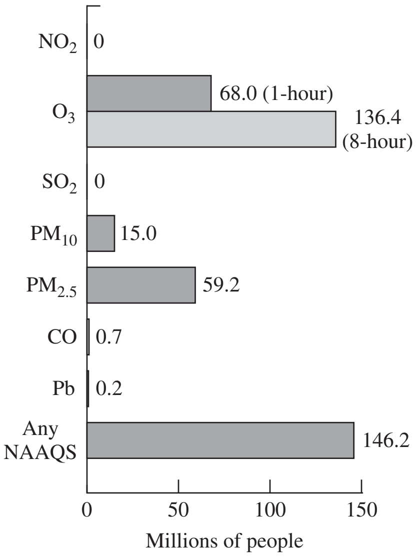


FIGURE 4 Numbers of people living in counties with pollution concentrations above NAAQS standards out of a 2002 U.S. population of 290 million.
(*Source:* U.S. EPA, *Air Trends Report,* 2003.)

5 Criteria Pollutants

Given the ongoing focus of the Clean Air Act, most of the monitoring of emissions, concentrations, and effects of air pollution has been directed toward the six criteria pollutants: ground-level ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), small particulates (PM_{10} and now $PM_{2.5}$), nitrogen dioxide (NO_2), and lead (Pb). The original ambient air quality standards for these pollutants were based on extensive documentation assembled and published by the EPA in a multivolume set of *Air Quality Criteria* documents, from which the name criteria pollutants originated.

Carbon Monoxide

Over two-thirds of the mass of all the pollutant emissions in the United States is the colorless, odorless, tasteless, poisonous gas, carbon monoxide (CO). It is produced when carbonaceous fuels are burned under less than ideal conditions. Incomplete combustion, yielding CO instead of CO_2, results when any of the following four variables are not kept sufficiently high: (1) oxygen supply, (2) combustion temperature, (3) gas residence time at high temperature, and (4) combustion chamber turbulence. These parameters are generally under much tighter control in stationary sources such as power plants than in motor vehicles, and CO emissions are correspondingly less. For example, power plants, which are designed and managed for maximum combustion-efficiency, produce less than 1 percent of all CO emissions in spite of the fact that they consume about 30 percent of our fossil fuel.

Over 80 percent of total CO emissions are from the transportation sector, and almost all of the CO in urban areas comes from motor vehicles. Hourly atmospheric concentrations of CO over our cities often reflect city driving patterns: peaks occur on weekdays during the morning and late afternoon rush hours, while on weekends there is typically but one lower peak in the late afternoon. Personal exposure to CO is very much determined by the proximity of motor vehicle traffic, with some occupational groups such as cab drivers, police, and parking lot attendants receiving far higher than average doses. Carbon monoxide emissions from motor vehicles decreased by 41 percent in the past 20 years (1983–2002) despite a doubling in vehicle-miles traveled. The impact of tightening emission controls, coupled with a drop in the fraction of older, more polluting cars still on the road, is clearly evident.

At levels of CO that occur in urban air, there are apparently no detrimental effects on materials or plants; those levels can, however, adversely affect human health. Carbon monoxide is an asphyxiant; it interferes with the blood's ability to carry oxygen from the lungs to the body's organs and tissues. When inhaled, it readily binds to hemoglobin in the bloodstream to form carboxyhemoglobin (COHb). Hemoglobin, in fact, has a much greater affinity for carbon monoxide than it does for oxygen, so that even small amounts of CO can seriously reduce the amount of oxygen conveyed throughout the body. With the bloodstream carrying less oxygen, brain function is affected and heart rate increases in an attempt to offset the oxygen deficit.

The usual way to express the amount of carboxyhemoglobin in the blood is as a percentage of the saturation level, %COHb. The amount of COHb formed in the blood is related to the CO concentration and the length of time exposed, as suggested in the following.

$$\%\mathrm{COHb} = \beta(1 - e^{\gamma t})[\mathrm{CO}] \qquad (6)$$

where

$\%COHb$ = carboxyhemoglobin as a percent of saturation
$[CO]$ = carbon monoxide concentration in ppm
γ = 0.402/hr
β = 0.15%/ppm CO
t = exposure time in hours

EXAMPLE 2 **Federal Standard for CO**

Estimate the %COHb expected for a one-hour exposure to 35 ppm (the federal standard).

Solution From (6):

$$\%COHb = 0.15\%/\text{ppm}[1 - \exp(0.402/\text{hr} \times 1\ \text{hr})][35\ \text{ppm}] = 1.7\ \text{percent}$$

Physiological effects can be noted at small percentages of COHb, increasing in severity as the concentration increases. The elderly, the fetus, and individuals who suffer from cardiovascular disease, particularly those with angina pectoris (a heart condition characterized by chest pain) or peripheral vascular disease, are the most sensitive to COHb since the heart must work harder in an attempt to offset the reduction in oxygen. Studies of patients with angina have shown an earlier-than-usual onset of pain during exercise when levels are as low as 2 percent COHb. In Example 3, it was calculated that an individual breathing CO at the federal ambient air quality standard of 35 ppm for one hour would be likely to reach 1.7 percent COHb. That is very close to the level at which health effects have been noted, and the federal standard has been criticized as a result. California's CO standard has been set lower (20 ppm) in an attempt to assure less than 2 percent COHb.

The reduction of oxygen in the bloodstream also affects the brain's ability to perceive and react. At 2.5 percent COHb, studies have shown an impairment in time-interval discrimination. (Subjects were less able to distinguish the duration of a tone signal.) Studies have also shown that at 5 percent, psychomotor response times may be affected, and patients with heart disease experience increased physiological stress. At 5 to 17 percent, manual dexterity, ability to learn, and performance in complex sensorimotor tasks such as driving are diminished. At 10 percent, headache is common, and at 30 percent COHb, most people will experience dizziness, headache, fatigue, and impaired judgment. Concentrations above 60 percent produce loss of consciousness and death if exposure continues.

Carbon monoxide concentrations near busy roadways frequently range from 5 to 50 ppm, and measurements made on congested highways indicate that drivers can be exposed to CO concentrations of 100 ppm. Carbon monoxide is an important pollutant in indoor air as well. Cigarette smoke contains more than 400 ppm CO, and smokers frequently have COHb levels between 5 and 10 percent. Tobacco smoke in bars and restaurants often raises indoor CO levels to 20 to 30 ppm, which is close to the one-hr ambient standard (Wadden and Scheff, 1983). Fortunately, COHb is removed from the bloodstream when clean air is breathed. Healthy subjects clear about half of the CO from their blood in three to four hours, so adverse effects are usually temporary.

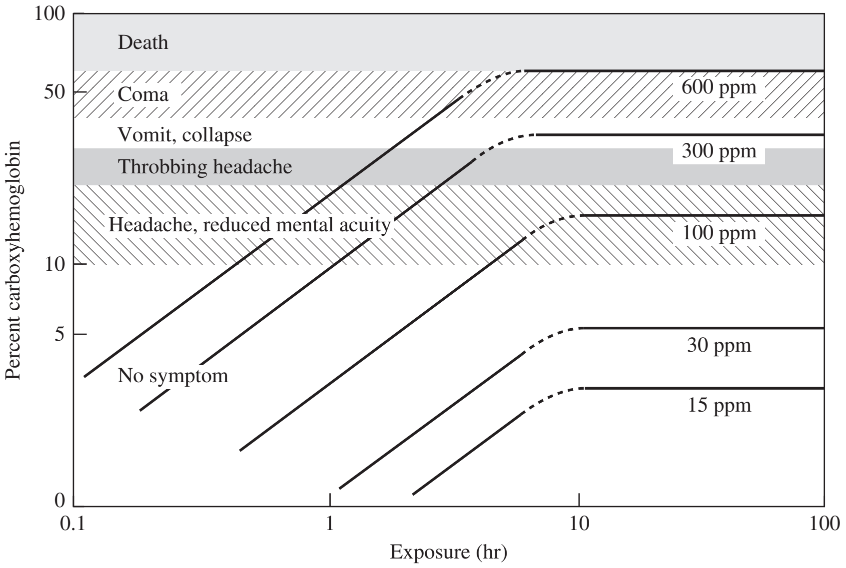


FIGURE 5 Effects of exposure to carbon monoxide.
(*Source:* Seinfeld, 1986.)

Figure 5 summarizes the relationship between exposure time and resulting percent carboxyhemoglobin, along with indications of some of the more severe impacts that can result. Notice the saturation effect for long exposures.

Oxides of Nitrogen

Although 7 oxides of nitrogen are known to occur, NO, NO_2, NO_3, N_2O, N_2O_3, N_2O_4, and N_2O_5—the only two that are important air pollutants are nitric oxide (NO) and nitrogen dioxide (NO_2). As mentioned earlier, there are two sources of nitrogen oxides (or NO_x) when fossil fuels are burned. *Thermal* NO_x is created when nitrogen and oxygen in the combustion air are heated to a high enough temperature (about 1,000 K) to oxidize the nitrogen. *Fuel* NO_x results from the oxidation of nitrogen compounds that are chemically bound in the fuel molecules themselves. Different fuels have different amounts of nitrogen in them, with natural gas having almost none and some coal having as much 3 percent by weight. Both thermal NO_x and fuel NO_x can be significant contributors to the total NO_x emissions, but fuel NO_x often is the dominant source. In 2003, transportation accounted for 57 percent of emissions, whereas stationary fuel combustion emitted 37 percent.

About 95 percent of anthropogenic emissions of NO_x are in the form of NO, which is a colorless gas that has no known adverse health effects at concentrations found in the atmosphere. However, NO readily oxidizes to NO_2, and NO_2 can irritate the lungs, cause bronchitis and pneumonia, and lower resistance to respiratory infections. NO_x also can react with volatile organic compounds in the presence of sunlight to form photochemical oxidants that have adverse health consequences as well.

Nitrogen dioxide has other environmental consequences besides those directly associated with human health. It reacts with the hydroxyl radical (OH) in the atmosphere to form nitric acid (HNO_3) that corrodes metal surfaces and contributes to the acid rain problem. It also can cause damage to terrestrial plants, and it is a significant cause of eutrophication, especially in nitrogen-limited estuaries such as Chesapeake Bay. Nitrogen dioxide is responsible for the reddish-brown color in the smog that hovers over cities like Los Angeles.

Reductions in NO_x emissions have been harder to come by than reductions in other criteria pollutants. In fact, emissions increased for several decades after the original Clean Air Act was passed, due in part to early automobile emission controls that focused on reducing CO emissions at the expense of increasing NO_x. Since then, better technology and more stringent standards applied to new cars and light-duty trucks are beginning to turn the corner. Emissions of NO_x from these vehicles peaked in 1994 and are now on the decline, but emissions from heavy-duty gasoline vehicles, diesel trucks, and nonroad vehicles are still increasing (U.S. EPA, 2003). Increasing transportation-sector emissions have been offset by reductions in emissions from power plants and other stationary sources.

Despite difficulties in controlling NO_x emissions, every monitoring location in the United States meets the NAAQS for NO_2, and a large majority of the sites have concentrations of less than half of the standards.

Volatile Organic Compounds (VOCs)

This class of compounds consists of volatile compounds that enter the atmosphere when solvents, fuels, and other organics evaporate, along with unburned and partially burned hydrocarbons that are emitted from tailpipes and smoke stacks when fossil fuels are not completely combusted. Regulation of VOCs is based on their role as precursors to the formation of photochemical smog. As such, the less-reactive hydrocarbon, methane, is often excluded, and the remainder is referred to as *Total Non-Methane Organic Compounds* (TNMOC).

In addition to anthropogenic sources (manmade, excluding wildfires and prescribed burnings), there are natural sources of reactive hydrocarbons such as oaks, spruce, eucalyptus, and citrus trees that emit isoprene; and maple, hickory, pine, spruce, and fir trees that emit monoterpenes. In some regions of the country, mostly in the east, these biogenic sources can be greater than anthropogenic VOC emissions.

The transportation sector is responsible for almost half of anthropogenic VOC emissions, and despite rapidly rising miles driven in motor vehicles, emissions decreased by 39 percent from 1993 to 2003. Reformulated gasoline with lower volatility coupled with the decline in the percentage of dirtier, older vehicles on the road account for much of the drop. As emissions from motor vehicles become less significant, other sources such as gasoline-powered lawnmowers, outboard motors, barbeque starter fluids, and oil-based paints begin to look much more important and are beginning to be regulated as well.

Industrial sources account for almost all of the other half of VOC emissions, with much of that again being caused by vaporization of hydrocarbons. Less than 2 percent of VOCs result from fossil-fuel combustion in power plants and industrial boilers.

In terms of direct human health effects of VOCs, it is our exposure to a long list of volatile and toxic compounds inside our buildings that provides our greatest risk.

Photochemical Smog and Ozone

When NO_x, VOCs, and sunlight come together, they can initiate a complex set of reactions that produce a number of secondary pollutants known as photochemical oxidants. Ozone (O_3) is the most abundant of the photochemical oxidants, and it is the one for which an ambient air quality standard has been written. Although it is responsible for many of the undesirable properties of photochemical smog, from chest constriction and irritation of the mucous membrane in people to the cracking of rubber products and damage to vegetation, it is not a cause of the eye irritation that is our most common complaint about smog. Eye irritation is caused by other components of photochemical smog, principally formaldehyde (HCHO), peroxybenzoyl nitrate (PBzN), peroxyacetyl nitrate (PAN), and acrolein (CH_2CHCOH).

In the very simplest of terms, we can express the formation of photochemical smog as

$$\text{VOCs} + NO_x + \text{Sunlight} \rightarrow \text{Photochemical smog} \tag{7}$$

The reaction in (7) only gives us the simplest overview. We will add a few details to give a sense of the key reactions involved, but a complete analysis is far beyond the scope of this book.

The $NO—NO_2—O_3$ Photochemical Reaction Sequence. Consider some of the important reactions involving NO_x without the complications associated with the added hydrocarbons. We can begin with the formation of NO during combustion (for simplicity, we just show the thermal NO_x reaction):

$$N_2 + O_2 \rightarrow 2\ NO \tag{8}$$

The nitric oxide thus emitted can oxidize to NO_2:

$$2\ NO + O_2 \rightarrow 2\ NO_2 \tag{9}$$

If sunlight is available, a photon with the right amount of energy can decompose NO_2 in a process called *photolysis*.

$$NO_2 + h\nu \rightarrow NO + O \tag{10}$$

where $h\nu$ represents a photon (with wavelength $\lambda < 0.39\ \mu m$). The freed atomic oxygen (O) can then combine with diatomic oxygen (O_2) to form ozone (O_3):

$$O + O_2 + M \rightarrow O_3 + M \tag{11}$$

where M represents a molecule (usually O_2 or N_2 since they are most abundant in air) whose presence is necessary to absorb excess energy from the reaction. Without M, the ozone would have too much energy to be stable, and it would dissociate back to O and O_2.

Ozone can then convert NO back to NO_2,

$$O_3 + NO \rightarrow NO_2 + O_2 \tag{12}$$

Notice the general tendency for NO_2 to create O_3 (10) and (11), whereas NO tends to destroy O_3 (12).

Equations (8) to (12) help explain the sequence of stages through which atmospheric NO, NO_2, and O_3 progress on a typical smoggy day. They suggest NO

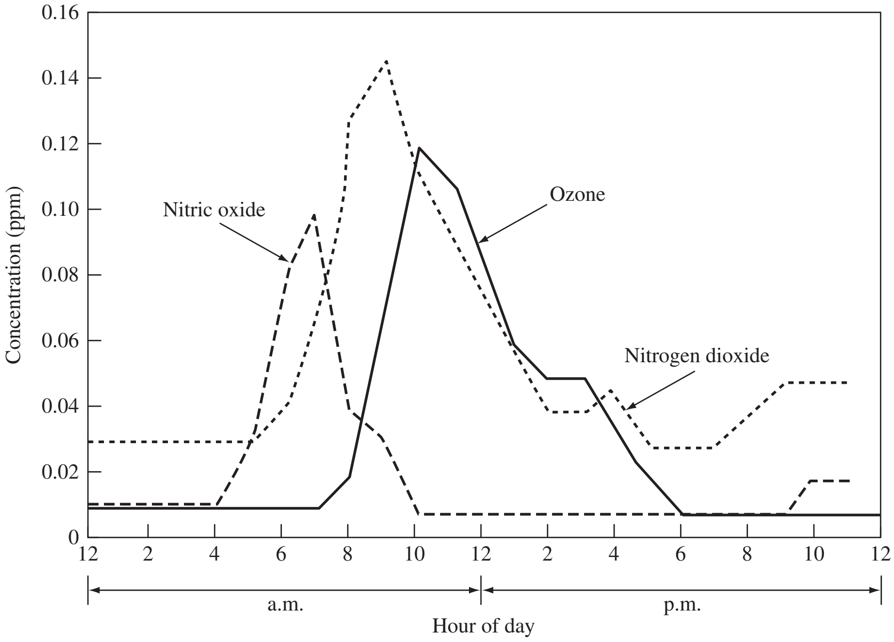


FIGURE 6 Diurnal variation of NO, NO_2, and O_3 concentrations in Los Angeles on July 19, 1965.
(*Source:* U.S. HEW, 1970.)

concentrations would rise as early morning traffic emits its load of NO. Then, as the morning progresses, we would expect to see a drop in NO and a rise in NO_2 as NO gets converted to NO_2. As the sun's intensity increases toward noon, the rate of photolysis of NO_2 increases; thus, NO_2 begins to drop while O_3 rises. Ozone is so effective in its reaction with NO (12) that as long as there is O_3 present, NO concentrations do not rise through the rest of the afternoon even though there may be new emissions.

The nitrogen dioxide photolytic cycle helps provide an explanation for the sequence NO to NO_2 to O_3 observable in both laboratory smog chambers and in cities such as Los Angeles (see Figure 6). A careful analysis of the reactions, however, predicts O_3 concentrations that are much lower than those frequently found on smoggy days. If only the NO_2 photolytic cycle is involved, O_3 cannot accumulate in sufficient quantity to explain actual measured data. By introducing certain types of hydrocarbons into the cycle, however, the balance of production and destruction of O_3 can be upset, which allows more O_3 to accumulate, closing the gap between theoretical and actual levels.

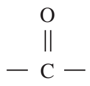
Hydrocarbons and NO_x. Expanding the nitrogen dioxide photolytic cycle to include hydrocarbons and other organics helps explain the increase in ozone above what would be predicted by the NO_x cycle alone. It also enables us to account for some of the other objectionable organic compounds found in photochemical smog.

The chemistry of photochemical smog is exceedingly complex. A multitude of organic chemicals are introduced to the atmosphere when fuels burn or volatiles evaporate, and many more are produced in the atmosphere as a result of chemical reactions. To help describe some of these reactions, let us begin with a very brief explanation of some of the nomenclature and notation used in organic chemistry.

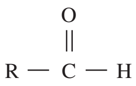
We begin by considering atoms or molecules with an odd number of electrons, called *free radicals*. Having an odd number of electrons means that one electron is not being used as a bonding electron to other atoms. For example, a carbon atom bonded to three hydrogens has one leftover electron. We can represent that molecule with the notation $CH_3^\bullet$, where the dot suggests the unpaired electron. Free radicals tend to be very reactive, and they are very important in the study of air pollution.

The *alkanes* are hydrocarbons in which each carbon forms single bonds with other atoms. The alkane series is the familiar sequence: methane (CH_4), ethane (C_2H_6), propane (C_3H_8), . . . ($C_nH_{2n} + 2$). If one of the hydrogens is removed from an alkane, the resulting free radical is called an *alkyl*. The alkyls then form a series beginning with methyl ($CH_3^\bullet$), ethyl ($C_2H_5^\bullet$), and so on. We could represent an alkyl with the general expression $C_nH_{2n+1}^\bullet$, but it is more convenient to call it simply $R^\bullet$.

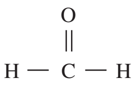
Another basic chemical unit that comes up over and over again in the study of photochemical smog is a *carbonyl,* a carbon atom with a double bond joining it to an oxygen as shown in the following diagram. A carbonyl with one bond shared with an alkyl, $R^\bullet$, and the other with a hydrogen atom, forms an *aldehyde*. Aldehydes can thus be written as RCHO. The simplest aldehyde is *formaldehyde* (HCHO), which corresponds to $R^\bullet$ being just a single hydrogen atom. A more complex aldehyde is acrolein, as shown. Both formaldehyde and acrolein are eye-irritating components of photochemical smog.


$-\overset{\overset{O}{\|}}{C}-$

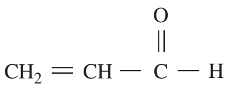
Carbonyl


$R-\overset{\overset{O}{\|}}{C}-H$

Aldehyde


$H-\overset{\overset{O}{\|}}{C}-H$

Formaldehyde


$CH_2=CH-\overset{\overset{O}{\|}}{C}-H$

Acrolein

Another important key to understanding atmospheric organic chemistry is the hydroxyl radical $OH^\bullet$, which is formed when atomic oxygen reacts with water:

$$O + H_2O \rightarrow 2\,OH^\bullet \tag{13}$$

The OH radical is extremely reactive, and its atmospheric concentration is so low that it has been difficult to detect. Nevertheless, it plays a key role in many reactions, including the oxidations of NO_2 to nitric acid and CO to CO_2.

$$OH^\bullet + NO_2 \rightarrow HNO_3 \tag{14}$$

$$OH^\bullet + CO \rightarrow CO_2 + H^\bullet \tag{15}$$

In fact, the hydroxyl radical is responsible for initiating the oxidation of VOCs and most of the other important atmospheric pollutants. It is therefore crucial to the natural removal of pollution from the atmosphere.

It was mentioned previously that the NO_2 photolytic cycle, by itself, underpredicts the observed concentrations of O_3. As that cycle is described in (10) to (12), the availability of NO_2 affects the rate of production of O_3, whereas the availability of NO affects the rate of destruction of O_3. The balance of O_3 production and destruction can be upset if there are other reactions that will enhance the rate of conversion of NO to NO_2. Any reactions that will help convert NO to NO_2 will

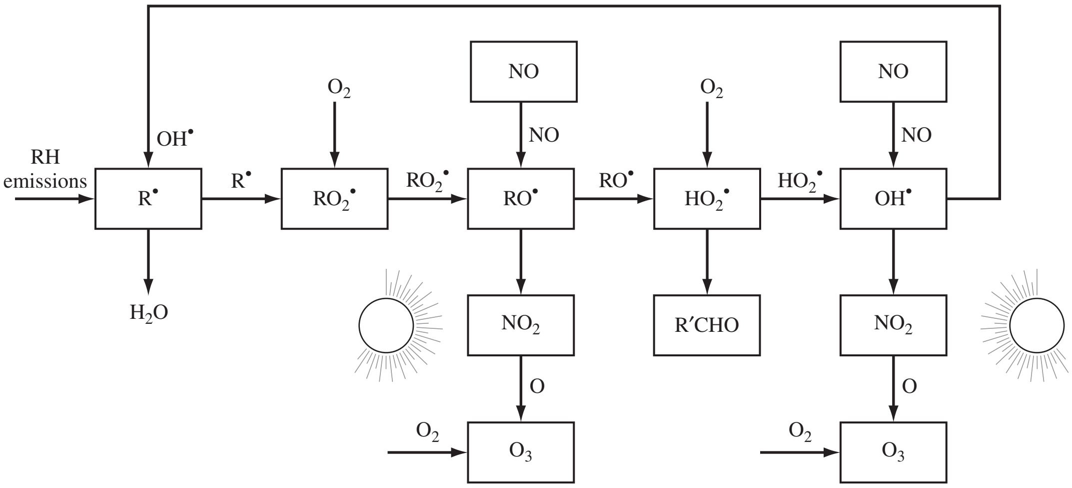


FIGURE 7 Showing one way that hydrocarbons can cause NO to convert to NO_2. Reducing NO slows the removal of O_3, while increasing NO_2 increases the production of O_3, so this cycle helps account for elevated atmospheric O_3 levels.

increase O_3 concentrations both by reducing the amount of NO available to destroy O_3, and increasing the amount of NO_2 available to create O_3.

The following three reactions provide one explanation for the way that hydrocarbons can enhance the rate of conversion of NO to NO_2 and hence increase O_3 concentrations. Starting with a hydrocarbon RH, we have

$$RH + OH^{\bullet} \rightarrow R^{\bullet} + H_2O \tag{16}$$

$$R^{\bullet} + O_2 \rightarrow RO_2^{\bullet} \tag{17}$$

$$RO_2^{\bullet} + NO \rightarrow RO^{\bullet} + NO_2 \tag{18}$$

These reactions are not complete, although they do show how hydrocarbons can help convert NO to NO_2, and thus increase O_3. As written, an already scarce hydroxyl $OH^{\bullet}$ is required to start the chain, and it appears to be destroyed in the process. Unless there is some way to rejuvenate that $OH^{\bullet}$, these reactions could not continue for long. The following pair of reactions shows one way that not only $OH^{\bullet}$ is regenerated, but also, in the process, how another NO is converted to NO_2. An aldehyde can also be formed from this pair of reactions:

$$RO^{\bullet} + O_2 \rightarrow HO_2^{\bullet} + R'CHO \tag{19}$$

$$HO_2^{\bullet} + NO \rightarrow NO_2 + OH^{\bullet} \tag{20}$$

where R′ is the hydrocarbon that balances (19).

The net effect of reactions (16) through (20) is that one hydrocarbon molecule converts two molecules of NO to NO_2 and produces an aldehyde R′CHO. The removal of NO by these reactions slows the rate at which O_3 is removed from the atmosphere, while the addition of NO_2 increases the rate at which it is produced, which allows higher levels of O_3 to accumulate in the air. These are summarized in Figure 7.

EXAMPLE 3 Ethane to Acetaldehyde

Suppose the hydrocarbon that begins reactions (16) through (20) is ethane (C_2H_6). Write the sequence of reactions.

Solution The hydrocarbon RH that appears in (16) is ethane, C_2H_6, so the free radical $R^\bullet$ is $C_2H_5^\bullet$. Reaction (16) thus becomes

$$C_2H_6 + OH^\bullet \rightarrow C_2H_5^\bullet + H_2O$$

Reactions 17 to 19 are

$$C_2H_5^\bullet + O_2 \rightarrow C_2H_5O_2^\bullet$$

$$C_2H_5O_2^\bullet + NO \rightarrow C_2H_5O^\bullet + NO_2$$

$$C_2H_5O^\bullet + O_2 \rightarrow HO_2^\bullet + CH_3CHO$$

Notice that R' in reaction 19 is thus $CH_3^\bullet$. Finally, (20) is

$$HO_2^\bullet + NO \rightarrow NO_2 + OH^\bullet$$

The CH_3CHO produced is an aldehyde called *acetaldehyde*. As we shall see, it plays a role in the formation of the eye irritant peroxyacetyl nitrate (PAN).

There are other ways that NO can be converted to NO_2. Carbon monoxide, for example, can do it too, in reactions that are similar to those given previously. Starting with CO and $OH^\bullet$, as in (15)

$$OH^\bullet + CO \rightarrow CO_2 + H^\bullet \tag{15}$$

The hydrogen atom then quickly combines with O_2 to form the hydroperoxyl radical, $HO_2^\bullet$:

$$H^\bullet + O_2 \rightarrow HO_2^\bullet \tag{21}$$

Now, going back to (20), NO is converted NO_2:

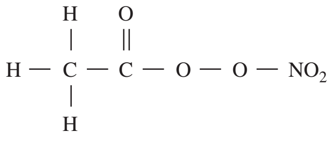
$$HO_2^\bullet + NO \rightarrow NO_2 + OH^\bullet \tag{20}$$

So we see one way that another of our criteria pollutants, CO, can contribute to the photochemical smog problem. By increasing the rate at which NO is converted to NO_2, CO aids in the accumulation of O_3.

We can also extend these relationships to show the formation of another of the eye irritants, peroxyacetyl nitrate (PAN).

$CH_3C(O)_2NO_2$

Peroxyacetyl nitrate (PAN)



$$H-\overset{\overset{\displaystyle H}{|}}{\underset{\underset{\displaystyle H}{|}}{C}}-\overset{\overset{\displaystyle O}{\|}}{C}-O-O-NO_2$$

Acetaldehyde, CH_3CHO, which Example 5 indicated can be formed from ethane emissions, can react with $OH^\bullet$:

$$CH_3CHO + O_2 + OH^\bullet \rightarrow CH_3C(O)O_2^\bullet + H_2O \tag{22}$$

The resulting acetylperoxy radical can react with NO_2 to create PAN:

$$CH_3C(O)O_2^\bullet + NO_2 \rightarrow CH_3C(O)O_2NO_2 \tag{23}$$

It must be pointed out once again that the preceding reactions are only a very limited description of the complex chemistry going on each day in the atmosphere over our cities.

Sources and Effects of Photochemical Oxidants. Since ozone and other photochemical oxidants are not primary pollutants, there is no direct way to specify emissions. An annual emissions inventory is kept by the EPA, however, for the precursors to photochemical oxidants, namely NO_x and VOCs. When all precursors are considered, especially if an adjustment is made for the sources' proximity to cities, the automobile is the dominant cause of photochemical smog. Conditions in Los Angeles, for example, are ideal for smog formation: the near total dependence on the automobile for transportation results in high hydrocarbon and NO_x emissions; long lasting atmospheric inversions restrict the vertical dispersion of pollutants; a ring of mountains nearly surrounds the city on three sides, reducing the horizontal dispersion; and an abundance of sunshine powers the photochemical reactions.

Photochemical smog is known to cause many annoying respiratory effects, such as coughing, shortness of breath, airway constriction, headache, chest tightness, and eye, nose, and throat irritation. These symptoms can be especially severe for asthmatics and others with impaired respiratory systems, but even healthy individuals who engage in strenuous exercise for relatively modest periods of time experience these symptoms at levels near the ambient air quality standard. Animal studies suggest that long-term exposures to ozone can lead to permanent scarring of lung tissue, loss of lung function, and reduced lung elasticity. As shown in Figure 4, 136 million people in the United States live in areas where ozone levels exceed national standards, far more than are exposed to any other criteria pollutant. As mentioned earlier, the Clean Air Act requires the EPA to set air quality standards that protect public health, without regard to cost. Accumulating evidence showed the two-decades-old 0.12 ppm ozone standard did not adequately protect some individuals, especially children and those with asthma. In 1997, after considerable public debate, the EPA strengthened the ozone standard to 0.08 ppm, contending that 1 million fewer incidences of decreased lung function in children each year could be realized if the standard is finally met.

Ozone has been shown to cause damage to tree foliage and to reduce growth rates of certain sensitive tree species. Ozone that enters plant leaves through stomatal openings produces byproducts that interfere with photosynthesis, which reduces crop yields for many important agricultural crops, such as corn, wheat, soybeans, and peanuts. Crop damage caused by motor-vehicle ozone alone has been estimated to cause losses of \$2 to \$3 billion per year (Murphy et al., 1999).

Particulate Matter

Atmospheric *particulate matter* consists of any dispersed matter, solid or liquid, in which the individual aggregates range from molecular clusters of 0.005 micrometers (μm) diameter to coarse particles up to about 100 μm (roughly the size of a human hair). As a category of criteria pollutant, particulate matter is extremely diverse and complex since size and chemical composition, as well as atmospheric concentration, are important characteristics.

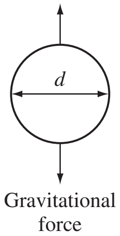
A number of terms are used to categorize particulates, depending on their size and phase (liquid or solid). The most general term is *aerosol,* which applies to any tiny particles, liquid or solid, dispersed in the atmosphere. Solid particles are called *dusts* if they are caused by grinding or crushing operations. Solid particles are called *fumes* if they are formed when vapors condense. Liquid particles may be called *mist*, or more loosely, *fog*. *Smoke* and *soot* are terms used to describe particles composed primarily of carbon that result from incomplete combustion. *Smog* is a term that was derived from smoke and fog, originally referring to particulate matter but now describing air pollution in general.

Particulate matter (PM) can be emitted directly as carbonaceous soot particles from incomplete combustion, or it can be formed in the atmosphere, as, for example, when gaseous SO_2 and NO_x emissions are transformed into liquid droplets of sulfates (sulfuric acid) and nitrates (nitric acid). In the eastern part of the United States, with its greater reliance on coal-fired power plants, sulfates (and associated ammonium) are the most important components of PM, whereas in California and other areas of the West, carbon (soot) and nitrates dominate.

Size and Chemical Composition. The original NAAQS for particulates did not take size into account. Larger particles could dominate the mass per unit volume measure but be unimportant in terms of human health risk. In 1987, however, the PM_{10} standard was introduced, and in 1997 the $PM_{2.5}$ standard was added. Particles smaller than 2.5 μm (i.e., $PM_{2.5}$) are referred to as *fine* particles, whereas those between 2.5 μm and 10 μm are the *coarse* fraction of PM_{10}. Coarse particles tend to settle quickly, so their spatial impact is limited to areas near their source, but fine particles have longer atmospheric lifetimes and are capable of traveling vast distances. For example, a well-studied dust storm that originated in the Gobi Desert in Asia on April 6, 2001, traveled eastward over the globe, impacting $PM_{2.5}$ measurements on the West Coast of the United States a week later and on the East Coast two weeks later.

Although particles may have very irregular shapes, their size can be described by an equivalent *aerodynamic diameter* determined by comparing them with perfect spheres having the same settling velocity. The particles of most interest have aerodynamic diameters in the range of 0.1 μm to 10 μm (roughly the size of bacteria). Particles smaller than these undergo random (Brownian) motion and, through coagulation, generally grow to sizes larger than 0.1 μm. Particles larger than 10 μm settle quickly; a 10 μm particle, for example, has a settling velocity of approximately 20 centimeters per minute.

We can use a fairly simple analysis to calculate the settling velocity of a spherical particle. When such a particle reaches its terminal velocity, the gravitational force pulling it down is balanced by a drag force that we can estimate. For particles that are larger than about 1 μm, with density much greater than air, we can use a simplified version of Stokes's law to approximate the drag force:



Gravitational force = drag force

$$mg = \frac{\pi}{6} d^3 \rho g = 3\pi\eta v d \tag{24}$$

$$v = \frac{d^2 \rho g}{18\eta}$$

where

m = mass of the particle (g)
g = gravitational acceleration = 9.80 m/s^2
d = particle diameter (m)
ρ = particle density (g/m^3)
η = viscosity of air = 0.0172 g/m·s
ν = settling velocity (m/sec)

EXAMPLE 4 Settling Velocity of a Spherical Particle

Find the settling velocity of a spherical droplet of water with diameter 2 μm, and estimate the residence time of such particles if they are uniformly distributed in the lower 1,000 m of atmosphere and their removal rate is determined by how fast they settle in still air.

Solution Using (24) with the density of water equal to 10^6g/m^3 gives

$$\nu = \frac{d^2 \rho g}{18\,\eta} = \frac{(2 \times 10^{-6}\text{ m})^2 \cdot (10^6\text{ g/m}^3) \cdot (9.8\text{ m/s}^2)}{18 \times 0.0172\text{ g/m}\cdot\text{s}}$$

$$\nu = 1.27 \times 10^{-4}\text{ m/s}$$

which is about 0.5 m/hr.

We can use a simple box model to estimate the residence time, τ, of N particles uniformly distributed in a box of atmosphere with height h (m). We need to know the rate of removal of particles:

$$\text{Rate of removal of particles} = \frac{N\text{ (particles)}}{h\text{ (m)}} \cdot \nu(\text{m/s}) = \frac{N\nu}{h}\text{ particles/s}$$

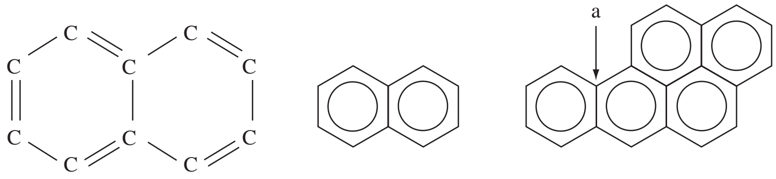
The residence time τ is the ratio of the number of particles in the box divided by the rate of removal of particles

$$\tau = \frac{\text{Particles in box}}{\text{Rate of removal of particles}} = \frac{N}{(N\nu/h)} = \frac{h}{\nu}$$

$$\tau = \frac{1{,}000\text{ m}}{1.27 \times 10^{-4}\text{ m/s}} = 7.9 \times 10^6\text{ s} \approx 91\text{ days}$$

In other words, these small particles in the atmosphere don't settle very quickly, and an estimate of residence time in terms of months is appropriate.

The black smoke, or soot, we see emitted from diesel engines and smokestacks consists mostly of solid particles made up of vast numbers of carbon atoms fused together in benzene rings. Although these particles themselves can irritate lungs, most often it is large organic molecules that stick to the surfaces of soot (adsorption) that are responsible for the most serious health effects. Of greatest concern is a class of benzene-like hydrocarbons called *polynuclear* (or polycyclic) *aromatic hydrocarbons* (PAHs). Polynuclear aromatic hydrocarbons consist of fused benzene rings, the



(*a*) Napthalene (*b*) Simplified (*c*) Benzo[a]pyrene

FIGURE 8 Polynuclear aromatic hydrocarbons consist of fused benzene rings. (*a*) Two fused benzene rings (napthalene) shown without their hydrogen bonds, and (*b*) its simplified representation; (*c*) benzo[a]pyrene, showing the "a" connection point.

simplest of which is napthalene, shown in Figure 8. PAHs are formed when carbon-containing materials are not completely oxidized during combustion. They are released as gases but quickly condense onto particles of soot. Common sources of PAHs include tobacco smoke, motor vehicle exhaust, the char on charcoal-broiled food, and smoke from wood and coal combustion. Probably the most dangerous of these PAHs is a substance called benzo[a]pyrene (BaP), which consists of five fused rings as shown in Figure 8c. BaP is a Category A human carcinogen known to cause lung and kidney cancer.

Impacts of Particulates on the Human Respiratory System. The ability of the human respiratory system to defend itself against particulate matter is, to a large extent, determined by the size of the particles. To help understand these defense mechanisms, consider the illustration in Figure 9. The upper respiratory system consists of the nasal cavity and the trachea, and the lower respiratory system consists of the bronchial tubes and the lungs. Each bronchus divides over and over again into smaller and smaller branches, terminating with a large number of tiny air sacs called alveoli.

Large particles that enter the respiratory system can be trapped by the hairs and lining of the nose. Once captured, they can be driven out by a cough or sneeze. Smaller particles that make it into the tracheobronchial system can be captured by mucus, worked back to the throat by tiny hairlike *cilia,* and removed by swallowing or spitting. Particles larger than about 10 μm are effectively removed in the upper respiratory system by these defense mechanisms. Smaller particles, however, are often able to traverse the many turns and bends in the upper respiratory system without being captured on the mucus lining. These particles may make it into the lungs, but depending on their size, they may or may not be deposited there. Some particles are so small that they tend to follow the air stream into the lungs and then right back out again. Other submicron-size particles are deposited by Brownian motion. Roughly between 0.5 μm and 10 μm, particles may be small enough to reach the lung and large enough to be deposited there by sedimentation. Sedimentation in the lungs is most effective for particles between 2 μm and 4 μm.

Particulates aggravate existing respiratory and cardiovascular disease and damage lung tissue, and some are carcinogenic. At concentrations near the pre-2006 national ambient standard, increases in the number of hospital visits for upper respiratory infections, cardiac disorders, bronchitis, asthma, pneumonia, and emphysema, and increased mortality rates have been observed. One study estimated

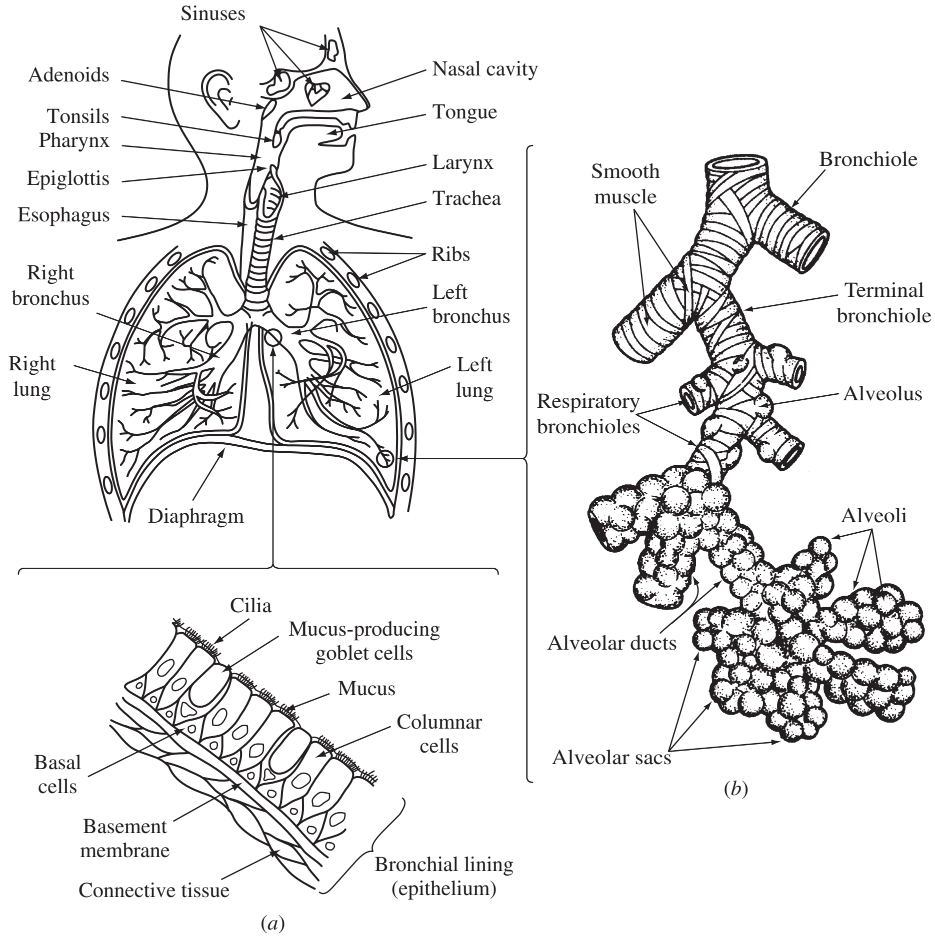


FIGURE 9 The human respiratory system. (*a*) The system as a whole and a cross-section of the bronchial lining showing the cilia, and (*b*) details of the lower respiratory system terminating in the alveoli.
(*Source:* Williamson, 1973.)

hospital admissions for congestive heart failure (an inability of the heart to pump out all the blood that returns to it) increased 8 percent for each 100 $\mu g/m^3$ increase in PM_{10} (Raloff, 1995). Evidence that tens of thousands of premature deaths are caused each year in the United States due to inhaled particular matter led the EPA to reevaluate its short-term $PM_{2.5}$ air quality standard in 2006. The new 24-hour standard reduction from 65 to 35 $\mu g/m^3$ is expected to result in annual reductions of 2,500 to 5,700 premature deaths in people with lung or heart disease, 2,600 cases of chronic bronchitis and 7,300 cases of acute bronchitis, 5,000 nonfatal heart attacks, 1,630 hospital admissions for cardiovascular or respiratory symptoms, 1,200 emergency room visits, and 350,000 days when people miss work or school. The EPA estimates the monetary value of these benefits at between \$17 billion and \$35 billion (U.S. EPA, 2006b).

Oxides of Sulfur

About 86 percent of the 15 million tons per year of anthropogenic sulfur oxide emissions are the result of fossil fuel combustion in stationary sources, and most of that is emitted by coal-fired power plants. Only about 5 percent comes from highway vehicles. The only significant noncombustion sources of sulfur emissions are associated with petroleum refining, copper smelting, and cement manufacture.

All fossil fuels as they are extracted from the ground contain some sulfur. Coal, which has the most, typically contains from 1 to 6 percent sulfur. About half of that is organic sulfur that is chemically bound to the coal. The other half is simply physically trapped in the noncarbon portion of coal, and much of that half can be removed by pulverizing and washing the coal before combustion. The amount of sulfur in petroleum tends to be less than a few percent, and if refined, almost all of that sulfur is removed during processing. Gasoline, for example, has much less than 1 ppm sulfur. Natural gas as it leaves the wellhead contains a considerable amount of sulfur in the form of highly toxic hydrogen sulfide (H_2S) that must be removed before the gas can be used. After natural gas is cleaned, however, negligible amounts of sulfur remain, which makes it a highly desirable replacement fuel for coal.

When these fuels are burned, the sulfur is released mostly as sulfur dioxide (SO_2) but also with small amounts of sulfur trioxide (SO_3). Sulfur dioxide, once released, can convert to SO_3 in a series of reactions, which, once again, involve a free radical such as $OH^\bullet$:

$$SO_2 + OH^\bullet \rightarrow HOSO_2^\bullet \tag{25}$$

$$HOSO_2^\bullet + O_2 \rightarrow SO_3 + HO_2^\bullet \tag{26}$$

The $HO_2^\bullet$ radical can then react with NO to return the initial $OH^\bullet$ as in (20). Sulfur trioxide reacts very quickly with H_2O to form sulfuric acid, which is the principal cause of acid rain.

$$SO_3 + H_2O \rightarrow H_2SO_4 \tag{27}$$

Sulfuric acid molecules rapidly become particles by either condensing on existing particles in the air or by merging with water vapor to form $H_2O—H_2SO_4$ droplets. Often, a significant fraction of particulate matter in the atmosphere consists of such sulfate (SO_4) aerosols. Sulfate particles in urban air typically have an effective size of less than 2 μm, with most of them being in the range of 0.2 to 0.9 μm. Their size is comparable to the wavelengths of visible light, and their presence greatly affects visibility. Their size also allows deep penetration into the respiratory system.

Acid Deposition. The transformation from SO_2 gas to sulfate particles (SO_4) is gradual, taking a matter of days. During that time, sulfur pollution may be deposited back onto the land or into water, either in the form of SO_2 or sulfate. In either form, sulfur pollution can be deposited by removal during precipitation (wet deposition), or by slow, continuous removal processes that occur without precipitation (dry deposition). Figure 10 suggests the effects of time and distance on the conversion and deposition of sulfur.

Figure 11 shows pH for wet deposition in the United States and Canada. Natural rainfall would have a pH value between 5 and 5.6, and anything less is loosely called "acid rain." About two-thirds of U.S. coal consumption, the source of

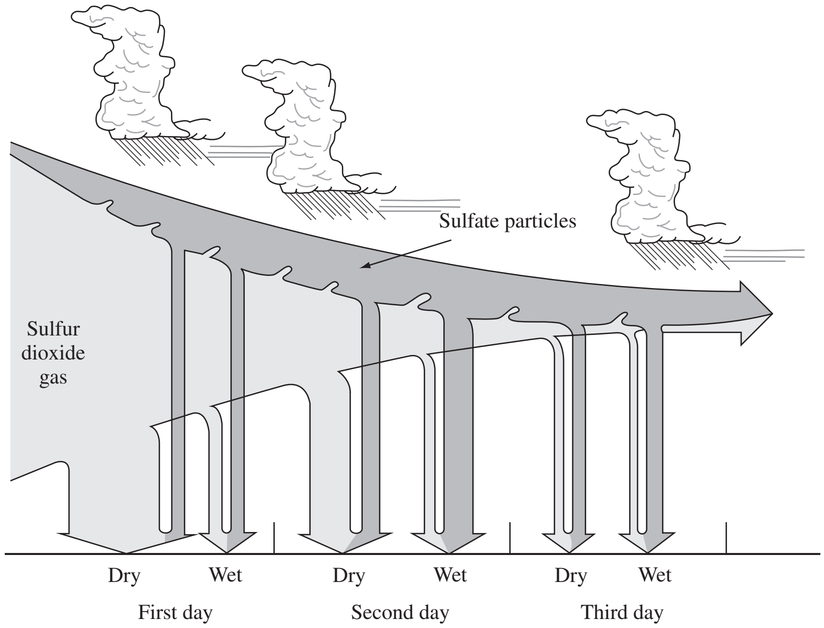


FIGURE 10 The effects of time and distance on conversion and deposition of sulfur pollution.
(*Source:* OTA, 1984.)

FIGURE 11 The pH of wet deposition shows that the worst acid rain occurs east of the Mississippi, where most of the coal-fired power plants are located.
(*Source:* National Atmospheric Deposition Program; 2004 data from *http://nadp.sws.uiuc.edu.*)

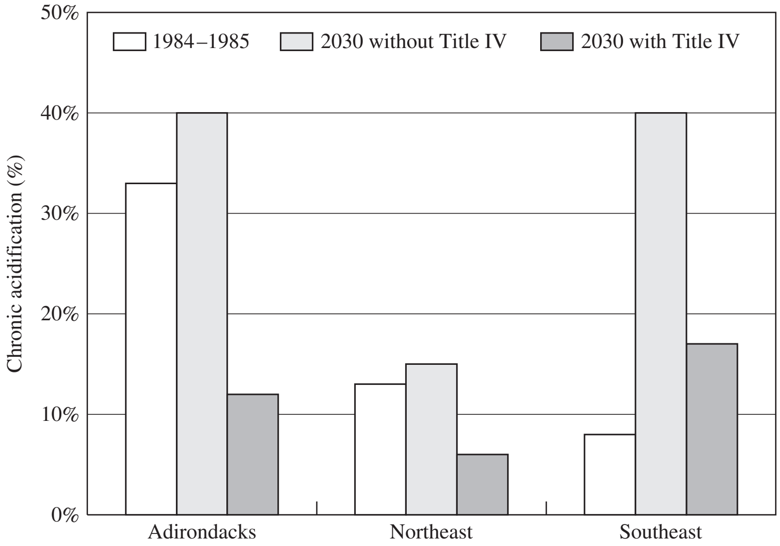


FIGURE 12 Fraction of lakes suffering from chronic acidification in three areas of the United States, comparing 1984–85 percentages with 2030 percentages with and without Title IV.
(*Source:* Based on Chestnut and Mills, 2005.)

most sulfur emissions, is east of the Mississippi, which correlates well with the acidity of rainfall in the eastern half of the country. Unfortunately, large areas of the eastern United States and Canada with the worst acid rain have lakes that are inherently the most sensitive to acidification.

Projected reductions in power-plant sulfur emissions driven by the Acid Rain Program (Title IV of the CAA) are projected to have significant impacts on the fraction of U.S. lakes suffering from chronic acidification. Figure 12 shows the results of one study on the fraction of lakes that are projected to be acidified in 2030 with and without Title IV.

Health and Welfare Impacts. Sulfur dioxide is highly water soluble, much more so than any of the other criteria pollutants. As a result, when inhaled, it is most likely to be absorbed in the moist passages of the upper respiratory tract, the nose and upper airways, where it does less long-term damage. Other gases, being less soluble, are more likely to reach the more critical terminal air sacs of the lungs. When sulfur is entrained in an aerosol, however, the aerodynamic properties of the particles themselves affect the area of deposition, and it is possible for sulfur oxides to reach far deeper into the lungs. The combination of particulate matter and sulfur oxides can then act synergistically, with the effects of both together being much more detrimental than either of them separately. In fact, in every major air pollution episode, the combination of sulfur oxides and particulates has been implicated as a cause of the excess mortality observed.

Sulfur oxides can damage trees, especially when trees are bathed in acid fog or clouds that tend to have very low pH levels. Acidification also damages plants by affecting their ability to extract nutrients from the soil. Nutrients are leached from soils more readily under acidic conditions, and low pH levels allow aluminum to solubilize, which interferes with the uptake of nutrients. Sulfurous pollutants can discolor paint, corrode metals, and cause organic fibers to weaken. Airborne sulfates significantly reduce visibility and discolor the atmosphere. Most of the visibility impairment in the eastern United States is caused by sulfates, whereas in the western United States, it is more often nitrogen oxides and dust.

Prolonged exposure to sulfates causes serious damage to building marble, limestone, and mortar, as the carbonates (e.g., limestone, $CaCO_3$) in these materials are replaced by sulfates. The reaction between limestone and sulfuric acid shows such a replacement:

$$CaCO_3 + H_2SO_4 \rightarrow CaSO_4 + CO_2 + H_2O \tag{28}$$

The calcium sulfate (gypsum, $CaSO_4$) produced by this reaction is water soluble and easily washes away, leaving a pitted, eroded surface. Many of the world's historic buildings and statues are rapidly being degraded due to this exposure. It is common now in such monuments as the Acropolis in Greece, for the original outdoor statuary to be moved into air-conditioned museums, leaving plaster replicas in their place.

Lead

Most lead emissions in the past were from motor vehicles burning gasoline containing the antiknock additive, tetraethyllead $Pb(C_2H_5)_4$. The antiknock properties of tetraethyllead were first discovered by Thomas Midgley and colleagues at the General Motors Research Laboratory in 1921; by coincidence, in the same laboratory, Midgley also developed the first chlorofluorocarbons (Thomas, 1995). Both discoveries were considered major advances at the time but later turned out to have terrible human health and environmental consequences.

In the United States (and just a handful of other developed countries), almost all lead emissions from gasoline have been eliminated so that by 1993, per capita emissions of lead in the United States were only 3 percent of the world average (Thomas, 1995). The decrease in the United States was originally motivated by the Clean Air Act Amendments of 1970 that dictated a 90 percent drop in CO, NO_x, and hydrocarbon emissions from automobiles. As will be described later in this chapter, the auto industry chose to use catalytic converters as their principal emission control system for those three pollutants. Catalytic converters, it turns out, are quickly rendered ineffective when exposed to lead, so beginning in the late 1970s, many cars were designed to only burn unleaded fuels. By the mid 1990s, after virtually all cars in the United States had catalytic converters, leaded gasoline was completely phased out. As a result, total U.S. lead emissions decreased 93 percent between 1982 and 2002 with the major remaining sources being industrial processes (particularly metal-processing plants) and leaded fuels used for aviation and nonroad vehicles (Figure 13). Unfortunately, in most of the rest of the world, leaded gasoline is still the predominant fuel and lead emissions remain very high.

Lead is emitted into the atmosphere primarily in the form of inorganic particulates. Much of this is removed from the atmosphere by settling in the immediate

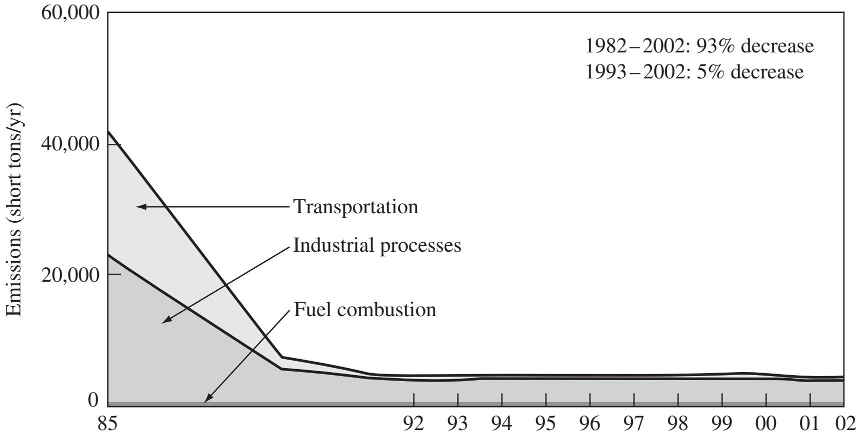


FIGURE 13 Lead emissions trend, 1985–2003.
(*Source:* U.S. EPA, 2003.)

vicinity of the source. Airborne lead may affect human populations by direct inhalation, in which case people living nearest to major sources such as highways or metals-processing plants are at greatest risk. Despite the elimination of most lead emissions from motor vehicles, the soil around highways is still heavily contaminated, and it can become airborne when disturbed. It also can be tracked into homes where it may end up embedded in carpeting ready to become airborne once again. Wiping your feet on a doormat is one easy and effective way to help reduce that hazard.

Inhalation of lead can occur indoors as well as outdoors. A major indoor source of lead is chipped and flaking particles of lead-based paints that were commonly used in the past. Paint containing $Pb_3(CO_3)_2(OH)_2$ was widely used in white paint, and red paints with Pb_3O_4 are still used for outdoor protection of metal surfaces. Those substances are no longer allowed for paints used indoors, but they can be used for exterior surfaces. When those paints chip, peel, or are sanded, lead particles become dusts that are easily ingested or inhaled. Chips of leaded paint are somewhat sweet and are all too often eaten by children living in older homes.

Although most human exposure to lead is from inhalation, it can also be ingested after airborne lead is deposited onto soil, water, and food crops such as leafy vegetables and fruits. It also can be ingested when water systems are contaminated. Lead once was used to make water pipes and was used until recently in solder to join copper water pipes, but those uses have been banned in new systems in the United States. Lead can still leach out of those older water systems, especially if the water is acidic or particularly soft. For such systems, running the water from a tap for a minute before drinking it is highly recommended.

Lead poisoning can cause aggressive, hostile, and destructive behavioral changes, as well as learning disabilities, seizures, severe and permanent brain damage, and even death. Measurements in actual communities suggest that an increase in airborne lead concentration of 1 $\mu g/m^3$ (the NAAQS is 1.5 $\mu g/m^3$) results in an increase of about 1 to 2 μg per deciliter ($\mu g/dL$) in human blood levels. Blood levels of 10 to 15 $\mu g/dL$ are associated with reduced intelligence and detrimental effects on growth and development of children, leading the U.S. Public Health Service to label lead as the greatest environmental health threat to our children (Thomas, 1995;

U.S. PHS, 1991). As lead was phased out of gasoline in the United States, the average concentration of lead in blood dropped from 16 to 3 μg/dL, which makes it one of the most successful of all environmental achievements.

6 Toxic Air Pollutants

The Clean Air Act Amendments of 1990 require the EPA to speed up efforts to control toxic air pollutants. As defined in the Act, these are pollutants that "are known to be, or may reasonably be anticipated to be, carcinogenic, mutagenic, teratogenic, neurotoxic, which cause reproductive dysfunction, or which are acutely or chronically toxic." There are literally hundreds of such chemicals, including 189 compounds that are listed in the 1990 Amendments. The EPA is charged with setting emission standards that achieve "the maximum degree of reduction in emissions" taking into account cost and other nonair-quality factors, and risk-based health standards that will assure no more than a 1-in-a-million risk for the most exposed individuals. In light of the years of effort that were needed to deal with the six criteria pollutants, the emission and air-quality standard setting that will be required for hundreds of additional chemicals is daunting.

Even before the 1990 Amendments were passed, the EPA had initiated National Emission Standards for Hazardous Air Pollutants (NESHAP) under Section 112 of the Act, but by 1988, only asbestos, benzene, beryllium, coke-oven emissions, inorganic arsenic, mercury, radionuclides, and vinyl chloride were regulated. Another important step in the regulation of air toxics was initiated by the Emergency Planning and Community Right to Know Act of 1986 (EPCRA) and expanded by the Pollution Prevention Act of 1990. Starting in 2000, the EPA began to focus on persistent bioaccumulative toxic (PBT) chemicals, which are of concern not only because they are toxic but also because their long life in the environment allows them to build up, or accumulate, in body tissue. The PBT chemicals include, for example, dioxin, lead, mercury, polycyclic aromatic compounds (PACs), polychlorinated biphenyls (PCBs), and certain pesticides.

Manufacturing facilities that use any of the now 650-some listed chemicals must submit annual reports to the EPA on their releases. These releases form a database called the Toxic Release Inventory (TRI) that has been in operation since 1987. The availability of the TRI to the public is intended to discourage use of these chemicals by manufacturers who want to avoid negative publicity. In 1993, reported air emissions totaled 1.7 billion pounds of toxic chemicals; in 2004, that had dropped only a few percent, down to 1.55 billion pounds.

7 Air Pollution in the World's Megacities

The world's population is becoming ever more urbanized, causing increasing environmental pressures, not the least of which is deterioration of air quality in many of the most rapidly growing cities. In 1970, there were only four cities in the world with populations over 10 million people (Tokyo, New York, Shanghai, and London); by 1990, there were 12; and by 2000, there were 24 such cities. Fully half of the world's population now live in urban areas. The World Health Organization (WHO) and the United Nations Environment Programme (UNEP) undertook a study of air quality in 20 megacities (cities with more than 10 million people), with a focus on the same six criteria pollutants that are regulated in the U.S. Clean Air Act except that in most cities it is total Suspended Particulate Matter (SPM) that is measured instead of PM_{10}. In many circumstances, the data being gathered are difficult to obtain and of questionable quality, but a preliminary assessment is beginning to emerge that shows increasing numbers of people exposed to increasingly unhealthy concentrations of air contaminants, especially in developing countries.

In most megacities, the principal source of air pollution is motor vehicles. In almost all cities in the developing world, leaded fuels are still being burned; a high percentage of the vehicles are diesel-powered trucks and buses with no emission controls; many streets are unpaved; and traffic congestion, which intensifies emissions, is overwhelming. The resulting concentrations of Pb, CO, NO_x, O_3, and TSP are often many times higher than air-quality guidelines defined by WHO. In addition, many countries have coal-fired power plants and other industrial facilities within city limits, so levels of SO_x, NO_x, and particulates are correspondingly high.

Mexico City used to be ranked by WHO as the world's most polluted city. Ozone concentrations above 0.10 ppm occur as many as 300 days per year (the U.S. 8-hr standard is 0.09 ppm), and during a 5-day period in November, 1997, each day's readings rose above 0.25 ppm. But now, Mexico City has lost that dubious distinction, with the 10 most polluted cities in the world now being in China.

A study of populations and air quality in 20 megacities showed that high levels of suspended particulate matter (SPM) is the most prevalent air-quality problem (WHO/UNEP, 1992). Of these 20 cities, 12 have serious problems with particulate matter, where "serious" is defined to mean WHO air-quality guidelines are exceeded by more than a factor of two (see Figure 14). Beijing, Mexico City, and Seoul have serious SPM combined with serious SO_x problems, which is a lethal combination that leads to increased mortality and morbidity. In a number of areas, including parts of Beijing, Calcutta, Delhi, Shanghai, and Seoul, combustion of coal and biomass fuels for cooking and heating leads to extremely high pollutant concentrations indoors where many people, women especially, spend most of their time. Indoor air quality will be described later in this chapter.

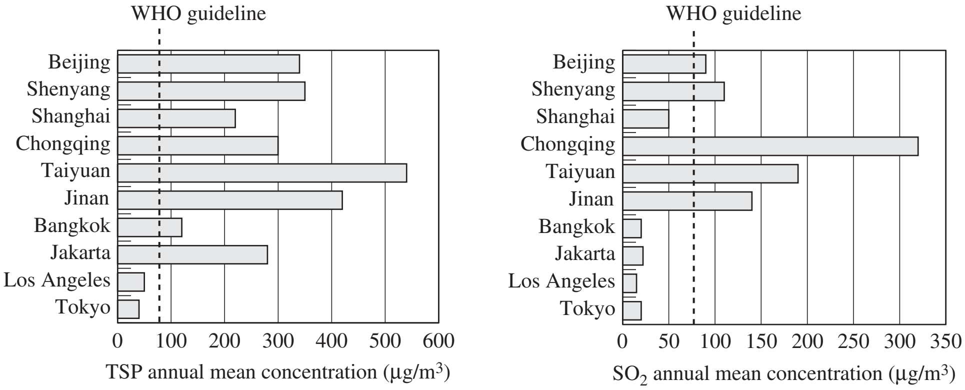


FIGURE 14 1995 annual mean concentrations of TSP and SO_2 in selected large cities. (*Source:* The World Bank, 1997.)

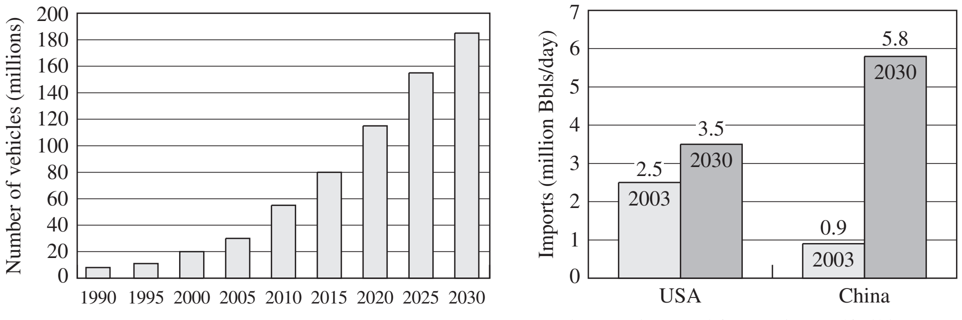
8 Motor Vehicle Emissions

The transportation sector in the United States accounts for a significant fraction of all the criteria pollutants except for sulfur oxides. Most of those emissions are from cars and trucks, which emit those pollutants right in the middle of crowded, urban populations. In many of the developing countries of the world, the growth rate in numbers of motor vehicles far outstrips the population growth rate, which is leading to urban congestion and very unhealthy air quality. China provides a particularly compelling example, with vehicle growth projected to rise from about 30 million in 2005 to over 180 million by 2030 (Figure 15). Vehicles are the principal cause of China's urban air pollution problems, which are estimated to be causing on the order of 400,000 respiratory attacks and as many as 50,000 infant deaths per year (Energy Foundation, 2005). Moreover, China's demand for imported oil for all of those cars and trucks is projected to rise rapidly, with much of that expected to come from the volatile Middle East. As Figure 15 shows, China's Persian Gulf oil imports will soon exceed those of the United States.

Emission Standards

The legislative history of auto emission controls began in California in 1959 with the adoption of state standards to control exhaust hydrocarbons (HC) and carbon monoxide (CO). It was supplemented in 1960 by standards to control emissions resulting from crankcase blowby. These standards, however, were not fully implemented until they were deemed technologically feasible in 1966. Federal standards for CO and HC began in 1968, and by 1970, the year in which the first major amendments to the Clean Air Act were enacted, the auto industry had reduced hydrocarbon emissions by almost three-fourths and carbon monoxide by about two-thirds. At that time, no controls were required for NO_x, and to some extent, improvements in HC and CO emissions were made at the expense of increased nitrogen oxide emissions.

The Clean Air Act Amendments of 1970 required that emissions from automobiles be reduced by 90 percent compared to levels already achieved by the 1970



(*a*) China's vehicle growth (*b*) Growing need for Persian Gulf oil imports

FIGURE 15 Projected growth in motor vehicles in China and estimated need for imported oil from the Persian Gulf.
(*Sources:* Vehicles: Energy Foundation, 2005; oil: EIA, *International Energy Outlook,* 2006.)

model year. These standards were to be reached by 1975 for CO and HC, and by 1976 for NO_x, with a one-year delay allowed if the industry could adequately prove that the technology was not available to meet them. The 1970 Amendments were unusual in that they were "technology forcing," that is, they mandated specific emission levels before technology was available to meet the standards. The automobile industry successfully argued the case for the one-year delay, in spite of the fact that the 1973 Honda, with a stratified charge engine, was able to meet the 1975 emission requirements while getting 40 mpg.

Subsequently, in response to the oil embargo of 1973 and concerns for the impact of emission standards on fuel economy, the Federal Energy Supply and Coordination Act of 1974 rolled the standards back again to 1978–79. Finally, the Clean Air Act Amendments of 1977 delayed the standards once more, setting the date for HC and CO compliance at 1981. At the same time, the 1977 Amendments eased the standard for NO_x from 0.4 g/mile to 1 g/mile, calling the 0.4 g/mile a "research objective" despite the fact that California, which is allowed to set its own auto emission standards as long as they are more stringent than federal requirements, stayed with the original 0.4 g/mile standard for NO_x. The 1990 Amendments to the Clean Air Act brought national standards to the level of California standards beginning with the 1996 models, including the 0.4 g/mile NO_x standard. Table 4 shows current (2006) Federal Motor Vehicle Emission Standards.

As usual, California leads the way in terms of emission reduction requirements. Table 4 shows California's Air Resources Board (CARB) requirements for cars (less than 3,750 lbs) to be designated as Low Emission Vehicles (LEVs), Ultra Low Emission Vehicles (ULEVs), and Super Ultra Low Emission Vehicles (SULEVs). Heavier vehicles have similar, but less stringent, emission labeling requirements. New vehicles sold in California must be at least LEVs. Compared to federal standards for each vehicle weight, an SULEV vehicle has emission reductions of 96 percent for HC, 70 percent for CO, and 95 percent for NO_x. California also has an Advanced Technology Partial Zero Emission Vehicle (AT-PZEV) designation for vehicles that meet SULEV emission standards but in addition have zero evaporate emissions and a 15-year/150,000-mile warranty. The fact that quite a number of

TABLE 4

Federal Motor Vehicle Emission Standards (g/mile)

Federal Standards	CO	NO_x	NMHC[a]
Passenger cars prior to controls (1967)	87.0	3.6	8.8
Passenger cars	3.4	0.4	0.25
Light-duty trucks (3,751–5,750 lbs.)	4.4	0.7	0.32
Light-duty trucks (>5,750 lbs.)	5.0	1.1	0.39
Motorcycles[b]	19.3	2.24[c]	
CARB Standards for vehicles <3,750 lbs			
Low Emission Vehicle (LEV)	3.4	0.2	0.075
Ultra Low Emissions Vehicle (ULEV)	1.7	0.2	0.040
Super Ultra Low Emissions Vehicle (SULEV)	1.0	0.02	0.010

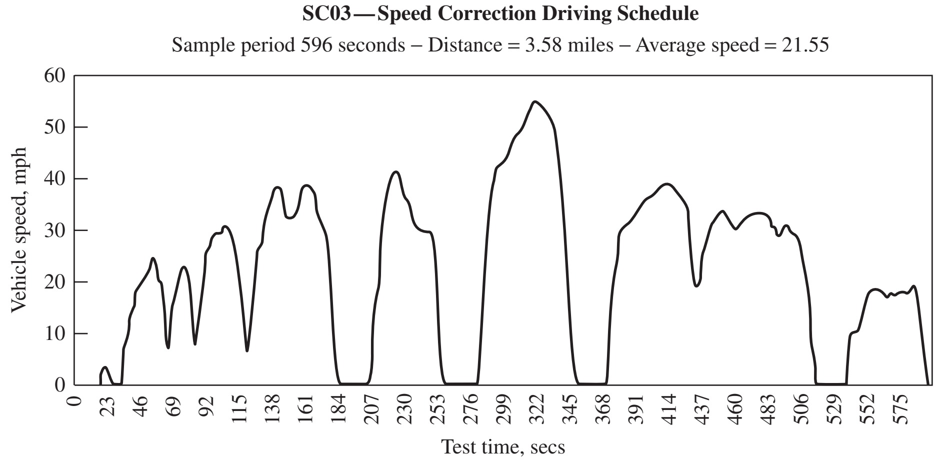
[a] NMHC stands for nonmethane hydrocarbons.

[b] Displacement 280 cc or greater.

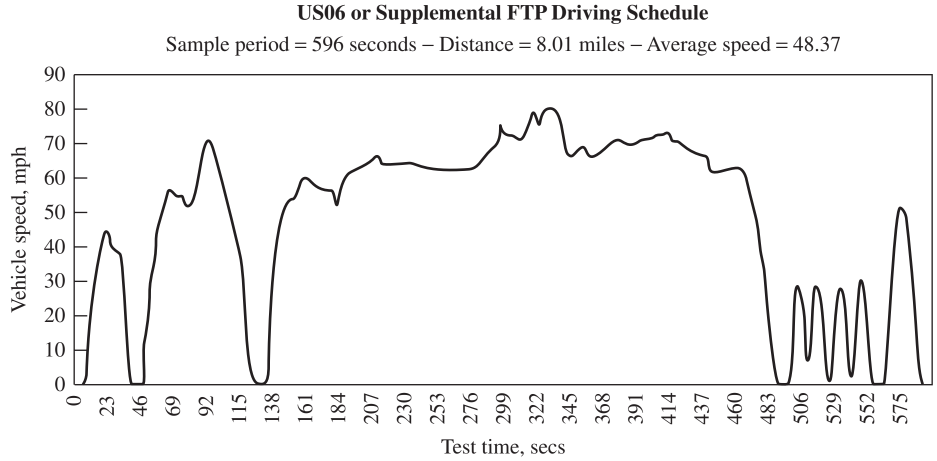
[c] NO_x + HC.

vehicles sold in California meet the AT-PZEV designation makes the current federal standards look rather modest.

Table 4 indicates that vehicle emission regulations are expressed in terms of grams of pollutant per mile of driving. Since emissions vary considerably as driving conditions change, it has been necessary to carefully define a "standard" driving cycle and to base emission regulations on that cycle. In 1996, the original Federal Test Procedure (FTP) was augmented with a Supplemental Federal Test Procedure (SFTP) designed to more accurately deal with more aggressive driving behavior (higher speeds, rapid acceleration, rapid speed fluctuations), impacts of air conditioners, and emissions after the engine is turned off. The two new drive cycles defined in the SFTP are shown in Figure 16.


SC03—Speed Correction Driving Schedule

Sample period 596 seconds – Distance = 3.58 miles – Average speed = 21.55


US06 or Supplemental FTP Driving Schedule

Sample period = 596 seconds – Distance = 8.01 miles – Average speed = 48.37

FIGURE 16 The new Supplemental Federal Test Procedure drive cycles. The SC03 represents driving immediately following vehicle startup, including air conditioning, whereas the US06 starts with a warmed-up engine and represents more aggressive driving patterns than the original FTP.

CAFE Fuel Economy Standards

The oil embargo of 1973 shifted Congress' attention from motor vehicle emissions to motor vehicle fuel efficiency. At the time, cars in the United States averaged about 14 miles per gallon (mpg). Corporate Average Fuel Economy (CAFE) standards were enacted in 1975, which required fuel efficiency for each manufacturer's fleet of new automobiles to average at least 27.5 mpg within a 10-year period. The FTP drive cycles defined for emissions were used to determine city and highway fuel efficiencies, which are then weighted 55 percent city and 45 percent highway to determine an overall fuel efficiency. A separate fuel efficiency standard was enacted to cover "light trucks." That standard started at 20.2 mpg and has been tightened to 22.2 mpg for model-year (MY) 2007 light trucks. Light trucks are defined, in part, by whether there is additional cargo-carrying capacity above conventional passenger cars, which includes readily detachable rear seats, or additional road clearance (e.g., axle clearance above 18 cm). That means most SUVs have been considered to be light trucks, subject to just the less-stringent mpg standard.

Another perverse incentive for SUVs was part of the Energy Tax Act of 1978, which established a gas-guzzler tax that applies only to passenger cars that get less than 22.5 mpg. The tax can be sizable; for example, the gas-guzzler tax for passenger cars getting less than 12.5 mpg is about $7,700. Most SUVs, being categorized as light trucks, are exempt from this tax.

The average CAFE fuel economy of new passenger cars sold in the United States did rise to the required 27.5 mpg by 1986, but the increasing popularity of SUVs and pick-up trucks, which now account for more than half of new vehicle sales, has tipped the scales and overall light-duty vehicle efficiency has dropped from its peak of 25.9 mpg in 1986–87 to 24 mpg in 2006. The American Council for an Energy Efficient Economy estimates raising CAFE standards by 5 percent annually until 2012 and by 3 percent per year thereafter could save 67 billion barrels of oil over the next 40 years. This is 10 to 20 times greater than the potential oil supply from the Arctic National Wildlife Refuge.

As most drivers know, the sticker CAFE fuel efficiency rating on their cars tends to overestimate actual mpg. A 2006 EPA report has generated more realistic fuel efficiency statistics that derate CAFE values by about 15 percent. They also exclude credits for alternative-fuel capability, which has allowed manufacturers to claim 27.5 mpg for vehicles capable of running on E85 (85 percent ethanol, 15 percent gasoline) even though there have been very few gas stations that offer this fuel. Using these more realistic estimates for "real world" fuel efficiency, the average fuel economy for all model year 2006 light-duty vehicles is estimated to be 21.0 mpg, the same value as achieved in 1994, but 5 percent lower than the peak value achieved in 1987–88. Figure 17 shows this adjusted fuel economy history.

Criticism of the CAFE program for light trucks resulted in new definitions that will apply on an optional basis to model year 2008–2010 light trucks and be required for model year 2011 vehicles. Under the new Reformed CAFE standards, the definition will be based on a measure of vehicle size called "footprint," which is a vehicle's wheelbase multiplied by its track width. Fuel economy standards will then be based on that footprint, with an overall light truck target of 24.1 mpg.

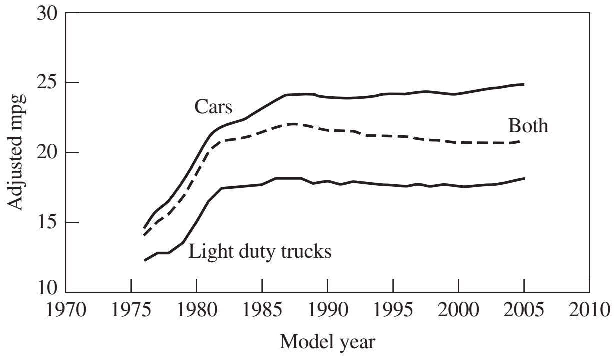


FIGURE 17 Adjusted fuel economy by model year (3-year moving average).
(*Source:* U.S. EPA, 2006, *Light-Duty Automotive Technology and Fuel Economy Trends: 1975 through 2006,* EPA420-S-06-003.)

EXAMPLE 5 Average Fuel Efficiency

Suppose a fuel efficiency test indicates a particular automobile gets 20 mpg in the city and 29 mpg on the highway. What would its CAFE fuel efficiency be?

If new cars get an average of 27.5 mpg, and new light trucks get 22.2 mpg, what would be the fleet average if 50 percent of sales are light trucks?

Solution The CAFE efficiency assumes 55 percent urban and 45 percent highway miles. If we want to imagine 1 mile of driving, the fuel efficiency would be

$$\text{MPG} = \frac{1 \text{ mile}}{\left(\dfrac{0.55 \text{ miles}}{20 \text{ miles/gal}} + \dfrac{0.45 \text{ miles}}{29 \text{ miles/gal}}\right)} = 23.2 \text{ mpg}$$

The fleet efficiency for 28 mpg cars and 22.2 mpg light trucks would be

$$\text{MPG} = 0.50 \times 27.5 \text{ mpg} + 0.50 \times 22.2 \text{ mpg} = 24.8 \text{ mpg}$$

Fuel efficiency is important for a number of reasons, not the least of which are national security issues related to our dependence on imported oil (approximately equal to the energy demands of our highway vehicles). While emissions of the conventional pollutants CO, HC, and NO_x are not necessarily directly related to fuel efficiency, emissions of the greenhouse gas carbon dioxide (CO_2) certainly are.

The Conventional Otto Cycle Engine

The most common internal combustion engine is a four-stroke, spark-ignited, piston engine invented around 1880 by a German engineer, Nicholas Otto. The operation of an Otto cycle, or four-stroke, engine is described in Figure 18. On the first, or *intake* stroke, the descending piston draws in a mixture of fuel and air through the open intake valve. The *compression* stroke follows, in which the rising piston

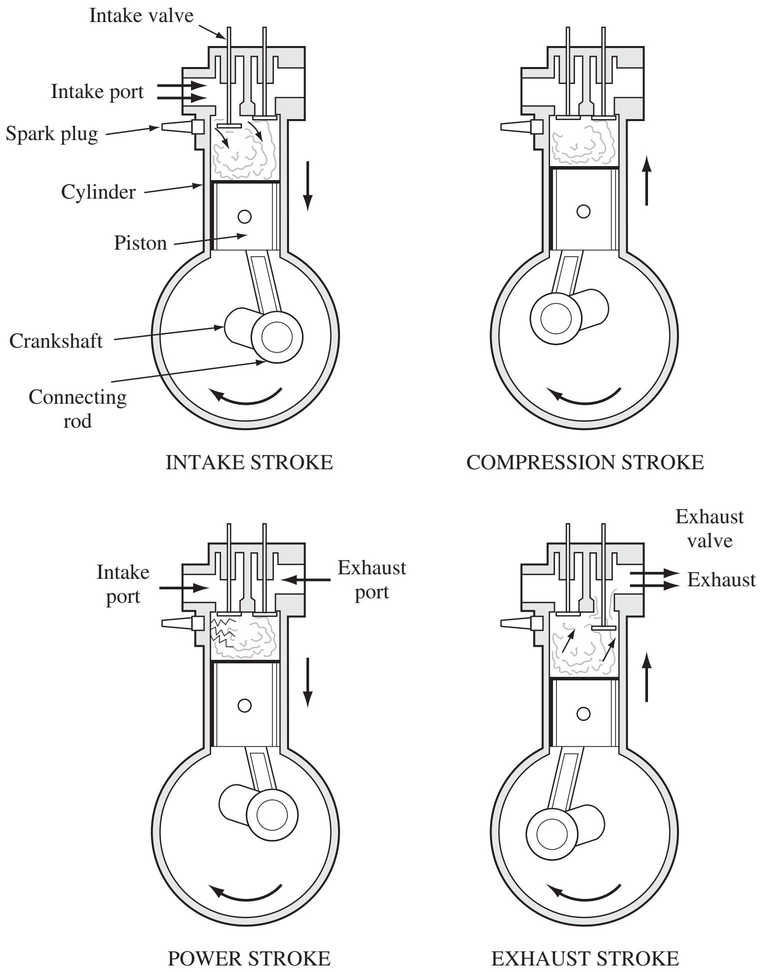


FIGURE 18 Schematic of a four-stroke, spark-ignited, internal combustion engine. *Intake:* Intake valve open, piston motion sucks in fresh air/fuel charge. *Compression:* Both valves closed, air/fuel mixture is compressed by rising piston, spark ignites mixture near end of stroke. *Power:* Air fuel mixture burns, increasing temperature and pressure, expansion of combustion gases drives pistons down. *Exhaust:* Exhaust valve open, spent gases are pushed out of cylinder by rising piston.
(*Source:* Powell and Brennan, 1988.)

compresses the air/fuel mixture in the cylinder against the now closed intake and exhaust valves. As the piston approaches the top of the compression stroke, the spark plug fires, igniting the mixture. In the *power* stroke, the burning mixture expands and forces the piston down, which turns the crankshaft and delivers power to the drive train. In the fourth, or *exhaust* stroke, the exhaust valve opens and the rising piston forces combustion products out of the cylinder, through the exhaust system, and into the air. Diesel engines operate with a similar cycle, but they compress the fuel more during the compression stroke, which makes the gases in the cylinder so hot that they ignite of their own accord without needing a spark plug.

The single most important factor in determining emissions from a four-stroke, internal combustion engine is the ratio of air to fuel in the mixture as it enters the cylinders during the intake stroke. To analyze that mixture ratio and its impact on emissions, let us begin with the stoichiometry of gasoline combustion. While modern gasolines are blends of various hydrocarbons, an average formulation can be represented as C_7H_{13}. We can write the following to represent its complete combustion in oxygen:

$$C_7H_{13} + 10.25\ O_2 \rightarrow 7\ CO_2 + 6.5\ H_2O \tag{29}$$

If we want to show complete combustion in *air,* we can modify this reaction to account for the fact that about 3.76 moles of N_2 accompany every mole of O_2 in air. Thus, $10.25 \times 3.76 = 38.54$ mol of N_2 can be placed on each side of the reaction, yielding

$$C_7H_{13} + 10.25\ O_2 + 38.54\ N_2 \rightarrow 7\ CO_2 + 6.5\ H_2O + 38.54\ N_2 \tag{30}$$

where any oxidization of nitrogen to nitrogen oxides has been neglected.

EXAMPLE 6 Stoichiometric Air-Fuel Ratio

Determine air-to-fuel ratio required for complete combustion of gasoline.

Solution For each mole of gasoline, 10.25 moles of O_2 and 38.54 moles of N_2 are required. Using (30), we can determine the masses of each constituent as

$$1\text{ mol } C_7H_{13} = 7 \times 12 + 13 \times 1 = 97\text{ g}$$

$$10.25\text{ mol } O_2 = 10.25 \times 2 \times 16 = 328\text{ g}$$

$$38.54\text{ mol } N_2 = 38.54 \times 2 \times 14 = 1{,}079\text{ g}$$

Considering air to be made up of only O_2 and N_2, the air-fuel ratio needed for complete oxidation of gasoline is

$$\frac{\text{Air}}{\text{Fuel}} = \frac{(328 + 1{,}079)\text{ g}}{97\text{ g}} = 14.5$$

This is known as the *stoichiometric ratio* for gasoline.

If the actual air/fuel mixture has less air than the stoichiometric ratio indicates is necessary for complete combustion, the mixture is said to be *rich*. If more air is provided than is necessary, the mixture is *lean*. A rich mixture encourages production of CO and unburned HC since there is not enough oxygen for complete combustion. On the other hand, a lean mixture helps reduce CO and HC emissions unless the mixture becomes so lean that misfiring occurs. Production of NO_x is also affected by the air/fuel ratio. For rich mixtures, the lack of oxygen lowers the combustion temperature, reducing NO_x emissions. In the other direction, beyond a certain point, lean mixtures may have so much excess air that the dilution lowers flame temperatures and also reduces NO_x production.

Figure 19 shows the relationship among CO, HC, and NO_x emissions, and the air/fuel ratio. Also shown is an indication of how the air/fuel ratio affects both power delivered and fuel economy. As can be seen, maximum power is obtained for a slightly rich mixture, while maximum fuel economy occurs with slightly lean

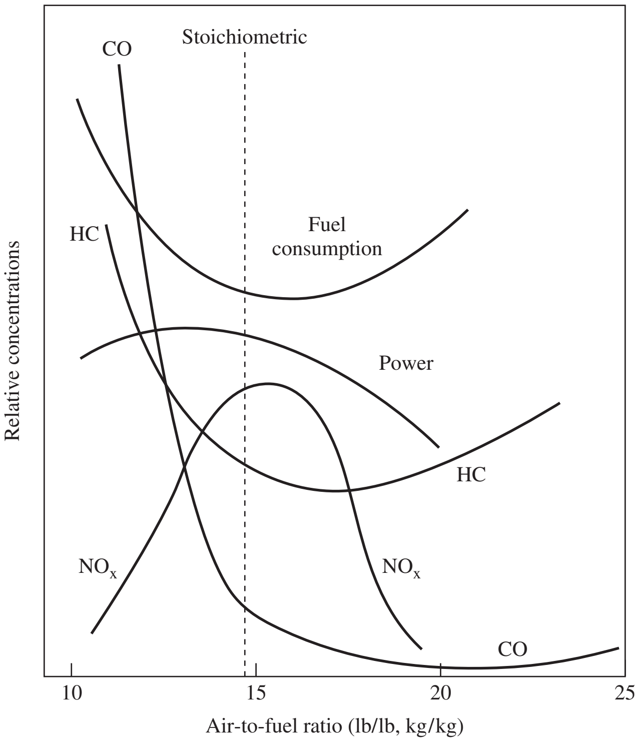
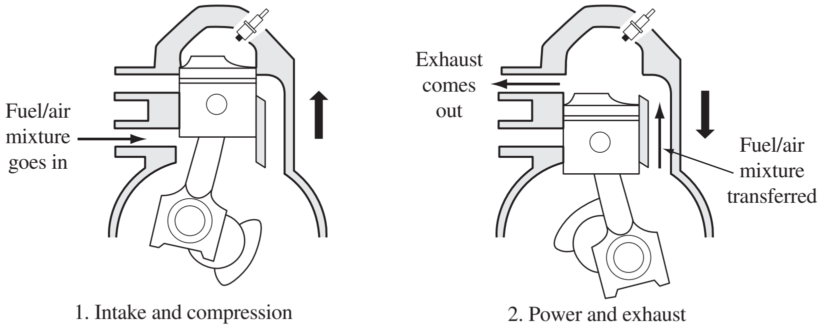


FIGURE 19 Effect of air/fuel ratio on emissions, power, and fuel economy.

mixtures. Before catalytic converters became the conventional method of emission control, most cars were designed to run slightly rich for better power and smoothness. When the first automobile emission limitations were written into law in the 1960s, only CO and HC were controlled, so manufacturers simply redesigned their engines to run on a less rich mixture. This simple change had the desired effect of reducing CO and HC, but at the same time NO_x emissions increased. Later, when the Clean Air Act Amendments of 1970 forced emission controls on all three pollutants, it was no longer possible to meet emission limitations by simply modifying the air/fuel ratio, and automotive engineers turned to the three-way catalytic converter.

The air/fuel ratio is, of course, not the only factor that influences the quantity of pollutants created during combustion. Other influencing factors include ignition timing, compression ratio, combustion chamber geometry, and very importantly, whether the vehicle is idling, accelerating, cruising, or decelerating. In addition, not all of the pollutants created in the combustion chamber pass directly into the exhaust system. Some find their way around the piston during the power and compression strokes, into the crankcase. This *blowby,* as it is called, used to be vented from the crankcase to the atmosphere, but now it is recycled back into the engine air intake system to give it a second chance at being burned and released in the exhaust stream. This control method is called *positive crankcase ventilation* (PCV), and the main component is a crankcase ventilator valve that adjusts the rate of removal of blowby gases to match the changing air intake requirements of the engine, making sure that the desired air/fuel ratio is not upset by these added gases.



1. Intake and compression 2. Power and exhaust

FIGURE 20 A two-stroke engine combines the intake and compression strokes as the piston rises and combines the power and exhaust strokes as the piston is driven downward. (*Source:* Powell and Brennan, 1992.)

Two-Stroke Engines

Two-stroke (or "two-cycle") engines are similar to conventional four-stroke engines in that they use a spark to ignite an air/fuel mixture to deliver power to a piston connected to a crankshaft. The key difference is that every other stroke is a power stroke rather than every fourth stroke as is the case with an Otto cycle engine. As suggested in Figure 20, the intake and compression strokes are combined into one, as are the power and exhaust strokes. On the upstroke, the piston compresses an air/fuel mixture that is about to be ignited while it draws the next cycle's fuel into the crankcase. The fuel is fired by a spark plug, driving the piston down. On that downstroke, the air/fuel mixture in the crankcase is forced around the piston into the combustion chamber, forcing the exhaust gases out the exhaust port. By doubling the number of power strokes, the two-cycle engine has more power for its size than a conventional four-stroke engine. It also is simpler because there are no moving valves. The net is a smaller, lighter, less expensive engine that has been popular for use in motorcycles, lawnmowers, and outboard motors.

The problem with two-strokes is that they emit much more pollution than four-stroke engines. As exhaust gases are being pushed out of the combustion chamber, some of the fresh fuel entering the chamber is pushed out as well. Also, two-stroke engines need oil to be mixed with the gasoline to lubricate moving parts in the crankcase, which leads to oily, smelly exhaust fumes as well as engine misfiring. While these engines constitute only a small fraction of the internal combustion engines in use, their emissions are disproportionately large. It has been estimated, for example, that a two-stroke outboard motor emits as much pollution in an hour as driving a late-model automobile 800 miles. In 1996, the EPA announced new rules that will virtually eliminate conventional two-stroke outboard motors.

Diesel Engines

At this time, the only real competition to the basic spark-ignition, internal combustion engine for motor vehicles is the *diesel*. Diesel engines have no carburetor since fuel is injected directly into the cylinder, and there is no conventional ignition system with plugs, points, and condenser since the fuel ignites spontaneously during the compression stroke. Diesels have much higher compression ratios than conventional

(Otto cycle) engines, and since they do not depend on spark ignition, they can run on very lean mixtures. Thus, they are inherently more fuel efficient. The increased fuel efficiency of diesels was responsible for their momentary surge in popularity just after the oil crises of the 1970s. They continue, however, to be popular in Europe.

Since diesels run with very lean mixtures, emissions of hydrocarbons and carbon monoxide are inherently very low. However, because high compression ratios create high temperatures, NO_x emissions can be relatively high. Moreover, since the fuel is burned with so much excess oxygen, conventional catalysts that require a lack of oxygen to chemically reduce NO_x are ineffective, and less efficient controls are used. In addition, diesels emit significant quantities of carbonaceous soot particles, and diesel engine exhaust has been labeled a probable human carcinogen containing over 40 cancer-related substances. Health risks from diesel exhaust are severe. The California Air Resources Board (CARB) estimates that 70 percent of the airborne cancer risk in the state is attributable to diesel particulate matter, and the combination of NO_x and PM is responsible each year for an estimated 2,900 premature deaths, 3,600 hospital admissions, 240,000 asthma attacks and respiratory symptoms, and 600,000 lost workdays. Advancing the technology needed for cleaner burning diesel engines has become a major public health priority.

The historic tradeoff for diesels—high torque and excellent fuel efficiency versus excessive NO_x and PM emissions as well as noise—have led to their on-again, off-again popularity. In Europe, where gasoline costs twice as much as in the United States, diesels outsell conventional spark-ignition vehicles. New emission controls announced in 2006 by several manufacturers suggest they may be on the verge of another comeback.

Automobile Emission Controls

Vehicular emissions are caused by both combustion and evaporation. During the exhaust stroke of an internal combustion engine, combustion gases are pushed through the exhaust manifold and out the tailpipe. In this exhaust system, most of the control of automobile emissions now occurs. In addition, fuel evaporation from the gas tank itself, from the engine while it is running, and during refueling at the gas station, all contribute significant amounts of hydrocarbon vapors to the atmosphere.

Exhaust emission controls have evolved considerably. An early approach to CO and HC control involved air injection into an enlarged exhaust manifold to encourage continued oxidation after these gases left the combustion chamber. Some degree of NO_x control was achieved by recirculating a portion of the exhaust gas back into the incoming air/fuel mixture. This relatively inert gas that is added to the incoming mixture absorbs some of the heat generated during combustion without affecting the air/fuel ratio. The heat absorbed by the recirculated exhaust gas helps reduce the combustion temperature and, hence, helps decrease the production of NO_x. Air injection as a method of control is no longer used, and exhaust gas recirculation is becoming less common.

The approach most favored by automobile manufacturers to achieve the emission standards dictated by the Clean Air Act has been the three-way catalytic converter ("three-way" just means it handles all three pollutants, CO, HC, and NO_x). A three-way converter is able to oxidize hydrocarbons and carbon monoxide to carbon dioxide, while reducing NO_x to N_2 all in the same catalyst bed. These catalytic converters are very effective in controlling emissions, and they have the advantage of allowing the engine to operate at near stoichiometric conditions where engine performance and efficiency are greatest. In fact, they *must* operate within a very narrow band of air/fuel ratios near the stoichiometric point or else their ability to reduce all three pollutants at once is severely compromised, as shown in Figure 21*a*. Maintaining that degree of control has required the development of

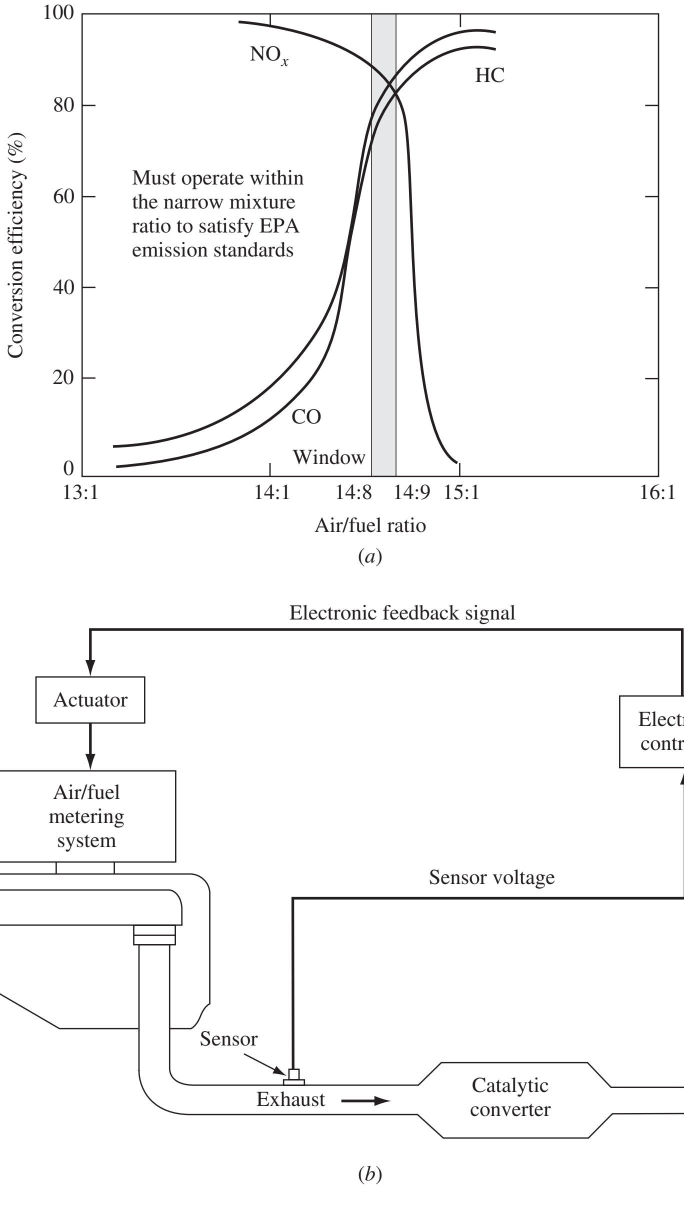


FIGURE 21 For a three-way catalyst to function correctly, the air/fuel ratio must be maintained within a very narrow band as shown in (*a*). To maintain that ratio, a closed-loop control system monitors the composition of exhaust gases and sends corrective signals to the air/fuel metering system.
(*Source:* After Powell and Brennan, 1988.)

precise electronic feedback control systems that monitor the composition of exhaust gas and feed that information to a microprocessor-controlled carburetor or fuel-injection system as shown in Figure 21*b*. Often it is a malfunction in this sophisticated control system that turns on your dashboard "engine" light.

Catalysts are quickly destroyed if leaded fuels are burned, resulting in no pollution control at all. In fact, it was the development of catalytic converters that led to the virtual elimination of tetraethyllead as an octane booster in gasoline. The enormous human health benefits resulting from reduced lead emissions are directly attributable to the development of catalytic converters. Lead, as well as other substances, including sulfur, zinc, manganese, silicone, and phosphorus, are said to "poison" the catalyst, rendering them ineffective by coating the surface of the catalyst, which keeps the exhaust gases from making contact.

Three-way converters are very effective once they are warmed up, but when they are cold, as well as when there are spurts of sudden acceleration or deceleration, they can release excessive amounts of pollution that may offset much of their perceived benefits. Those shortcomings are being addressed, and low-emission vehicles now usually have a separate early catalytic converter in the system to reduce startup emissions by burning off the hydrocarbons from the extra-rich mixture used in a cold engine. Another problem with catalytic converters is their rate of emission of nitrous oxide (N_2O, also known as laughing gas), which is relatively harmless in terms of human health, but is a potent greenhouse gas.

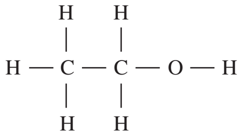
The particulate (soot) emissions coupled with the inherently high rate of NO_x production in diesel engines make control of their tailpipe emissions more difficult than is the case for conventional spark-ignition engines. A two-way oxidation catalyst is commonly used to oxidize CO and HC to CO_2 and H_2O. These catalysts also help reduce soot, but since diesel fuel is burned with so much excess oxygen (very lean), conventional catalysts that require a lack of oxygen to control NO_x are ineffective. Control technologies include NO_x adsorbers, selective catalytic reduction (SCR), and diesel particulate filters (DPFs). DPFs are problematic since they can clog and lose their effectiveness over time.

A promising new approach to diesel engine design is called the *homogeneous-charge compression-ignition* (HCCI) engine. By carefully controlling the concentration and temperature of fuel and air, the entire mixture begins to burn simultaneously during compression. HCCI engines provide more complete combustion at lower temperatures, which greatly reduces the formation of NO_x, and thereby increases the likelihood that tailpipe emission controls can achieve U.S. standards.

Cleaner Gasoline

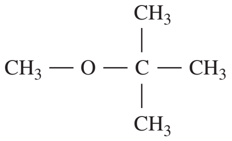
Motor vehicle hydrocarbon emissions are not only the result of inefficient combustion; they are also caused by vaporization of the fuel itself. Vaporization rates are very dependent on temperature, so evaporation is especially high just after the engine is turned off, when the fans, cooling systems, and airflow due to vehicle movement are no longer operative. To help control these evaporative emissions, automobiles are equipped with vapor-recovery systems, but these are not totally effective. A complementary approach is to reduce the volatility of the gasoline itself, but there are tradeoffs. In cold, winter conditions, vehicles are harder to start unless fuels vaporize readily at low temperatures, but in the summer, that higher volatility would lead to excessive evaporative emissions. Refineries then must formulate their fuels differently depending on location and season.

Gasoline can also be reformulated to help reduce tailpipe emissions of CO and HC by adding oxygenates that encourage more complete combustion. The Clean Air Act Amendments of 1990 have since 1992 required 2 percent oxygen (by weight) in fuels used in areas that have high levels of CO pollution. The additives most often used to create these "oxygenated" fuels have been ethanol (C_2H_5OH) and methyl tertiary butyl ether (MTBE).



$$H-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-O-H$$

Ethanol

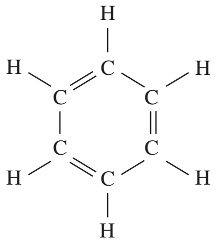


$$CH_3-O-\overset{CH_3}{\underset{CH_3}{C}}-CH_3$$

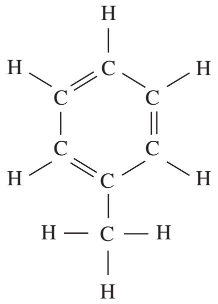
Methyl tertiary butyl ether (MTBE)

MTBE has been the preferred oxygenate because it has a higher octane number and is not as volatile as ethanol. On the other hand, MTBE is water soluble and does not biodegrade easily. Unfortunately, leaking underground fuel tanks, as well as fuel spills, have allowed MTBE to contaminate groundwaters and reservoirs, creating a potential drinking water hazard. Studies with rats and mice suggest drinking MTBE may cause gastrointestinal irritation, liver and kidney damage, and nervous system effects. There is some evidence that inhalation while pumping gasoline can cause nausea, dizziness, mental confusion, and headaches. The full health effect impacts of MTBE have not yet been fully assessed. As a result of these concerns, the EPA has been wrestling with banning MTBE and perhaps removing the 2 percent oxygen requirement in reformulated gasoline that motivated its use in the first place. California banned use of MTBE as of 2003.

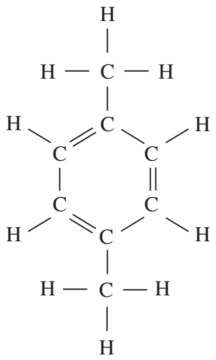
Another controversy over the formulation of gasoline has to do with the octane enhancers that have replaced tetraethyllead. Until recently, the antiknock properties of tetraethyllead have been achieved in the United States by increasing the proportion in gasoline of certain benzene derivatives such as toluene, $C_6H_5CH_3$, and xylene, $C_6H_4(CH_3)_2$. Along with benzene, these are collectively known as *BTX* (*b*enzene, *t*oluene, *x*ylene). BTX boosts the octane rating of gasoline, but BTX hydrocarbons are more reactive than normal constituents of gasoline, so they can increase formation of photochemical smog.



Benzene



Toluene



Xylene

There is another octane enhancer, methylcyclopentadienyl manganese tricarbonyl, better known as MMT, which was used in the United States until 1977 when health concerns led to its withdrawal. Combustion products of MMT contain toxic manganese, and the EPA opposed its use until health risks could be evaluated. The manufacturer of MMT argued that it had been used for many years in Canada without apparent health effects, and in 1995, a Federal appeals court ruled in their favor and use of MMT began again in the United States in 1996.

Alternative Fuels

A number of alternatives to gasoline are being investigated as possible fuels for the future. These include ethanol, methanol, compressed natural gas (CNG), propane, hydrogen, and electricity. The motivation for developing vehicles capable of running on these fuels is a combination of the potential to reduce air pollution and, perhaps more importantly, to reduce our demand for petroleum and the negative environmental, nonrenewable resource depletion, and national security issues associated with its use.

The Alternative Motor Fuel Act of 1988 (AMFA). One of the incentives provided by Congress in AMFA is a provision that allows manufacturers to earn "credits" toward meeting their fleet CAFE obligations whenever they sell a vehicle that has the capability of running on an alternative fuel—whether or not that vehicle ever does so. Several million *flexible-fuel (FFV),* or *dual-fuel,* vehicles are on the road today because of that provision.

The CAFE fuel-efficiency of a flex-fuel vehicle is based on a weighted average of the fuel efficiency while running on gasoline (or diesel) and the efficiency while running on the alternative fuel. To provide the motivation for flex-fuel vehicles, the mpg calculation allows 1 gallon of alternative fuel to be counted as being equivalent to 0.15 gallons of gasoline. It also assumes 50 percent of the miles driven are powered by the alternative fuel, whether or not such a fuel is ever used (U.S. DOT, 2002). Example 6 shows an example calculation.

EXAMPLE 6 CAFE Fuel Efficiency for a Flex-Fuel Vehicle

A flex-fuel vehicle gets 24 mpg while running on gasoline and 16 mpg while running on pure ethanol (the energy density of ethanol is only about two-thirds that of gasoline). Find its CAFE fuel efficiency rating.

Solution Using the standard calculation method with 50 percent of the miles on gasoline and 50 percent on ethanol, with 1 gallon of ethanol counted as 0.15 gallons of gasoline, gives

$$\text{CAFE mpg} = \frac{1 \text{ mile}}{\left[\dfrac{0.5 \text{ mile}}{24 \text{ mile/gal gas}} + \left(\dfrac{0.5 \text{ mile}}{16 \text{ mile/gal alcohol}} \times \dfrac{0.15 \text{ gal gas}}{1 \text{ gal alcohol}}\right)\right]} = 39.2 \text{ mpg}$$

So, the manufacturer gets to sell a 24 mpg (gasoline) car and count it as a 39.2 mpg vehicle in their fleet of vehicles sold that year, even if the buyer never uses ethanol fuel.

While the flex-fuel CAFE credits have resulted in slightly lower average fuel economy for fleets of vehicles certified as meeting the 27.5 mpg standard, the credits do help break the "chicken and egg" problem of which comes first: cars that can run on alternative fuels, or gas stations that will provide those fuels.

Ethanol. Ethanol (CH_3CH_2OH), also known as grain alcohol, has long been used as an oxygenate to reduce CO emissions, in which case the gasoline/ethanol mixture is usually referred to as *gasohol*. But it is now being much more widely discussed as a potential fuel to help reduce our dependence on imported oil.

Ethanol can be blended with gasoline with a range of mixture ratios. The two mixtures common in the United States are E10 (10 percent ethanol, 90 percent gasoline, by volume) and E85 (85 percent ethanol, 15 percent gasoline). Motivated by AMFA, several million flex-fuel vehicles capable of running on any range of mixtures from pure gasoline all the way to E85, have been part of the U.S. motor vehicle fleet for many years. In fact, Henry Ford's original Model T's and later Model A's were flex fuel vehicles capable of running on ethanol or gasoline.

The main modifications required to make a gasoline vehicle into an ethanol flex-fuel vehicle include a sensor that detects the type of fuel being pumped to the engine, higher flow-rate fuel injectors needed to accommodate the lower energy density of ethanol relative to gasoline, and electronic and software modifications to ensure proper operation of the fuel injection system. In addition, certain modifications to the materials used in fuel lines, fuel tanks, and fuel pumps are needed. These modifications are so inexpensive that manufacturers, motivated in part by the CAFE benefits, often offer their flex-fuel vehicles at no extra cost.

Most owners of these flex-fuel vehicles have no idea that they are ethanol capable, and even if they did, there are so few gas stations that offer E85 that most have never tried to use it. To help inform buyers, it has been suggested that flex-fuel vehicles be equipped with some distinguishing feature, such as a yellow gas cap (symbolizing the color of corn, from which most U.S. ethanol is made). There will be certain differences for drivers using E85. For example, the energy density of E85 is 29 percent less than that of gasoline (Table 5), which suggests E85 would need to sell for 29 percent less per gallon to be equivalent in energy content. For example, if gasoline costs $3.00 per gallon, E85 would need to cost no more than 0.71 × $3.00 or $2.13 to be cost effective. The vehicle's driving range on a tank of E85 would be similarly decreased.

TABLE 5

Energy Density of Fuels (LHV)

	(Btu/Gallon)	(MJ/L)	% of Gasoline
Gasoline	115,400	32.2	100
Ethanol	75,670	21.1	66
E85	81,630	22.8	71
Methanol	56,600	15.8	49
M85	65,400	18.2	57
Petro-diesel	137,680	36.4	113
Biodiesel (avg)	121,600	33.9	105

Source: http://bioenergy.ornl.gov/papers/misc/energy_conv.html.

In Brazil, large amounts of ethanol are fermented from sugarcane, which supplies roughly 40 percent of its nondiesel vehicle fuel. In fact, all gasoline sold in Brazil now must have at least 20–25 percent ethanol. In the United States, about 90 percent of ethanol comes from corn, but it is also produced using a petroleum-based feedstock, ethylene. While corn and sugarcane have been the traditional sources of ethanol, there is considerable excitement these days for the potential to derive ethanol from cellulosic (woody) materials ranging from agricultural waste to fast-growing perennial grasses such as switchgrass and miscanthus. Cellulosic ethanol wouldn't necessarily compete with food crops for agricultural land and its energy balance ratio (energy derived divided by energy needed in production) would be much higher than that for corn or sugarcane.

In terms of CO_2 emissions, there has been some controversy over whether or not corn-based ethanol is better than gasoline. Burning biomass fuels has the inherent advantage that the carbon released to the atmosphere is equivalent to the carbon that was removed during photosynthesis. That simple balance is somewhat offset, however, by fossil fuels needed to power the farm equipment, manufacture the chemicals and fertilizers, process and transport the biomass fuel, and so forth. Similarly, the drilling, pumping, refining, and transporting required to deliver gasoline to your tank also involves a carbon investment that must be accounted for when comparing gasoline to ethanol. With careful accounting, greenhouse gas emissions of corn-based ethanol are similar to gasoline, sugarcane is a better source, and cellulosic ethanol promises to be far better than either corn or sugarcane (Farrell et al., 2006). All three are much less petroleum intensive than gasoline, which is a national security benefit.

While there are clear carbon advantages to ethanol, those may be offset by the potential for ethanol to make conventional air pollution worse. A recent study (Jacobson, 2007) concludes vehicles fueled by E85 would have significantly higher emissions of organic gases, especially formaldehyde (HCHO) and acetaldehyde (CH_3CHO), which could lead to an increase in ozone-related mortality, hospitalizations, and emergency-room visits, especially in sunny, smoggy areas like Los Angeles.

Methanol. Methanol (CH_3OH), also known as wood alcohol, has long been the fuel of choice in high-performance racing cars. It has a much higher octane rating than gasoline, which enables engines to be designed with higher compression ratios for increased power. It burns with a lower flame temperature than does either gasoline or diesel fuel, so NO_x emissions are reduced. It burns more completely, as well, so that HC and CO emissions are also reduced.

Methanol is not without its problems. Its emissions are much higher in formaldehyde (HCHO), an eye irritant and a suspected carcinogen. The low volatility of pure methanol makes it difficult to start engines in cold conditions, and it is highly toxic. It burns without a visible flame, which can lead to especially dangerous conditions in an accident involving fire. Moreover, its energy content is only about half that of gasoline, so either much larger fuel tanks would be required or the driving range would be considerably reduced.

Some of the problems of pure methanol can be overcome by using a blend of methanol and gasoline. A mixture of 85 percent methanol and 15 percent gasoline (M85) eliminates the cold start problem and yields a visible flame as well. It also tastes much worse than pure methanol, which should discourage ingestion. The blend has an octane rating of 102 (versus 87–92 for gasoline) and a volumetric energy content of a little more than half that of gasoline. While the increase in combustion efficiency does slightly offset the lower energy density, the driving range for M85 is still only about 60 percent of that for gasoline. Eventually, if methanol becomes widely available, it would be possible to take advantage of the increased octane level with more efficient engines dedicated to M85 use.

With regard to emissions of the greenhouse gas, carbon dioxide, methanol derived from natural gas produces slightly less CO_2 than gasoline. On the other hand, if methanol is produced from coal, total CO_2 emissions released during coal processing and methanol combustion could be double that of gasoline (Moyer, 1989).

Biodiesel. Biodiesel can be created from vegetable oils, animal fats, or recycled restaurant greases. It is a biodegradable, domestic, renewable energy fuel that has the potential to help reduce the need for petroleum-based transportation fuels as well as helping farmers by providing a market for excess soybean oil. As a renewable fuel whose production requires little or no fossil fuel, its net CO_2 emissions are only about one-fourth that of standard diesel (U.S. DOE, 2006). The oxygen in biodiesel enables more complete combustion to take place, and the lower sulfur content reduces sulfate emissions. In fact, biodiesel meets the EPA 2006 regulations for ultra-low sulfur diesel fuel. Compared with conventional diesel, biodiesel substantially reduces emissions of unburned hydrocarbons, carbon monoxide, sulfates, polycyclic aromatic hydrocarbons, and particulate matter. NO_x emissions, however, are increased.

Biodiesel can be blended with conventional diesel in any mixture strength, with the most common blend B2 (2 percent biodiesel, 98 percent standard diesel) being popular because of its superior lubricity compared to new ultra-low-sulfur diesel fuels. Other common blends include B5 and B20. Conventional diesel engines can run on B20 without any modifications, but higher concentrations may require certain engine modifications to avoid maintenance or operational problems. A fleet of recycling and garbage trucks in San Jose, CA, are currently running on B100. Biodiesel has a high Cetane number (good for cold-start fuel ignition) and superior lubricity, which can help extend engine life. The energy density of B100 (118,170 Btu/gal) is slightly lower than diesel No. 2 (typically around 129,000 Btu/gal), but in the low concentrations normally used, that decrease should be unnoticeable. Some reports that biodiesel may cause problems with certain types of elastomeric materials have been reported. B100, and similarly high concentrations, may be inappropriate for cold climates due to potential gelling of the fuel. Winterization packages, such as fuel heaters, can mitigate this problem, however, as was demonstrated by a successful test of B100 used in vehicles in Yellowstone National Park.

Biodiesel is produced using a process called *transesterification* in which the unwanted glycerin in vegetable oil or fat is removed by chemical reactions involving an alcohol, such as methanol or ethanol, and a catalyst. The catalyst is usually sodium or potassium hydroxide. The biodiesel end product consists of long chain methyl esters. Though some have tried it, running straight vegetable oil or recycled grease in diesel engines can cause multiple engine problems and should be avoided.

Biodiesel can also be used for nontransportation applications, such as boilers or home heating oil systems. In these applications, sulfur emissions can be dramatically reduced since home heating oil tends to have high sulfur content, and biodiesel has almost none. B20, for example, can reduce sulfur emissions by almost 20 percent. NO_x emissions are reduced by roughly the same fraction (U.S. DOE, 2006).

Compressed Natural Gas (CNG). This is already used in fleets of trucks, delivery vans, and buses, especially in Canada and New Zealand. CNG is a very clean fuel, and very low emissions of reactive hydrocarbons, carbon monoxide, particulates, and toxics are possible. NO_x emissions can be higher than gasoline-fueled vehicles, but only because earlier NO_x emission-control systems were not as well developed as those designed for gasoline. Heavy fuel tanks are required to contain the highly pressurized gas, and at 3,600 pounds per square inch of pressure, it takes three times as much volume to store an equivalent amount of energy compared with gasoline. Methane is a much more potent greenhouse gas than carbon dioxide, so leakage could make methane worse for global warming than gasoline. CNG vehicles qualify for the alternative-fuels CAFE benefits. For example, a vehicle that gets 25 miles per 100 ft^3 of (STP) natural gas has a CAFE rating of over 200 mpg (see Problem 17).

To counter the problem of finding a CNG fueling station, one manufacturer has introduced a home refueling appliance called Phill (Figure 22). The wall-mounted unit has a small compressor that can deliver up to 3,600 psi of natural gas, which allows a CNG vehicle to refuel overnight directly from a home natural gas supply line.

Liquefied Petroleum Gas (LPG). LPG (propane) is another fuel that has been used for vehicle fleets such as school buses, trucks, fork lifts, taxis, and farm vehicles. Some vehicles have been designed with dual tanks so that either gasoline or LPG can be burned, but the added flexibility of fuel sources is offset somewhat by the added

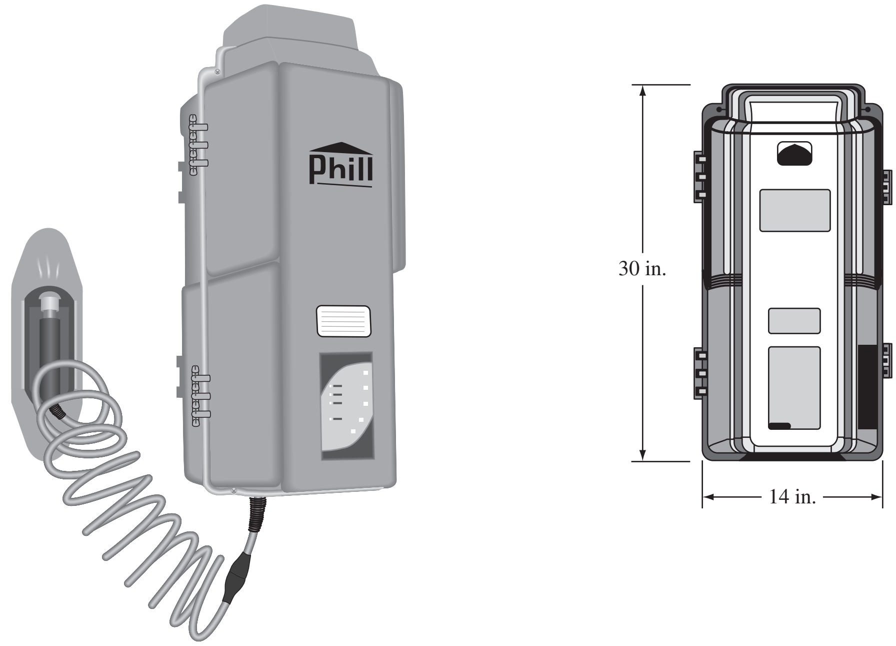


FIGURE 22 Phill is a home CNG refueling appliance for natural-gas powered vehicles. (*Source:* Courtesy FuelMaker Corporation.)

space, weight, and complication of such systems. Compared with gasoline, carbon monoxide emissions are somewhat lower; hydrocarbon emissions are roughly the same but are less reactive; NO_x emissions tend to be higher but could be lower if appropriate control systems are developed; and carbon dioxide may be lower or higher depending on whether the LPG is derived from natural gas or petroleum.

Electric-Drive Vehicles

Some of the most promising ways to reduce motor vehicle air pollution are based on augmenting, or even replacing, internal-combustion engines with electric motors. By now, everyone is familiar with Hybrid-Electric Vehicles (HEVs), which were described in the previous edition of this book as "A Car for the Future." Well, the future arrived very quickly and HEVs are now quite common. A new generation of HEVs called Plug-in Hybrid-Electric Vehicles (PHEVs) are now being introduced, and in 2006, Tesla Motors introduced an all-electric roadster with phenomenal performance characteristics, which has reinvigorated electric vehicle (EV) enthusiasts. Finally, some are optimistic about hydrogen-powered, fuel-cell vehicles as the car of the future.

Hybrid Electric Vehicles (HEVs). HEVs have both an internal combustion engine (ICE) and an onboard electric power system made up of a generator, electric motor, and battery storage. Two approaches to the system design are used. In a *serial* design, the ICE sends all of its power directly to a generator. The generator then either charges the batteries or powers an electric motor that drives the wheels. The engine never directly powers the car. In the more popular *parallel* design, the ICE and the electric motor can both supply power to the wheels, switching back and forth as driving conditions vary.

Figure 23 shows one version of a parallel HEV system under varying operating conditions. Notice the dual function of the motor/generator. When electric power is put into a motor/generator, it acts as an electric motor delivering mechanical shaft power; when shaft power is put in, it acts as a generator delivering electricity. For very slow-speed driving, the car can run entirely on its batteries. When cruising along, the ICE and motor/generator (acting as an electric motor) provide power, with perhaps some shaft power driving a small additional generator to recharge the batteries. When bursts of power are needed, the battery pack is used to provide additional power to the electric motor. An added advantage of electric drive is the ability to do regenerative braking; that is, when coasting or when the brakes are applied the electric motor becomes an electric generator that recharges the batteries with the energy that would normally be dissipated by brake pads. Thus, the energy storage device is charged partly from the onboard electric power system and partly from regenerative braking.

Plug-In Hybrid Electric Vehicles (PHEVs). The traction batteries in a conventional HEV have two purposes. They provide energy for acceleration, and they provide a place to store energy captured during slowing or braking. As such, they act primarily as an energy buffer, not as a source of energy to power the vehicle very far, or very fast, on electricity alone. A PHEV, on the other hand, is an HEV with enough extra battery capacity added to allow the car to be driven a considerable

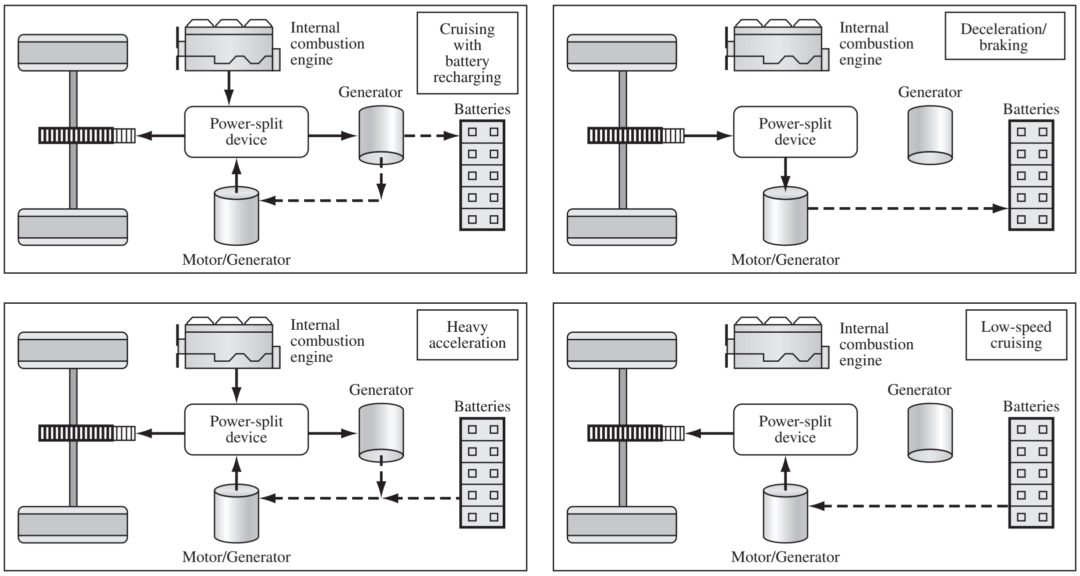


FIGURE 23 Showing energy flows through a parallel HEV under four different driving conditions. The figure is modeled after the Toyota Prius Hybrid Synergy Drive System.

distance on battery power alone. The intention would be to charge those batteries, most likely at night and at home, by plugging the car into the utility grid, so that the next day, the first 30 or so miles would be powered by the batteries, and after that, the car would operate as a conventional HEV running on gasoline. With typical driving patterns, probably half or more of the miles driven would be on electricity instead of gasoline, which could easily result in PHEVs that could travel over 100 miles per gallon of gasoline (plus the electrical input).

To the extent that PHEVs operate as electric vehicles, there would be virtually no tailpipe emissions, so pollution would shift from the city where vehicles and people are concentrated, to the countryside where power plants tend to be located. Not only would the location of emissions shift, but the type of emissions would change as well. If the power plants burn fossil fuels, then HC and CO emissions would be reduced, NO_x could still be a problem, and particulate and SO_x emissions could greatly increase. Photochemical smog problems in the city would be eased, but acid rain and other industrial smog problems could be exacerbated. Of course, to the extent that electric power production might shift from fossil fuels, especially coal, to renewable energy sources or nuclear power, there would be an obvious air quality advantage even in the countryside. The additional benefit of driving cars on domestic fuels rather than imported oil is also very important (oil is used to generate less than 2 percent of U.S. electricity).

Assuming PHEVs are charged at night, when roughly half of the nation's generators are idle (and power is cheap), with proper planning and incentives, a sizeable fleet could operate without building new power plants. In fact, the only additional infrastructure needed for a PHEV is an extension cord. Figure 24 shows one scenario that indicates roughly 40 percent of the vehicle-miles in California could be electric-drive vehicles supplied by plants currently idle at night.

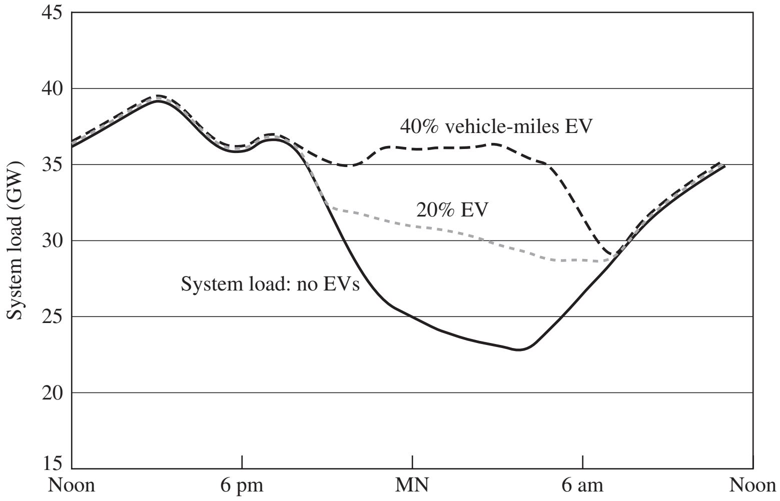


FIGURE 24 Showing a typical day's load curve for California and the impact of 20 percent and 40 percent of automobile vehicle-miles supplied by off-peak power.

As the following example illustrates, the "fuel" cost for electricity in an efficient PHEV is roughly one-fourth that of gasoline. The current hesitation to embrace PHEVs is based on a concern for the additional cost of batteries and their likely longevity. Assuming these will be overcome, PHEVs could well be the quickest and easiest way to ease our dependence on foreign oil and reduce urban air pollution.

EXAMPLE 7 Cost of Electricity for a PHEV

Suppose a PHEV gets 45 mpg while running on gasoline that costs \$3.00/gallon. If it takes 0.25 kWh to drive 1 mile on electricity, compare the cost of fuel for gasoline and electricity. Assume electricity is purchased at an off-peak rate of 6¢/kWh.

Solution On a per-mile basis, the cost of gasoline is

$$\text{Gasoline} = \frac{\$3.00/\text{gal}}{45\ \text{mi/gal}} \times \frac{100¢}{\$} = 6.67¢/\text{mile}$$

The cost to run on electricity is

$$\text{Electricity} = 0.25\ \text{kWh/mi} \times 6¢/\text{kWh} = 1.5¢/\text{mile}$$

Electric Vehicles (EVs). The old image of electric vehicles with golf-cart performance was shattered in 2006 when a small start-up company called Tesla Motors introduced an all-electric roadster with a 250-mile range that accelerates from 0 to 60 mph in about 4 seconds. Compared to a PHEV, or even a conventional ICE, an EV is especially simple since the drive-power system consists almost entirely of just a motor/generator, battery bank, transmission, and power electronics. The extra weight of the batteries is somewhat offset by weight reductions gained by having such a simple drive system. The 90 percent efficient, 185-kW (250 hp) motor/generator in the Tesla,

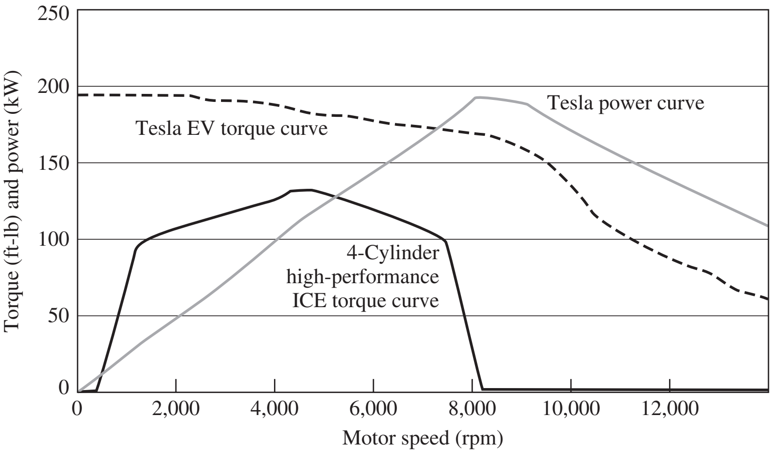


FIGURE 25 Torque and power curves for the Tesla EV versus a high-performance ICE engine. Electric motors show high torque across the range of motor speeds.

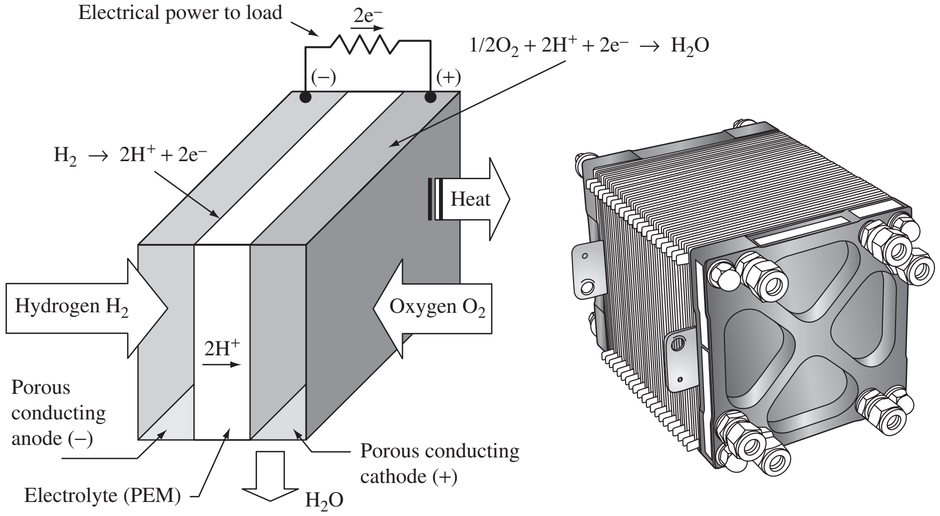
for example, is much smaller, more efficient, and lighter than a conventional high-performance ICE (it weighs only about 70 pounds and is the size of a big watermelon).

Electric vehicles have inherent performance benefits as well as reduced emissions, when compared to conventional ICEs. The torque curve for electric motors is quite different from that of ICEs (see Figure 25), and it is the shaft-twisting foot-pounds of torque that determines the ability of a car to accelerate quickly. Electric motors have their highest torque at zero rpm, and it remains high throughout most of their motor-speed curves. That means they can accelerate quickly, especially from a standing start. ICEs, on the other hand, have very little torque at low rpm, and for peak performance, their rpm has to be maintained within a very narrow range, which translates to frequent gear changes to maintain optimum torque.

Fuel-Cell Vehicles (FCVs). In the eyes of some, the ideal future vehicle would be one that converts hydrogen and oxygen in a fuel cell to produce electricity, which could then power an electric-drive system. The only emissions from the vehicle itself would be pure water. A number of different fuel cell technologies for both electric vehicles and stationary applications have been under development for a number of years. The technology most appropriate for vehicles is called a proton-exchange-membrane (PEM) cell (sometimes called a polymer-electrolyte membrane fuel cell).

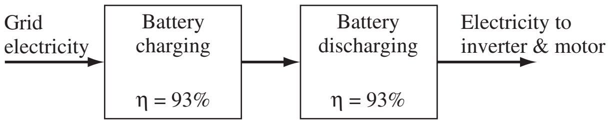
As shown in Figure 26, a PEM cell consists of two porous gas diffusion electrodes separated by a thin membrane that allows positive hydrogen ions to pass through it, while blocking electrons or neutral gases. The catalyst encourages the entering hydrogen gas to dissociate into protons and electrons. The protons pass through the electrolyte and the electrons flow through the external circuit delivering electrical power to the load. The hydrogen ions and electrons recombine in the cathode, forming water. Each cell produces only about 1/2 volt, so many cells are connected in series to form a fuel cell stack with sufficient voltage to charge batteries or power the electric motor in the vehicle.

The source of hydrogen for a fuel cell can be any of a number of hydrocarbon fuels, including natural gas, and even gasoline, or it may be hydrogen gas obtained by electrolysis of water. In the usual circumstance, natural gas would be stripped of its hydrogen using steam and a catalyst in a fuel reformer. The remaining CO_2 is

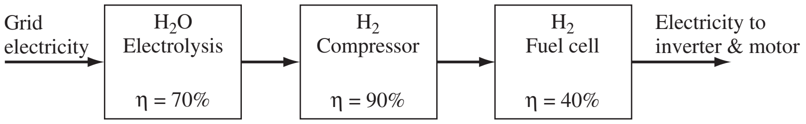


(*a*) A single cell (*b*) A fuel-cell stack

FIGURE 26 Basic configuration of an individual cell in a PEM fuel cell. A fuel cell stack consists of multiple series-connected cells.



(*a*) Battery-electric car: grid-to-motor efficiency = 86%



(*b*) Electrolysis/Fuel cell car: grid-to-motor efficiency = 25%

FIGURE 27 The efficiency from grid electricity to electric power delivered to the motor for a battery-electric vehicle is more than three times as efficient as using electrolysis to provide hydrogen for a fuel-cell vehicle.
(*Source:* Courtesy of Tesla Motors.)

released into the atmosphere, so there are greenhouse gas emissions. The second approach, electrolysis, involves passing an electric current through water to break apart the H_2O molecules, releasing H_2 and O_2 gases. Essentially no greenhouse gas emissions would be produced if the electricity for electrolysis comes from a solar, wind, or nuclear source. The overall efficiency from electricity to hydrogen and back to electricity generated by the fuel cell is very low, however. As Figure 27 illustrates, that conversion is only about one-third as efficient as simply charging and discharging batteries in an EV.

Hydrogen can be pressurized and stored as a compressed gas, but even at 8,000 psi, it would take nearly five times the volume to store the same energy as gasoline. At least one manufacturer is trying to develop a −253°C, liquefied hydrogen storage system. Even as a liquid, the energy density of hydrogen is still only about one-fourth that of gasoline. Hydrogen can also be stored in the form of metal-hydrides in which hydrogen gas is absorbed into the crystalline structure of a metal such as titanium or magnesium. Metal-hydrides have been used to store hydrogen for use in experimental fuel-cell-powered buses, but weight and volume issues are still significant. To store the same energy in metal hydrides as a 15-gallon, 150-pound gasoline tank would require a 45-gallon tank weighing approximately 600 pounds.

Fuel-cell cars have numerous hurdles to overcome before they can be considered a viable technology. Besides the technical problems of complexity, weight, range, and acceleration for the vehicle itself, there are questions to be resolved about where the hydrogen will come from in the first place, as well as how it will be transported and distributed to fueling stations. The "chicken-and-egg" problem of which comes first, fuel-cell-powered cars or hydrogen fueling stations, is problematic.

Well-to-Wheels Greenhouse Gas Emissions. A "well-to-wheels" analysis of CO_2 emissions starts at the source of fuel and ends with power delivered to the wheels. For a conventional gasoline-powered car, the analysis is reasonably straightforward. It starts with an estimate of the efficiency with which petroleum in the ground is extracted, processed into gasoline, and transported to the local gas station (around 82 percent). Combining that with an assumption that all of the carbon in gasoline is released as CO_2, along with a miles-per-gallon estimate, leads to emissions per mile. For PHEVs, EVs, and FCVs, we need to estimate CO_2 emissions per kWh of electricity, which depends on the mix of fuels and efficiencies for the source of that power. And we need an estimate of miles per kWh. For PHEVs, which burn gasoline and run on electricity, we need to assume some ratio of miles driven on each energy source. In other words, any comparison between vehicle types is very dependent on the assumptions made.

Figure 28 shows a well-to-wheels analysis in which carbon from electricity is the average from the U.S. grid (including 96 percent-efficient transmission). For the PHEV, half of the miles driven are powered by electricity at 0.25 kWh/mile, and half are powered by gasoline at 45 mpg (a little lower than the 50 mpg HEV to account for the extra weight of batteries). Other assumptions and key data are shown in the figure. Compared to the 50 mpg HEV, the PHEV has 9 percent lower carbon, and the EV is 29 percent lower. It is interesting to note that the fuel-cell vehicle is responsible for 8 percent more carbon than a conventional 50-mpg car and 52 percent more than the battery-electric vehicle.

The carbon emissions associated with electricity for the PHEV and EV in Figure 28 are assumed to come from the average U.S. grid. The overall efficiency of that power is less than one-third, and more than half of that is generated with high-carbon-intensity, coal-fired power plants. If we imagine the power coming from the newest and best 60 percent-efficient, natural-gas-fired, combined-cycle (NGCC) plants, and with a 96 percent-efficient grid delivery system, the carbon emissions for the EV drop to about 76 gCO_2/mile, which is one-third that of a 50-mpg HEV (see Problem 23).

There is another possibility for reducing emissions of greenhouse gases and criteria pollutants from our automobile fleet. Both PHEVs and EVs could get their electricity indirectly from the sun using grid-connected photovoltaics (PVs). Under a scheme called *net metering,* the owner of a PV array can sell electricity to the grid whenever more power is being produced than is needed, and buy it back again at

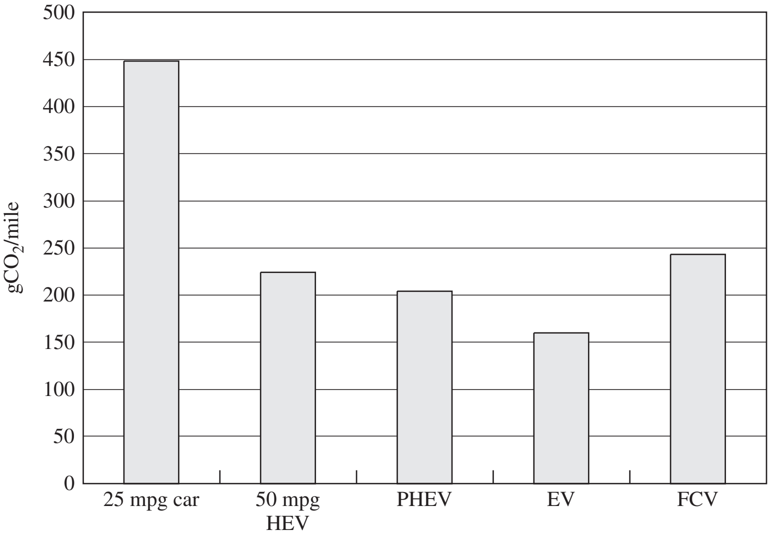


FIGURE 28 Well-to-wheels carbon analysis of five vehicles.
Assumptions: Well-to-wheels gasoline 11.2 $kgCO_2$/gal; Average grid electricity 640 gCO_2/kWh (including 96 percent efficient grid); HEV: 50 miles/gallon; PHEV: Half miles on gasoline @ 45 mpg, half at 0.25 kWh/mi; EV: 0.25 kWh/mi; FCV: 0.36 miles/MJ, Natural gas @ 14.4 gC/MJ, 61 percent NG-to-H_2 efficiency.

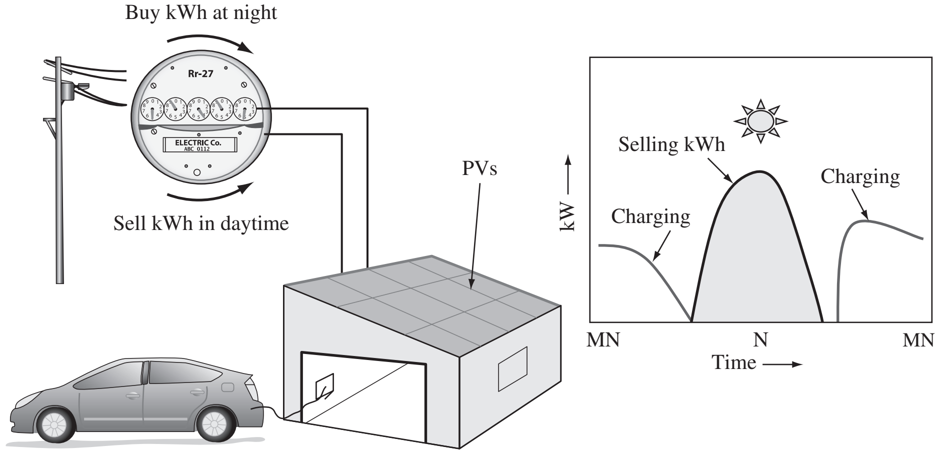


FIGURE 29 Photovoltaics on the roof of a single-car garage in a reasonably sunny location could provide enough emission-free solar power for a PHEV or EV to travel more than 10,000 miles per year.

any time. Figure 29 illustrates the concept. As Example 8 suggests, a PV array roughly the size of the roof of a single-car garage could generate enough electricity to drive an efficient PHEV or EV over 10,000 miles per year on pollution-free solar electricity. The economics are interesting as well. As Example 7 demonstrated, electricity used to displace gasoline is worth about $0.25/kWh, which is roughly the current cost of PV electricity without any subsidies.

EXAMPLE 8 Driving on Sunlight

In "full sun" (defined as 1 kW/m^2 of solar radiation), a 15 percent-efficient PV array can deliver 0.15 kW/m^2 of DC power under standard test conditions. After accounting for dirt, temperature effects, and conversion from DC to AC, about 75 percent of that will be delivered to the grid. For a location with an annual average insolation (solar radiation) equivalent of 5 hours/day of full sun, how many kWh per day could be generated? If an EV or PHEV requires 0.25 kWh of power from the grid to travel 1 mile, how many miles could be driven on the electricity delivered by a 15 m^2 (161 ft^2) PV array?

Solution The average daily energy delivered by the PV array would be

$$\begin{aligned}\text{PV daily energy} &= 15\ \text{m}^2 \times 0.150\ \text{kW/m}^2 \times 0.75 \times 5\ \text{hr/day}\\ &= 8.44\ \text{kWh/day}\end{aligned}$$

At 0.25 kWh/mile, the vehicle could be driven

$$\text{Annual miles} = \frac{8.44\ \text{kWh/day}}{0.25\ \text{kWh/mile}} \times 365\ \text{day/yr} = 12{,}320\ \text{miles/yr}$$

9 Stationary Sources

Nontransportation fossil-fuel combustion is responsible for 90 percent of the SO_x and almost half of the NO_x and PM_{10} emitted in the United States. Most of that is released at electric power plants, and most of the power plant emissions result from the combustion of coal. Both the regulatory and technological approaches to reducing those emissions are very different from what we just described for motor vehicles. Although most stationary source emissions are caused by combustion of fossil fuels, other processes such as evaporation of volatile organic substances, grinding, and forest fires, can be important as well. Our focus, however, will be on combustion.

Since most air pollutants are produced during combustion, one of the most important, but most overlooked, approaches to reducing emissions is simply to reduce the consumption of fossil fuels. There are three broad approaches that can be taken to reduce fossil fuel consumption: (1) increase the conversion efficiency from fuel to energy, (2) increase the efficiency with which energy is used, and (3) substitute other, less polluting, energy sources for fossil fuels. Progress has been made on all three fronts. The conversion efficiency from fuel to electricity for power plants has improved eight-fold from the 5 percent efficiency of Edison's original plants a century ago to nearly 40 percent for a modern coal-fired power plant, and almost 60 percent for a new natural-gas-fired, combined-cycle (NGCC) power plant. Encouraging gains have been made in improving the energy efficiency of end uses, and some estimates suggest that improvements in electric motors and motor controls, better lighting systems, more efficient manufacturing processes, and so forth can cut the per capita electricity demands in half. The third approach is to reduce our reliance on fossil fuels by increasing power production from solar, wind, hydroelectric, geothermal, and nuclear sources.

To the extent that fossil fuels continue to be used, there are three general approaches that can be used to reduce emissions:

1. *Precombustion controls* reduce the emission potential of the fuel itself. Examples include switching to fuels with less sulfur or nitrogen content in power plants. In some cases, fossil-fuels can be physically or chemically treated to remove some of the sulfur or nitrogen before combustion.
2. *Combustion controls* reduce emissions by improving the combustion process itself. Examples include new burners in power plants that reduce NO_x emissions, and new fluidized bed boilers that reduce both NO_x and SO_x.
3. *Postcombustion controls* capture emissions after they have been formed but before they are released to the air. On power plants, these may be combinations of particulate collection devices and flue-gas desulfurization techniques, used after combustion but before the exhaust stack.

Coal-Fired Power Plants

Coal-fired power plants generate over half of U.S. electricity. They do so at the lowest cost of any power source (though wind is quickly closing that gap). Moreover, the United States has a rather abundant coal resource base to supply those power plants for decades into the future. Those benefits are somewhat offset by the environmental harm caused by coal. Not only are coal plants responsible for emitting great quantities of sulfur oxides, nitrogen oxides, and particulates, they also emit twice as much CO_2 per kWh generated than new natural-gas-fired plants. Most early emission controls for coal were designed to reduce particulates using postcombustion equipment such as baghouses and electrostatic precipitators. More recently, especially as a result of increased awareness of the seriousness of acid rain and the conversion of gaseous SO_x and NO_x into sulfate and nitrate particles, attention has shifted somewhat toward redesigned combustors for NO_x reductions and scrubbers for SO_x control.

Before going into any of the details on these and other emission control techniques, it is useful to introduce a typical conventional coal-fired power plant, as shown in Figure 30. In this plant, coal that has been crushed in a pulverizer is burned to make steam in a boiler for the turbine-generator system. The steam is then condensed, in this case using a cooling tower to dissipate waste heat to the atmosphere, and the condensate is then pumped back into the boiler. The flue gas from the boiler is sent to an electrostatic precipitator, which adds a charge to the particulates in the gas stream so that they can be attracted to electrodes that collect them. Next, a wet scrubber sprays a limestone slurry over the flue gas, precipitating the sulfur and removing it in a sludge of calcium sulfite or calcium sulfate, which then must be treated and disposed of.

All of those emission controls shown in Figure 30 are costly and now account for close to 40 percent of the construction cost of a new coal-fired power plant. Not only have emission controls been costly in monetary terms, they also use up about 5 percent of the power generated by the plant, reducing the plant's overall efficiency.

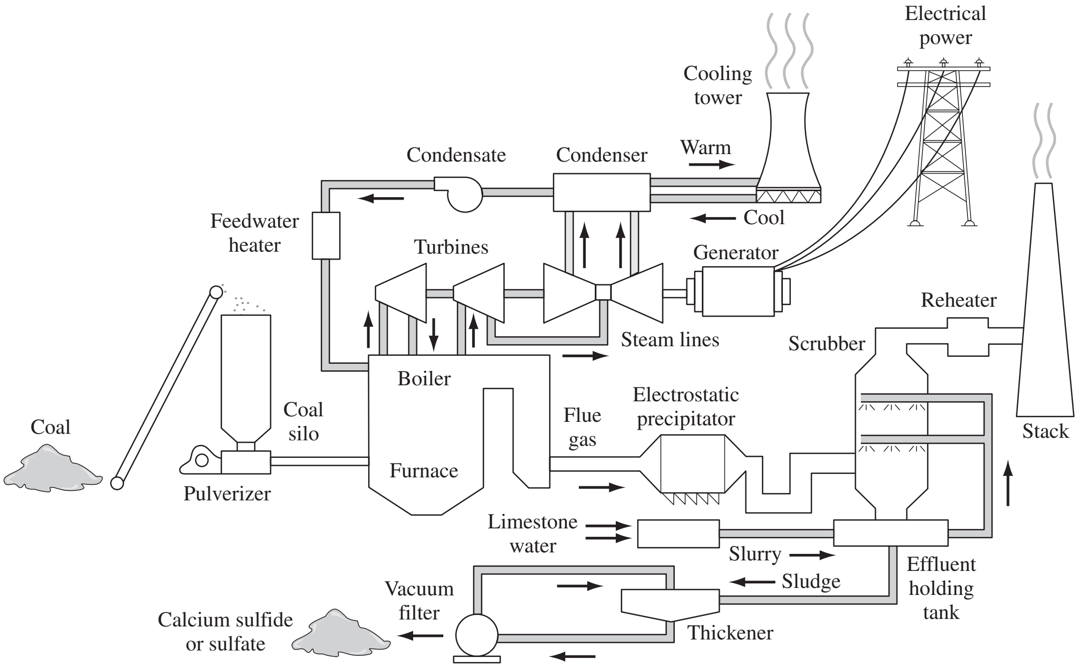


FIGURE 30 Typical modern coal-fired power plant using an electrostatic precipitator for particulate control and a limestone-based SO_2 scrubber. A cooling tower is shown for thermal pollution control.

Precombustion Controls

One of the two precombustion control techniques used on coal-fired power plants to control sulfur emissions is *fuel-switching*. As the name suggests, it involves either substituting low sulfur coal for fuel with higher sulfur content or, perhaps, blending the two. Fuel switching can reduce SO_2 emissions by anywhere from 30 to 90 percent, depending on the sulfur content of the fuel currently being burned. Often, fuel switching can be a temporary measure taken only during periods of time when atmospheric conditions are particularly adverse.

The sulfur content of coal is sometimes expressed as a percent, or, quite often, in terms of mass of SO_2 emissions per unit of heat energy delivered. Power plant coal typically has sulfur content between 0.2 to 5.5 percent (by weight), corresponding to an uncontrolled SO_2 emission rate of about 0.4 to 10 lb/MBtu (0.17 to 4.3 g/MJ). New Source Performance Standards (NSPS) for coal plants restrict emissions to no more than 1.2 lb/MBtu (0.5 g/MJ). Coal that can meet the NSPS without controls is sometimes referred to as "compliance" coal, or more loosely, as "low sulfur" coal.

In round numbers, just under half of U.S. coal reserves are in the eastern half of the United States (principally Illinois, West Virginia, Kentucky, Pennsylvania, and Ohio), and just over half are in the western United States (Montana, Wyoming, Colorado, and Texas). While that distribution is nearly even, the location of low sulfur coal reserves is very unevenly divided. Almost all of our low sulfur coal, about 85 percent, is located in the western states (the only notable exceptions are small reserves in West Virginia and Kentucky). But two-thirds of our coal consumption and three-fourths of coal production occurs east of the Mississippi. Increasing reliance on fuel switching, then, would greatly increase the cost to transport the coal from mine to power plant. It would also significantly affect the economies of a number of states if eastern mines decrease production while western ones increase theirs. In addition, western coal often has different ash, moisture, and energy content, which may require modifications to existing power plants to allow fuel switching. Such costs, however, are relatively minor compared to the increased cost of low sulfur fuel.

The other precombustion approach currently used to reduce sulfur emissions is coal cleaning. Sulfur in coal is either bound into organic coal molecules themselves, in which case precleaning would require chemical or biological treatment (at the research stage now), or it can be in the form of inorganic pyrite (FeS_2), which can readily be separated using physical treatment. Pyrite has a specific gravity that is 3.6 times greater than coal, and that difference allows various coal "washing" steps to separate the two. Such physical cleaning not only reduces the sulfur content of coal, but it also reduces the ash content, increases the energy per unit weight of fuel (which reduces coal transportation and pulverization costs), and creates more uniform coal characteristics that can increase boiler efficiency. These benefits can offset much of the cost of coal cleaning. It has been estimated that this fairly simple coal cleaning can reduce sulfur emissions by about 10 percent.

Fluidized-Bed Combustion (FBC)

Fluidized-bed combustion (FBC) is one of the most promising clean coal technologies. In an FBC boiler, crushed coal mixed with limestone is held in suspension (fluidized) by fast-rising air injected from the bottom of the bed. Sulfur oxides formed during combustion react with the limestone ($CaCO_3$) to form solid calcium sulfate ($CaSO_4$), which falls to the bottom of the furnace and is removed. Sulfur removal rates can be higher than 90 percent.

In an FBC boiler, the hot, fluidized particles are in direct contact with the boiler tubes. This enables much of the heat to be transferred to the boiler tubes by conduction, which is much more efficient than convection and radiation heat transfer occurring in conventional boilers. The increase in heat transfer efficiency enables the boilers to operate at around 800°C, which is about half the temperature of conventional boilers and is well below the 1,400°C threshold at which thermal NO_x is formed. In addition to efficient SO_x and NO_x control, FBC boilers are less sensitive to variations in coal quality. Coal with higher ash content can be burned without fouling heat exchange surfaces since the lower combustion temperature is below the melting point of ash. In fact, fluidized bed combustors can efficiently burn almost any solid fuel, including municipal waste, cow manure, and agricultural wastes such as tree bark.

Utility-scale FBC applications are just beginning, the first unit having gone into operation in 1986. Even though the technology is still being modified and improved, the fuel flexibility advantages and the inherent control of emissions have enabled FBC to cross the commercial threshold, and in 2005, most boiler manufacturers offered them as a standard package.

Integrated Gasification Combined Cycle

Another promising technology, called integrated gasification combined cycle (IGCC), offers the combination of increased combustion efficiency with reduced

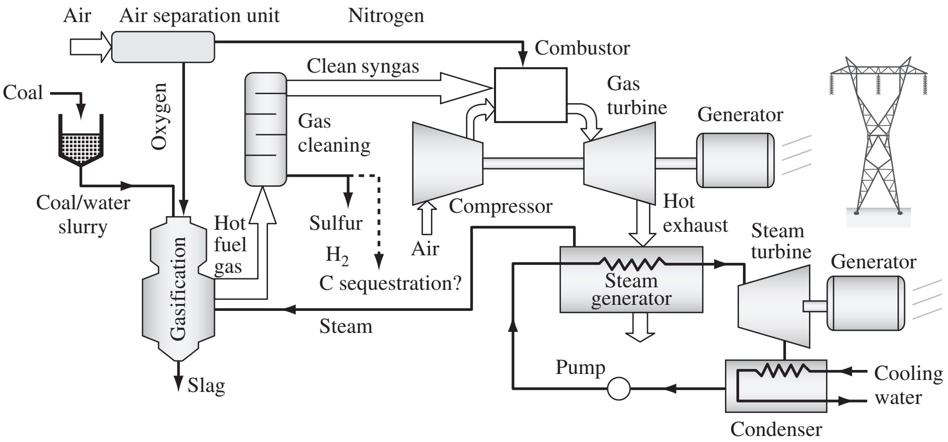


FIGURE 31 Integrated gasification combined cycle. Gasified coal fuels an efficient gas turbine; waste heat from the gas turbine powers a steam cycle. CO_2 is removed during gas cleaning, which offers the opportunity to eventually sequester carbon to help reduce global warming impacts.

emissions (Figure 31). In the first stage of an IGCC, a coal-water slurry is brought into contact with steam to form a fuel gas, or *syngas,* consisting mostly of carbon monoxide (CO) and hydrogen (H_2):

$$C + H_2O \rightarrow CO + H_2 \tag{31}$$

The syngas is cleaned up, removing most of the particulates, mercury, and sulfur, and then burned in a high-efficiency gas turbine that spins a generator. The air separation unit provides N_2 to cool the gas turbine and O_2 to increase syngas combustion efficiency. Waste heat from the gas turbine runs a conventional steam cycle, generating more electricity. Despite energy losses in the gasification process, the combination of gas turbine and steam turbine offsets those losses, and the overall efficiency of fuel to electricity can approach 45 percent. This is considerably higher than the best pulverized coal plants, which are usually less than 40 percent.

IGCC plants are more expensive than pulverized coal, and they have trouble competing economically with natural-gas-fired combined cycle plants. As of 2006, there were only four such plants in the world.

IGCC with Carbon Sequestration

Most of the interest in IGCC is based on the potential to control carbon emissions from coal-fired power plants. By first converting coal to a syngas, it is possible to design the gas cleaning stage in such a way that the carbon can be extracted before combustion. By doing so, it may be possible to build “clean coal” power plants that would be able to take advantage of the relative abundance of coal without contributing so much to global warming.

The key, of course, is finding a way to store all of that carbon, essentially forever. One of the more promising approaches is permanent CO_2 storage in geologic formations such as deep brine aquifers. Such formations consist of highly porous rock, similar to those containing oil and gas, but without the hydrocarbons that produced our fossil fuels. Instead, they are filled with water containing high concentrations of salts dissolved out of the surrounding rocks. When these formations are capped with impermeable rock, they might be viable for CO_2 sequestration for the indefinite future.

Controlling NO_x Emissions

Recall that nitrogen oxides are formed partly by the oxidation of nitrogen in the fuel itself (fuel NO_x) and partly by the oxidation of nitrogen in the combustion air (thermal NO_x). Coal-fired power plants, which are responsible for about one-fourth of total NO_x emissions in the United States, emit roughly twice as much fuel NO_x as thermal NO_x. Modifications to the combustion processes described next are designed to reduce both sources of NO_x. In one technique called *low excess air,* the amount of air made available for combustion is carefully controlled at the minimum amount required for complete combustion. Low excess air technology can be retrofitted onto some boilers at a modest cost, yielding from 15 to 50 percent lower NO_x emissions.

There is a new, second-generation *low* NO_x *burner* technology that promises even greater NO_x removal efficiencies, and it can be retrofitted onto more existing furnaces. Low NO_x burners employ a staged combustion process that delays mixing the fuel and air in the boiler. In the first stage of combustion, the fuel starts burning in an air-starved environment, causing the fuel-bound nitrogen to be released as nitrogen gas, N_2, rather than NO_x. The following stage introduces more air to allow complete combustion of the fuel to take place. Potential NO_x reductions of 45 to 60 percent seem likely. Another combustion modification incorporates a staged burner for NO_x control combined with limestone injection for SO_2 control. This *limestone injection multistage burner* (LIMB) technology is still under development but looks promising.

While improved fuel-combustion technologies are now the primary method of controlling emissions, new postcombustion controls may augment or replace low-NO_x burners. One such technology, *Selective Catalytic Reduction* (SCR), is already widely used in Europe and Japan but has only recently been applied to high-sulfur-content U.S. coal. With the SCR process, ammonia (NH_3) is injected into boiler flue gas, and the mix is passed through a catalyst bed in which the NO_x and NH_3 react to form nitrogen and water vapor. The primary reactions are

$$4\,NO + 4\,NH_3 + O_2 \rightarrow 4\,N_2 + 6\,H_2O \tag{32}$$

$$2\,NO_2 + 4\,NH_3 + O_2 \rightarrow 3\,N_2 + 6\,H_2O \tag{33}$$

It has been estimated that SCR for 90 percent NO_x removal could add \$66/kW (1996 dollars) to the capital cost of a 250-MW coal-fired power plant, which translates to a 30-year, levelized extra cost of about 0.3¢/kWh (U.S. DOE and Southern Company, 1997).

Flue Gas Desulfurization (Scrubbers)

Flue-gas desulfurization (FSD) technologies can be categorized as being either *wet* or *dry* depending on the phase in which the main reactions occur, and either *throw-away* or *regenerative,* depending on whether or not the sulfur from the flue gas is discarded or recovered in a usable form. Most scrubbers operating in the United States use wet, throwaway processes.

In most wet scrubbers, finely pulverized limestone ($CaCO_3$) is mixed with water to create a slurry that is sprayed into the flue gases. The flue gas SO_2 is absorbed by the slurry, producing a calcium sulfite ($CaSO_3$) or a calcium sulfate ($CaSO_4$) precipitate. The precipitate is removed from the scrubber as a sludge. Though the chemical reactions between SO_2 and limestone involve a number of steps, an overall relationship resulting in the production of inert calcium sulfite dihydrate is

$$CaCO_3 + SO_2 + 2\,H_2O \rightarrow CaSO_3 \cdot 2\,H_2O + CO_2 \tag{34}$$

About 90 percent of the SO_2 can be captured from the flue gas using limestone in wet scrubbers.

Wet scrubbers sometimes use lime (CaO) instead of limestone in the slurry, and the following overall reaction applies:

$$CaO + SO_2 + 2\,H_2O \rightarrow CaSO_3 \cdot 2\,H_2O \tag{35}$$

Lime slurries can achieve greater SO_2 removal efficiencies, up to 95 percent. However, lime is more expensive than limestone, so it is not widely used. Dry scrubbers must use lime, and that increased cost is one reason for their relative lack of use.

Although wet scrubbers can capture very high fractions of flue gas SO_2, they have been accepted by the utilities with some reluctance. They are expensive, costing on the order of $200 million for a large power plant. If they are installed on older plants, with less remaining lifetime, their capital costs must be amortized over a shorter period of time, and annual revenue requirements increase accordingly. Scrubbers also reduce the net energy delivered to the transmission lines. The energy to run scrubber pumps, fans, and flue gas reheat systems requires close to 5 percent of the total power produced by the plant. Scrubbers are also subject to corrosion, scaling, and plugging problems, which may reduce overall power plant reliability.

Scrubbers also use large amounts of water and create similarly large volumes of sludge that has the consistency of toothpaste. A large, 1,000-MW plant burning 3-percent sulfur coal can produces enough sludge each year to cover a square mile of land to a depth of over one foot (Shepard, 1988). Sludge treatment often involves oxidation of calcium sulfite to calcium sulfate, which precipitates easier thickening and vacuum filtration. Calcium sulfate (gypsum) can be reused in the construction industry.

Particulate Control

A number of gas-cleaning devices can be used to remove particulates. The most appropriate device for a given source will depend on such factors as particle size, concentration, corrosivity, toxicity, volumetric flow rate, required collection efficiency, allowable pressure drops, and costs.

Cyclone Collectors. For relatively large particles, the most commonly used control device is the centrifugal, or *cyclone,* collector. As shown in Figure 32,

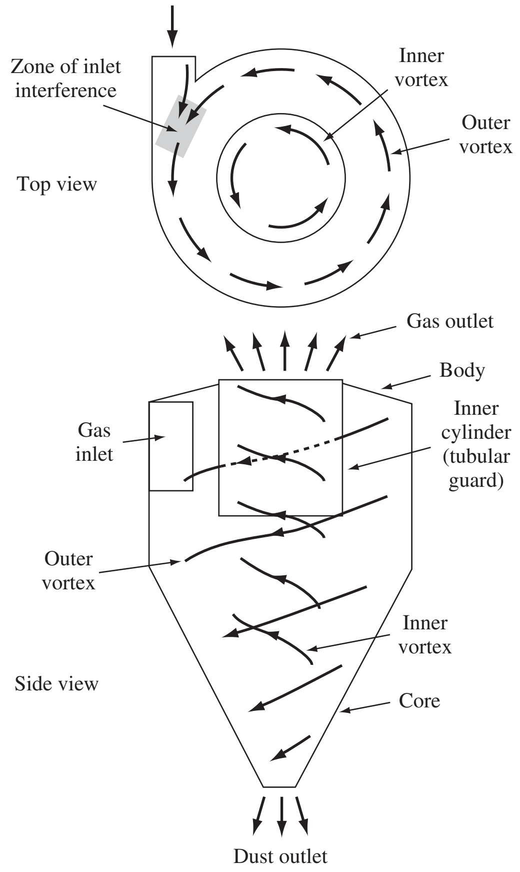


FIGURE 32 Conventional reverse-flow cyclone.
(*Source:* U.S. HEW, 1969.)

particle-laden gas enters tangentially near the top of the cyclone. As the gas spins in the cylindrical shell, centrifugal forces cause the particles to collide with the outer walls, and then gravity causes them to slide down into a hopper at the bottom. The spiraling gases then exit the collector from the top. Efficiencies of cyclones can be above 90 percent for particles larger than 5 μm, but that efficiency drops off rapidly for the small particle sizes that are of greatest concern for human health. While they are not efficient enough to meet emission standards, they are relatively inexpensive and maintenance free, which makes them ideal as precleaners for more expensive and critical final control devices, such as baghouses and electrostatic precipitators.

Electrostatic Precipitators. To collect really small particles, most utility power plants use electrostatic precipitators. Figure 33 shows one configuration for an electrostatic precipitator consisting of vertical wires placed between parallel collector plates. The plates are grounded, and the wires are charged up to a very high (negative) voltage of perhaps 100,000 V. The intense electric field created near the wires causes a corona discharge, ionizing gas molecules in the air stream. The negative ions and free electrons thus created move toward the grounded plates, and

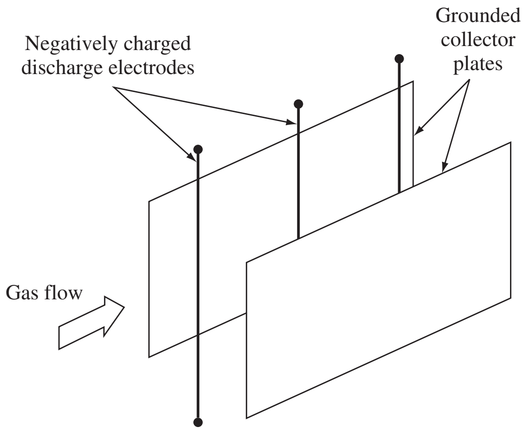


FIGURE 33 Schematic representation of a flat-surface type electrostatic precipitator. Particles in the gas stream acquire negative charge as they pass through the corona and are then attracted to the grounded collecting plates.

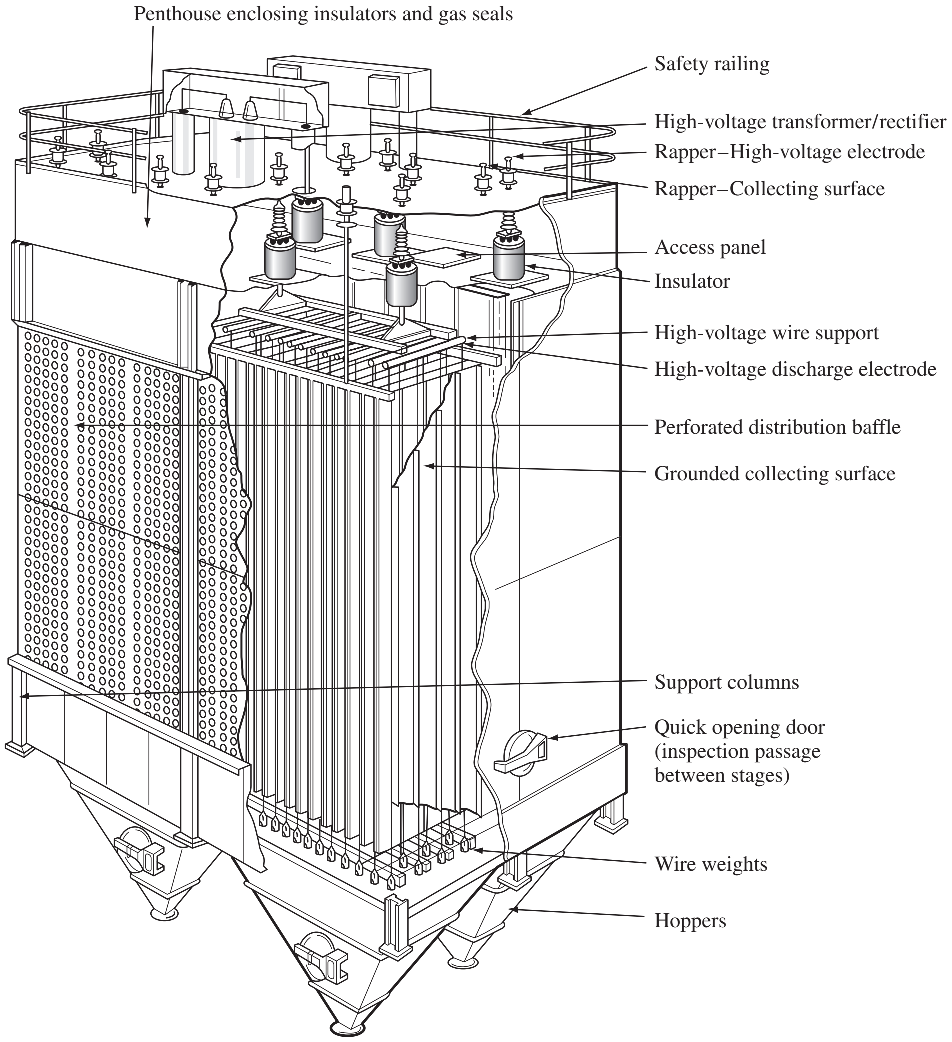


FIGURE 34 Cutaway view of a flat surface-type electrostatic precipitator. (*Source:* U.S. HEW, 1969.)

along the way, some attach themselves to passing particulate matter. The particles now carry a charge, which causes them to move under the influence of the electric field to a grounded collecting surface. They are removed from the collection electrode either by gravitational forces, by rapping, or by flushing the collecting plate with liquids. Actual electrostatic precipitators, such as the one shown in Figure 34, may have hundreds of parallel plates, with total collection areas measured in the tens of thousands of square meters.

Electrostatic precipitators can easily remove more than 98 percent of the particles passing through them, including particles of submicron size. Some have efficiencies even greater than 99.9 percent. They can handle large flue-gas flow rates with little pressure drop, and they have relatively low operation and maintenance costs. They are quite versatile, operating on both solid and liquid particles. Precipitators are very efficient, but they are expensive and take a lot of space. Area requirements increases nonlinearly with collection efficiency, in accordance with the following relationship, known as the Deutsch-Anderson equation:

$$\eta = 1 - e^{-wA/Q} \tag{36}$$

where η is the fractional collection efficiency, A is the total area of collection plates, Q is the volumetric flow rate of gas through the precipitator, and w is a parameter known as the effective drift velocity. The *effective drift velocity* is the terminal speed at which particles approach the collection plate under the influence of the electric field. It is determined from either pilot studies or previous experience with similar units. Any consistent set of units can be used for these quantities.

EXAMPLE 9 Electrostatic Precipitator Area

An electrostatic precipitator with 6,000 m^2 of collector plate area is 97 percent efficient in treating 200 m^3/s of flue gas from a 200-MW power plant. How large would the plate area have to be to increase the efficiency to 98 percent and to 99 percent?

Solution Rearranging (36) to solve for the drift velocity, w, gives

$$w = -\frac{Q}{A}\ln(1 - \eta) = -\frac{200\ \text{m}^3/\text{s}}{6{,}000\ \text{m}^2}\ln(1 - 0.97) = 0.117\ \text{m/s}$$

To achieve 98 percent efficiency, the area required would be

$$A_{98} = -\frac{Q}{w}\ln(1 - \eta) = -\frac{200\ \text{m}^3/\text{s}}{0.117\ \text{m/s}}\ln(1 - 0.98) = 6{,}690\ \text{m}^2$$

To achieve 99 percent, the area required would be

$$A_{99} = -\frac{200\ \text{m}^3/\text{s}}{0.117\ \text{m/s}}\ln(1 - 0.99) = 7{,}880\ \text{m}^2$$

That is about the area of two football fields! As these calculations suggest, the additional collector area required to achieve incremental improvements in collection efficiency goes up rapidly. To increase from 97 to 98 percent required 690 m^2 of added area, whereas the next 1 percent increment requires 1,190 m^2.

Baghouses. The major competition that electrostatic precipitators have for efficient collection of small particles is *fabric filtration*. Dust-bearing gases are passed through fabric filter bags, which are suspended upside-down in a large chamber, called a baghouse, as shown in Figure 35. A baghouse may contain thousands of bags, which are often distributed among several compartments. This allows individual compartments to be cleaned while others remain in operation.

Part of the filtration is accomplished by the fabric itself, but a more significant part of the filtration is caused by the dust that accumulates on the inside of the bags. Efficiencies approach 100 percent removal of particles as small as 1 μm; substantial quantities of particles as small as 0.01 μm are also removed. They have certain disadvantages, however. As is the case for precipitators, baghouses are large and expensive. They can be harmed by corrosive chemicals in the flue gases, and they cannot operate in moist environments. There is also some potential for fires or explosions if the dust is combustible. The popularity of baghouse filters is rising, however, and they now rival precipitators in total industrial sales.

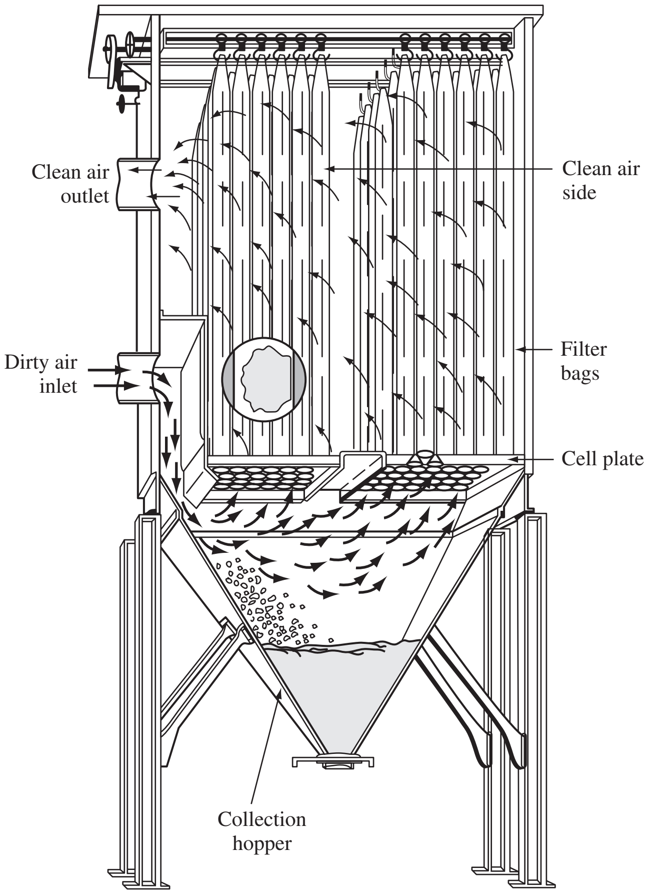


FIGURE 35 Typical simple fabric filter baghouse.
(*Source:* Courtesy Wheelabrator Air Pollution Control.)

Combined Heat and Power Systems

The traditional approach to controlling emissions from stationary sources has been focused on technologies applied to conventional, large, centralized power stations. For these plants, only about one-third of the input energy ends up as electricity, which is then shipped to customers who may be located hundreds of miles away. The other two-thirds of the input energy ends up as low temperature waste heat that can't very easily be delivered to customers who, not surprisingly, don't want to live very close to one of these giant power plants. If, instead, power plants were smaller and cleaner, customers could imagine having them located on their side of the electric meter, in which case the waste heat might be captured and used for space heating, space cooling, and/or process heat for industrial processes. Less fuel would need to be burned, which could translate into lower emissions.

A number of technologies can be used for these *combined heat-and-power* (CHP) systems, including fuel cells, microturbines, Stirling engines, and new, high-efficiency reciprocating engines. For a number of reasons, fuel cells are the most interesting. They are quiet, emit almost no pollution, are modular so they can be small enough to power a laptop computer or large enough to provide utility-scale power, and have the potential to be more efficient than electricity generated by burning fuels. Given these benign characteristics, fuel cells can be sited in locations where no conventional power plant would be allowed, in which case the waste heat from the fuel cell can often be used for on-site needs. By doing on-site cogeneration of heat and electricity, the overall efficiency of fuel cells can climb toward 80 or 85 percent.

Figure 36 shows the three key processes in a complete fuel-cell system, including a reformer that strips hydrogen from natural gas, the fuel-cell stack itself, and a power conditioner to convert the fuel cell's DC power into AC. For high-temperature fuel cells, such as solid-oxide (SOFC) and perhaps molten-carbonate cells (MCFC), the reforming can be done internally without a separate fuel processor. Solid polymer PEM cells have electrical efficiencies around 35 to 40 percent, whereas SOFC and MCFC systems can exceed 50 percent electrical efficiency.

To get a sense of the efficiency improvements that CHP systems offer, Figure 37 shows an example in which a unit with 35 percent electrical efficiency and

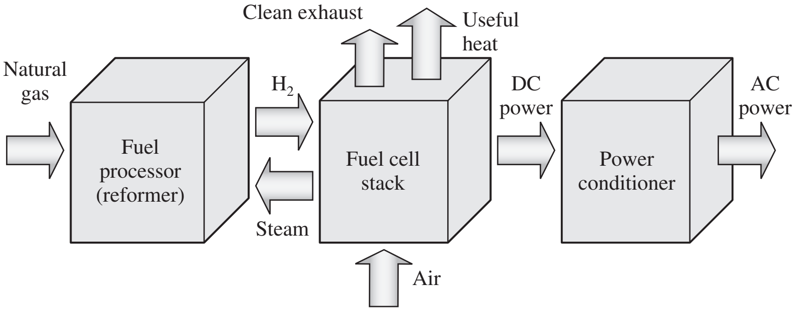


FIGURE 36 A fuel-cell system includes a reformer to liberate H_2 from natural gas, the fuel-cell stack, and a DC-to-AC power conditioner.

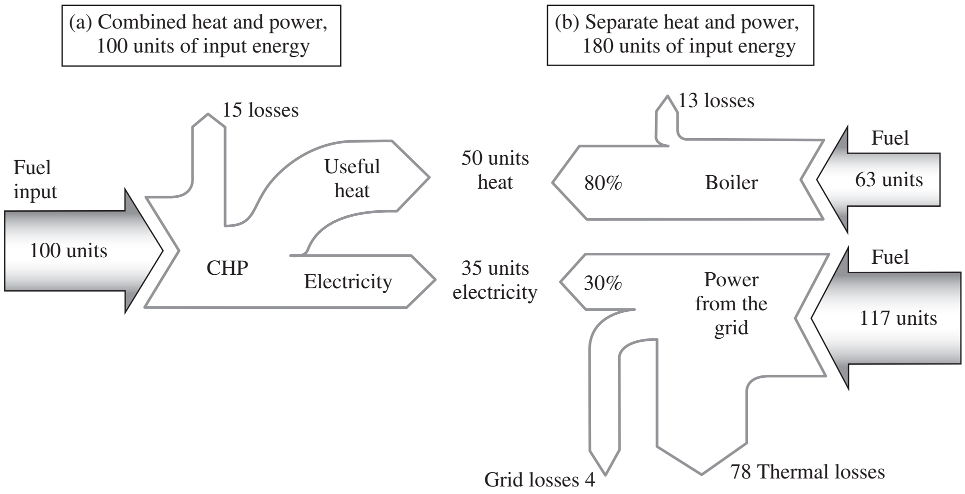


FIGURE 37 In this example, the CHP system provides the same heat and electricity with 100 units of fuel instead of the 180 units that would be needed in a separated system.

50 percent thermal efficiency is compared with a separate boiler for heat plus the utility grid for electricity. The separated system would require 180 units of fuel energy to deliver the same end-use heat and electricity as would be provided by just 100 units into the CHP system. That is a fuel savings of 44 percent.

10 Air Pollution and Meteorology

Obviously, air quality at a given site can vary tremendously from day to day even though emissions may remain relatively constant. The determining factors have to do with the weather: how strong are the winds, in what direction are they blowing, what is the temperature profile, how much sunlight is available, and how long has it been since the last strong winds or precipitation were able to clear the air? Air quality is dependent on the dynamics of the atmosphere, the study of which is called *meteorology.*

Adiabatic Lapse Rate

The ease with which pollutants can disperse vertically into the atmosphere is largely determined by the rate of change of air temperature with altitude. For some temperature profiles, the air is *stable,* that is, air at a given altitude has physical forces acting on it that make it want to remain at that elevation. Stable air discourages the dispersion and dilution of pollutants. For other temperature profiles, the air is *unstable*. In this case, rapid vertical mixing takes place that encourages pollutant dispersal and increases air quality. Obviously, vertical stability of the atmosphere is an important factor that helps determine the ability of the atmosphere to dilute emissions; hence, it is crucial to air quality.

Let us investigate the relationship between atmospheric stability and temperature. It is useful to imagine a "parcel" of air being made up of a number of air molecules with an imaginary boundary around them. If this parcel of air moves upward in the atmosphere, it will experience less pressure, causing it to expand and cool. On the other hand, if it moves downward, more pressure will compress the air, and its temperature will increase. This heating or cooling of a gas as it is compressed or expanded should be a familiar concept. Pumping up a bicycle tire, for example, warms the valve on the tire; releasing the contents of a pressurized spray can allows the contents to expand and cool, cooling your finger on the button as well.

As a starting point, we need a relationship for the rate of change of temperature of a parcel of air as it moves up or down in the atmosphere. As it moves, we can imagine its temperature, pressure, and volume changing, and we might imagine its surroundings adding or subtracting energy from the parcel. If we make small changes in these quantities, and apply both the ideal gas law and the first law of thermodynamics, it is relatively straightforward to derive the following expression (see Problem 28):

$$dQ = C_p\,dT - V\,dP \tag{37}$$

where

dQ = heat added to the parcel per unit mass (J/kg)
C_p = specific heat of air at constant pressure; that is, the amount of heat required to raise the temperature of 1 kg of air by 1°C while holding its pressure constant (= 1,005 J/kg-K)
dT = incremental temperature change (K)
V = volume per unit mass (m^3/kg)
dP = incremental pressure change in the parcel (Pa)

Let us make the quite accurate assumption that as the parcel moves, there is no heat transferred across its boundary; that is, this process is *adiabatic*. This means that $dQ = 0$, so we can rearrange (37) as

$$\frac{dT}{dP} = \frac{V}{C_p} \tag{38}$$

Equation (38) gives us an indication of how atmospheric temperature would change with air pressure, but what we are really interested in is how it changes with altitude. To do that, we need to know how pressure and altitude are related.

Consider a static column of air with cross-section A, as shown in Figure 38. A horizontal slice of air in that column of thickness dz and density ρ will have mass

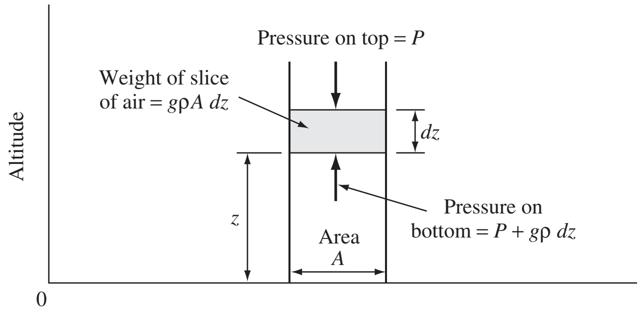


FIGURE 38 A column of air in static equilibrium used to determine the relationship between air pressure and altitude.

$\rho A\, dz$. If the pressure at the top of the slice due to the weight of air above it is $P(z + dz)$, then the pressure at the bottom of the slice, $P(z)$, will be $P(z + dz)$ plus the added weight per unit area of the slice itself:

$$P(z) = P(z + dz) + \frac{g\rho A\, dz}{A} \tag{39}$$

where g is the gravitational constant. We can write the incremental pressure, dP, for an incremental change in elevation, dz, as:

$$dP = P(z + dz) - P(z) = -g\rho\, dz \tag{40}$$

Expressing the rate of change in temperature with altitude as a product and substituting in (38) and (40) gives

$$\frac{dT}{dz} = \frac{dT}{dP} \cdot \frac{dP}{dz} = \left(\frac{V}{C_p}\right)(-g\rho) \tag{41}$$

However, since V is volume per unit mass, and ρ is mass per unit volume, the product $V\rho = 1$, and (41) simplifies to

$$\frac{dT}{dz} = \frac{-g}{C_p} \tag{42}$$

The negative sign indicates that temperature decreases with increasing altitude. Substituting the constant $g = 9.806\ \text{m/s}^2$, and the constant-volume specific heat of dry air at room temperature, $C_p = 1{,}005$ J/kg-K, into (42) yields

$$\frac{dT}{dz} = \frac{-9.806\ \text{m/s}^2}{1{,}005\ \text{J/kg-K}} \times \frac{1\ \text{J}}{\text{kg-m}^2/\text{s}^2} = -0.00976\ \text{K/m} \tag{43}$$

This is a very important result. When its sign is changed to keep things simple, $-dT/dz$ is given a special name, the *dry adiabatic lapse rate,* Γ_d.

$$\Gamma_d = -\frac{dT}{dz} = 9.76°\text{C/km} = 5.4°\text{F}/1{,}000\ \text{ft} \tag{44}$$

Equation 44 tells us that moving a parcel of dry air up or down will cause its temperature to change by 9.76°C/km, or roughly 10°C/km. This temperature profile will be used as a reference against which actual ambient air temperature gradients will be compared. As we shall see, if the actual air temperature decreases faster with increasing elevation than the adiabatic lapse rate, the air will be unstable and rapid mixing and dilution of pollutants will occur. Conversely, if the actual air temperature drops more slowly than the adiabatic lapse rate, the air will be stable and air pollutants will concentrate.

Equation (44) was derived assuming our parcel of air could be treated as an ideal gas that could be moved about in the atmosphere without any heat transfer between it and its surroundings. Both assumptions are very good. It was also assumed that the air was dry, but this may or may not be such a good assumption. If air has some water vapor in it, C_p changes slightly from the value assumed but not enough to warrant a correction. On the other hand, if enough water vapor is present that condensation occurs when the parcel is raised and cooled, latent heat will be released. The added heat means a saturated air parcel will not cool as rapidly as a dry one. Unlike the dry adiabatic rate, the *saturated adiabatic lapse rate* (Γ_s) is not

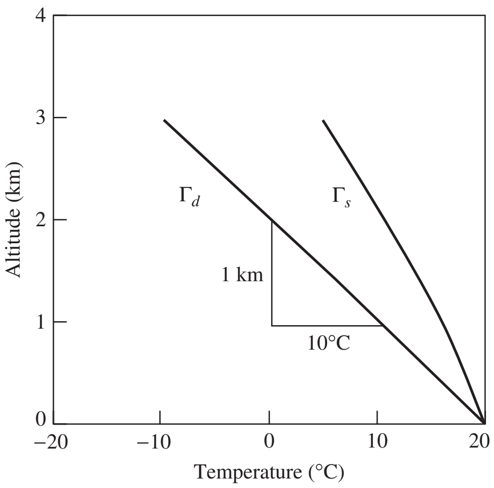


FIGURE 39 The dry adiabatic lapse rate Γ_d is a constant 10°C/km, but the saturated adiabatic lapse rate Γ_s varies with temperature. In the troposphere, Γ_s is approximately 6°C/km.

a constant since the amount of moisture that air can hold before condensation begins is a function of temperature. A reasonable average value of the moist adiabatic lapse rate in the troposphere is about 6°C/km. Figure 39 shows the dry and saturated adiabatic rates.

EXAMPLE 10 Air Conditioning for a High-Altitude Aircraft?

An aircraft flying at an altitude of 9 km (30,000 ft) draws in fresh air at −40°C for cabin ventilation. If that fresh air is compressed to the pressure at sea level, would the air need to be heated or cooled to be delivered to the cabin at 20°C?

Solution As the air is compressed, it warms up. As it warms up, it is even easier for the air to hold whatever moisture it may have had, so there is no condensation to worry about, and the dry adiabatic lapse rate can be used. At 10°C per km, compression will raise the air temperature by $10 \times 9 = 90$°C, making it $-40 + 90$°C $= 50$°C (122°F). It needs to be air conditioned!

Atmospheric Stability

Our interest in lapse rates is based on the need to understand atmospheric stability since it is such a crucial factor in the atmosphere's ability to dilute pollution. There are a number of factors, such as windspeeds, sunlight, and geographical features, that cause the actual *ambient lapse rate* (Γ) in the real atmosphere to differ from the 1°C/100 m dry adiabatic lapse rate just calculated. Differences between the ambient lapse rate and the adiabatic lapse rate determine the stability of the atmosphere.

Consider a parcel of air at some given altitude. It has the same temperature and pressure as the air surrounding it. Our test for atmospheric stability will be based on the following mental experiment. If we imagine raising the parcel of air slightly, it will experience less atmospheric pressure, so it will expand. Because it will have done work on its environment (by expanding), the internal energy in the parcel will be reduced, so its temperature will drop. Assuming the parcel is raised fast enough to be able to ignore any heat transfer between the surrounding air and the parcel, the cooling will follow the adiabatic lapse rate. After raising the parcel, note its temperature, and compare it with the temperature of the surrounding air. If the parcel, at this higher elevation, is now colder than its surroundings, it will be denser than the surrounding air and will want to sink back down again. That is, whatever it was that caused the parcel to start to move upward will immediately be opposed by conditions that make the parcel want to go back down again. The atmosphere is said to be *stable*.

If, however, raising the parcel causes its temperature to be greater than the surrounding air, it will be less dense than the surrounding air and will experience buoyancy forces that will encourage it to keep moving upward. The original motion upward will be reinforced, and the parcel will continue to climb. This is an *unstable* atmosphere.

Consider Figure 40*a*, which shows an ambient temperature profile for air that cools more rapidly with altitude than the dry adiabatic lapse rate. In this case, the ambient air temperature is said to be *superadiabatic*. Imagine a 20°C parcel of air at 1 km to be just like the air surrounding it. If that parcel is raised to 2 km, it will cool adiabatically to 10°C. The 10°C parcel of air at 2 km is now warmer than the surrounding air (0°C in the figure), so it is less dense, more buoyant, and wants to keep rising ("warm air rises"). In other words, shoving the parcel of air up some in a superadiabatic atmosphere creates forces that want to keep it moving upward. Conversely, a parcel of air at 1 km that is nudged downward becomes colder than the surrounding air, and so it keeps sinking. It doesn't matter whether the parcel of air is nudged up or down; in either case, it keeps on going, and the atmosphere is said to be unstable.

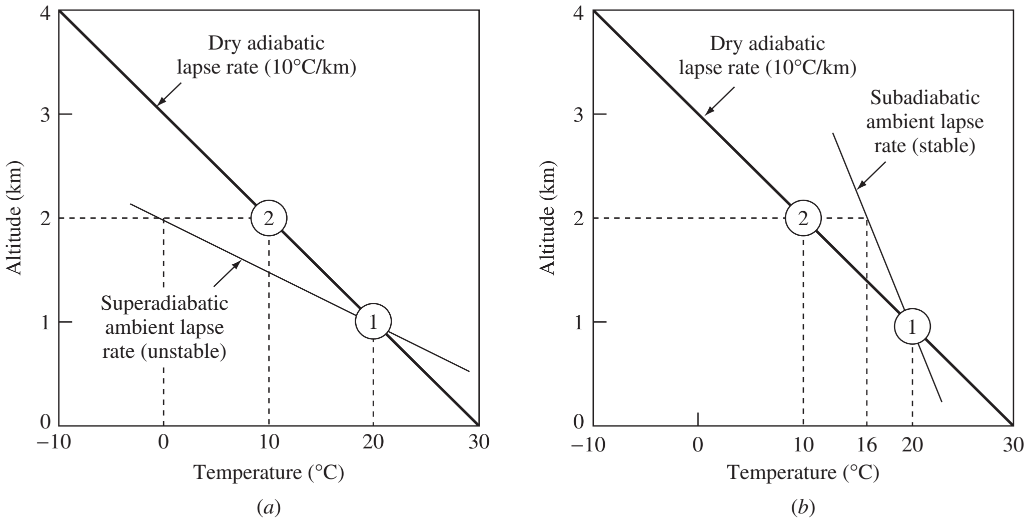


FIGURE 40 Demonstrating atmospheric stability in a dry atmosphere. When a 20°C parcel of air at 1 km (position 1) moves up to 2 km (position 2), its temperature drops to 10°C (following the dry adiabatic lapse rate). In (*a*) the parcel of air raised to 2 km is warmer than the surrounding ambient air so the parcel keeps rising (unstable). In (*b*) the parcel at 2 km is colder than ambient, so it sinks back down (stable).

In Figure 40*b*, the ambient temperature profile is drawn for the subadiabatic case in which the ambient temperature cools less rapidly than the adiabatic lapse rate. If we once again imagine a parcel of air at 1 km and 20°C that is, for one reason or another, nudged upward, it will find itself colder than the surrounding air. At its new elevation, it experiences the same pressure as the air around it, but it is colder, so it will be denser and will sink back down. Conversely, a parcel starting at 1 km and 20°C that starts moving downward will get warmer, and less dense than its surroundings, so buoyancy forces will push it back up again. In other words, anything that starts a parcel of air moving up or down will cause the parcel to experience forces that will want it to return to its original altitude. The subadiabatic atmospheric profile is therefore stable.

If the ambient lapse rate is equal to the adiabatic lapse rate, moving the parcel upward or downward results in its temperature changing by the same amount as its surroundings. In any new position, it experiences no forces that either make it continue its motion or make it want to return to its original elevation. The parcel likes where it was, and it likes its new position too. Such an atmospheric is said to be *neutrally stable*.

Figure 40 was drawn for a dry atmosphere; that is, as a parcel moves upward, the air remains unsaturated, there is no condensation of moisture, and it cools at 10°C/km. If condensation does occur, the adiabatic lapse rate drops to around 6°C/km and that slope determines the condition of neutral stability. Figure 41 shows a summary of how lapse rates determine atmospheric stability. If the ambient lapse rate shows cooling at a faster rate than the dry adiabatic lapse rate, the atmosphere is *absolutely unstable*. The air will always want to move to some new altitude so vertical dispersion of pollution is enhanced. For ambient temperatures that cool less rapidly than the saturated adiabatic lapse rate, the atmosphere is *absolutely stable*. That air wants to stay where it is. For ambient lapse rates between the dry and wet adiabatic lapse rates, the atmosphere may be stable or it may be unstable;

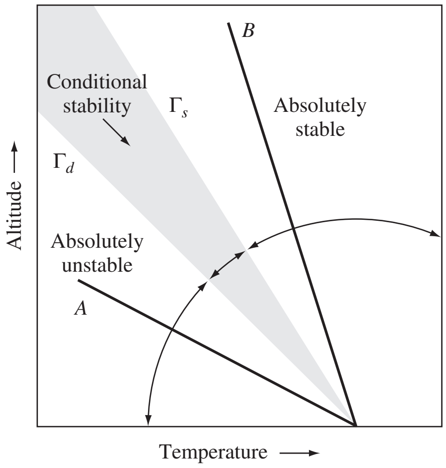


FIGURE 41 Atmospheres in which the ambient temperature profile cools more rapidly than the dry adiabatic lapse rate (Γ_d), such as line *A*, are absolutely unstable. If ambient temperatures drop more slowly than the saturated adiabatic lapse rate (Γ_s), such as *B*, the atmosphere is absolutely stable. Between the two lies a region of conditional stability.

we can't tell without knowing the actual ambient temperature profile and the actual adiabatic profile. That region is labeled as having *conditional stability*.

An extreme case of a subadiabatic lapse rate is one in which ambient temperatures increase with altitude. Such *temperature inversions* yield a very stable air mass and pollution tends to stay trapped there. The heating caused by absorption of incoming solar energy by oxygen and ozone in the upper atmosphere creates a stratospheric temperature inversion. That inversion causes the stratosphere to be extremely stable, often trapping pollutants for many years.

Temperature Inversions

Temperature inversions represent the extreme case of atmospheric stability, creating a virtual lid on the upward movement of pollution. There are several causes of inversions, but the two that are the most important from an air quality standpoint are *radiation inversions* and *subsidence inversions*. Radiation inversions are caused by nocturnal cooling of the Earth's surface, especially on clear winter nights. The second, subsidence inversions, are the result of the compressive heating of descending air masses in high pressure zones. There are other, less important, causes of inversions such as *frontal inversions*. A frontal inversion is created when a cold air mass passes under a warm air mass, but these are short lived and tend to be accompanied by precipitation that cleanses the air. There are also inversions associated with geographical features of the landscape. Warm air passing over a cold body of water, for example, creates an inversion. There are also inversions in valleys when cold air rolls down the canyons at night under warmer air that might exist aloft.

Radiation Inversions. The surface of the Earth cools down at night by radiating energy toward space. On a cloudy night, the Earth's radiation tends to be absorbed by water vapor, which in turn reradiates some of that energy back to the ground. On a clear night, however, the surface more readily radiates energy to space, and thus ground cooling occurs much more rapidly. As the ground cools, the temperature of the air in contact with the ground also drops. As is often the case on clear winter nights, the temperature of this air just above the ground becomes colder than the air above it, creating an inversion.

Radiation inversions begin to form at about dusk. As the evening progresses, the inversion extends to a higher and higher elevation, reaching perhaps a few hundred meters before the morning sun warms the ground again, breaking up the inversion. Figure 42 shows the development of a radiation inversion through the night, followed by the erosion of the inversion that takes place the next day.

Radiation inversions occur close to the ground, mostly during the winter, and lasts for only a matter of hours. They often begin at about the time traffic builds up in the early evening, which traps auto exhaust at ground level and causes elevated concentrations of pollution for commuters. Without sunlight, photochemical reactions can't take place, so the biggest problem is usually the accumulation of carbon monoxide (CO). In the morning, as the sun warms the ground and the inversion begins to break up, pollutants that have been trapped in the stable air mass are suddenly brought back to Earth in a process known as *fumigation*. Fumigation can cause short-lived, high concentrations of pollution at ground level.

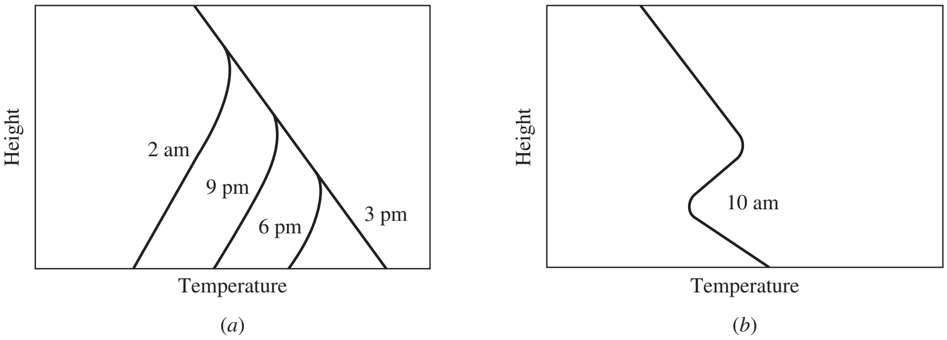


FIGURE 42 Development of a radiation inversion (*a*), and the subsequent erosion of the inversion (*b*). The times are representative only. The breakup of the inversion in the morning leads to a process called fumigation during which pollution at ground level can rapidly increase.

Radiation inversions are important in another context besides air pollution. Fruit growers in places like California have long known that their crops are in greatest danger of frost damage on winter nights when the skies are clear and a radiation inversion sets in. Since the air even a few meters up is warmer than the air at crop level, one way to help protect sensitive crops on such nights is simply to mix the air with large motor-driven fans.

Subsidence Inversions. While radiation inversions are mostly a short-lived, ground-level, wintertime phenomenon, the other important cause of inversions, subsidence, creates quite the opposite characteristics. Subsidence inversions may last for months on end, occur at higher elevations, and are more common in summer than winter.

Subsidence inversions are associated with high-pressure weather systems, known as *anticyclones*. Air in the middle of a high pressure zone is descending, while on the edges, it is rising. Air near the ground moves outward from the center, while air aloft moves toward the center from the edges. The result is a massive vertical circulation system. As air in the center of the system falls, it experiences greater pressure and is compressed and heated. If its temperature at elevation z_1 is T_1, then as it falls to elevation z_2, it will be heated adiabatically to $T_2 = T_1 + \Gamma_d(z_1 - z_2)$, as shown in Figure 43. As is often the case, this compressive heating warms the descending air to a higher temperature than the air below, whose temperature is dictated primarily by conditions on the ground.

Since subsiding air gets warmer as it descends, it is better able to hold whatever water vapor it may have without condensing. Unless there are sources of new moisture, its relative humidity drops, and there is little chance that clouds will form. The result is that high-pressure zones create clear, dry weather with lots of sunshine during the day and clear skies at night. Clear skies during the day allow solar warming of the Earth's surface, which often creates superadiabatic conditions under the inversion. The result of subsidence and surface warming is an inversion located anywhere from several hundred meters above the surface to several thousand meters, that lasts as long as the high-pressure weather system persists. At night, the surface can cool quickly by radiation, which may result in a radiation inversion located under the subsidence inversion, as shown in Figure 44.

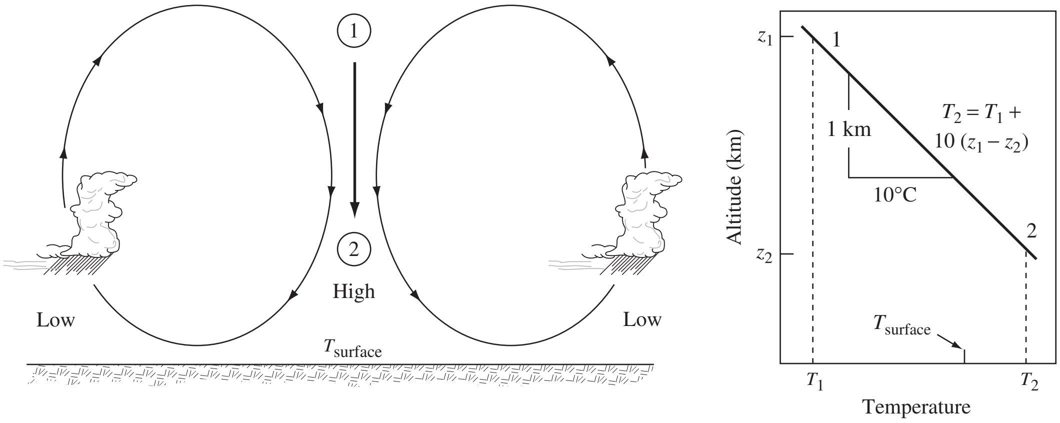
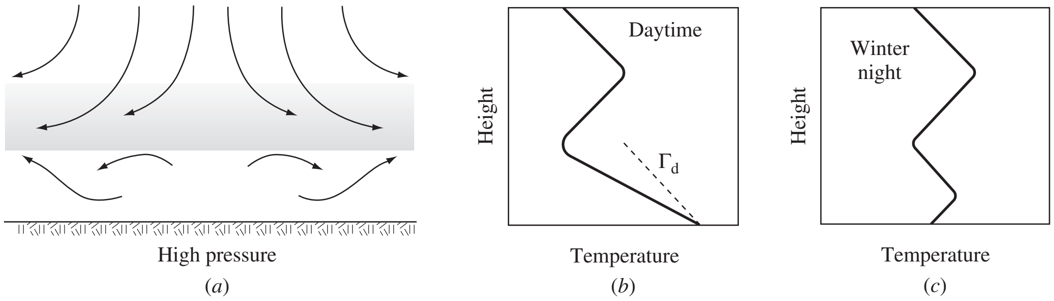


FIGURE 43 Descending air is warmed 10°C/km by compression. It is common to have this subsidence heating create warmer conditions aloft than near the surface, which causes a temperature inversion. Clear skies are typical in a region of high pressure, while cloudiness and precipitation are common in areas with low atmospheric pressure.



(*a*) (*b*) (*c*)

FIGURE 44 Descending air causes subsidence inversions (*a*). During the day, air under the inversion may be unstable due to solar warming of the surface (*b*). Radiation inversions may form under the subsidence inversion when nights are clear, especially in winter (*c*).

Some anticyclones and their accompanying subsidence inversions drift across the continents (west to east in northern midlatitudes) so that at any given spot they may appear to come and go with some frequency. On the other hand, other anticyclones are semipermanent in nature and can cause subsidence inversions to last for months at a time. These semipermanent highs are the result of the general atmospheric circulation patterns shown in Figure 45.

At the equinox, the equator is directly under the sun, and the air there will be heated, become buoyant, and rise. As that air approaches the top of the troposphere (roughly 10 to 12 km), it begins to turn, some heading north and some south. An eighteenth century meteorologist, George Hadley, postulated that the air would continue to the poles before descending. In actuality, it descends at a latitude of about 30° and then returns to the equator, forming what is known as a *Hadley cell*. In similar fashion, though not as distinct, there are other cells, linked as in a chain, between 30° and 60° latitude, and between 60° and the poles. The descending air at

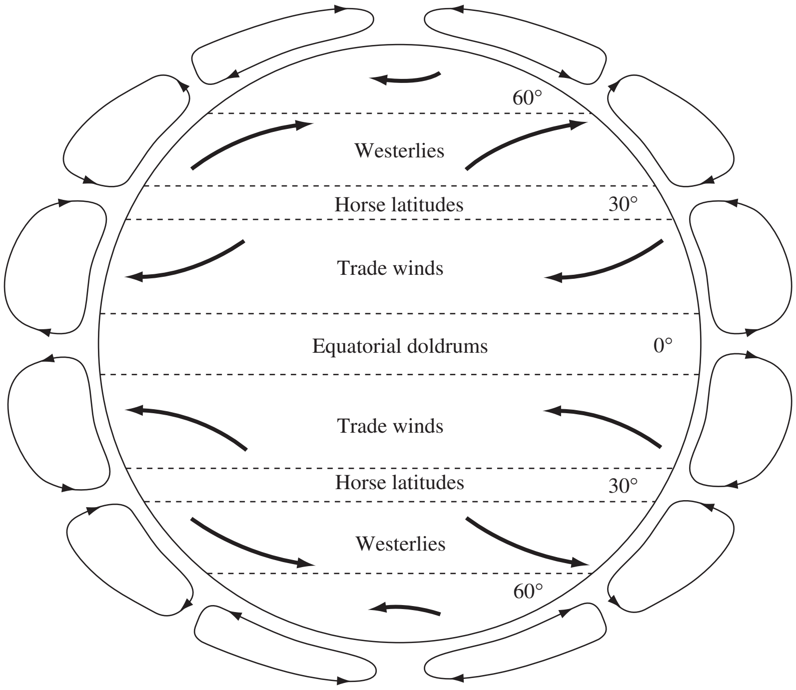


FIGURE 45 Idealized general air circulation patterns, drawn at the equinox. High-pressure zones are common around 30° latitude producing clear skies and subsidence inversions.

30° creates a persistent high-pressure zone with corresponding lack of clouds or rainfall that is a contributing factor to the creation of the world's great deserts. The deserts of southern California, the southwestern United States, the Sahara, the Chilean desert, the Kalahari in South Africa, and the great deserts in Australia, are all located at roughly 30° latitude for this reason. Conversely, rising air near the equator and near 60° latitude tends to be moist after passing over oceans. As it cools, the moisture condenses, causing clouds and rainfall. For this reason, some of the wettest regions of the world are in those bands of latitude.

The air that moves along the surface of the Earth to close a Hadley cell is affected by Coriolis forces, causing it to curve toward the west if it is moving toward the equator or toward the east if it is moving toward the poles. The resulting winds between roughly 30° and 60° are known as the *westerlies,* and between the 30° and the equator they are the *trade winds*. Near the equator, there is little wind since the air is mostly rising. That band is called the *equatorial doldrums*. Similarly, the surface air is relatively calm around 30° latitude, forming a band called the *horse latitudes*. (Apparently early explorers were sometimes forced to lighten the load by throwing horses overboard to avoid being becalmed as they sailed to the New World.)

All of these important bands of latitudes, the doldrums, horse latitudes, high-pressure and low-pressure zones, move up and down the globe as the seasons change. Figure 45 shows global circulation patterns at an equinox, when the sun is directly over the equator. In Northern Hemisphere summer, the sun moves northward, as do the persistent high- and low-pressure zones associated with these bands. There is, for example, a massive high-pressure zone off the coast of California that moves into place over Los Angeles (latitude 34°) and San Francisco (latitude 38°) in the spring, that remains there until late fall. That is the principal reason for California's sunny climate, as well as its smog. Clear skies assure plenty of sunlight to power photochemical reactions, the lack of rainfall eliminates that atmospheric cleansing mechanism, and prolonged subsidence inversions concentrate the pollutants. In the case of Los Angeles, there is also a ring of mountains around the city that tends to keep winds from blowing the smog away. In the Northern Hemisphere winter, the sun is directly overhead somewhere in the Southern Hemisphere, and the bands move southward.

The global circulation patterns of Figure 45 are, of course, idealized. The interactions between sea and land, the effects of storms and other periodic disturbances, and geographical features such as high mountain ranges all make this model useful only on a macroscale. Even a simplified global model, however, helps explain a number of significant features of the world's climate and some aspects of regional air pollution problems.

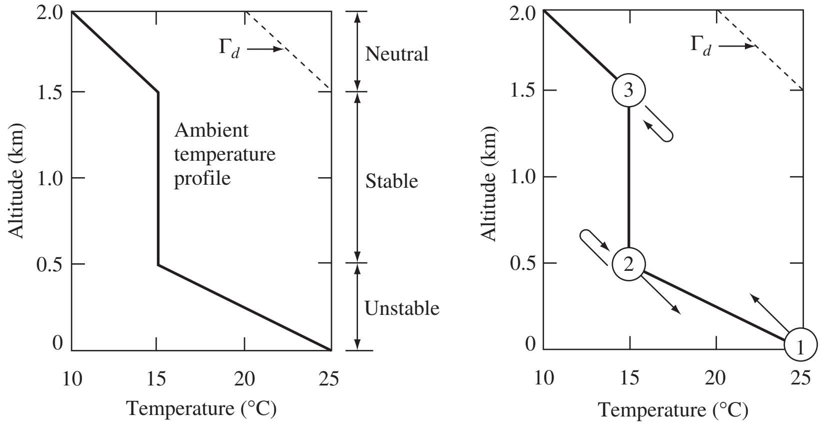
Atmospheric Stability and Mixing Depth

Atmospheric stability is important because it determines the ability of pollutants to disperse vertically into the atmosphere. We already know that layers of the atmosphere have stability characteristics that depend on their local lapse rate in comparison with adiabatic lapse rates. That picture is incomplete, however, and we must introduce another conceptual experiment before we can be sure of the consequences of local lapse rates.

Consider the temperature profile in Figure 46*a* and, for the moment, assume that the air is dry so we don't have to worry about the difference between dry and saturated lapse rates. If we look only at the lapse rates, the layer of air from ground up to 0.5 km is superadiabatic, and should be labeled "unstable"; the layer from 0.5 to 1.5 km is subadiabatic and labeled "stable"; and the layer above 1.5 km has a lapse rate equal to the dry adiabatic lapse rate 10°C/km and labeled "neutral." Call these "local" stability classifications.

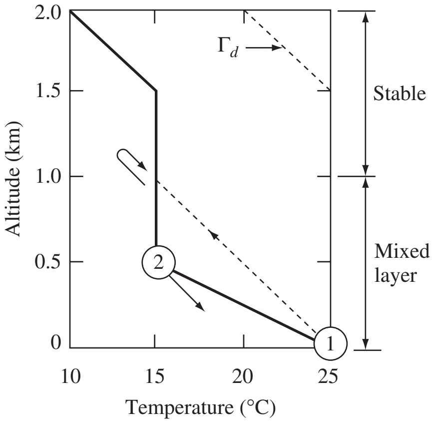
Now consider three parcels of air, labeled 1, 2, and 3, in Figure 46*b*. If for some reason one of these parcels starts to move up or down, its temperature will change by 10°C/km. After imagining that the parcel has moved, compare its temperature with the temperature of the surrounding air at the new altitude. If the parcel is warmer, it will be buoyant and will rise; if it is colder than the surrounding air, it will be denser and will sink. With this in mind, consider parcel 1. If for some reason parcel 1 starts to move slightly upward, it will cool at the dry adiabatic lapse rate of 10°C/km, making it warmer than the surrounding air, so the parcel will continue rising on its own. If parcel 2 starts to rise, however, it will find itself colder than the surrounding air, so it will sink back to where it started from. If it sinks below its initial level, it will be colder than surrounding air and will continue to sink. If parcel 3 starts down, it will be warmer than the surrounding air and will rise back to where it was.

So, how far up will parcel 1 go? It will keep rising as long as it is warmer than the surrounding air. As shown in 46*c*, it will rise to 1 km and then stop. If it went higher than 1 km, it would find itself colder than its surroundings and would sink



(*a*) Local stability

(*b*) Moving parcels



(*c*) Mixed layer

FIGURE 46 Illustrating the difference between local stability and stability based on mixing layers. (*a*) Local stability determined by lapse rates; (*b*) parcels 1 and 2 will move; (*c*) parcel 1 will rise to 1 km creating turbulence and mixing from ground-level to 1 km.

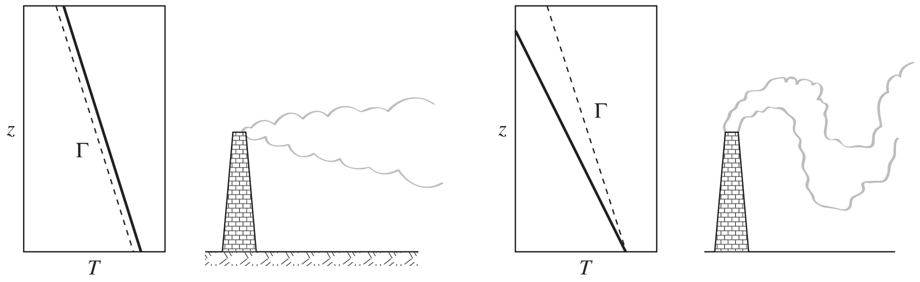
back to 1 km. In other words, air from ground-level to 1 km will actually experience turbulent mixing despite the "stable" lapse rate above 0.5 km.

The process illustrated in Figure 46 helps provide a sense of the ability of the atmosphere to provide dilution of air pollutants. By projecting the ground temperature upwards at the dry adiabatic lapse rate until it crosses the actual ambient temperature profile, a region of turbulent mixing is identified. The altitude of the top of that mixed layer is called the *mixing depth* (or, sometimes, the *mixing height*).

The product of mixing depth and average windspeed within the mixing depth is sometimes used as an indicator of the atmosphere's dispersive capability. This product is known as the *ventilation coefficient* (m^2/s). Values of ventilation coefficient less than about 6,000 m^2/s are considered indicative of high air pollution potential (Portelli and Lewis, 1987).

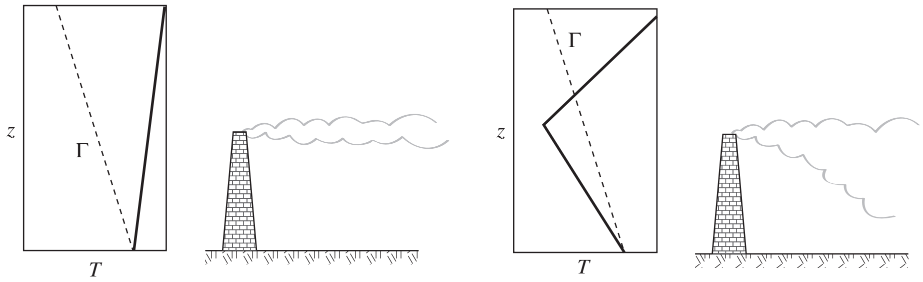
Smokestack Plumes and Adiabatic Lapse Rates

The atmospheric temperature profile affects the dispersion of pollutants from a smokestack, as shown in Figure 47. If a smokestack were to emit pollutants into a neutrally stable atmosphere, we might expect the plume to be relatively symmetrical, as shown in Figure 47*a*. The term used to describe this plume is *coning*. In Figure 47*b*, the atmosphere is very unstable, and there is rapid vertical air movement, both up and down, producing a *looping* plume. In Figure 47*c*, a *fanning* plume results when a stable atmosphere greatly restricts the dispersion of the plume



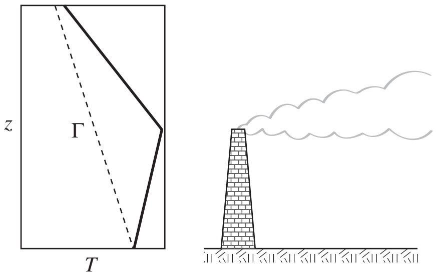
Coning
(*a*)

Looping
(*b*)



Fanning
(*c*)

Fumigation
(*d*)



Lofting
(*e*)

FIGURE 47 Effect of atmospheric lapse rates and stack heights on plume behavior. The dashed line is the dry adiabatic lapse rate for reference.

in the vertical direction, although it still spreads horizontally. In Figure 47*d*, when a stack is under an inversion layer, emissions move downward much more easily than upward. The resulting *fumigation* can lead to greatly elevated downwind, ground-level concentrations.

When the stack is above an inversion layer, as in Figure 47*e*, mixing in the upward direction is uninhibited, but downward motion is greatly restricted by the inversion's stable air. Such *lofting* helps keep the pollution high above the ground, reducing exposure to people living downwind. In fact, a common approach to air pollution control in the past has been to build taller and taller stacks to emit pollutants above inversions. An unfortunate consequence of this approach, however, has been that pollutants released from tall stacks are able to travel great distances and may cause unexpected effects, such as acid deposition, hundreds of miles from the source.

11 The Point Source Gaussian Plume Model

The Clean Air Act specifies new source emission standards, and it specifies ambient air quality standards. The connecting link between the two is the atmosphere. How do pollutants behave once they have been emitted, and how may we predict their concentrations in the atmosphere? How can the limitations imposed by a Prevention of Significant Deterioration policy be shown to be satisfied for new sources in an attainment area? How can we predict the improvement in air quality that must be achieved when new sources are proposed for a nonattainment area? To help answer questions like these, computer models that use such information as predicted emissions, smokestack heights, wind data, atmospheric temperature profiles, ambient temperatures, solar insolation, and local terrain features have been developed and will be discussed in this section.

At the heart of almost every computer program that attempts to relate emissions to air quality is the assumption that the time-averaged pollutant concentration downwind from a source can be modeled using a normal, or Gaussian, distribution curve. The basic Gaussian dispersion model applies to a single *point source* such as a smokestack, but it can be modified to account for *line sources* (such as emissions from motor vehicles along a highway), or *area sources* (these can be modeled as a large number of point sources).

To begin, consider just a single point source such as that shown in Figure 48. The coordinate system has been set up to show a cross-section of the plume, with z representing the vertical direction and x the distance directly downwind from the source. If we were to observe the plume at any particular instant, it might have some irregular shape, such as the outline of the looping plume shown. A few minutes later, however, the plume might have an entirely different boundary. If we were to set up a camera and leave the shutter open for a while, we could imagine getting a photograph of a time-averaged plume envelope such as that shown in the figure.

Since stack emissions have some initial upward velocity and buoyancy, it might be some distance downwind before the plume envelope might begin to look symmetrical about a centerline. The centerline would be somewhat above the actual

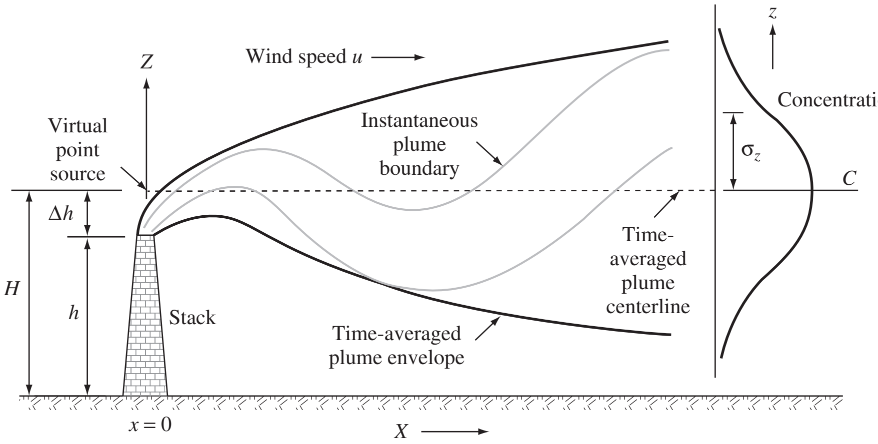


FIGURE 48 The instantaneous plume boundary and a time-averaged plume envelope.

stack height. The highest concentration of pollution would be along this centerline, with decreasing concentrations as the distance from this centerline increases. The Gaussian plume model assumes that the pollutant concentration follows a normal distribution about this centerline in both the vertical plane, as shown in the figure, and in the horizontal direction, not shown. It also treats emissions as if they came from a virtual point source along the plume centerline, at an *effective stack height H*.

The Gaussian point-source dispersion equation relates average, steady-state pollutant concentrations to the source strength, windspeed, effective stack height, and atmospheric conditions. Its form can be derived from basic considerations involving gaseous diffusion in three-dimensional space. The derivation, however, is beyond the scope of this book (see, for example, Seinfeld and Pandis, 1998). It is important to note the following assumptions that are incorporated into the analysis:

- The rate of emissions from the source is constant.
- The windspeed is constant both in time and with elevation.
- The pollutant is conservative; that is, it is not lost by decay, chemical reaction, or deposition. When it hits the ground, none is absorbed, and all is reflected.
- The terrain is relatively flat, open country.

The three-dimensional coordinate system established in Figure 49 has the stack at the origin, with distance directly downwind given by x, distance off the downwind axis specified by y, and elevation given by z. Since our concern is going to be only with receptors (people and ecosystems) at ground-level, the form of the Gaussian plume equation given here is less general than it can be, and applies only for $z = 0$:

$$C(x, y) = \frac{Q}{\pi\, u_{\mathrm{H}} \sigma_y \sigma_z} \exp\left(\frac{-H^2}{2\sigma_z^2}\right) \exp\left(\frac{-y^2}{2\sigma_y^2}\right) \tag{45}$$

where

$C(x, y)$ = concentration at ground-level at the point (x, y), μg/m^3
x = distance directly downwind, m

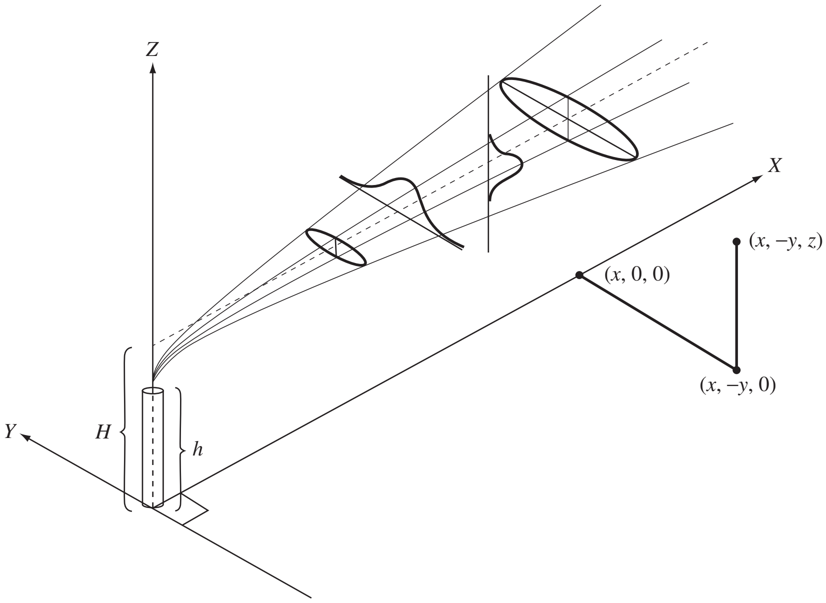


FIGURE 49 Plume dispersion coordinate system, showing Gaussian distributions in the horizontal and vertical directions.
(*Source:* Turner, 1970.)

y = horizontal distance from the plume centerline, m
Q = emission rate of pollutants, μg/s
H = effective stack height, m ($H = h + \Delta h$, where h = actual stack height, and Δh = plume rise)
u_H = average windspeed at the effective height of the stack, m/s
σ_y = horizontal dispersion coefficient (standard deviation), m
σ_z = vertical dispersion coefficient (standard deviation), m

Before we get into the details of (45), several features are worth noting. Ground-level pollution concentration is directly proportional to the source strength Q, so it is easy to determine how much source reduction is necessary to achieve a desired decrease in downwind concentration. The units of Q, by the way, have been given in micrograms so concentrations will be in the usual micrograms per cubic meter (μg/m^3). The ground-level pollution decreases when taller stacks are used, although the relationship is not linear. Also note that there is no explicit relationship between emission rate, Q, and downwind distance, x. The downwind distance will be brought into the equation when σ_x and σ_y are evaluated.

Equation (45) suggests that downwind concentration appears to be inversely proportional to the windspeed, which is what might be expected intuitively. In actuality, the inverse relationship is slightly modified by the dependence of plume rise Δh on windspeed. Higher windspeeds reduce the effective height of the stack H, which keeps ground-level pollution from dropping quite as much as a simple inverse relationship would imply.

Finally, and most important, although the Gaussian plume equation is based on both theory and actual measured data, it is still, at best, a crude model. Predictions based on the model should be assumed to be accurate to within perhaps ±50 percent. In spite of that uncertainty, however, it is still very useful since it has almost universal acceptance, is easy to use, and allows comparison between the estimates made by different modelers in varying situations.

Windspeed Changes with Elevation

The windspeed to be used in (42), u_H, is the windspeed at the effective stack height. Usually windspeed is measured with an anemometer that is set up at a height of 10 meters above the ground, so we need some way to relate windspeed at the anemometer height with windspeed at the effective stack height. The following power law expression is frequently used for elevations less than a few hundred meters above the ground:

$$\left(\frac{u_H}{u_a}\right) = \left(\frac{H}{z_a}\right)^p \tag{46}$$

where

u_H = windspeed at the elevation H
u_a = windspeed at the anemometer height
H = effective height of the plume
z_a = anemometer height above ground
p = a dimensionless parameter that depends on surface roughness and atmospheric stability

Table 6 gives values for p that are recommended by the EPA when there are rough surfaces in the vicinity of the anemometer (Peterson, 1978). For smooth terrain such as flat fields, or near bodies of water, the values of p given in Table 6 should be multiplied by 0.6. The *stability class* indicators in the table will be further clarified in Table 7, when we compute the characteristics of smokestack plumes.

TABLE 6

Wind Profile Exponent *p*, for Rough Terrain[a]

Stability Class	Description	Exponent p
A	Very unstable	0.15
B	Moderately unstable	0.15
C	Slightly unstable	0.20
D	Neutral	0.25
E	Slightly stable	0.40
F	Stable	0.60

[a] For smooth terrain, multiply p by 0.6; see Table 7 for further descriptions of the stability classifications used here.

Source: Peterson, 1978.

TABLE 7

Atmospheric Stability Classifications

Surface Windspeed[a](m/s)	Day Solar Insolation			Night Cloudiness[e]	
	Strong[b]	Moderate[c]	Slight[d]	Cloudy (≥4/8)	Clear (≤3/8)
<2	A	A–B[f]	B	E	F
2–3	A–B	B	C	E	F
3–5	B	B–C	C	D	E
5–6	C	C–D	D	D	D
>6	C	D	D	D	D

[a] Surface windspeed is measured at 10 m above the ground.
[b] Corresponds to clear summer day with sun higher than 60° above the horizon.
[c] Corresponds to a summer day with a few broken clouds, or a clear day with sun 35–60° above the horizon.
[d] Corresponds to a fall afternoon, or a cloudy summer day, or clear summer day with the sun 15–35° above the horizon.
[e] Cloudiness is defined as the fraction of sky covered by clouds.
[f] For A–B, B–C, or C–D conditions, average the values obtained for each.

Note: A, Very unstable; B, moderately unstable; C, slightly unstable; D, neutral; E, slightly stable; F, stable. Regardless of windspeed, class D should be assumed for overcast conditions, day or night.

Source: Turner, 1970.

EXAMPLE 11 **Windspeed at the Stack**

Suppose an anemometer at a height of 10 m above ground measures the windspeed at 2.5 m/s. Estimate the windspeed at an elevation of 300 m in rough terrain if the atmosphere is slightly unstable.

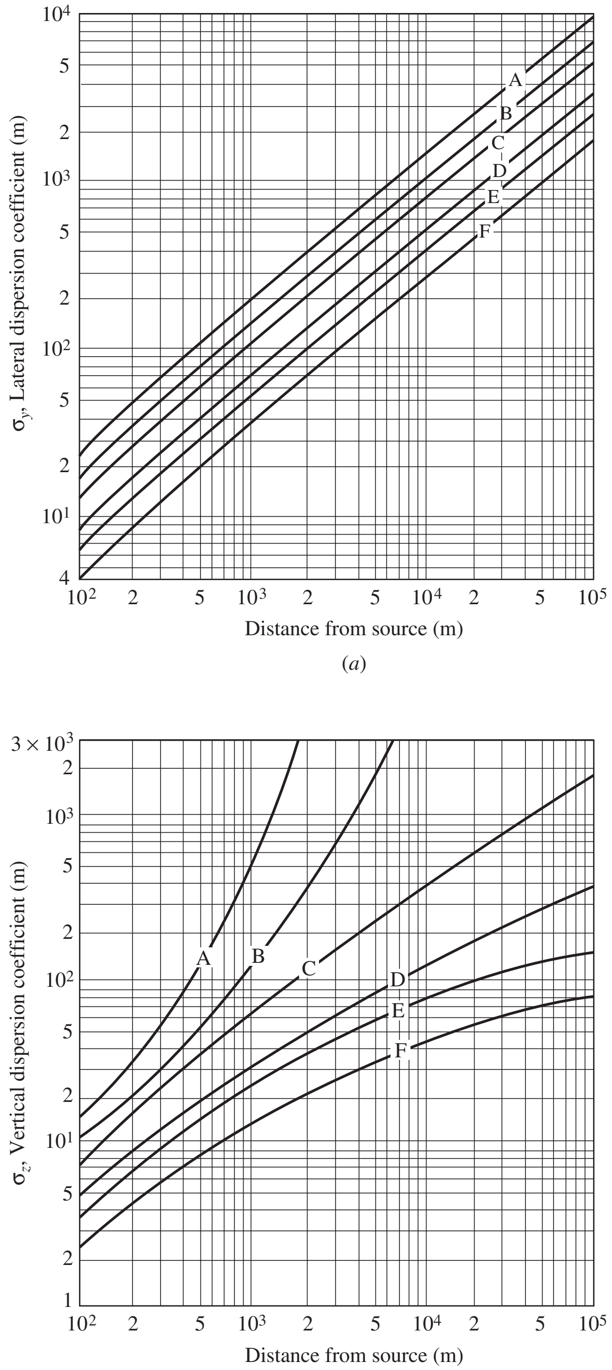
Solution From Table 6, we find the wind profile exponent for a slightly unstable atmosphere (stability class C) is 0.20. Rearranging (46) gives

$$u_{\mathrm{H}} = u_{\mathrm{a}}\left(\frac{H}{z_{\mathrm{a}}}\right)^{p} = 2.5 \cdot \left(\frac{300}{10}\right)^{0.2} = 4.9 \text{ m/s}$$

The Gaussian Dispersion Coefficients

The two dispersion coefficients in (45), σ_y and σ_z, need explanation. These are really just the standard deviations of the horizontal and vertical Gaussian distributions, respectively (about 68 percent of the area under a Gaussian curve is within $\pm 1\sigma$ of the mean value). Smaller values for a dispersion coefficient mean the Gaussian curve is narrower, with a higher peak, and larger values mean the opposite. The further downwind we go from the source, the larger these coefficients become. This causes the Gaussian curves to spread further and further. These coefficients are not only a function of downwind distance, but they also depend, in a complex way, on atmospheric stability.

The most common procedure for estimating the dispersion coefficients was introduced by Pasquill (1961), modified by Gifford (1961), and adopted by the U.S. Public Health Service (Turner, 1970); it is presented here as Figure 50. The parameters A through F in Figure 50 represent stability classifications based on



(*a*)

(*b*)

FIGURE 50 Gaussian dispersion coefficients as a function of distance downwind (*a*) horizontal coefficient, σ_y, (*b*) vertical coefficient, σ_z.
(*Source:* Turner, 1970.)

qualitative descriptions of prevailing environmental conditions. Table 7 describes these parameters. For example, a clear summer day, with the sun higher than 60° above the horizon, having windspeeds less than 2 m/s (at an elevation of 10 m), creates a *very unstable* atmosphere with stability classification A. The opposite extreme is classification F, which is labeled *stable* and corresponds to a clear night (less than 3/8 of the sky covered by clouds), with winds less than 3 m/s.

The graphical presentation of dispersion coefficients in Figure 50 is inconvenient when doing calculations with a computer. A reasonable fit to those graphs can be obtained using the following equations (Martin, 1976):

$$\sigma_y = a\,x^{0.894} \tag{47}$$

and

$$\sigma_z = cx^d + f \tag{48}$$

where the constants a, c, d, and f are given in Table 8 for each stability classification. The downwind distance x must be expressed in kilometers to yield σ_y and σ_z in meters. For convenience, a few values for the dispersion coefficients calculated using (47) and (48) are listed in Table 9.

TABLE 8

Values of the Constants *a, c, d,* and *f* for Use in (47) and (48)

		$x \le 1$ km			$x \ge 1$ km		
Stability	a	c	d	f	c	d	f
A	213	440.8	1.941	9.27	459.7	2.094	−9.6
B	156	106.6	1.149	3.3	108.2	1.098	2.0
C	104	61.0	0.911	0	61.0	0.911	0
D	68	33.2	0.725	−1.7	44.5	0.516	−13.0
E	50.5	22.8	0.678	−1.3	55.4	0.305	−34.0
F	34	14.35	0.740	−0.35	62.6	0.180	−48.6

Note: The computed values of σ will be in meters when x is given in kilometers.
Source: Martin, 1976.

TABLE 9

Dispersion Coefficients (m) for Selected Distances Downwind (km), Computed with (47) and (48)

Distance x	Stability Class and σ_y						Stability Class and σ_z					
(km)	A	B	C	D	E	F	A	B	C	D	E	F
0.2	51	37	25	16	12	8	29	20	14	9	6	4
0.4	94	69	46	30	22	15	84	40	26	15	11	7
0.6	135	99	66	43	32	22	173	63	38	21	15	9
0.8	174	128	85	56	41	28	295	86	50	27	18	12
1	213	156	104	68	50	34	450	110	61	31	22	14
2	396	290	193	126	94	63	1953	234	115	51	34	22
4	736	539	359	235	174	117		498	216	78	51	32
8	1367	1001	667	436	324	218		1063	406	117	70	42
16	2540	1860	1240	811	602	405		2274	763	173	95	55
20	3101	2271	1514	990	735	495		2904	934	196	104	59

Downwind Ground-Level Concentration

The ground-level concentration of pollution directly downwind of the stack is of interest since pollution will be highest along that axis. With $y = 0$, (45) simplifies to

$$C(x, 0) = \frac{Q}{\pi u_{\mathrm{H}} \sigma_y \sigma_z} \exp\left(\frac{-H^2}{2\sigma_z^2}\right) \tag{49}$$

The following example illustrates the use of the Gaussian plume equation.

EXAMPLE 12 **A Power Plant Plume**

A 40 percent-efficient, 1,000 MW (10^6 kW) coal-fired power plant emits SO_2 at the legally allowable rate of 0.6 lb SO_2 per million Btu of heat into the plant. The stack has an effective height of 300 m. An anemometer on a 10-m pole measures 2.5 m/s of wind, and it is a cloudy summer day. Predict the ground-level concentration of SO_2 4 km directly downwind.

Solution We know the output power of the plant, 10^6 kW, but emission standards are written in terms of input power. Using the conversion 3,412 Btu = 1 kWh, the input power is

$$\text{Input power} = \frac{\text{Output power}}{\text{Efficiency}} = \frac{1.0 \times 10^6 \text{ kW}}{0.40} \times \frac{3{,}412 \text{ Btu}}{\text{kWh}}$$
$$= 8{,}530 \times 10^6 \text{ Btu/h}$$

So the SO_2 emission rate would be

$$Q = \frac{8{,}530 \times 10^6 \text{ Btu}}{\text{h}} \times \frac{0.6 \text{ lb } SO_2}{10^6 \text{ Btu}} \times \frac{\text{kg}}{2.2 \text{ lb}} \times \frac{10^9\ \mu\text{g}}{\text{kg}} \times \frac{\text{h}}{3{,}600 \text{ s}}$$
$$= 6.47 \times 10^8\ \mu\text{g } SO_2\text{/s}$$

For the anemometer windspeed and solar conditions given, Table 7 indicates that the appropriate stability classification is C. Note that Table 7 uses the windspeed as measured at the standard 10-m anemometer height. The windspeed in the Gaussian plume equation, however, requires that we estimate the wind at the effective stack height. We did that in Example 11 and found it to be 4.9 m/s at 300 meters.

At 4 km downwind, Table 9 indicates the dispersion coefficients are $\sigma_y = 359$ m and $\sigma_z = 216$ m. Plugging these into (49) gives

$$C(4, 0) = \frac{6.47 \times 10^8\ \mu\text{g/s}}{\pi \times 4.9 \text{ m/s} \times 359 \text{ m} \times 216 \text{ m}} \exp\left[\frac{-(300)^2}{2 \times (216)^2}\right] = 206\ \mu\text{g/m}^3$$

Using 24.465×10^{-3} m³/mol (at 1 atm, 25°C) along with the molecular weight of SO_2 ($32 + 2 \times 16 = 64$ g/mol), we can convert units to the more commonly used ppm:

$$C = \frac{206 \times 10^{-6} \text{ g/m}^3 \times 24.465 \times 10^{-3} \text{ m}^3\text{/mol}}{46 \text{ g/mol}} = 0.109 \times 10^{-6} = 0.109 \text{ ppm}$$

So the power plant in Example 12 would add 0.11 ppm to whatever SO_2 level is already there from other sources. For perspective, let us compare these emissions to ambient air-quality standards. Table 1 indicates that the annual average SO_2 concentration must be less than 0.03 ppm, which means this power plant by itself would cause pollution to greatly exceed that standard if these atmospheric conditions prevailed over the full year. The 24-hour SO_2 standard is 0.14 ppm, so under the conditions stated, this plant would not violate that standard, but it would make it difficult to allow any other sources in the area.

Example 12 opens the door to many interesting questions. How does the concentration vary as distance downwind changes? What would be the effect of changes in windspeed or stability classification? How would we utilize statistical data on windspeed, wind direction, and atmospheric conditions to be sure that all air-quality standards will be met? If they will not be met, what are the alternatives? Some examples come to mind. We might raise the stack height (adding to the acid deposition problem); we might increase the efficiency of the scrubber to clean flue gases, or we might recommend using coal with a lower sulfur content. Perhaps energy conservation efforts might reduce the size of the power plant required to meet projected needs. Clearly, to do a proper siting analysis for a new source, such as the power plant in Example 12, would require a complex study. Though we will not carry out such calculations here, we do have the crucial starting point for such a study—namely the Gaussian plume model.

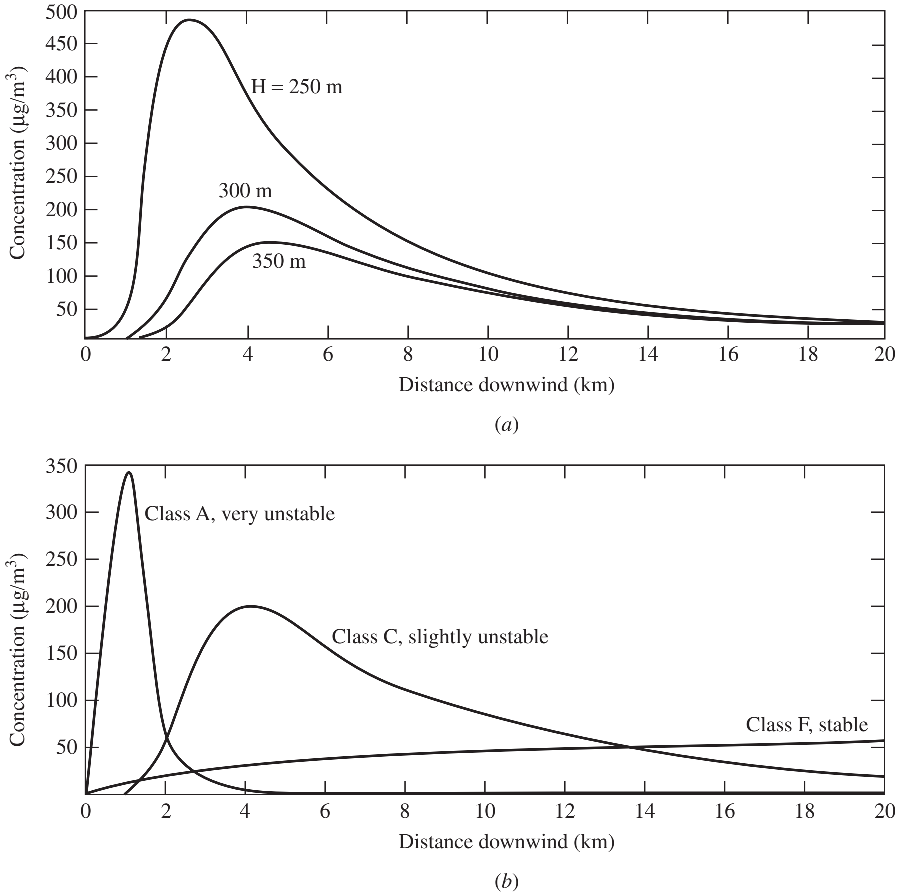
Peak Downwind Concentration

Plotting (49) by hand is tedious, especially if we want to do a sensitivity analysis to see how the results change with changing stack heights and atmospheric conditions. It is much easier to work this out with a computer, and it is especially simple if a spreadsheet program with graphics capability is used. Using Example 12 as a base case, the effect of changes in effective stack height on downwind concentration has been plotted in Figure 51*a*. The effect of changing stability class, while keeping stack height constant at 300 m, is shown in Figure 51*b*.

The downwind concentration of pollution is sensitive to changes in effective stack height, as can be seen in Figure 51*a*. Raising the effective stack height from 250 to 300 m reduces the peak concentration by more than half. That ability of tall stacks to cut ground-level pollution significantly has led to some tremendously tall stacks. In fact, the tallest stack in the world is for a smelter in Subdury, Ontario, and it is as tall as the Empire State Building in New York (380 m). The effective stack height would be even higher.

The impact of changing stability classification shown in Figure 51*b* is perhaps unexpected. The highest peak downwind concentration occurs when the atmosphere is very unstable rather than stable. The turbulence in an unstable atmosphere brings the looping plume to Earth very quickly, resulting in high peak values near the stack. Downwind, however, concentrations drop off very quickly. Having a high peak concentration near the stack may be a satisfactory situation as long as any populations or ecosystems that might be damaged by the pollution are more than a few kilometers away.

The stable atmosphere shown in Figure 51*b*, on the other hand, causes a relatively low peak concentration. The diagram of a fanning plume in Figure 47*c* may help you to understand that conclusion. The plume fans out horizontally, but



(*a*)

(*b*)

FIGURE 51 The effect of variations in key parameters on SO_2 plume for the coal plant in Example 12. (*a*) Impact of changes in the effective stack height for a constant stability classification, and (*b*) effect of stability classification given a constant effective stack height.

not much reaches the ground until the plume is some distance downwind. As shown in Figure 51*b*, the stable atmosphere is worse than the unstable atmospheres beyond about 14 km. It should be noted, however, that the concentrations shown in Figure 51 do not account for the fact that effective stack height depends on the stability classification. If that complication is included, the high peaks shown for unstable atmospheres are reduced modestly.

An obvious question is how can the peak downwind concentration be predicted from (49)? Unfortunately, it is not possible to derive a mathematical solution. One way to predict the peak is to simply plot curves of the sort shown in Figure 51 using a computer. With the ready availability of PCs and spreadsheet programs, that has become an easy enough way to deal with the problem. For hand calculations, however, this approach would be far too tedious. Turner (1970) has derived curves that use the stability classification and effective stack height as parameters to find the distance downwind to the maximum concentration ($x_{\max}$).

The same process yields a normalized concentration $(Cu_H/Q)_{max}$, from which the maximum concentration can be found using the following:

$$C_{max} = \frac{Q}{u_H}\left(\frac{Cu_H}{Q}\right)_{max} \tag{50}$$

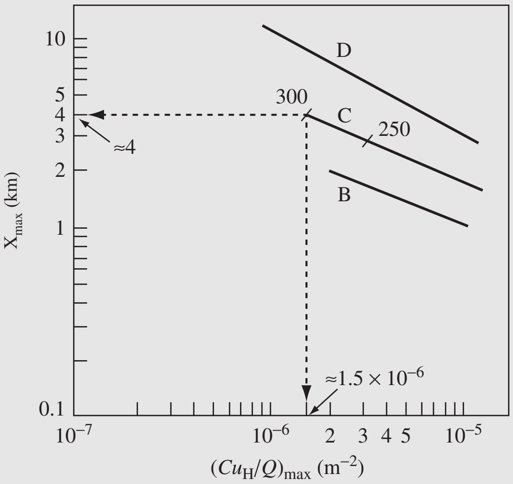
The curves are presented in Figure 52, and the following example illustrates their use.

EXAMPLE 13 Peak Downwind Concentration

For the 1,000 MW, coal-fired power plant of Example 12, use Figure 52 to determine the distance downwind to reach the maximum SO_2 concentration. Then find that concentration.

Solution The stability classification is C, and the effective stack height is 300 m. From Figure 52, the distance downwind, x_{max}, is about 4 km (which agrees with Figure 51). Note the scales on Figure 52 are logarithmic, so be careful when reading values.

Reading across from the point on the figure corresponding to stability class C and effective stack height H = 300 m, $(Cu_H/Q)_{max}$ looks to be about $1.5 \times 10^{-6}\ m^{-2}$.



Windspeed at the effective height of the stack was found in Example 11 to be $u_H = 4.9$ m/s, and Q was $6.47 \times 10^8\ \mu g/s$, so that

$$C_{max} = \frac{Q}{u_H}\left(\frac{Cu_H}{Q}\right)_{max}$$

$$= \frac{6.47 \times 10^8\ \mu g/s}{4.9\ m/s} \times \frac{1.5 \times 10^{-6}}{m^2} = 198\ \mu g/m^3$$

which is fairly close to the 206 $\mu g/m^3$ calculated at 4 km in Example 12. It is quite difficult to read Figure 52, so high accuracy should not be expected.

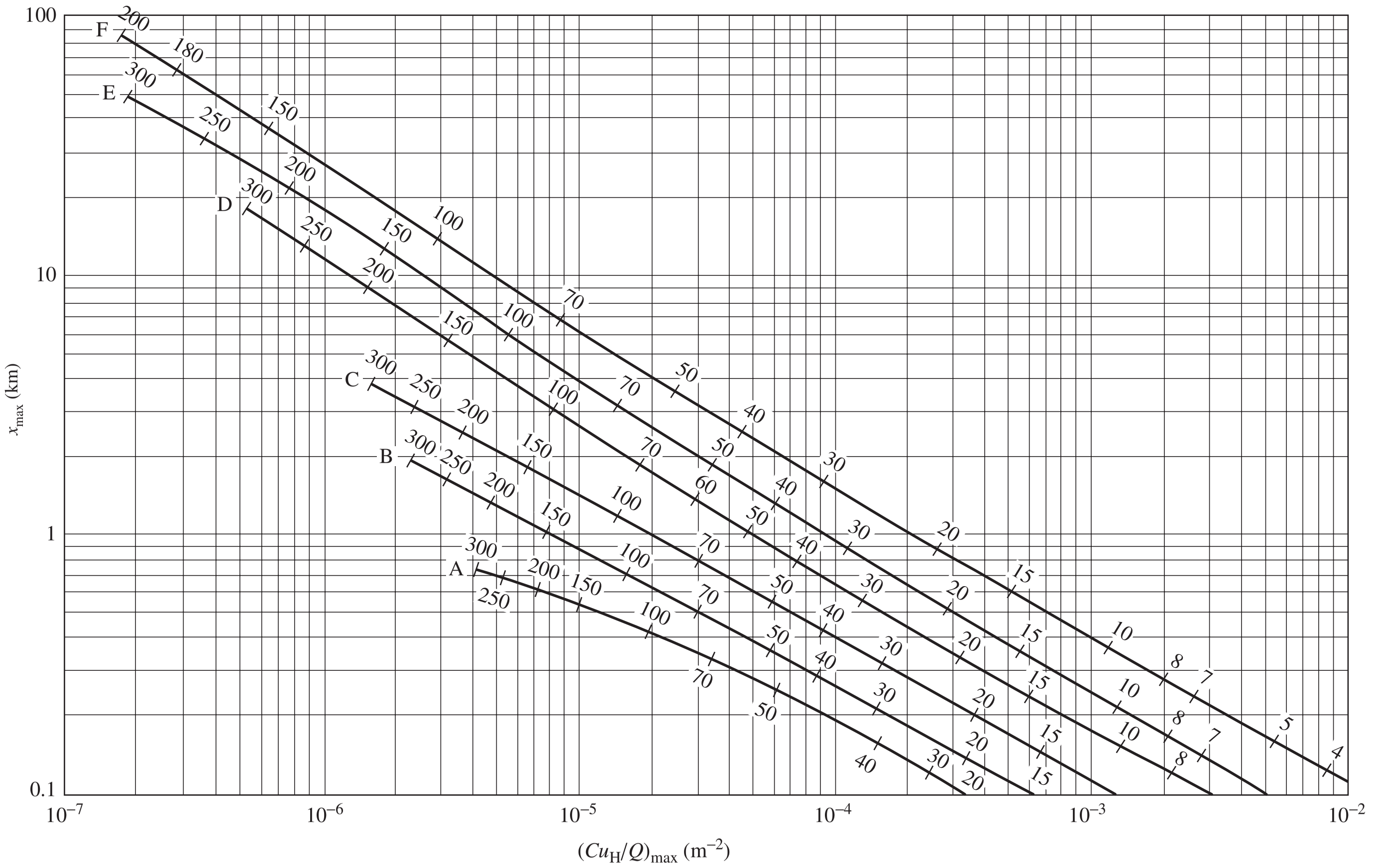


FIGURE 52 To determine the downwind concentration peak, enter the graph at the appropriate stability classification and effective stack height (numbers on the graph, in meters) and then move across to find the distance to the peak, and down to find a parameter from which the peak concentration can be found.
(*Source:* Turner, 1970.)

Plume Rise

So far, we have dealt only with effective stack height H in the calculations. The difference between the actual stack height h and the effective height H is called the plume rise Δh. Plume rise is caused by a combination of factors, the most important ones being the buoyancy and momentum of the exhaust gases, and the stability of the atmosphere itself. Buoyancy results when exhaust gases are hotter than the ambient air, or when the molecular weight of the exhaust is lower than that of air (or a combination of both factors). Momentum is caused by the mass and velocity of the gases as they leave the stack.

A number of techniques have been proposed in the literature for dealing with plume rise, and they tend to yield very different results. The EPA recommends a model based on work by Briggs (1972), and so we will use it. Plume rise depends on momentum and buoyancy. The buoyancy assumed in this analysis is due to the temperature of the stack gases being higher than the surrounding ambient, but one could also include differences in molecular weight of the exhaust gas versus air.

The following plume rise equation can be used for *stable* conditions (stability categories E and F):

$$\Delta h = 2.6 \left(\frac{F}{u_h S} \right)^{1/3} \tag{51}$$

The quantity F is called the *buoyancy flux parameter* (m^4/s^3)

$$F = g r^2 v_s \left(1 - \frac{T_a}{T_s} \right) \tag{52}$$

where

Δh = plume rise, m
g = gravitational acceleration, 9.8 m/s^2
r = inside radius of the stack, m
u_h = windspeed at the height of the stack, m/s
v_s = stack gas exit velocity, m/s
T_s = stack gas temperature, K
T_a = ambient temperature, K

The quantity S is a *stability parameter* with units of s^{-2} given by

$$S = \frac{g}{T_a} \left(\frac{\Delta T_a}{\Delta z} + 0.01 \text{ °C/m} \right) \tag{53}$$

The quantity $\Delta T_a / \Delta T_z$ represents the actual rate of change of ambient temperature with altitude in °C/m (note that a positive value means temperature is increasing with altitude).

For *neutral* or *unstable* conditions in the atmosphere (stability categories A–D), the following equation can be used to estimate plume rise:

$$\Delta h = \frac{1.6\, F^{1/3} x_f^{2/3}}{u_h} \tag{54}$$

where

x_f = distance downwind to point of final plume rise, m

Since (54) is used when conditions are neutral or unstable, it may be difficult to define the distance downwind at which the plume centerline stops rising. The following is sometimes used:

$$\text{Use } x_f = 120\,F^{0.4} \quad \text{if} \quad F \geq 55\ \text{m}^4/\text{s}^3$$

$$x_f = 50\,F^{5/8} \quad \text{if} \quad F < 55\ \text{m}^4/\text{s}^3$$

EXAMPLE 14 **Plume Rise**

A large power plant has a 250-m stack with inside radius 2 m. The exit velocity of the stack gases is estimated at 15 m/s, at a temperature of 140°C (413 K). Ambient temperature is 25°C (298 K), and winds at stack height are estimated to be 5 m/s. Estimate the effective height of the stack if (a) the atmosphere is stable with temperature increasing at the rate of 2°C/km, and (b) the atmosphere is slightly unstable, class C.

Solution First, find the buoyancy parameter F from (52)

$$F = gr^2 v_s \left(1 - \frac{T_a}{T_s}\right)$$

$$= 9.8\ \text{m/s}^2 \times (2\ \text{m})^2 \times 15\ \text{m/s} \times \left(1 - \frac{298}{413}\right) = 164\ \text{m}^4/\text{s}^3$$

(a) With the atmosphere stable, we need to use (51) and (53),

$$S = \frac{g}{T_a}\left(\frac{\Delta T_a}{\Delta z} + 0.01°\text{C/m}\right)$$

$$= \frac{9.8\ \text{m/s}^2}{298\ \text{K}}(0.002 + 0.01)\ \text{K/m} = 0.0004/\text{s}^2$$

$$\Delta h = 2.6\left(\frac{F}{u_h S}\right)^{1/3}$$

$$= 2.6\left(\frac{164\ \text{m}^4/\text{s}^3}{5\ \text{m/s} \times 0.0004/\text{s}^2}\right)^{1/3} = 113\ \text{m}$$

So, the effective stack height is $H = h + \Delta h = 250 + 113 = 363$ m

(b) With an unstable atmosphere, class C, we need to use (54). Since $F > 55\ \text{m}^4/\text{s}^3$, the distance downwind to the point of final plume rise that should be used is

$$x_f = 120\,F^{0.4} = 120 \times (164)^{0.4} = 923\ \text{m}$$

$$\Delta h = \frac{1.6\,F^{1/3} x_f^{2/3}}{u_h} = \frac{1.6(164)^{1/3}(923)^{2/3}}{5} = 166\ \text{m}$$

and the effective stack height is $H = 250 + 166 = 413$ m.

Downwind Concentration Under a Temperature Inversion

The Gaussian plume equation, as presented thus far, applies to an atmosphere in which the temperature profile is a simple straight line. If, as is often the case, there is an inversion above the effective stack height, then the basic Gaussian equation must be modified to account for the fact that the vertical dispersion of pollutants is limited by the inversion.

If the pollutants are assumed to reflect off the inversion layer, just as they were assumed to reflect off the ground in the basic Gaussian equation, then an estimate of the concentration at any point downwind would require an analysis of these multiple reflections. That complexity can be avoided if we are willing to restrict our predictions of plume concentration to distances far enough downwind that the summation of these multiple reflections converges into a closed form solution. Beyond that distance, the air is considered to be completely mixed under the inversion, with uniform concentrations from ground level to the bottom of the inversion layer.

Turner (1970) suggests the following modified Gaussian equation to estimate ground-level concentrations downwind under an inversion. It is derived based on the assumption that the downwind distance from the source is at least twice the distance to where the plume first interacts with the inversion layer:

$$C(x, 0) = \frac{Q}{(2\pi)^{1/2} u_{\mathrm{H}}\, \sigma_y\, L} \quad \text{for } x \geq 2X_L \tag{55}$$

where

L = elevation of the bottom of the inversion layer (m)

X_L = the distance downwind where the plume first encounters the inversion layer

Notice that (55) is applicable only for distances $x \geq 2X_L$ (see Figure 53). That distance, X_L, occurs at the point where the vertical dispersion coefficient, σ_z, is equal to

$$\sigma_z = 0.47(L - H) \quad \text{at } x = X_L \tag{56}$$

After σ_z is found from (56), the distance X_L can be estimated using Figure 50*b*, or using (48) and Table 7.

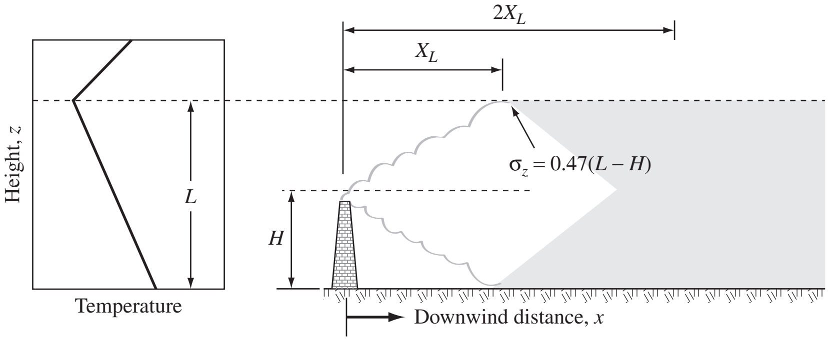


FIGURE 53 Plume dispersion under an elevated inversion. Equation (55) applies for distances greater than $2X_L$ downwind, where X_L occurs at the point where $\sigma_z = 0.47(L - H)$.

For distances $x \leq X_L$, the standard Gaussian plume equation (42) can be used to estimate downwind concentrations. For $X_L \geq 2X_L$, (55) applies. For distances between X_L and $2X_L$, concentrations can be estimated by interpolating between the values computed for $x = X_L$ and $x = 2X_L$.

EXAMPLE 15 **Concentration Under an Inversion Aloft**

Consider a stack with effective height 100 m, emitting SO_2 at the rate of $2 \times 10^8\ \mu g/s$. Windspeed at 10 m is 5.5 m/s, and at 100 m, it is 10 m/s. It is a clear summer day with the sun nearly overhead and an inversion layer starting at 300 m. Estimate the ground-level SO_2 concentration at a distance downwind twice that where reflection begins to occur from the inversion.

Solution Equation (56) gives σ_z at the distance X_L,

$$\sigma_z = 0.47(L - H) = 0.47 \times (300 - 100) = 94 \text{ m}$$

To find X_L, we can use Figure 50*b*, but first we need the stability classification below the inversion. From Table 7, a clear summer day with 5.5 m/s windspeed corresponds to stability class C. We need to find X_L in order to find σ_Y in (55). Entering Figure 50*b* on the vertical axis at $\sigma_z = 94$ m, then going across to the class C line and dropping down to the horizontal axis, leads to a very rough estimate for X_L of about 1,600 m.

To find the concentration at $x = 2X_L = 2 \times 1{,}600 \text{ m} = 3{,}200$ m, we need to estimate σ_y at that point. Using Figure 47*a* at $x = 3{,}200$ m and class C, we can estimate σ_y to be about 300 m (we could have used Eq. 47 with Table 8). Using (55), the concentration directly downwind at ground-level is

$$C(x, 0) = \frac{Q}{(2\pi)^{1/2} u_H\, \sigma_y\, L}$$

$$= \frac{2 \times 10^8\ \mu g/s}{(2\pi)^{1/2} \times 10 \text{ m/s} \times 300 \text{ m} \times 300 \text{ m}} = 90\ \mu g/m^3$$

A Line-Source Dispersion Model

In some circumstances, it is appropriate to model sources distributed along a line as if they formed a continuously emitting, infinite line source. Examples of line sources that might be modeled this way include motor vehicles traveling along a straight section of highway, agricultural burning along the edge of a field, or a line of industrial sources on the banks of a river. For simplicity, we will consider only the case of an infinite-length source at ground-level, with winds blowing perpendicular to the line as shown in Figure 54.

Under these specialized circumstances, the ground-level concentration of pollution at perpendicular distance x from the line source can be described by the following:

$$C(x) = \frac{2q}{\sqrt{2\pi}\, \sigma_z u} \tag{57}$$

where

q = emission rate per unit of distance along the line (g/m-s).

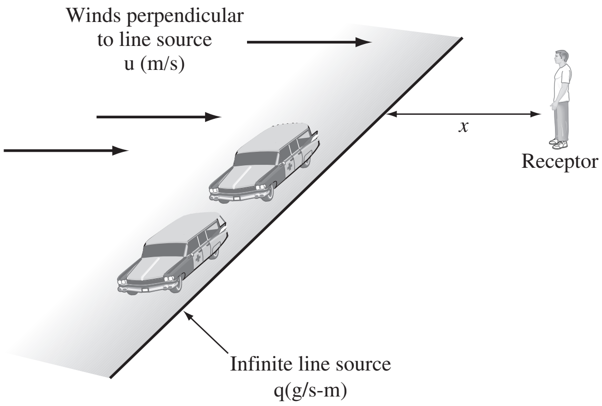


FIGURE 54 Showing the geometry of a line source, such as a straight section of highway, and the receptor located at a distance *x* away from the line.

EXAMPLE 16 **CO Near a Freeway**

The federal emission standards for CO from new vehicles is 3.4 g/mile. Suppose a highway has 10 vehicles per second passing a given spot, each emitting 3.4 g/mile of CO. If the wind is perpendicular to the highway and blowing at 5 mph (2.2 m/s) on an overcast day, estimate the ground-level CO concentration 100 m from the freeway.

Solution We need to estimate the CO emission rate per meter of freeway,

$$q = 10 \text{ vehicles/s} \times 3.4 \text{ g/veh-mile} \times 1 \text{ mile/1,609 m} = 0.021 \text{ g/s-m}$$

and we need the vertical dispersion coefficient σ_z. Since this is an overcast day, the footnotes of Table 7 indicate we should use stability class D. Checking Table 9 for σ_z doesn't help since it doesn't give values for such short distances. So, with (48) and values of *c*, *d*, and *f* from Table 8, we can compute σ_z:

$$\sigma_z = cx^d + f = 33.2(0.1)^{0.725} - 1.7 = 4.6 \text{ m}$$

Substituting these values into (57) gives

$$C(0.1 \text{ km}) = \frac{2q}{\sqrt{2\pi}\,\sigma_z u} = \frac{2 \times 0.021 \text{ g/m} \cdot \text{s} \times 10^3 \text{ mg/g}}{(2\pi)^{1/2} \times 4.6 \text{ m} \times 2.2 \text{ m/s}} = 1.7 \text{ mg/m}^3$$

The eight-hour CO air quality standard is 10 mg/m^3, so this estimate is well below the standard.

Area-Source Models

For distributed sources, a number of approaches can be taken to estimate pollutant concentrations. If there are a modest number of point sources, it is reasonable to use the point-source Gaussian plume equation for each source to predict its individual

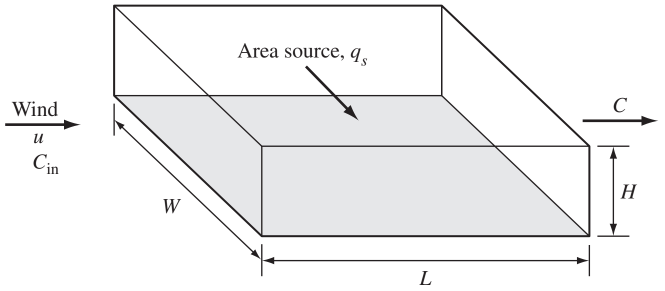


FIGURE 55 Box model for an airshed over a city. Emissions per unit area are given by q_s, pollutants are assumed to be uniformly mixed in the box with concentration C, and upwind of the box, the concentration is C_{in}.

contribution. Then, by superposition, find the total concentration at a given location by summing the individual contributions. Multiple use of the Gaussian line-source equation is another approach. By dividing an area into a series of parallel strips and then treating each strip as a line source, the total concentration on any strip can be estimated.

A much simpler, more intuitive approach can be taken to estimate pollutant concentrations over an area (such as a city) by using the box model concepts introduced in Section 1.3. Consider the airshed over an urban area to be represented by a rectangular box, such as is shown in Figure 55, with base dimensions L and W and height H. The box is oriented so that wind, with speed u, is normal to one side of the box. The height of the box is determined by atmospheric conditions, and we could consider it to be just the mixing depth. Emissions per unit area will be represented by q_s (g/m^2-s).

Consider the air blowing into the box on the upwind side to have pollutant concentration C_{in}, and, for simplicity, assume that no pollution is lost from the box along the sides parallel to the wind or from the top. We will also assume that the pollutants are rapidly and completely mixed in the box, creating a uniform average concentration C. Finally, we will treat the pollutants as if they are conservative, that is, they do not react, decay, or fall out of the air stream. All of these restrictions can be modified in more sophisticated versions of a box model.

Working with pollutant mass, the amount of pollution in the box is the volume of the box times the concentration, $LWHC$. The rate at which air is entering and leaving the box is the area of either end times the windspeed, WHu, so the rate at which pollution is entering the box is $WHuC_{\text{in}}$. The rate that it leaves the box is $WHuC$. If we assume the pollutant is conservative, then we can then write the following mass balance for the box:

$$\begin{pmatrix}\text{Rate of change of}\\ \text{pollution in the box}\end{pmatrix} = \begin{pmatrix}\text{Rate of pollution}\\ \text{entering the box}\end{pmatrix} - \begin{pmatrix}\text{Rate of pollution}\\ \text{leaving the box}\end{pmatrix}$$

or

$$LWH\frac{dC}{dt} = q_sLW + WHuC_{\text{in}} - WHuC \tag{58}$$

where

C = pollutant concentration in the airshed, mg/m^3
C_{in} = concentration in the incoming air, mg/m^3
q_s = emission rate per unit of area, mg/s-m^2
H = mixing height, m
L = length of airshed, m
W = width of airshed, m
u = average windspeed against one edge of the box, m/s

The steady-state solution to (58) can be obtained by simply setting $dC/dt = 0$, so that

$$C(\infty) = \frac{q_s L}{uH} + C_{\text{in}} \tag{59}$$

which looks reasonable. If the air entering the box is clean, the steady-state concentration is proportional to the emission rate and inversely proportional to the ventilation coefficient (the product of mixing depth and windspeed). If it is not clean, we just add the effect of the incoming concentration. We can also solve (58) to obtain the time-dependent increase in pollution above the city. Letting $C(0)$ be the concentration in the airshed above the city (the box) at time $t = 0$, the solution becomes

$$C(t) = \left(\frac{q_s L}{uH} + C_{\text{in}}\right)(1 - e^{-ut/L}) + C(0)e^{-ut/L} \tag{60}$$

If we assume the incoming wind blows no pollution into the box, and if the initial concentration in the box is zero, then (60) simplifies to

$$C(t) = \frac{q_s L}{uH}(1 - e^{-ut/L}) \tag{61}$$

When $t = L/u$, the exponential function becomes e^{-1} and the concentration reaches about 63 percent of its final value. That value of time has various names. It is called the *time constant,* the *ventilation time,* the *residence time,* or the *e-folding time*.

EXAMPLE 17 **Evening Rush Hour Traffic**

Suppose within a square city, 15 km on a side, there are 200,000 cars on the road, each being driven 30 km between 4 PM and 6 PM, and each emitting 3 g/km of CO. It is a clear winter evening with a radiation inversion that restricts mixing to 20 m. The wind is bringing clean air at a steady rate of 1.0 m/s along an edge of the city. Use a box model to estimate the CO concentration at 6 PM if there was no CO in the air at 4 PM, and the only source of CO is cars. Assume that CO is conservative and that there is complete and instantaneous mixing in the box.

Solution The emissions per m^2, q_s, would be

$$q_s = \frac{200{,}000 \text{ cars} \times 30 \text{ km/car} \times 3 \text{ g/km}}{(15 \times 10^3 \text{ m})^2 \times 3{,}600 \text{ s/hr} \times 2 \text{ hr}} = 1.1 \times 10^{-5} \text{ g/s-m}^2$$

Using (61), the concentration after 2 hours (7,200 s) would be

$$C(t) = \frac{q_s L}{uH}(1 - e^{-ut/L})$$

$$C(2\ hr) = \frac{1.1 \times 10^{-5}\ \text{g/s-m}^2 \times 15 \times 10^3\ \text{m}}{1.0\ \text{m/s} \times 20\ \text{m}}\left[1 - \exp\left(\frac{-1.0\ \text{m/s} \times 7{,}200\ \text{s}}{15{,}000\ \text{m}}\right)\right]$$

$$= 3.2 \times 10^{-3}\ \text{g/m}^3 = 3.2\ \text{mg/m}^3$$

which is considerably below both the 1-hr NAAQS for CO of 40 mg/m^3 and the eight-hour standard of 10 mg/m^3. Any CO that was already in the air at 4 PM would, of course, increase this estimate. The time constant, $L/u = 15{,}000\ \text{m}/(1\ \text{m/s}) = 15{,}000\ \text{s} = 4.2$ hours, suggests that in these 2 hours, the concentration is well below what it would become if these conditions were to continue.

12 Indoor Air Quality

This chapter began with the rather startling assertion that, in terms of human exposure to respirable particulate matter, a 2 percent reduction in emissions of *environmental tobacco smoke* (ETS) would be equivalent to eliminating all the coal-fired power plants in the United States (Smith, 1993b). The key, of course, goes back to the theme of Chapter 5 that health effects are the result of exposure, not emissions. We are exposed to chemicals and particulates in the air we breathe, and the air we breathe is mostly inside of buildings where we spend almost all of our time.

Some of the indoor pollutants are the same ones that we are now familiar with from our study of ambient air quality. For example, combustion that takes place inside of homes and other buildings to cook, heat water, and provide space heating and cooling can produce elevated levels of carbon monoxide, nitrogen oxides, hydrocarbons, and other respirable particulates. Cigarette smoke emits carbon monoxide, benzene, acrolein, and other aldehydes, and particulates, as well as about 4,000 other chemicals. Some photocopying machines emit ozone. Building materials such as particleboard, plywood, urea-formaldehyde foam insulation, various adhesives, and other building materials emit formaldehyde. Chipped and peeling paint containing lead becomes airborne toxic dust. A long list of volatile organic compounds are emitted from household cleaning products, paints, carpeting, and a variety of other chemicals we use in our homes.

Some pollutants are somewhat unique to the indoor environment, such as asbestos used for fireproofing and insulation, radon gas that seeps out of the soil and collects in houses, and biological pollutants such as bacteria, molds, animal dander, dust mites, pollen, and viruses. One indoor bacterium, *Legionella,* has caused both Legionnaires' disease and Pontiac fever. Many pollutants, such as cigarette smoke or radon gas, if they are emitted outdoors, have adequate dilution air so people tend not to be exposed to hazardous levels of contamination. Indoors, however, these pollutants can be concentrated, leading all too often to harmful exposure levels.

Air pollution in homes is largely under our control, but in offices and other workplaces, we depend on the builder's choice of materials used in construction and

TABLE 10

Sources and Exposure Guidelines of Indoor Air Contaminants

Pollutant and Indoor Sources	Guidelines, Average Concentrations
Asbestos and other fibrous aerosols	
Friable asbestos; fireproofing, thermal and acoustic insulation, decoration	0.2 fibers/mL for fibers longer than 5 μm
Hard asbestos: vinyl floor and cement products	
Carbon monoxide	
Kerosene and gas space heaters, gas stoves, wood stoves, fireplaces, smoking	10 mg/m^3 for 8 hr, 40 mg/m^3 for 1 hr
Formaldehyde	
Particleboard, paneling, plywood, carpets, ceiling tile, urea-formaldehyde foam insulation, other construction materials	120 μg/m^3
Inhalable particulate matter	
Smoking, vacuuming, wood stoves, fireplaces	55–110 μg/m^3 annual, 150–350 μg/m^3 for 24 hr
Nitrogen dioxide	
Kerosene and gas space heaters, gas stoves	100 μg/m^3 annual
Ozone	
Photocopying machines, electrostatic air cleaners	235 μg/m^3/hr once a year
Radon and radon progeny	
Diffusion from soil, groundwater, building materials	0.01 working levels annual
Sulfur dioxide	
Kerosene space heaters	80 μg/m^3 annual, 365 μg/m^3 24-hr
Volatile organics	
Cooking, smoking, room deodorizers, cleaning sprays, paints, varnishes, solvents, carpets, furniture, draperies	None available

Source: Nagda, Rector, and Koontz, 1987.

the proper design and maintenance of building air filtration and ventilation systems. It is not at all unusual these days for occupants of new and remodeled buildings to experience a bewildering array of symptoms that characterize what is now called *sick-building syndrome*. Sneezing or coughing; watery eyes; headaches; eye, nose, and throat irritation; dry or itchy skin; nausea and dizziness; fatigue; difficulty in concentrating; and general malaise arise without any clearly identifiable causes. Studies suggest that nearly 25 percent of U.S. office workers perceive air quality problems in their work environments, and 20 percent believe their work is impaired by reactions to indoor pollution (Kreiss, 1990). The importance of this field of study has finally been recognized, and a considerable body of research is underway (Ott, Steinemann, and Wallace, 2006).

Table 10 summarizes some of the sources and exposure guidelines for a number of pollutants that are commonly found in indoor environments.

Environmental Tobacco Smoke (ETS)

One category of indoor air pollution stands out from all others, and that is tobacco smoke. It is estimated that about 26 percent of the U.S. adult population are smokers who consume more than 500 billion cigarettes annually. Tobacco smoke contains more than 4,000 chemicals, including more than 40 that are known to cause cancer in humans or animals, and many are strong respiratory irritants. Smoking is thought to be responsible for between 80 and 90 percent of the lung cancer deaths in the United States. The combined death rate from lung cancer, emphyzema, and cardiovascular disease attributable to smoking is over one-half million per year. Unfortunately, it is not only smokers who suffer the ill effects of tobacco.

First, let's consider some terminology. Smokers inhale what is referred to as *mainstream* smoke. *Sidestream* smoke, emitted from smoldering cigarettes, mixed with smoke exhaled by smokers is known as *environmental tobacco smoke* (ETS) or *secondhand smoke*. Breathing air with ETS is called *involuntary* or *passive smoking*.

In early 1993, the EPA released a report (*Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders*) that concluded that ETS causes lung cancer in adult nonsmokers and has serious effects on the respiratory system of children. ETS has been listed as a Group A, known human carcinogen. The EPA concludes that secondhand smoke is causing an estimated 3,000 lung cancer deaths annually in the United States as well as 150,000 to 300,000 cases of pneumonia or bronchitis among children under 18 months of age each year. It worsens the condition of up to 1 million asthmatic children. Smokers have nearly 10 times as much of the carcinogen, benzene, in their blood as nonsmokers, and a pregnant smoker undoubtedly passes benzene to her developing fetus. That correlates with the fact that children of smokers die of leukemia at several times the rate of children of nonsmokers. An even more recent study concludes that the incidence rate of sudden infant death syndrome increases as exposure to tobacco smoke increases.

Asbestos

Another somewhat special indoor air-quality problem can be caused by asbestos-containing materials. Asbestos used to be a common building material found in structural fireproofing, heating-system insulation, floor and ceiling tiles, and roofing felts and shingles. It has also been used in consumer products such as fireplace gloves, ironing board covers, and certain hair dryers.

As some asbestos-containing materials age, or if they are physically damaged in some way during their use, microscopic fibers may be dispersed into the indoor air environment. Inhalation of these fibers can lead to a number of life-threatening diseases, including asbestosis, lung cancer, and mesothelioma. Asbestos fibers have long been known to be human carcinogens; in fact, asbestos was one of the first substances categorized as a hazardous air pollutant under Section 112 of the Clean Air Act. It was identified by the EPA as a hazardous pollutant in 1971, and in 1973, the Asbestos National Emissions Standards for Hazardous Air Pollutants (NESHAP) were initiated. An especially deadly combination is asbestos exposure along with tobacco smoking, which elevates the lung cancer risk by approximately fivefold.

There are two broad categories of asbestos-containing materials: friable and nonfriable. Friable materials are those that can be crumbled, pulverized, or reduced to powder by hand pressure. Obviously, these are the most dangerous since asbestos fibers are easily released into the environment. Nonfriable materials, such as vinyl asbestos floor tiles and asphalt roofing products, do not usually require special handling even during demolition and renovation of buildings. Friable substances, on the other hand, are subject to NESHAP during such work. Since disturbing asbestos that is in good condition may create more of a risk than simply sealing it in place, it is often best not to attempt to remove asbestos-containing materials in a building unless there is known to be a problem or major construction work is to ensue.

Radon

One of the most publicized indoor air pollution problems is caused by the accumulation of radon gas in some homes in some parts of the country. Radon gas and its radioactive daughters are known carcinogens and may be the second leading cause of lung cancer, after smoking.

Radon is a radioactive gas that is part of a natural decay chain beginning with uranium and ending with lead. A simplified description of the sequence, along with half-lives and alpha, gamma, and beta radiation emitted, is shown in Figure 56. Radon gas formed in pore spaces between mineral grains in soil can work its way to the surface where it can enter buildings through the floor. Radon itself is inert, but its short-lived decay products—polonium, lead, and bismuth—are chemically active and easily become attached to inhaled particles that can lodge in the lungs. In fact, it is the alpha-emitting polonium, formed as radon decays, that causes the greatest lung damage.

Radon can also be emitted from some earth-derived building materials such as brick, concrete, and tiles. It can be captured in groundwater, to be released when

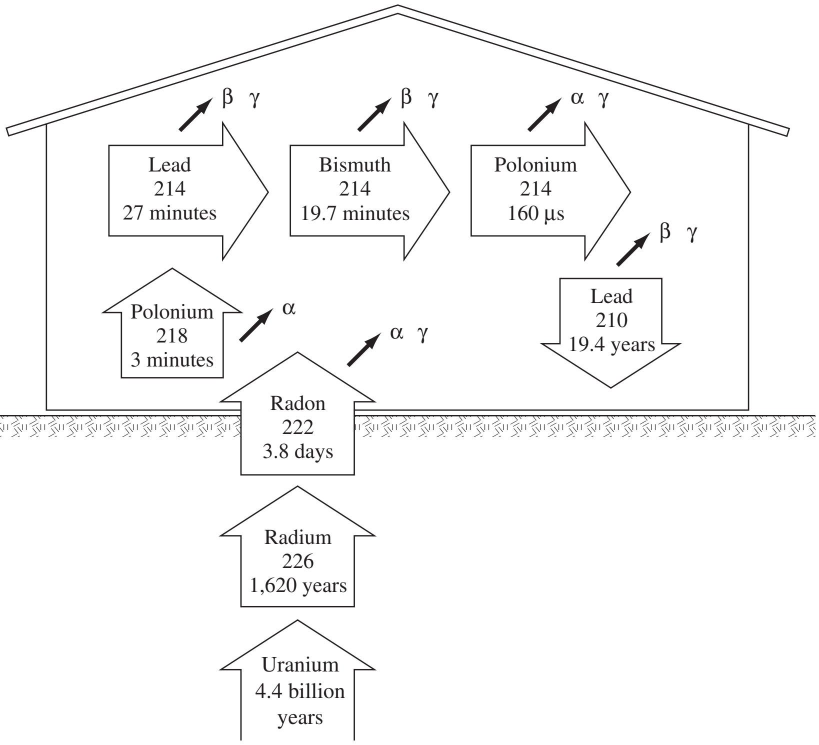


FIGURE 56 Simplified uranium decay series, with half-lives and emissions. Radon gas that seeps out of soils can decay inside of buildings.

TABLE 11

Radon Measurement Units

Curie	1 Ci	$= 3.7 \times 10^{10}$ radioactive decays per second (1 g radium)
Picocurie	1 pCi	= 2.2 radioactive decays per minute
Becquerel	1 Bq	= 1 radioactive decay per second
Picocurie per liter	1 pCi/L	= 37 Bq/m^3
EPA criterion	4 pCi/L	= 150 Bq/m^3 (0.0000028 parts per trillion of radon)
Working level[a]	1 WL	≈ 200 pCi/L = 7400 Bq/m^3 of radon
Working-level month	1 WLM	= 1 working-level of exposure for 173 hours

[a] 1 WL is defined as 100 pCi/L of radon in equilibrium with its progeny.

that water is aerated, for example, during showers (the radon risk in water is from inhalation of the released gas, not from drinking the water itself). Radon has also been detected in some natural gas supplies, so modest amounts may be released during cooking. By far, the most important source of radon, however, is through soils under homes.

The units of measurement used in describing radon are somewhat unusual and are summarized in Table 11. The decay rate of 1 gram of radium, 3.7×10^{10} disintegrations per second, is the origin of the *Curie*. A more intuitive unit is the *Becquerel,* which corresponds to 1 radioactive decay per second. The *working-level-month* (WLM) is based on the exposure a miner received working 173 hours per month. Indoor concentrations of radon have traditionally been expressed in *picocuries per liter* of air (pCi/L). It is estimated that the average house in the United States has 1.25 pCi/L, and around 6 million homes (6 percent of the total) have levels above 4 pCi/L, which is the EPA's level of concern. Retrofitting those 6 million U.S. homes with radon control measures could cost on the order of $50 billion.

Early estimates of the lung cancer risk associated with inhalation of radon and its progeny were based in large part on epidemiologic studies of uranium miners who had been exposed to high radon levels and for whom the incidence rate of lung cancer was much higher than that of the general population. Translating data from miners, however, to risks in residences introduces large uncertainties. Working miners, for example, would be expected to have higher breathing rates than people resting at home, they would be exposed to much greater concentrations of particulate matter along with the radon, and they have much higher exposure to radon.

A number of recent studies have found evidence of elevated lung cancer rates at the comparatively low radon levels found in homes. A 1994 Swedish report showed progressively higher lung cancer rates as home radon exposure increased. At exposure levels of 3.8 to 10.8 picocuries per liter (pCi/L), lung cancer risks increased 30 percent over those who lived in homes with less than 1.4 pCi/L. Above 10.8 pCi/L, the incidence rate was 80 percent higher. The National Academy of Science's Biological Effects of Ionizing Radiation (BEIR) VI Report in 1999 concluded that radon causes between 15,000 and 22,000 lung-cancer deaths per year in the United States, with most of those being among smokers (NRC, 1999). Two confirming studies in the United States and in Europe in 2005 led the U.S. Surgeon General to issue an indoor radon national health advisory.

Much of the danger associated with radon is caused by radon daughters that attach themselves to small particulate matter that is inhaled deep into the lungs. The

TABLE 12

Estimated Annual Lung Cancer Deaths in the United States Attributable to Radon Exposure, 1986 Data

Smoking History	Population (Millions)	Lung Cancer Deaths All Causes	 Radon-Attributable
Never smoked	145	5,000	500
Former smoker	43	57,000	6,400
Current, light smoker[a]	38	37,600	4,500
Current, heavy smoker	14	30,800	4,200
Total	241	130,400	15,700

[a] Fewer than 25 cigarettes per day.
Source: Nazaroff and Teichman, 1990.

combination of cigarette smoke and radon is, as a result, an especially deadly combination. According to data summarized in Table 12, only about 3 percent of the radon mortality rate is estimated to occur among individuals who have never smoked. If these data are true in general, well over 90 percent of the lung-cancer risk associated with radon could be controlled by eliminating smoking without any change in radon concentrations (Nazaroff and Teichman, 1990).

Homes built over radon-rich, high-permeability soils are most vulnerable to elevated radon levels. The emanation rate from soil seems to range from about 0.1 pCi/m^2-s to over 100 pCi/m^2-s, with a value of 1 pCi/m^2-s being fairly typical. This large variation in emission rates is the principal cause of variations in indoor radon concentrations measured across the country. When emission rates are low, as is the case for most of the United States, homes can be tightly constructed for energy efficiency without concern for elevated levels of radon. In areas where soil radon emissions are high, tight building construction techniques can elevate indoor radon concentrations. Smokers, of course, would be most affected.

The EPA, the U.S. Geological Survey, and Lawrence Berkeley National Labs have attempted to evaluate the radon potential across the United States, which has led to the development of radon maps such as the one shown in Figure 57. Such maps help identify regions in which testing of existing homes for radon, as well as recommendations for preventative radon control measures in new construction, are most highly recommended. The usual threshold for action is 4 pCi/L. Since significant variations in radon are often found between one neighborhood and another within the same city, decisions about remediation measures are best made after thorough testing of individual homes.

Available techniques to help reduce indoor radon concentrations depend somewhat on what type of floor construction has been employed. Many houses in the United States are built over basements that may or may not be heated; other homes, especially in the west United States, are built over crawl spaces or on concrete slabs. Occupied basements present the greatest risk since the wall and floor areas exposed to the soil are large, while the usual ventilation rate in basements is low. With every type of construction, mitigation begins with efforts to seal any and all cracks and openings between the floor and the soil beneath. For houses with basements or crawl spaces, increased ventilation of those areas, either by natural or mechanical means, can be very effective. If the basement is heated, a heat recovery

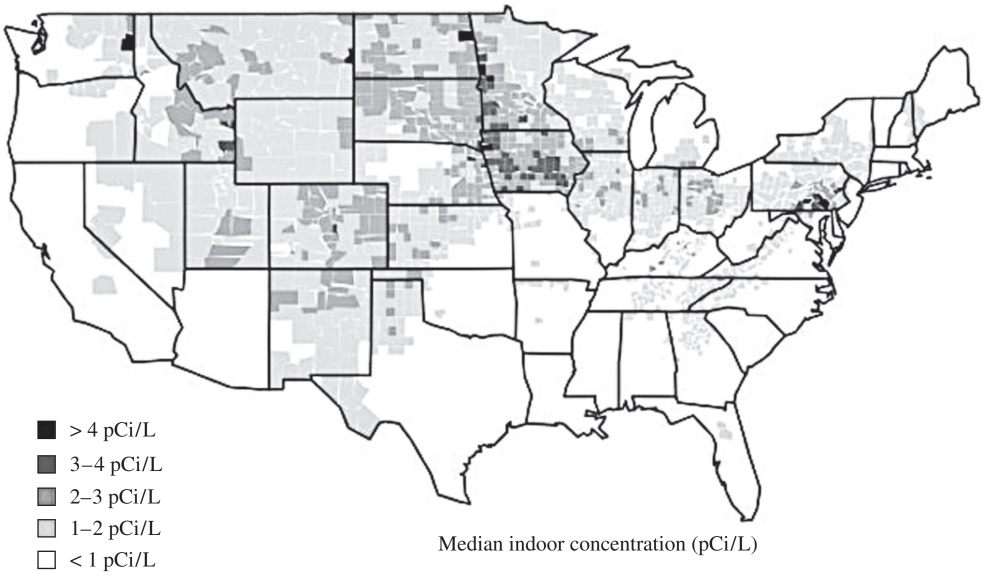


FIGURE 57 Median, long-term, living-area concentrations of radon for countries in the contiguous United States.
(*Source:* http://www.stat.columbia.edu/radon/.)

ventilator may be called for to increase ventilation without wasting excessive amounts of energy. One of the most widely used radon reduction techniques uses subslab suction as shown in Figure 58. Radon exhaust pipes pass through the floor into the aggregate below. Fans suck the radon gas out of the aggregate and exhaust it into the air above the house.

Exposure Assessment

Toxic chemicals affect human health when people are exposed to them. Perhaps it is because this simple notion is so obvious that its importance is so often overlooked. To be exposed to pollutants, they have to be in the air we breathe, the water we drink, or the food we eat, so it makes sense to approach health risk assessment by evaluating those pathways. Such an approach, called the *total exposure assessment methodology* (TEAM), was launched by the EPA's Office of Research and Development in 1979, and considerable data on actual exposures have now been collected (Ott, Steinemann, and Wallace, 2006).

From the perspective of air pollution, the starting point for a total environmental assessment is to determine the amount of time spent indoors, in vehicles, and outdoors. Large-scale field studies based on carefully kept diaries suggests that Americans are inside of buildings 89 percent of the time, in vehicles 6 percent, and are outdoors 74 minutes per day, or 5 percent of the time (Figure 59). Perhaps surprisingly, Californians, who might think they are outdoors much more than national averages would suggest, are only slightly different, 6 percent outdoors rather than 5 (Jenkins, 1992). Also shown in Figure 59 are estimates of the time spent indoors in less developed countries (LDCs). As countries develop, a greater fraction of their populations live in urban areas and spend more time indoors.

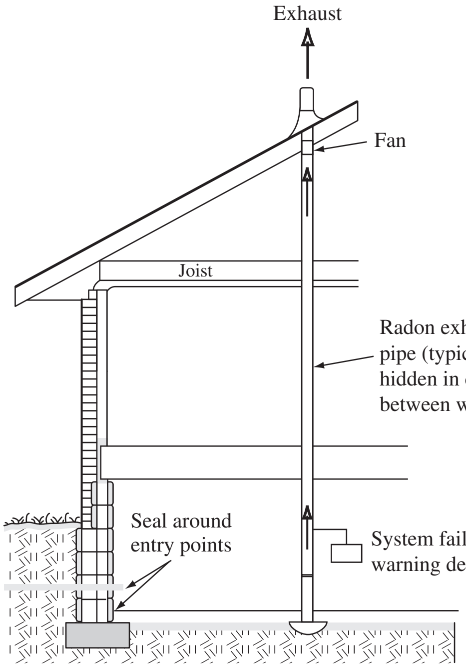
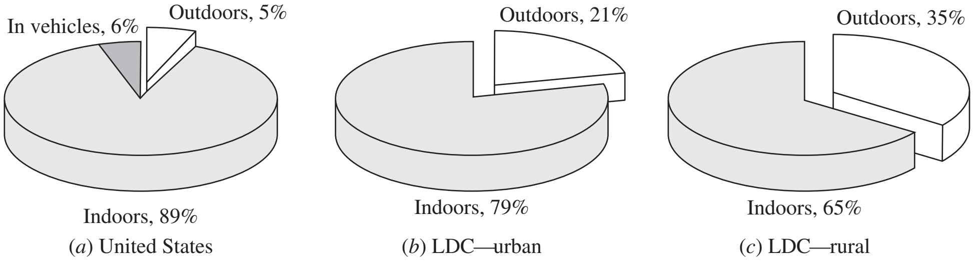


FIGURE 58 A subslab suction radon mitigation system.
(*Source:* U.S. EPA, 1993.)



(*a*) United States (*b*) LDC—urban (*c*) LDC—rural

FIGURE 59 (*a*) In the United States, 89 percent of our time is spent indoors (Robinson, Thomas, and Behan, 1991). Percentage of time spent indoors and outdoors in less developed countries: (*b*) urban and (*c*) rural.
(*Source:* Smith, 1993a.)

The importance of the amount of time spent indoors is dramatized by studies of human exposure to a number of especially important air pollutants. These TEAM studies depend on actual exposure measurements made with personal air quality monitors that accompany participants during their normal daily activities. Exposure measurements are supplemented with blood gas analyses to verify the actual uptake of pollutants. One such study (Smith, 1993b) estimates that per capita emissions of particulates from coal-fired power plants in the United States are about 1.6 kg/person, whereas emissions of environmental tobacco smoke (ETS) are roughly 0.050 kg/person. That is, coal plant particulate emissions are about 30 times higher than ETS emissions. But coal plant emissions are released into a large reservoir, the

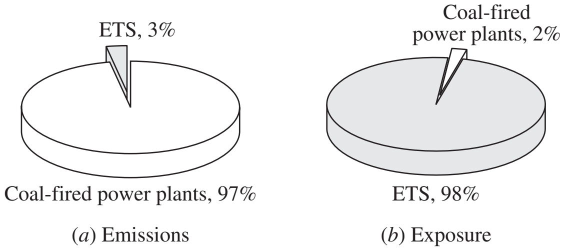


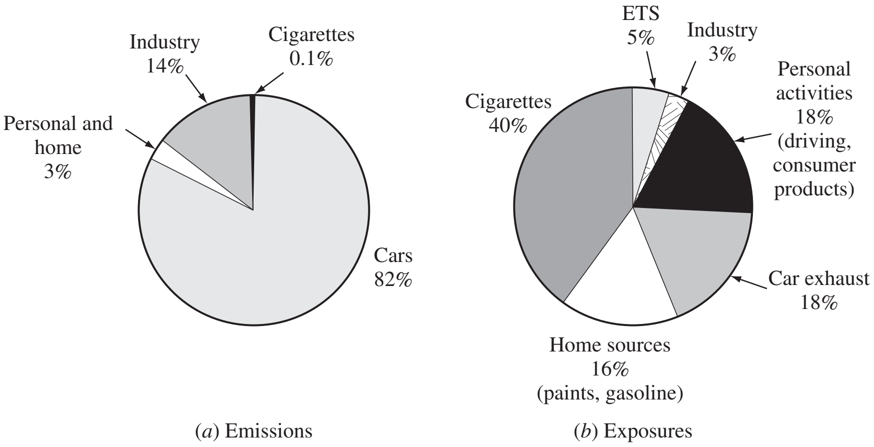
FIGURE 60 Particulate emissions and exposure for environmental tobacco smoke (ETS) and U.S. coal-fired power plants. Of the total, (*a*) ETS is 3 percent of emissions, but (*b*) 98 percent of exposure.
(*Source:* Based on Smith, 1993b.)

atmosphere, whereas almost all ETS particulates are released indoors, where there is less dilution and where people spend most of their time. In terms of particulate matter inhaled, it has been estimated that human exposure to ETS is 60 times the exposure from coal-fired power plants even though ETS constitutes only 3 percent of emissions. These data, summarized in Figure 60, suggest that major reductions in coal plant emissions would have far lower overall effect on public health than what could be realized with modest reductions in ETS.

Similar conclusions can be reached for the Class-A human carcinogen, benzene. Combustion of gasoline accounts for 82 percent of benzene emissions in the United States, whereas industrial sources account for most of the rest, 14 percent. Cigarette smoke is responsible for only 0.1 percent of all benzene emissions, but on the average, 40 percent of all benzene exposure is direct inhalation by smokers, and 5 percent is exposure to ETS. In comparison, all industrial sources combined are responsible for less benzene exposure (3 percent) than passive smoking (5 percent). Figure 61 shows these data (Wallace, 1995).

Even in communities located next to industrial facilities, indoor exposure dominates. One such study focused on the large petroleum storage and loading facility at the Alyeska Marine Terminal located just 3 miles from a residential community in Valdez, Alaska. The Marine Terminal accounted for 11 percent of benzene exposure, whereas personal activities such as smoking, driving, consumer products used in homes, and gasoline vapors in attached garages accounted for the remaining 89 percent (Yokum, Murray, and Mikkelsen, 1991). Another major study in northern New Jersey, which has one of the greatest concentrations of oil refineries and chemical plants in the nation, measured exposure to 25 different VOCs. No difference in exposure was seen between residents near the plants and farther away (Wallace, 1995).

Cigarettes emit not only benzene but a host of other toxic chemicals including styrene, toluene, xylenes, ethylbenzene, formaldehyde, and 1,3-butadiene. Newly dry-cleaned clothes emit tetrachloroethylene for days after putting them into your closet. Moth repellents and spray room air fresheners emit p-dichlorobenzene, which causes cancer in rats and mice. Hot showers, clothes, and dishwashing are major sources of airborne chloroform, another carcinogen. Cooking adds polynuclear aromatic hydrocarbons (PAHs) to indoor air. Pesticides stored in garages and



(*a*) Emissions (*b*) Exposures

FIGURE 61 (*a*) Benzene emissions are mainly due to car exhaust and to industry, while all cigarettes smoked in the United States provide only 0.1 percent. (*b*) Benzene exposures are overhwhelmingly due to cigarettes, mostly from active smoking. ETS contributes more exposure than all industrial sources combined.
(*Source:* Wallace, 1995.)

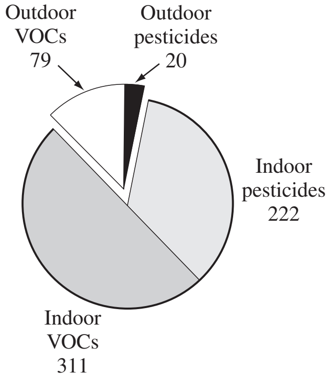


FIGURE 62 TEAM studies of 33 carcinogens commonly found in homes show an upper-bound cancer risk of about 6×10^{-4}. This is similar in magnitude to risk from radon and passive smoking. Numbers are lifetime cancer risks per million people.
(*Source:* Wallace, 1995.)

homes are major sources of many other carcinogens found indoors. Pesticides that are used outdoors, and lead particles that settled onto the soil, are tracked into homes on dirt from shoes. All told, TEAM studies of human exposure to 33 VOCs and pesticides in the United States show that indoor exposures account for about 85 percent of total cancer risk from these chemicals (Figure 62).

Clearly, if we want to reduce the hazards of air pollution, the place to start is at home. Simple things like using a floor mat, or removing shoes at the entrance, can reduce the amount of outside pollution that is brought into a home on our shoes.

Dry-cleaning clothes less often, or at least airing them out for a day or two before putting them in your closet will help. Exhaust fans over ranges can reduce particulate and PAH exposure, and bathroom fans can help clear out chloroform from chlorinated shower water. Be sure carbon monoxide and other combustion products formed in water heaters and space heaters are properly vented, and certainly don't try to heat a house with a charcoal grill or an unvented kerosene heater. Be sure paints, cleaning fluids, and other sources of VOCs are sealed properly before storage, and provide plenty of ventilation when you use them. And, if you smoke, do so outside if at all possible. Simple, no-cost measures such as these can significantly reduce your exposure to toxic air pollutants.

In the less-developed countries of the world, indoor exposure to hazardous air pollutants is much more severe. Traditional methods of cooking often involve "dirty" fuels such as coal, wood, animal dung, charcoal, and kerosene, which are often burned in homes or cooking huts without chimneys or proper ventilation. Women and their infant children are often exposed to extremely high concentrations of particulates and other products of combustion for lengthy periods of time. One result is elevated levels of acute respiratory infections (ARI) in children that a number of studies correlate with indoor exposure to smoke. Acute respiratory infections, such as pneumonia, are the main cause of childhood deaths worldwide, killing approximately 4.3 million children per year. That is 30 percent more deaths than are caused by the number two killer, diarrhea. A number of studies show two to six times the incidence rate of ARI among children living in homes where cooking is done on an open biomass stove, compared to those using more modern fuels, such as LPG, with chimneys. Similar elevated incidence rates for lung cancer have been observed in urban homes in China that burn coal for cooking or heating compared with homes that use natural gas (Smith, 1993a). The enormity of the indoor air quality problem in developing countries is truly staggering.

Infiltration, Ventilation, and Air Quality

Just as with outdoor air, the amount of air available to dilute pollutants is an important indicator of likely contaminant concentrations. Indoor air can be exchanged with outdoor air by any combination of three mechanisms: infiltration, natural ventilation, and forced ventilation. *Infiltration* is the term used to describe the natural air exchange that occurs between a building and its environment when doors and windows are closed. That is, it is leakage that occurs through various cracks and holes that exist in the building envelope. *Natural ventilation* is the air exchange that occurs when windows or doors are purposely opened to increase air circulation, whereas *forced ventilation* occurs when mechanical air-handling systems induce air exchange using fans or blowers.

Large amounts of energy are lost when conditioned air (heated or cooled) that leaks out of buildings is replaced by outside air that must be mechanically heated or cooled to maintain desired interior temperatures. It is not uncommon, for example, for one-third of a home's space heating and cooling energy requirements to be caused by unwanted leakage of air, that is, infiltration. Nationwide, about 10 percent of total U.S. energy consumption is accounted for by infiltration, which translates into tens of billions of dollars worth of wasted energy each year. Since infiltration is quite easily, and cheaply, controlled, it is not surprising that tightening buildings has become a popular way to help save energy. Unfortunately, in the process of saving energy, we may exacerbate indoor air quality problems unless we simultaneously reduce the sources of pollution.

Air leaks in and out of buildings through numerous cracks and openings in the building envelope. The obvious cracks around windows and doors are the usual ones we try to plug with caulk and weatherstripping, but there are many less obvious, but potentially more important, leakage areas such as those created when plumbing, ducts, and electrical wiring penetrate walls, ceilings, and floors; fireplaces without dampers; ceiling holes created around recessed light fixtures, attic access hatches, and any other bypasses created between heated spaces and the attic; gaps where foundations are connected to walls; exhaust vents in bathrooms and kitchens; and, perhaps most importantly, leaky ductwork in homes with forced-air heating systems. Figure 63 shows some of these infiltration sites.

Infiltration is driven by pressure differences between the inside of the building and the outdoor air. These pressure differences can be caused by wind or by inside-to-outside temperature differences. Wind blowing against a building creates higher pressure on one side of the building than the other, inducing infiltration through cracks and other openings in the walls. Temperature induced infiltration (usually referred to as the *stack effect*) is influenced less by holes in the walls than by various openings in the floors and ceilings. In the winter, warm air in a building wants to rise, exiting through breaks in the ceiling and drawing in colder air through floor openings. Thus, infiltration rates are influenced not only by how fast the wind is blowing and how great the temperature difference is between inside and outside but

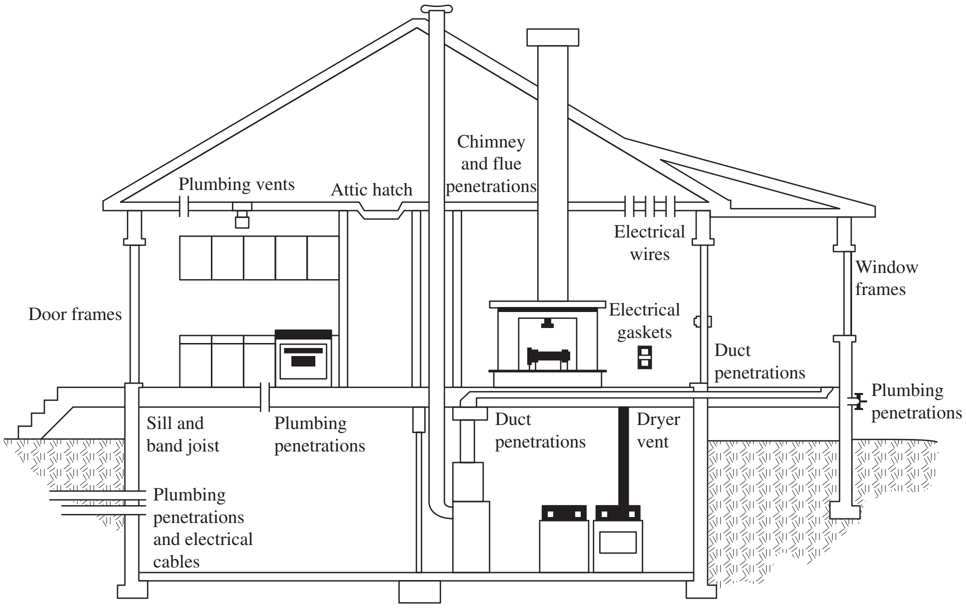
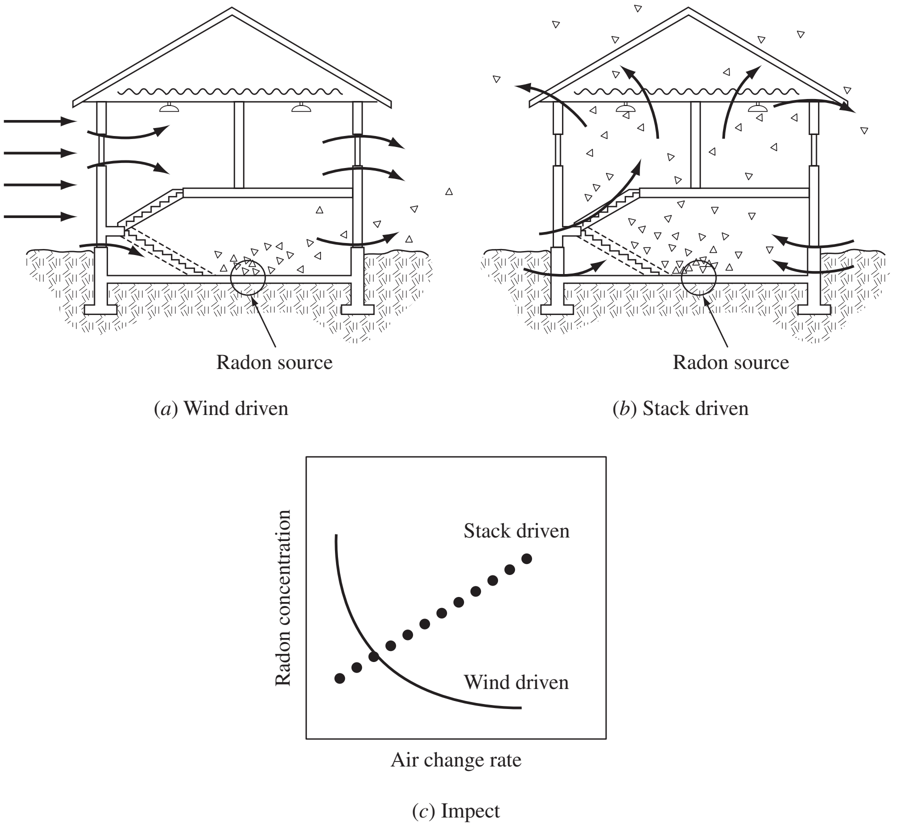


FIGURE 63 Infiltration sites in homes.
(*Source:* National Association of Home Builders Web site at ***http://www.nahbrc.org/greenguidelines/userguide_energy_prescriptive.html***).

also by the locations of the leaks in the building envelope. Greater leakage areas in the floor and ceiling encourage stack-driven infiltration, whereas leakage areas in vertical surfaces encourage wind-driven infiltration.

Moreover, while it is usually assumed that increasing the infiltration rate will enhance indoor air quality, that may not be the case in one important circumstance: for radon that is emitted from the soil under a building. For radon, wind-driven infiltration helps reduce indoor concentrations by allowing radon-free fresh air to blow into the building. Stack-driven infiltration, which draws air through the floor, may actually encourage new radon to enter the building, negating the cleaning that infiltration usually causes. Figure 64 illustrates these important differences.

Infiltration rates may be expressed in units such as m^3/hr or cubic feet per minute (cfm), but more often the units are given in air changes per hour (ach). The air exchange rate in ach is simply the number of times per hour that a volume of air equal to the volume of space in the house is exchanged with outside air. Typical average infiltration rates in American homes range from about 0.5 ach to 1 ach, with newer houses being more likely to have rates at the lower end of the scale while older homes are closer to the top end. Some very poorly built houses have rates as high as



(*a*) Wind driven

(*b*) Stack driven

(*c*) Impect

FIGURE 64 Wind-driven infiltration helps reduce radon, whereas stack-driven infiltration may cause it to increase.
(*Source:* Reece, 1988.)

3–4 ach. Carefully constructed new homes today can quite easily be built to achieve infiltration rates that are as low as 0.1 ach by using continuous plastic sheet "vapor barriers" in the walls, along with careful application of foam sealants and caulks to seal cracks and holes. At such low infiltration rates, moisture and pollutant build-up can be serious enough to require extra ventilation, and the trick is to get that ventilation without throwing away the heat that the outgoing stale air contains.

One way to get extra ventilation with minimal heat loss is with a mechanical *heat-recovery ventilator* (HRV), in which the warm, outgoing stale air transfers much of its heat to the cold, fresh air being drawn into the house. Another, simpler and cheaper approach is to provide mechanical ventilation systems that can be used intermittently in the immediate vicinity of concentrated sources of pollutants. Exhaust fans in bathrooms and range hoods over gas stoves, for example, can greatly reduce indoor pollution, and by using them only as necessary, heat losses can be modest.

An Indoor Air Quality Model

It is quite straightforward to apply the box model concepts developed earlier to the problem of indoor air quality. The simple model we will use treats the building as a single, well-mixed box, with sources and sinks for the pollutants in question. If necessary, the simple model can be expanded to include several boxes, each characterized by uniform pollutant concentrations. A two-box model, for example, is sometimes used for radon estimates, where one box is used to model radon concentrations within the living space of a dwelling, and the other models the air space beneath the house.

Consider the simple, one-box model of a building shown in Figure 65. There are sources of pollution within the building that can be characterized by various emission rates. In addition, ambient air entering the building may bring new sources of pollution, which adds to whatever may be generated inside. Those pollutants may be removed from the building by infiltration or ventilation, or they may be nonconservative and decay with time. In addition, if there is a mechanical air-cleaning system, some pollutants may be removed as indoor air is passed through the cleaning system and returned. To help keep the model simple, we will ignore such mechanical filtration.

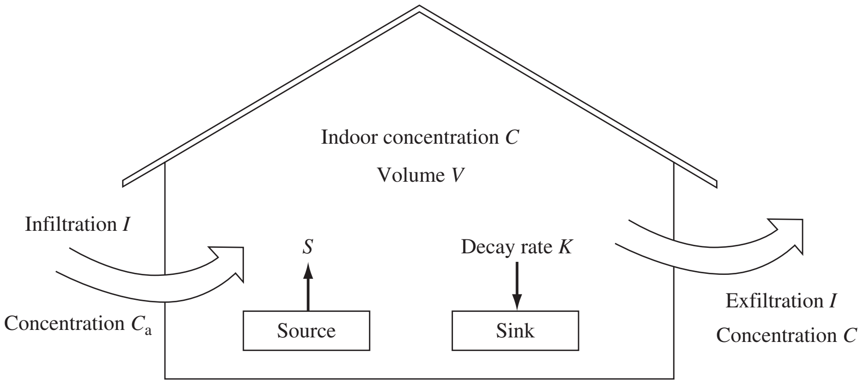


FIGURE 65 Box model for indoor air pollution.

TABLE 13

Some Measured Pollutant Emission Rates for Various Sources

Source	Pollutant Emission Rate, (mg/hr) CO	NO_x[a]	SO_2	HCHO
Gas range				
Oven	1900	52	0.9	23
One top burner	1840	83	1.5	16
Kerosene heater[b]				
Convective	71	122	—	1.1
Radiant	590	15	—	4.0
One cigarette				
(sidestream smoke)[c](mg)	86	0.05	—	1.44

[a] NO_x reported as N.
[b] New portable heaters, warm-up emissions not included, SO_2 not measured.
[c] *Source:* National Research Council, 1981.
Sources: Traynor et al., 1981, 1982.

TABLE 14

Decay Constants or Reactivities, *K*

Pollutant	*K* (1/hr)
CO	0.0
NO	0.0
NO_x (as N)	0.15
HCHO	0.4
SO_2	0.23
Particles ($<0.5\ \mu$m)	0.48
Radon	7.6×10^{-3}

Source: Traynor et al., 1981.

A basic mass balance for pollution in the building, assuming well-mixed conditions, is

$$\begin{pmatrix}\text{rate of increase}\\ \text{in the box}\end{pmatrix} = \begin{pmatrix}\text{rate of pollution}\\ \text{entering the box}\end{pmatrix} - \begin{pmatrix}\text{rate of pollution}\\ \text{leaving the box}\end{pmatrix} - \begin{pmatrix}\text{rate of decay}\\ \text{in the box}\end{pmatrix}$$

$$V\frac{dC}{dt} = (S + C_anV) - CnV - KCV \qquad (62)$$

where

V = volume of conditioned space in building (m^3/air change)
n = number of air changes per hour (ach)
S = source emission rate (mg/hr)
C = indoor concentration (mg/m^3)
C_a = ambient concentration (mg/m^3)
K = pollutant decay rate or reactivity (1/hr)

An easy way to find the steady-state solution is to set $dC/dt = 0$ to yield

$$C(\infty) = \frac{(S/V) + C_a n}{n + K} \tag{63}$$

A general solution is

$$C(t) = \left[\frac{(S/V) + C_a n}{n + K}\right]\left[1 - e^{-(n+K)t}\right] + C(0)e^{-(n+K)t} \tag{64}$$

where

$C(0)$ = the initial concentration in the building

Some of the pollutants that we might want to model, such as CO and NO, can be treated as if they were conservative—that is, they do not decay with time or have significant reactivities, and thus $K = 0$. Equation (64) simplifies considerably for the basic condition where there is no ambient concentration of the pollutant ($C_a = 0$); there is no initial concentration of pollution in the building [$C(0) = 0$]; and the pollutant is conservative ($K = 0$). Under these circumstances,

$$C(t) = \left(\frac{S}{nV}\right)(1 - e^{-nt}) \tag{65}$$

Examples of emission rates for various sources are given in Table 13, and estimates of appropriate decay rates for common indoor pollutants are given in Table 14.

EXAMPLE 18 A Portable Kerosene Heater

An unvented, portable, radiant heater, fueled with kerosene is tested under controlled laboratory conditions. After running the heater for two hours in a test chamber with a 46.0 m^3 volume and an infiltration rate of 0.25 ach, the concentration of carbon monoxide (CO) reaches 20 ppm. Initial CO in the lab is 0, and the ambient CO level is negligible throughout the run. Treating CO as a conservative pollutant, find the rate at which the heater emits CO. If the heater were to be used in a small home to heat 120 m^3 of space having 0.4 ach, predict the steady-state concentration.

Solution Begin by converting the volumetric concentration of 20 ppm to mg/m^3 at an assumed 1 atm of pressure and 25°C:

$$C(\text{mg/m}^3) = \frac{\text{ppm} \times \text{mol wt}}{24.465} = \frac{20 \times (12 + 16)}{24.465} = 22.9 \text{ mg/m}^3$$

Rearranging (65) to solve for the heater's emission rate S gives

$$S = \frac{nVC(t)}{(1 - e^{-nt})} = \frac{0.25 \text{ ac/hr} \times 46.0 \text{ m}^3/\text{ac} \times 22.9 \text{ mg/m}^3}{[1 - e^{-0.25/\text{hr} \times 2\text{hr}}]} = 669 \text{ mg/hr}$$

In the small home, the steady-state concentration using (63) would be

$$C(\infty) = \frac{S}{nV} = \frac{669 \text{ mg/hr}}{0.4 \text{ ac/hr} \times 120 \text{ m}^3/\text{ac}} = 13.9 \text{ mg/m}^3 \text{ (12.1 ppm)}$$

which exceeds the eight-hr ambient standard of 9 ppm for CO.

Portable kerosene heaters, such as the one introduced in Example 18, became quite popular soon after the energy crises of the 1970s, and they continue to be used indoors, especially in developing countries, despite warning labels. There are two types of kerosene heaters: radiant heaters, which heat objects in a direct line-of-sight by radiation heat transfer, and convective heaters that warm up the air, which, in turn, heats objects in the room by convection. Radiant heaters operate at lower temperatures so their NO_x emissions are reduced, but their CO emissions are higher than the hotter running convective types. Both types of kerosene heaters have high enough emission rates to cause concern if used in enclosed spaces such as a home or trailer.

The following example shows how we can estimate the radon concentration in a home using this simple indoor air pollution model.

EXAMPLE 19 **Indoor Radon Concentration**

Suppose the soil under a single-story house emits 1.0 pCi/m^2-s of radon gas. As a worst case, assume that all of this gas finds its way through the floor and into the house. The house has 250 m^2 of floor space, an average ceiling height of 2.6 m, and an air change rate of 0.9 ach. Estimate the steady-state concentration of radon in the house, assuming the ambient concentration is negligible.

Solution We can find the steady-state concentration using (62) with $C_a = 0$. The decay rate of radon is given in Table 14 as $K = 7.6 \times 10^{-3}$/hr, which we could have calculated using (3.8) along with radon's 3.8-day half-life:

$$C(\infty) = \frac{(S/V)}{n + K}$$

$$= \frac{\left(\dfrac{1\ \text{pCi/m}^2\text{s} \times 3{,}600\ \text{s/hr} \times 250\ \text{m}^2}{250\ \text{m}^2 \times 2.6\ \text{m}}\right)}{0.9/\text{hr} + 7.6 \times 10^{-3}/\text{hr}}$$

$$= 1.5 \times 10^3\ \text{pCi/m}^3 = 1.5\ \text{pCi/L}$$

which is fairly typical for homes in the United States.

PROBLEMS

1 Convert the following (eight-hour) indoor air quality standards established by the U.S. Occupational Safety and Health Administration (OSHA) from ppm to mg/m^3 (at 25°C and 1 atm), or vice versa.

(a) Carbon dioxide (CO_2), 5,000 ppm

(b) Formaldehyde (HCHO), 3.6 mg/m^3

(c) Nitric oxide (NO), 25 ppm

2 Consider a new 38 percent-efficient 600-MW power plant burning 9,000 Btu/lb coal containing 1 percent sulfur. If a 70 percent-efficient scrubber is used, what would be the emission rate of sulfur (lb/hr)?

3 In the power plant in Problem 2, if all of the sulfur oxidizes to SO_2, how many pounds per hour of SO_2 would be released if the plant is equipped with a 90 percent-efficient scrubber? How many pounds of SO_2 per kilowatt-hour of electricity generated would be released?

4 A new coal-fired power plant has been built using a sulfur emission control system that is 70 percent efficient. If all of the sulfur oxidizes to SO_2 and if the emissions of SO_2 are limited to 0.6 lb SO_2 per million Btu of heat into the power plant, what maximum percent sulfur content can the fuel have?

(a) If 15,000 Btu/lb coal is burned?

(b) If 9,000 Btu/lb coal is burned?

5 Compliance coal releases no more than 1.2 lb of SO_2 per 10^6 Btu of heat released, without controls. What maximum percentage sulfur could 12,000 Btu/lb compliance coal contain if all of the sulfur oxidizes to SO_2 during combustion?

6 What AQI descriptor (Good, Moderate, etc.) should be reported for air quality on the following days?

Pollutant	Day 1	Day 2	Day 3
O_3, 1-hr (ppm)	0.15	0.22	0.12
CO, 8-hr (ppm)	12	15	8
$PM_{2.5}$, 24-hr ($\mu g/m^3$)	130	150	10
PM_{10}, 24-hr ($\mu g/m^3$)	180	300	100
SO_2, 24-hr (ppm)	0.12	0.20	0.05
NO_2, 1-hr (ppm)	0.4	0.7	0.1

7 The OSHA standard for worker exposure to 8 hours of CO is 50 ppm. What percentage COHb would result from this exposure?

8 An indoor "tractor pull" competition resulted in a CO concentration of 436 ppm. What percent COHb would result for a spectator who is exposed to one hour at that level? How long would it take to reach 10 percent COHb, which is a level at which most people will experience dizziness and headache?

9 What hydrocarbon, RH, reacting with the $OH^\bullet$ radical in (16), would produce formaldehyde, HCHO, in (19)?

10 Suppose propene, $CH_2 = CH—CH_3$, is the hydrocarbon (RH) that reacts with the hydroxyl radical $OH^\bullet$ in reaction (16). Write the set of chemical reactions that end up with an aldehyde. What is the final aldehyde?

11 Find the settling velocity of a 20-μm diameter particle with density 1,500 kg/m^3. If this particle had been hurled to a height of 8,000 m during a volcanic eruption, estimate the time required to reach the ground (sea level). You many assume that the viscosity of air does not change enough to worry about. If winds average 10 m/s, how far away would it blow on its way down?

12 Find the residence time in the atmosphere for a 10-μm particle with unit density (i.e., the density of water, 10^6 g/m^3) at 1,000 m elevation.

13 Equation (24) is valid for Reynolds numbers much less than 1:

$$\text{Re} = \frac{\rho_{\text{air}} d\nu}{\eta}$$

where ρ_{air} is air density (1.29×10^3 g/m^3), d is the particle diameter, ν is its velocity, and η is viscosity. Find the settling velocity and Reynolds numbers for particles having the density of water droplets, with diameters

(a) 1 μm

(b) 10 μm

(c) 20 μm

14 For the following potential oxygenates for gasoline, find the percentage by weight of oxygen. Then find the fraction (by weight) of oxygenate needed to provide 2 percent oxygen to the resulting blend of gasoline (which has no oxygen) and oxygenate.

(a) Ethanol, CH_3CH_2OH

(b) Methyl tertiary butyl ether (MTBE), $CH_3OC(CH_3)_3$

(c) Ethyl tertiary butyl ether (ETBE), $CH_3CH_2OC(CH_3)_3$

(d) Tertiary amyl methyl ether (TAME), $CH_3CH_2C(CH_3)_2OCH_3$

15 The specific gravity of the oxygenate ethanol (CH_3CH_2OH) is 0.791, whereas for pure gasoline it is 0.739. A blend of gasoline (with no oxygen of its own) and ethanol needs to provide 2 percent oxygen by weight to the resulting fuel. What volumetric fraction of the resulting blend needs to be ethanol?

16 Calculate the CAFE fuel efficiency for flex-fuel cars that get

(a) 18 mpg on gasoline and 12 mpg on ethanol.

(b) 22 mpg on gasoline and 15 mpg on ethanol.

(c) 27 mpg on gasoline and 18 mpg on ethanol.

17 The CAFE credit that is provided for natural-gas-powered vehicles is similar to that for alcohol fuels (Example 6). The AMFA rules say 100 ft^3 of natural gas is equivalent to 0.823 gallons of gasoline, and each equivalent gallon counts as 0.15 gallons of gasoline. All of the miles driven are assumed to be powered by natural gas. Show that a vehicle that achieves 25 miles per 100 ft^3 of natural gas would have a CAFE rating of 203 mpg.

18 Using the energy content of the fuel as the criterion, what would the "break-even" price of E85 have to be if gasoline sells for \$3.50/gallon?

19 On an energy-content basis, which would be cheaper?

(a) E85 at \$2/gallon or gasoline at \$3/gallon

(b) E85 at \$2.50/gallon or gasoline at \$3.30/gallon

(c) E85 at \$2.75/gal or gasoline at \$4/gal

20 An automobile with a 15-gallon fuel tank gets 25 mpg when powered by pure gasoline. Using energy density values from Table 6, estimate the fuel efficiency (mpg) and range (miles) for that car when it runs on

(a) gasohol (E10).

(b) E85.

(c) half a tank of gasoline and half a tank of E85.

21 Suppose a 50-mpg hybrid-electric vehicle (HEV) is retrofitted with extra batteries, making it a plug-in hybrid (PHEV). The batteries give the PHEV a 30-mile/day range on electricity, with a grid-to-wheels efficiency of 0.25 kWh/mile. After those 30 miles, the extra weight of batteries means the car only gets 45 mpg when running on gasoline. The PHEV is driven 50 round-trip miles per day back and forth to work, 5 days per week. On weekends, it is driven 25 miles per day for errands.

(a) At $3.50 per gallon of gasoline and $0.08/kWh for electricity, what is the cost of fuel expressed as a ¢/mile average? Compare it to the original HEV.

(b) If the car hadn't been converted, what would be the annual cost of fuel and how much would be saved per year by making this a PHEV?

(c) If the conversion costs $3,000, what is the simple payback period for this investment? How many miles do the batteries have to last to make this worthwhile?

22 Use the footnotes from Figure 28 to verify the height of the well-to-wheels CO_2/mile bars for the following vehicles:

(a) 25-mpg conventional vehicle

(b) 50-mpg hybrid-electric vehicle (HEV)

(c) Plug-in hybrid electric vehicle (PHEV)

(d) Battery-powered electric vehicle (EV)

(e) Fuel-cell vehicle (FCV)

23 Suppose the 0.25 kWh/mi EV and PHEV in Figure 28 get their electricity from a 60 percent-efficient natural-gas-fired combined-cycle (NGCC) power plant delivering electricity through a 96 percent-efficient grid. Natural gas emits 14.4 gC/MJ and 1 kWh = 3.6 MJ.

(a) Find the gCO_2/mile for this EV.

(b) Find the gCO_2/mile for this PHEV.

24 Los Angeles has an average solar radiation equivalent to 5.5 hours per day of full sun hitting a south-facing, sloped garage roof. How big would a 17 percent-efficient grid-connected PV array need to be to power a 0.25 kWh/mile EV or PHEV an average of 30 miles per day? Assume a 75 percent conversion factor from rated-PV power to AC power delivered to the grid.

25 Consider a 50 percent-efficient solid-oxide fuel cell (SOFC) used in a combined heat-and-power (CHP) system. The SOFC converts 50 percent of its natural gas fuel energy into electricity and 20 percent into useful heat. Compare its overall energy efficiency with a separate system that provides the same heat and power with 30 percent-efficient grid electricity and an 80 percent efficient boiler.

26 Compare the carbon emissions for a natural-gas-fired combined heat-and-power (CHP) system with the following separately generated electricity and heat systems. For carbon emissions, assume natural gas emits 14.4 gC/MJ, and for the grid assume 175 gC/kWh. The joule equivalent of one kWh of electricity is 3.6 MJ.

(a) CHP with 36 percent electrical efficiency and 40 percent thermal efficiency versus an 85 percent-efficient gas boiler for heat and the grid for electricity.

(b) CHP with 50 percent electrical efficiency and 20 percent thermal efficiency versus a 33 percent-efficient, 280 g/kWh, coal-fired power plant for electricity and an 80 percent-efficient gas-fired boiler for heat.

27 In 1989, 0.39×10^{12} g of particulates were released when 685 million (2,000 lb) tons of coal were burned in power plants that produced 1,400 billion kWh of electricity. Assume

that the average heat content of the coal is 10,000 Btu/lb. What must have been the average efficiency (heat to electricity) of these coal plants? How much particulate matter would have been released if all of the plants met the New Source Performance Standards that limit particulate emissions to 0.03 lb per 10^6 Btu of heat?

28 In this chapter, the following expression was used in the derivation of the dry adiabatic lapse rate: $dQ = C_p\,dT - V\,dP$ (Eq. 37). Derive that expression starting with a statement of the first law of thermodynamics: $dQ = dU + dW$, where $dU = C_v\,dT$ is the change in internal energy when an amount of heat, dQ, is added to the gas, raising its temperature by dT and causing it to expand and do work $dW = P\,dV$. C_v is the specific heat at constant volume. Then use the ideal gas law, $PV = nRT$, where n is moles (a constant) and R is the gas constant, along with the definition of the derivative of a product, $d(PV) = P\,dV + V\,dP$, to find another expression for dQ. Finally, using the definition $C_p = (dQ/dT)$ with the pressure held constant, show that $C_p = C_v + nR$, and you're about there.

29 Suppose the following atmospheric altitude versus temperature data have been collected.

Altitude (m)	Temp (°C)
0	20
100	18
200	16
300	15
400	16
500	17
600	18

(a) What would be the mixing depth?

(b) How high would you expect a plume to rise if it is emitted at 21°C from a 100-m stack if it rises at the dry adiabatic lapse rate? Would you expect the plume to be looping, coning, fanning, or fumigating?

30 For the temperature profile given in Problem 29, if the daytime surface temperature is 22°C, and a weather station anemometer at 10 m height shows winds averaging 4 m/s, what would be the ventilation coefficient? Assume stability class C, and use the wind at the height halfway to the mixing depth.

31 A tall stack and a nearby short stack have plumes as shown in Figure P31. Which atmospheric temperature profile would be most likely to cause that pair of plumes? The dotted lines represent the dry adiabatic lapse rate.

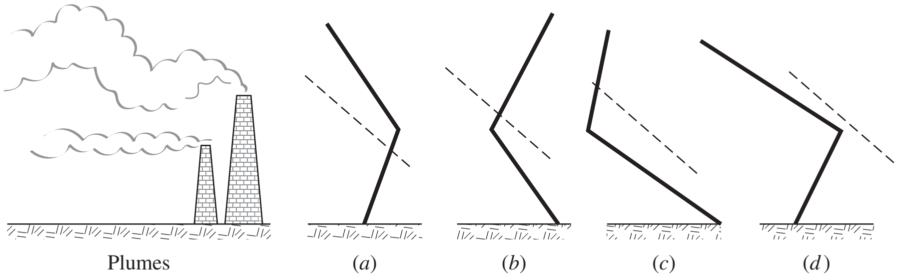


FIGURE P31

32 A point source Gaussian plume model for a power plant uses 50 m as the effective stack height (see Figure P32). The night is overcast. (*Note:* This is not the same as "cloudy" in this model; check the footnotes in Table 8.) Your concern is with ground-level pollution at two locations, A and B, which are 1.2 and 1.4 km directly downwind from the stack.

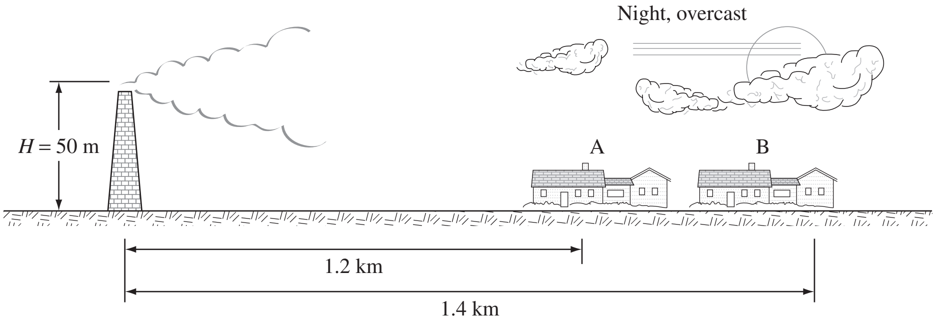


FIGURE P32

(a) At what distance will the maximum concentration of pollution occur? Which location (A or B) would have the higher level of pollution?

(b) Suppose the sky clears up, and the windspeed stays less than 5 m/s. Will the location downwind at which the maximum concentration occurs move? If so, will it move closer to the stack or further away from the stack? (*H* remains 50 m.)

(c) Under the new conditions in part b, which house would experience the most pollution?

33 Suppose a bonfire emits CO at the rate of 20 g/s on a clear night when the wind is blowing at 2 m/s. If the effective stack height at the fire is 6 m, (a) what would you expect the ground-level CO concentration to be at 400 m downwind? (b) Estimate the maximum ground-level concentration.

34 A coal-fired power plant with effective stack height of 100 m emits 1.2 g/s of SO_2 per megawatt of power delivered. If winds are assumed to be 4 m/s at that height and just over 3 m/s at 10 m, how big could the plant be (MW) without having the ground level SO_2 exceed 365 μg/m^3? (First decide which stability classification leads to the worst conditions.)

35 A 35 percent-efficient coal-fired power plant with effective height of 100 m emits SO_2 at a the rate of 0.6 lb/10^6 Btu into the plant. If winds are assumed to be 4 m/s at the stack height and just over 3 m/s at 10 m, how large could the plant be (MW) without having the ground level SO_2 exceed 365 μg/m^3?

36 A stack emitting 80 g/s of NO has an effective stack height of 100 m. The windspeed is 4 m/s at 10 m, and it is a clear summer day with the sun nearly overhead. Estimate the ground-level NO concentration

(a) directly downwind at a distance of 2 km.

(b) at the point downwind where NO is a maximum.

(c) at a point located 2 km downwind and 0.1 km off the downwind axis.

37 For stability class C, the ratio of σ_y/σ_z is essentially a constant, independent of distance x. Assuming it is a constant, take the derivative of (49) and

(a) show that the distance downwind from a stack at which the maximum concentration occurs corresponds to the point where $\sigma_z = H/\sqrt{2} = 0.707\,H$.

(b) show that the maximum concentration is

$$C_{\max} = \frac{Q}{\pi \sigma_y \sigma_z u e} = \frac{0.117Q}{\sigma_y \sigma_z u}$$

(c) show that $C_{\max}$ is inversely proportional to H^2.

38 The world's tallest stack is on a copper smelter in Sudbury, Ontario. It stands 380 m high and has an inner diameter at the top of 15.2 m. If 130°C gases exit the stack at 20 m/s while the ambient temperature is 10°C and the winds at stack height are 8 m/s, use the Briggs model to estimate the effective stack height. Assume a slightly unstable atmosphere, class C.

39 Repeat Problem 38 for a stable, isothermal atmosphere (no temperature change with altitude).

40 A power plant has a 100-m stack with an inside radius of 1 m. The exhaust gases leave the stack with an exit velocity of 10 m/s at a temperature of 120°C. Ambient temperature is 6°C, winds at the effective stack height are estimated to be 5 m/s, surface windspeed is 3 m/s, and it is a cloudy summer day. Estimate the effective height of this stack.

41 A 200-MW power plant has a 100-m stack with radius 2.5 m, flue gas exit velocity 13.5 m/s, and gas exit temperature 145°C. Ambient temperature is 15°C, windspeed at the stack is 5 m/s, and the atmosphere is stable, class E, with a lapse rate of 5°C/km. If it emits 300 g/s of SO_2, estimate the concentration at ground level at a distance of 16 km directly downwind.

42 A source emits 20 g/s of some pollutant from a stack with effective height 50 m in winds that average 5 m/s. On a single graph, sketch the downwind concentration for stability classifications A, C, and F, using Figure 52 to identify the peak concentration and distance.

43 A source emits 20 g/s of some pollutant in winds that average 5 m/s, on a class C day. Use Figure 52 to find the peak concentrations for effective stack heights of 50, 100, and 200 m. Note whether the concentration is roughly proportional to $(1/H^2)$, as Problem 37 suggests.

44 A paper plant is being proposed for a location 1 km upwind from a town. It will emit 40 g/s of hydrogen sulfide, which has an odor threshold of about 0.1 mg/m^3. Winds at the stack may vary from 4 to 10 m/s blowing toward the town. What minimum stack height should be used to assure concentrations are nor more than 0.1 times the odor threshold at the near edge of town on a class B day? To be conservative, the stack will be designed assuming no plume rise. If the town extends beyond the 1-km distance, will any buildings experience higher concentrations than a residence at the boundary under these conditions?

45 A stack with effective height of 45 m emits SO_2 at the rate of 150 g/s. Winds are estimated at 5 m/s at the stack height, the stability class is C, and there is an inversion at 100 m. Estimate the ground-level concentration at the point where reflections begin to occur from the inversion and at a point twice that distance downwind.

46 A point source with effective stack height of 50 m emits 80 g/s of SO_2 on a clear summer day with surface winds at 4 m/s. Winds at 50 m are 5 m/s. An inversion layer starts at an elevation of 250 m.

(a) Estimate the ground-level SO_2 concentration at a distance of 4 km downwind from the stack.

(b) If there had been no inversion layer, estimate the concentration 4 km downwind.

47 A long line of burning agricultural waste emits 0.3 g/m-s of particulate matter on a clear fall afternoon with winds blowing 3 m/s perpendicular to the line. Estimate the ground-level particulate concentration 400 m downwind from the line.

48 A freeway has 10,000 vehicles per hour passing a house 200 m away. Each car emits an average of 1.5 g/mi of NO_x, and winds are blowing at 2 m/s across the freeway towards the house. Estimate the NO_x concentration at the house on a clear summer day near noon (assuming that NO_x is chemically stable).

49 Consider an area-source box model for air pollution above a peninsula of land (see Figure P49). The length of the box is 15 km, its width is 80 km, and a radiation inversion restricts mixing to 15 m. Wind is blowing clean air into the long dimension of the box at 0.5 m/s. Between 4 and 6 PM there are 250,000 vehicles on the road, each being driven 40 km and each emitting 4 g/km of CO.

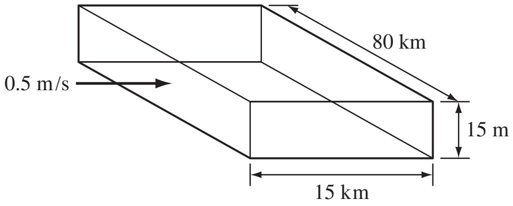


FIGURE P49

(a) Find the average rate of CO emissions during this two-hour period (g CO/s per m^2 of land).

(b) Estimate the concentration of CO at 6 PM if there was no CO in the air at 4 PM. Assume that CO is conservative and that there is instantaneous and complete mixing in the box.

(c) If the windspeed is 0, use (58) to derive a relationship between CO and time and use it to find the CO over the peninsula at 6 PM.

50 Consider a box model for an air shed over a city 1×10^5 m on a side, with a mixing depth of 1,200 m. Winds with no SO_2 blow at 4 m/s against one side of the box. SO_2 is emitted in the box at the rate of 20 kg/s. If SO_2 is considered to be conservative, estimate the steady-state concentration in the air shed.

51 With the same air shed and ambient conditions as given in Problem 50, assume that the emissions occur only on weekdays. If emissions stop at 5 PM on Friday, estimate the SO_2 concentration at midnight. If they start again on Monday at 8 AM, what would the concentration be by 5 PM.

52 If steady-state conditions have been reached for the city in Problem 50 and then the wind drops to 2 m/s, estimate the concentration of SO_2 two hours later.

53 If the wind blowing in to the air shed in Problem 50 has 5 μg/m^3 of SO_2 in it, and the SO_2 concentration in the air shed at 8 AM. Monday is 10 μg/m^3; estimate the concentration at noon, assuming that emission are still 20 kg/s.

54 With the same airshed and ambient conditions as given in Problem 50, if SO_2 is not conservative and in fact has a reactivity of 0.23/hr, estimate its steady-state concentration over the city.

55 Consider use of a tracer gas to determine the air exchange rate in a room. By injecting a stable gas into the room and then monitoring the decay in concentration with time, we can estimate I (ach). The governing equation is

$$C = C_0 e^{-nt}$$

Taking the log of both sides gives: ln ln$C$ = ln $C_0 - nt$

Thus, if you plot ln C versus time, you should get a straight line with negative slope equal to the infiltration rate n (see Figure P55).

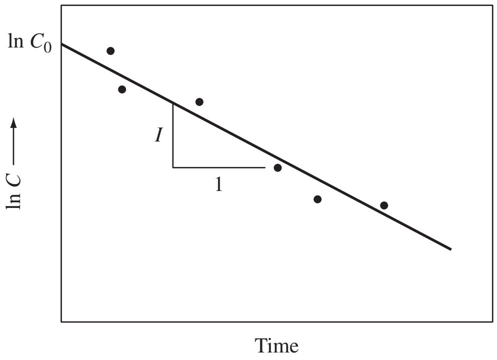


FIGURE P55

Suppose you gather the following data:

Time (hr)	Concentration (ppm)
0	10.0
0.5	8.0
1.0	6.0
1.5	5.0
2.0	3.3

Plot ln C versus time; find the slope and thus the infiltration rate I (ach).

56 A single-story home with infiltration rate of 0.5 ach has 200 m^2 of floor space and a total volume of 500 m^3. If 0.6 pCi/m^2-s of radon is emitted from the soil and enters the house, estimate the steady-state indoor radon concentration.

57 If the house in Problem 56 were to have been built as a two-story house with 100 m^2 on each floor and the same total volume, what would the estimated radon concentration be?

58 If an individual lived for 30 years in the house given in Problem 56, estimate the cancer risk from the resulting radon exposure.

59 Consider a "tight" 300 m^3 home with 0.2 ach infiltration rate. The only source of CO in the home is the gas range, and the ambient concentration of CO is always zero. Suppose there is no CO in the home at 6 PM, but then the oven and two burners are on for one hour. Assume that the air is well mixed in the house, and estimate the CO concentration in the home at 7 PM and again at 10 PM.

60 A convective kerosene heater is tested in a well-mixed 27 m^3 chamber having an air exchange rate of 0.39 ach. After one hour of operation, the NO concentration reached 4.7 ppm. Treating NO as a conservative pollutant,

(a) Estimate the NO source strength of the heater (mg/hr).

(b) Estimate the NO concentration that would be expected in the lab one hour after turning off the heater.

(c) If this heater were to be used in the home described in Problem 59, what steady-state concentration of NO would you expect to be caused by the heater?

61 Find the settling velocity of 2.5-micron particles having density 1.5×10^6 g/m^3. In a room with 2.5-meter-high ceilings, use a well-mixed box model to estimate the residence time of these particles.

62 Consider a 100-MW (100,000 kilowatt), 33.3 percent-efficient coal-fired power plant that operates at the equivalent of full power 70 percent of the time and no power 30 percent of the time (i.e., its capacity factor is 0.70).

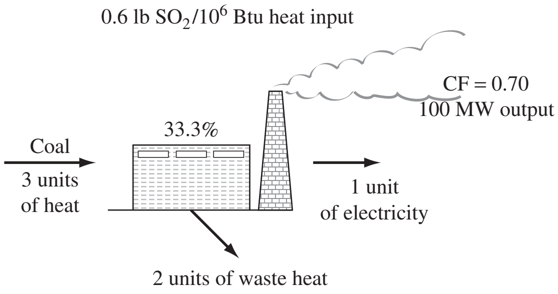


FIGURE P62

(a) How much electricity (kWhr/yr) would the plant produce each year?

(b) How may Btu of heat per year would be needed to generate that much electricity? (*Note:* At 33.3 percent efficiency, 3 kWhr of heat, at 3,412 Btu/kWhr, are needed to produce 1 kWhr of electricity, as shown in Figure P62.)

(c) Suppose this coal plant has been emitting 0.6 lb of SO_2 per million Btu of heat input, and suppose it has enough SO_2 allowances for it to continue to do so. If the power company decides to shut down this coal plant and replaces it with a natural-gas-fired plant that emits no SO_2, how many (2,000-lb) tons of SO_2 emissions would be saved in one year? If SO_2 allowances are trading at $400 each, how much could this power company receive by selling one year's worth of allowances?

REFERENCES

American Lung Association, American Medical Association, U.S. Consumer Product Safety Commission, U.S. Environmental Protection Agency, 1994, *Indoor Air Pollution: An Introduction for Health Professionals,* Environmental Protection Agency, Washington, DC.

Briggs, G. A., 1972, Dispersion of chimney plumes in neutral and stable surroundings, *Atmospheric Environment,* 6(1).

California Air Resources Board (CARB), 2004, *Initial Statement of reasons for proposed rulemaking, public hearing to consider adoption of regulation to control greenhouse gas emissions from motor vehicles,* August.

California Energy Commission, 1989, *News and Comment, CEC Quarterly Newsletter,* No. 23, Fall.

Chestnut, L. G., and D. M. Mills, 2005, A fresh look at the benefits and costs of the U.S. Acid Rain Program, *Journal of Environmental Management,* 77(3):252–266.

DeCicco, J., and M. Ross, 1993, *An Updated Assessment of the Near-Term Potential for Improving Automotive Fuel Economy,* American Council for an Energy-Efficient Economy, Washington, DC.

Diamond, R. C., and D. T. Grimsrud, 1984, *Manual on Indoor Air Quality,* prepared by Lawrence Berkeley Laboratory for the Electric Power Research Institute, Berkeley, CA.

Energy Foundation, *China's Energy Challenge, 2004 Annual Report,* 2005.

Energy Information Administration (EIA), 2006, *International Energy Outlook, 2006,* Washington, DC.

Farrell, A. E., R. J. Plevin, B. T. Turner, A. D. Jones, M. O'Hare, and D. M. Kammen, 2006, Ethanol can contribute to energy and environmental goals, *Science,* 311(27):506–508.

Gifford, F. A., 1961, Uses of routine meteorological observations for estimating atmospheric dispersion, *Nuclear Safety,* 2(4).

Interagency Taskforce on Acid Precipitation, 1983, *Annual Report 1983 to the President and Congress,* Washington, DC.

Jacobson, M. Z., 2007, *Effects of Ethanol (E85) Versus Gasoline Vehicles on Cancer and Mortality in the United States* (in publication).

Jenkins, P. L., T. J. Phillips, E. Mulberg, and S. P. Hui, 1992, Activity patterns of Californians: Use of and proximity to indoor pollution sources, *Atmospheric Environment,* 26A(2): 2141–2148.

Kreiss, K., 1990, The sick building syndrome: Where is the epidemiologic basis?, *American Journal of Public Health,* 80:1172–1173.

Lovins, A., and H. Lovins, 1995, Reinventing the wheels, *The Atlantic Monthly,* January.

Martin, D. O., 1976, The change of concentration standard deviation with distance, *Journal of the Air Pollution Control Association,* 26(2).

McCosh, D., 1994, Emerging technologies for the supercar, *Popular Science,* June, 95–101.

Moyer, C. B., 1989, *Global Warming Benefits of Alternative Motor Fuels,* Acurex Corporation, Mountain View, CA.

Murphy, J. J., M. A. Delucchi, D. R. McCubbin, and H. J. Kim, 1999, The cost of crop damage caused by ozone air pollution from motor vehicles, *Journal of Environmental Management,* 55:273–289.

Nagda, N. L., H. E. Rector, and M. D. Koontz, 1987, *Guidelines for Monitoring Indoor Air Quality,* Hemisphere, Washington, DC.

National Research Council, 1981, *Indoor Pollutants,* National Academy Press, Washington, DC.

National Research Council, 1999, *Health Effects of Exposure to Radon (BEIR VI),* National Academy Press, Washington, DC.

Nazaroff, W. W., and K. Teichman, 1990, Indoor radon, exploring U.S. federal policy for controlling human exposures, *Environmental Science and Technology,* 24:774–782.

NEDO. 1994, *NEDO Creates New Energy,* New Energy and Industrial Technology Development Organization, Tokyo.

OTA, 1984, *Acid Rain and Transported Air Pollutants, Implications for Public Policy,* Office of Technology Assessment, Washington, DC.

Ott, W. R., A. C. Steinemann, and L. A. Wallace, 2006, *Exposure Analysis,* CRC Press, Boca Raton, FL.

Pasquill, F., 1961, The estimation of the dispersion of windborne material, *Meteorological Magazine,* 90: 1063.

Perkins, H. C., 1974, *Air Pollution,* McGraw-Hill, New York.

Pershagen, G., G. Akerblom, O. Axelson, B. Clavensjo, L. Damber, G. Desai, A. Enflo, F. Lagarde, H. Mellamder, and M. Svartengren, 1994, Residential radon exposure and lung cancer in Sweden, *New England Journal of Medicine,* 330(3):159–164.

Peterson, W. B., 1978, *User's Guide for PAL—A Gaussian Plume Algorithm for Point, Area, and Line Resources,* U.S. Environmental Protection Agency, Research Triangle Park, NC.

Portelli, R. V., and P. J. Lewis, 1987, Meteorology, in *Atmospheric Pollution,* E. E. Pickett (ed.), Hemisphere, Washington, DC.

Powell, J. D., and R. P. Brennan, 1988, *The Automobile, Technology and Society,* Prentice Hall, Englewood Cliffs, NJ.

Powell, J. D., and R. P. Brennan, 1992, *Automotive Technology,* Stanford University, Stanford, CA.

Raloff, J., 1995, Heart-y risks from breathing fine dust, *Science News,* 148, July 1.

Reece, N. S., 1988, Indoor air quality, *Homebuilding & Remodeling Resource,* November/December.

Robinson, J. P., J. Thomas, and J. V. Behar, 1991, Time spent in activities, locations, and microenvironments: A California-national comparison. Report under Contract No. 69-01-7324, Delivery Order 12, Exposure Assessment Research Division, National Exposure Research Center, U.S. Environmental Protection Agency, Las Vegas, NV.

Seinfeld, J. H., and S. N. Pandis, 1998, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change,* John Wiley & Sons, Inc., Hoboken, New Jersey.

Shepard, M., 1988, Coal technologies for a new age, *EPRI Journal,* Electric Power Research Institute, Jan./Feb.

Smith, K. R., 1993a, Fuel combustion, air pollution exposure, and health: The situation in the developing countries, *Annual Review of Energy and Environment,* 18:529–566, Annual Reviews Inc., Palo Alto, CA.

Smith, K. R., 1993b, Taking the true measure of air pollution, *EPA Journal,* U.S. Environmental Protection Agency, October–December, 6–8.

SMUD, 1993, *Fuel Cell Power Plants,* Sacramento Municipal Utility District, Sacramento, CA.

The World Bank, 1997, *Clear Water, Blue Skies: China's Environment in the New Century*, Washington, DC.

Thomas, V. M., 1995, The elimination of lead in gasoline, *Annual Review of Energy and the Environment,* Annual Reviews, Inc., Palo Alto.

Traynor, G. W., D. W. Anthon, and C. D. Hollowell, 1981, *Technique for Determining Pollutant Emissions from a Gas-Fired Range,* Lawrence Berkeley Laboratory, LBL-9522, December.

Traynor, G. W., J. R. Allen, M. G. Apte, J. F. Dillworth, J. R. Girman, C. D. Hollowell, and J. F. Koonce, Jr., 1982, *Indoor Air Pollution from Portable Kerosene-Fired Space Heaters, Wood-Burning Stoves, and Wood-Burning Furnaces,* Lawrence Berkeley Laboratory, LBL-14027, March.

Turner, D. B., 1970, *Workbook of Atmospheric Dispersion Estimates,* U.S. Environmental Protection Agency, Washington, DC.

U.S. Department of Energy (U.S. DOE), 2006, *Biodiesel Handling and Use Guidelines, 3rd ed.*, DOE/GO-102006-2358, September.

U.S. Department of Transportation, 2002, *Effects of the Alternative Motor Fuels Act CAFE Incentives Policy: Report to Congress,* March, Washington, DC.

U.S. DOE and Southern Company, 1997, *Control of Nitrogen Oxide Emissions: Selective Catalytic Reduction (SCR)*, Clean Coal Technology Topical Report Number 9, July.

U.S. EPA, 1993, *Radon, The Threat with a Simple Solution,* EPA 402-K-93-008, Environmental Protection Agency, Washington, DC.

U.S. EPA, 1994a, *Acid Rain Program: Conservation and Renewable Energy Reserve Update,* EPA 430-R-94-010, Environmental Protection Agency, Washington, DC.

U.S. EPA, 1994b, *National Air Quality and Emissions Trends Report, 1993,* EPA-454/R-94-026, Environmental Protection Agency, Washington, DC.

U.S. EPA, 1995a, *Air Quality Trends,* EPA-454/F-95-003, Environmental Protection Agency, Washington, DC.

U.S. EPA, 1995b, *1993 Air Toxic Release Inventory, Public Data Release,* EPA 745-R-95-019, Environmental Protection Agency, Washington, DC.

U.S. EPA, 2003, *Air Trends Report,* Washington, DC.

U.S. EPA, 2006a, *NO_x Budget Trading Program 2005 Program Compliance and Environmental Results,* September, retrieved from http://www.epa.gov/airmarkets.

U.S. EPA, 2006b, *September 2006 Revisions to the National Ambient Air Quality Standards for Particle Pollution,* Washington, DC.

U.S. EPA, 2006c, *Light-Duty Automotive Technology and Fuel Economy Trends: 1975 through 2006,* EPA 420-S-06-003, Washington, DC.

U.S. HEW, 1969, *Control Techniques for Particulate Air Pollutants,* National Air Pollution Control Administration, Washington, DC.

U.S. HEW, 1970, *Air Quality Criteria for Carbon Monoxide,* AP-62, National Air Pollution Control Administration, Washington, DC.

U.S. PHS, 1991, *Preventing Lead Poisoning in Young Children,* U.S. Department of Health and Human Services, Washington, DC.

Wadden, R. A., and P. A. Scheff, 1983, *Indoor Air Pollution, Characterization, Prediction, and Control,* Wiley-Interscience, New York.

Wallace, L., 1993, The TEAM studies, *EPA Journal,* U.S. Environmental Protection Agency, October–December, 23–24.

Wallace, L., 1995, Human exposure to environmental pollutants: A decade of experience, *Clinical and Experimental Allergy,* 25:4–9.

Wark, K., and C. F. Warner, 1981, *Air Pollution, Its Origin and Control,* Harper & Row, New York.

Williamson, S. J., 1973, *Fundamentals of Air Pollution,* Addison-Wesley, Reading, MA.

WHO, UNEP, 1992, *Urban Air Pollution in Megacities of the World,* World Health Organization and the United Nations Environment Programme, Blackwell Publishers, Oxford, UK.

Yeager, K. E., 1991, Powering the second electrical century, in *Energy and the Environment in the 21st Century,* J. W. Tester (ed.), MIT Press, Boston.

Yokum, J. E., D. R. Murray, and R. Mikkelsen, 1991, Valdez Air Health Study—Personal, Indoor, Outdoor, and Tracer Monitoring, Paper No. 91–172.12, 84th Annual Meeting of the Air and Waste Management Association, Vancouver, BC.

This page intentionally left blank

Global Atmospheric Change

"The observed increase [in temperatures] could be largely due to natural variability; alternatively, this variability and other man-made factors could have offset a still larger man-made greenhouse warming."

—*IPCC 1990*

"The balance of evidence suggests a discernible human influence on global climate."

—*IPCC 1995*

"There is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities."

—*IPCC 2001*

"Most of the observed increase in globally averaged temperatures since the mid-twentieth century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations."

—*IPCC 2007*

"Be Worried. Be Very Worried"

—*Time, April 3, 2006*

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

1 Introduction

The preceding four quotes from the Intergovernmental Panel on Climate Change (IPCC) demonstrate a cautious evolution in the degree of certainty expressed by climate scientists about the role played by human activities in causing the past half-century's rapid rise in global average temperature. The *Time* magazine cover story in April 2006, with the provocative title "Be Worried. Be *Very* Worried" well captured the growing awareness among the general public that dealing with global warming is perhaps the most important environmental challenge of the twenty-first century.

While the atmosphere is made up almost entirely of nitrogen and oxygen, other gases and particles existing in very small concentrations determine to a large extent the habitability of our planet. In this chapter, we will focus on several of these other gases, including carbon dioxide (CO_2), nitrous oxide (N_2O), methane (CH_4), and ozone (O_3), as well as a category of man-made gases called *halocarbons* that includes chlorofluorocarbons (CFCs), hydrochlorofluorocarbons (HCFCs), hydrofluorocarbons (HFCs), carbon tetrachloride (CCl_4), methylchloroform (CH_3CCl_3), and halons. The two problems of greenhouse effect enhancement, leading to global climate change, and stratospheric ozone depletion, which increases our exposure to life-threatening ultraviolet radiation, are linked to changes in these trace gases and are the subject of this chapter.

The problems of global warming and stratospheric ozone depletion are linked by their shared dependence on minute changes in our global atmosphere, and they are, by their very nature, problems whose solutions require international cooperation. The IPCC was established in 1988 by the World Meteorological Organization (WMO) and the United Nations Environment Program (UNEP) to (1) assess available scientific information on climate change, (2) assess the environmental and socioeconomic impacts of climate change, and (3) formulate appropriate response strategies (IPPC, 1995). The IPCC provided the technical documentation that led to the United Nations Framework Convention on Climate Change (UNFCCC), which was signed by 150 nations, including the United States, at the UN Conference on Environment and Development in Rio de Janeiro in 1992. The UNFCCC is a climate treaty that laid the groundwork for nations to stabilize greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous interference with the climate system. It took effect on March 21, 1994, after 50 countries ratified the treaty. The subsequent Kyoto Protocol had its origins in UNFCCC.

Another effort by WMO and UNEP to organize the scientific community around the issue of stratospheric ozone depletion led to the signing of the Montreal Protocol on Substances That Deplete the Ozone Layer in 1987. The Montreal Protocol, and subsequent Amendments framed in London (1990), Copenhagen (1992), and Vienna (1995), has been extraordinarily important both in terms of its success in reducing emissions of ozone depleting substances and its clear demonstration that nations can come together to address global environmental problems.

Much of this chapter is based on the multivolume scientific assessment reports on climate change and ozone depletion that have been written as part of the WMO and UNEP efforts.

2 The Atmosphere of Earth

When the Earth was formed, some 4.6 billion years ago, it probably had an atmosphere made up of helium and compounds of hydrogen such as molecular hydrogen, methane, and ammonia. That early atmosphere is thought to have escaped into space, after which our current atmosphere slowly began to form. Through volcanic activity, gases such as carbon dioxide, water vapor, and various compounds of nitrogen and sulfur were released over time. Molecular oxygen (O_2) eventually began to form both as a result of photodissociation of water vapor and by photosynthesis by plants that were evolving underwater where life was protected from the sun's intense, biologically damaging ultraviolet radiation. As atmospheric oxygen levels gradually increased, more and more ozone (O_3) was formed in the atmosphere. It is thought that the absorption of incoming ultraviolet radiation by that ozone provided the protection necessary for life to begin to emerge onto the land.

Table 1 shows the composition of the Earth's atmosphere as it exists now, expressed in volumetric fractions. The values given are for "clean," dry, air and don't include the relatively small but extremely important amounts of water vapor and particulate matter. While most of the values in the table are essentially unchanging, that is not the case for the principal greenhouse gases carbon dioxide CO_2, methane CH_4, and nitrous oxide N_2O, which are rising.

It is convenient to think of the atmosphere as being divided into various horizontal layers, each characterized by the slope of its temperature profile. Starting at the Earth's surface, these layers are called the *troposphere, stratosphere, mesosphere,* and *thermosphere*. The troposphere and mesosphere are characterized by decreasing temperatures with altitude, while the stratosphere and thermosphere show increasing temperatures. The transition altitudes separating these layers are called the

TABLE 1

Composition of Clean Dry Air (Fraction by Volume in Troposphere, 2006)

Constituent	Formula	Percent by Volume	Parts per Million
Nitrogen	N_2	78.08	780,800
Oxygen	O_2	20.95	209,500
Argon	Ar	0.93	9,300
Carbon dioxide	CO_2	0.038	380
Neon	Ne	0.0018	18
Helium	He	0.0005	5.2
Methane	CH_4	0.00017	1.7
Krypton	Kr	0.00011	1.1
Nitrous oxide	N_2O	0.00003	0.3
Hydrogen	H_2	0.00005	0.5
Ozone	O_3	0.000004	0.04

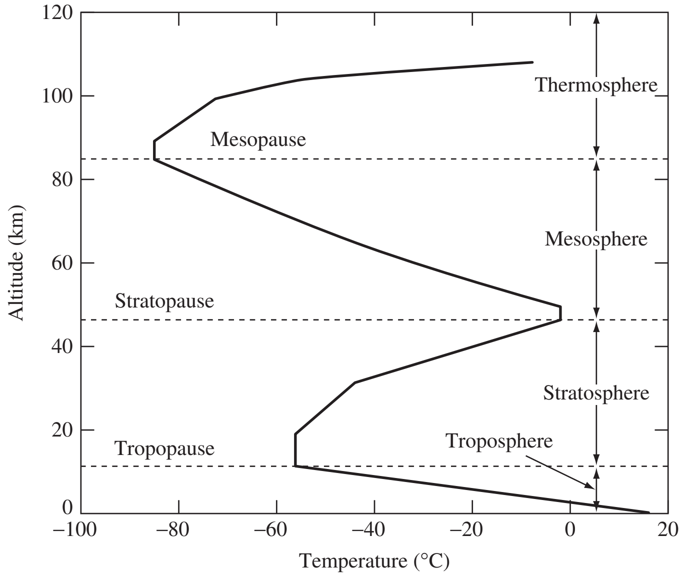


FIGURE 1 The U.S. Standard Atmosphere, showing the four major layers.

tropopause, stratopause, and *mesopause*. Obviously, the conditions in the actual atmosphere of Earth vary with time and location, but a useful, idealized temperature profile, known as the *U.S. Standard Atmosphere,* provides a convenient starting point for atmospheric studies (Figure 1).

More than 80 percent of the mass of the atmosphere and virtually all of the water vapor, clouds, and precipitation occur in the troposphere. At mid-latitudes, the troposphere extends up to 10 or 12 km (about the altitude of a typical commercial airline flight). At the poles, it may be only about 5 to 6 km, while at the equator it is about 18 km. In the troposphere, temperatures typically decrease at 5 to 7°C per km, which is essentially the wet adiabatic lapse rate corresponding to the rate of change of temperature as water-saturated air rises. The troposphere is usually a very turbulent place; that is, strong vertical air movements lead to rapid and complete mixing. This mixing is good for air quality since it rapidly disperses pollutants.

Above the troposphere is a stable layer of very dry air called the stratosphere. Pollutants that find their way into the stratosphere may remain there for many years before they eventually drift back into the troposphere, where they can be more easily diluted and ultimately removed by settling or precipitation. In the stratosphere, short-wavelength ultraviolet energy is absorbed by ozone (O_3) and oxygen (O_2), causing the air to be heated. The resulting temperature inversion is what causes the stratosphere to be so stable. The troposphere and stratosphere combined account for about 99.9 percent of the mass of the atmosphere. Together they extend only about 50 km above the surface of the Earth, a distance equal to less than 1 percent of the Earth's radius.

Beyond the stratosphere lies the mesosphere, another layer where air mixes fairly readily, and above that the thermosphere. The heating of the thermosphere is due to the absorption of solar energy by atomic oxygen. Within the thermosphere is a relatively dense band of charged particles, called the *ionosphere.* (Before satellites, the ionosphere was especially important to worldwide communications because of its ability to reflect radio waves back to Earth.)

3 Global Temperature

The common definition of climate suggests it is the prevailing or average weather of a place as determined by temperature and other meteorological conditions over a period of years. Average temperature, then, is but one measure of climate; many others, including precipitation, winds, glaciation, and frequency of extreme events such as typhoons and hurricanes, are also important parameters. Long-term variations in average temperature, however, are the single most important attribute of climate change.

Climatologists have used a number of clues to piece together past global temperatures, including evidence gathered from historical documents, tree rings, changes in ice volume and sea level, fossil pollen analysis, and geologic observations related to glacial movements. One of the most fruitful approaches involves the analysis of concentrations of various stable isotopes of hydrogen and oxygen as found in ice cores and sea-floor sediments.

Isotopes and Temperature

Recall that atoms of a particular element all have the same number of protons but may have varying numbers of neutrons. The sum of the protons and neutrons, called the mass number, identifies which isotope of that element we are talking about. Some isotopes are radioactive, such as radon-222 (^{222}Rn), and their decay rate can be used to date various geological events. Other isotopes are stable, which means they can be used to track the movement of substances in the environment over geologic time.

Consider water, H_2O. Hydrogen, with one proton, has two stable isotopes, ^{1}H and ^{2}H (also called deuterium, D). Oxygen, with eight protons, has three stable isotopes, ^{16}O, ^{17}O, and ^{18}O. Thus we can imagine many different combinations of atoms making up molecules of water, including the two most important “heavier water” molecules $^{1}H^{2}H^{16}O$ and $^{1}H^{1}H^{18}O$.

When water evaporates from the oceans, it contains a mix of isotopes of hydrogen and oxygen. Heavier water containing ^{18}O or deuterium (^{2}H) evaporates less readily and condenses more quickly than light water, which means those heavy isotopes get left behind as water vapor from the ocean travels toward the poles. That means snow falling in the Arctic and Antarctic will have lower concentrations of deuterium (D or ^{2}H) and ^{18}O than seawater. This separation of light isotopes from heavier ones is temperature dependent. As it gets colder, less and less of the heavier deuterium and ^{18}O make it to the poles. The $^{2}H/^{1}H$ and $^{18}O/^{16}O$ ratios in precipitation that forms glaciers and ice sheets therefore decrease when it gets colder and increase when it gets warmer, thus providing a very important historic measure of global temperature.

The relative abundance of ^{2}H and ^{18}O in seawater is so important that standardized concentrations have been established called the *Vienna Standard Mean Ocean Water* (VSMOW) ratios. The VSMOW ratio of $^{2}H/^{1}H$ is 0.00015575 (1 deuterium per 6,422 atoms of conventional hydrogen) and the standardized $^{18}O/^{16}O$ ratio is 0.0020052 (1 part ^{18}O to 499 parts ^{16}O). The shifts in isotope ratios are compared to the standard using a parts-per-thousand $\delta X(^{0}/_{00})$ notation. For example, the $\delta^{18}O(^{0}/_{00})$ concentration is expressed as

$$\delta^{18}O(^{0}/_{00}) = \left[\frac{(^{18}O/^{16}O)\text{ sample}}{(^{18}O/^{16}O)\text{ standard}} - 1\right] \times 10^{3} \quad (1)$$

Negative values of δ^{18}O or δD represent isotope ratios below the standard, while positive numbers correspond to higher ratios. For example, $\delta^{18}\mathrm{O}(^0/_{00}) = -10$ means the concentration of ^{18}O is lower than the standard by 10 parts per thousand ("10 per mil"), or 1 percent. Changes in δD and δ^{18}O in precipitation seem to be linearly related to temperature. For δ^{18}O, the change is about 0.7 per °C, while for deuterium, it is about 6 per mil per °C (McGuffie and Henderson-Sellers, 2005).

EXAMPLE 1 Deuterium in an Ice Core

Suppose a deuterium measurement taken near the surface of an ice core yields an isotope ratio of $^2\mathrm{H}/^1\mathrm{H} = 8.753 \times 10^{-5}$.

a. Find the corresponding $\delta\mathrm{D}(^0/_{00})$ using the VSMOW standard for deuterium.
b. Suppose an ice sample from a deeper, and older, depth in the core shows $\delta\mathrm{D}(^0/_{00}) = -445$. Using the estimate of 5.6 $\delta\mathrm{D}(^0/_{00})$ per °C, how much colder was it at the time that ice layer was created?

Solution

a. Using the VSMOW standard of 0.00015575 for deuterium in (1) gives

$$\delta\mathrm{D}(^0/_{00}) = \left[\frac{(^2\mathrm{H}/^1\mathrm{H})\text{ sample}}{(^2\mathrm{H}/^1\mathrm{H})\text{ standard}} - 1\right] \times 10^3$$

$$\delta\mathrm{D}(^0/_{00}) = \left[\frac{8.753 \times 10^{-5}}{0.00015575} - 1\right] \times 10^3 = -438$$

b. The 7 per mil drop in δD from -438 to -445 translates to about $7/5.6 \times 1°\mathrm{C} = 1.25°\mathrm{C}$ colder than the present.

As the world's ice volume increases, it selectively removes ^{16}O from the hydrologic cycle and concentrates the remaining ^{18}O in the decreasing volume of the oceans. Hence, marine organisms that build their shells out of calcium carbonate in seawater will have a higher ratio of ^{18}O to ^{16}O in their shells when it is cold and more of the world's water is locked up in glaciers and ice. By dating marine sediments extracted from deep sea cores, and observing the ratio of the two oxygen isotopes in their carbonates, a historic record of the volume of ice storage on Earth can be created. Figure 2 shows a reconstruction of climate based on δ^{18}O in benthic

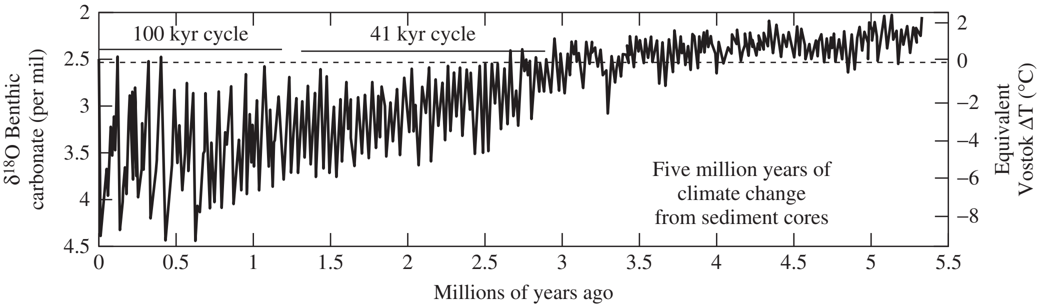


FIGURE 2 Five million years of climate change from deep sea sediment cores.
(*Source:* Lisiecki and Raymo, 2005.)

carbonates over the past 5 million years. The temperature scale calibration is based on a comparison with ice core data. Notice the isotope record in ocean sediments moves in the opposite direction to the isotope record in glacial ice. That is, for example, while warmer temperatures cause *decreases* in ^{18}O in oceans and sediments, warmer temperatures cause *increases* in ^{18}O in that year's layer of glacial ice.

The Antarctic and Greenland Ice Cores

Analysis of ice cores taken in Greenland and Antarctica have provided a remarkable picture of the climate and composition of the Earth's atmosphere over hundreds of thousands of years of history. Isotopic analysis of glacial ice provides the temperature record, while analysis of the composition of the air bubbles captured in the ice as it was formed gives a corresponding record of the concentrations of various atmospheric gases as well as wind-blown dust and ash.

The Vostok station in Antarctica is the site from which a number of significant cores have been drilled, the deepest of which extends 3,623-meters into the ice. The Vostok core provides a climate record that extends back some 420,000 years and includes four past glacial cycles. The record-holding ice core in terms of the length of its historical record, as opposed to its depth, is the 3,270-m *European Project for Ice Coring in Antarctica* (EPICA) core drilled into an ice summit called Dome C located several hundred miles away from Vostok. The EPICA core goes back in time by more than 720,000 years and covers 8 previous glacial cycles.

Figure 3 shows a very strong correlation between atmospheric carbon dioxide and methane concentrations taken from air bubbles in the Antarctic cores compared

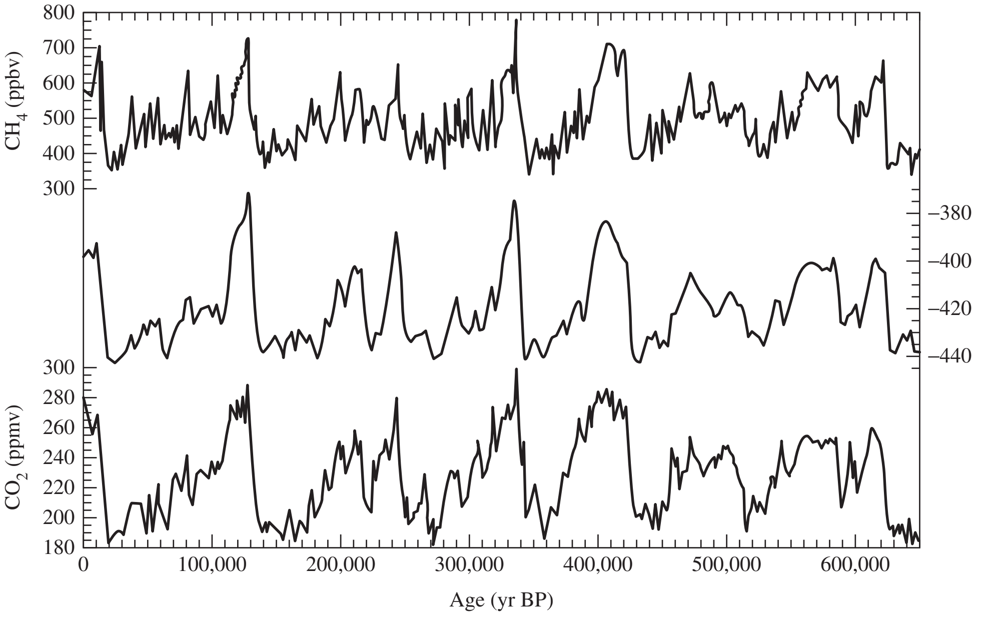


FIGURE 3 Antarctic carbon dioxide and methane concentrations correlate well with the climate proxy δD. The current CO$_2$ concentration (390 ppm in 2007) is far higher than it has been over the past 650,000 years.
(*Source:* Data from Siegenthaler et al., 2005, and Spahni et al., 2005.)

with the δD temperature surrogate. During glacial periods, the greenhouse gases CO_2 and CH_4 are low; during the warmer interglacial periods, they are high. Perhaps the most startling feature in the figure is how much higher the current CO_2 concentration is (390 ppm in 2007) than it has been in over 650,000 years. What the data do not make clear is whether it is changing concentrations of greenhouse gases that cause temperature changes or vice versa. Switches from glacial to interglacial periods may well be triggered by other phenomena, such as natural variations in the Earth's orbit, coupled with positive feedbacks between warming and increased greenhouse gas emissions.

In the opposite hemisphere, a number of cores have been drilled into the the 3-km-thick glacial cap in Greenland. The Greenland data show cycles of gradual cooling, followed by rapid warming, during the last glacial period. These cycles, which have a period of between 500 and 2,000 years, are called *Dansgaard-Oeschger* events after their discoverers. They seem to be linked to another phenomenon, called *Heinrich* events, in which it is thought that enormous flotillas of icebergs periodically drifted across the North Atlantic; their melting deposited unusual layers of sediments on the seafloor that were used to identify and date these events. One of the most startling observations from Greenland ice cores is the speed with which large, *regional* temperature jumps can occur. During the last glacial period, 10,000 to 110,000 years ago, the temperature of Greenland showed jumps of 5 to 7°C within just a decade or two. Global temperatures, on the other hand, do not seem to have changed nearly that rapidly in the past. Over the past 10,000 years (the present interglacial period), it seems unlikely that global mean temperatures have ever changed faster than 1°C per century (IPCC, 1995, 3.6.3).

More Recent Global Temperature

While ice cores and seafloor sediments provide indirect methods of estimating temperatures over many hundreds of thousands of years of Earth history, ever since the mid-nineteenth century, the temperature record has been based on actual measurements. Most of the data are from actual thermometer readings taken at weather stations across the globe, including ships at sea, augmented more recently with satellite data. Adjustments have been made to discount the impact of warming caused by the growth of cities around meteorological stations. Dark surfaces in urban areas absorbing heat from the sun, along with waste heat from energy consumption, cause elevated temperatures referred to as the urban heat island effect. To offset the bias caused by urban warming, nighttime satellite images are used to identify stations outside of brightly lit urban areas. Greater reliance is then placed on data sets taken from those unlit areas and other rural areas across the globe. After these adjustments, the remaining urban influence on the global record is estimated to be less than 0.1°C (Hansen et al., 2001).

Figure 4 shows the Goddard Institute for Space Sciences (GISS) estimate of the mean annual global temperature expressed as a deviation, or anomaly, from a fixed reference temperature. In this graph, the reference is the 1951–1980 average global temperature, estimated to be 14°C. Eleven of the years between 1995 and 2006 rank among the 12 warmest years in the instrumental record of global surface temperature (IPCC, 2007). The current average temperature is probably the highest it has been since the beginning of the current interglacial period, which began some

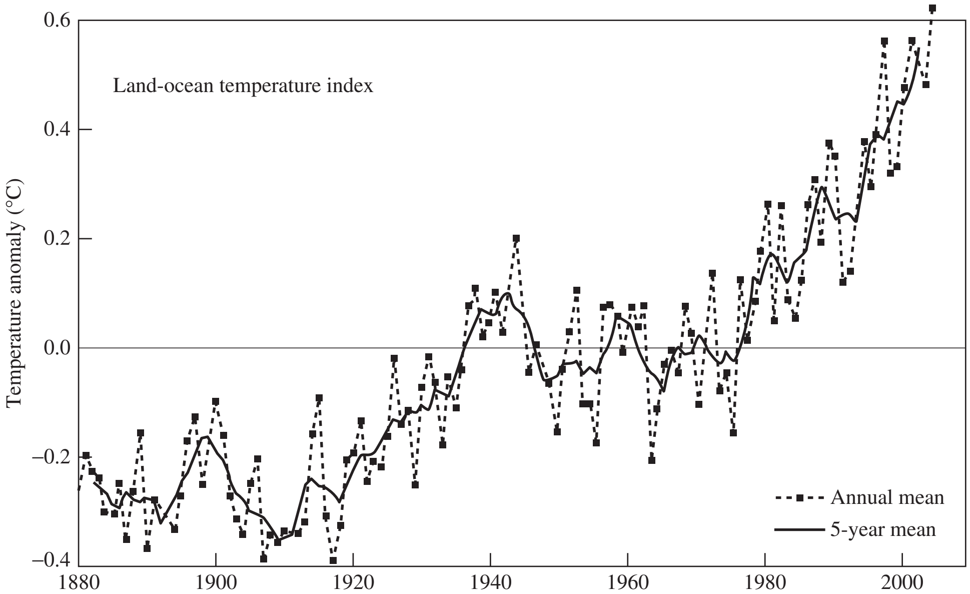


FIGURE 4 Global mean annual temperature anomaly referenced to the 14°C estimated average for 1951–1980. The Earth has warmed about 0.8°C since 1990.
(*Source:* NASA Goddard Institute for Space Studies Web site, 2006.)

12,000 years ago. While the overall temperature trend has been steeply upward, the period between about 1940 and 1980 shows an actual decline in temperature. That cooling is attributed to particulates, especially sulfates, caused by rapidly growing fossil fuel combustion during a period of time when emission controls were not yet in widespread use. Those aerosols reflecting sunlight back into space are thought to have masked the ongoing global warming caused by greenhouse gas emissions. When power plant emissions began to be controlled, global warming reemerged and the post-1980 temperature increases have averaged about 0.28°C per decade. The more recent 1992–1994 cool spell is attributed to the increased reflection of sunlight off the atmosphere caused by the eruption of Mount Pinatubo in the Philippines, in June 1991.

Attempts have been made to combine the measured temperature anomaly with other temperature proxies such as historical documents, tree rings, corals, ice cores, and lake sediments to reconstruct patterns of the past thousand or so years of temperature change. These attempts focus on the Northern Hemisphere mean temperature since there isn't much data available for the rest of the globe. One such study is shown in Figure 5. Several periods of time are sometimes noted, including a "Medieval Warm Period" from roughly 900 to 1300, during which some regions of the globe experienced unusual warmth, and the "Little Ice Age" from about 1500 to 1900, when the opposite occurred.

The overall shape of the 1,000-year profile shown in Figure 5 has picked up the nickname of the temperature "hockey stick." The hockey stick has been a source of some controversy, with some arguing that the current temperature and rate of change of temperature might not be without precedent in the past 1,000 years, and therefore the warming we are now observing may be a natural phenomenon. The

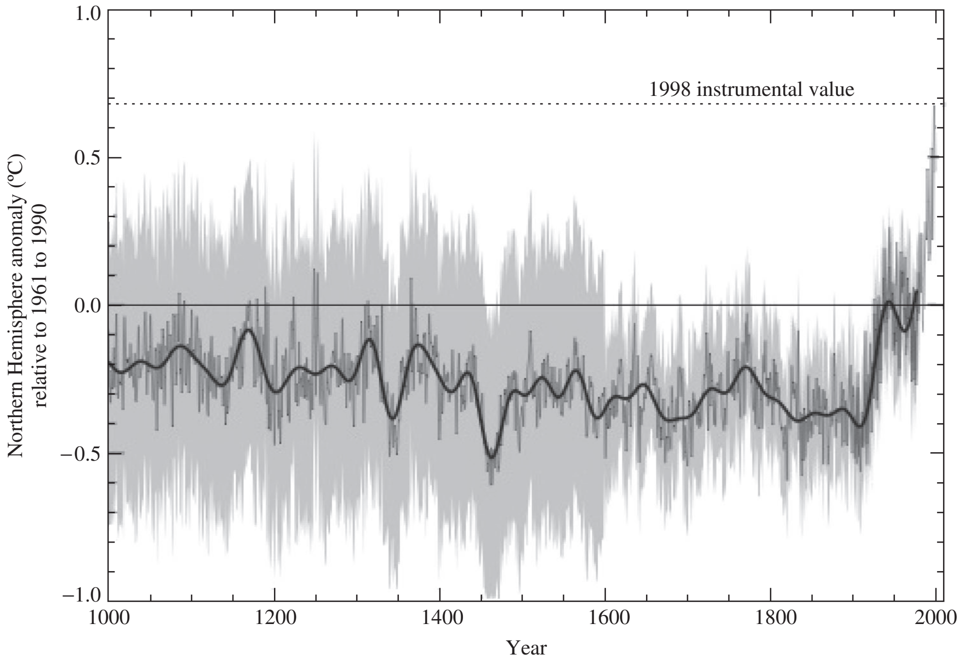


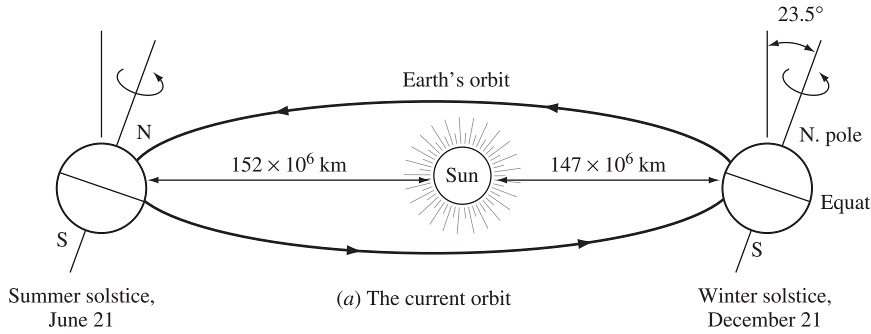
FIGURE 5 The Northern Hemisphere temperature anomaly, sometimes referred to as the "hockey stick."
(*Source:* IPCC, 2001.)

data set from which the graph has been drawn has enough uncertainty in it to keep this controversy alive, but the National Research Council took up the issue in 2006 and concluded that there is a high level of confidence that the global mean surface temperature during the past few decades is higher than at any time over the preceding four centuries, but beyond that there is less confidence. Whether or not the warming is unprecedented, the current rapid increase in anthropogenic greenhouse gas emissions, the well understood physics and chemistry of the atmosphere, and direct measurements of greenhouse gas impacts, all tell us the climate is changing, and we have a lot to do with causing that change.

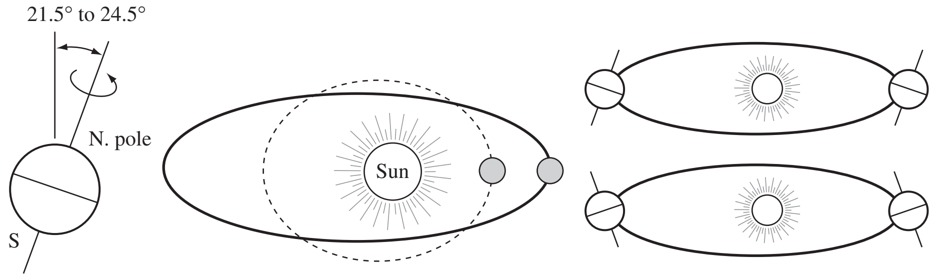
Orbital Variations and Sunspots

The history and future of the Earth's orbit around the sun can be calculated precisely. The connection between orbital variations and climate were first proposed in the 1930s by an astronomer, Milutin Milankovitch, and the orbital cycles are now referred to as *Milankovitch* oscillations. Changes in the orbit affect the amount of sunlight striking the Earth as well as the distribution of sunlight both geographically and seasonally. Those variations are thought to be quite influential in the timing of the coming and going of ice ages and interglacial periods.

There are three primary orbital cycles. The shape of the Earth's orbit oscillates from elliptical to more nearly circular with a period of 100,000 years (*eccentricity*). The Earth's tilt angle with respect to its orbit fluctuates from 21.5° to 24.5° with a period of 41,000 years (*obliquity*). Finally, there is a 23,000-year period associated



(*a*) The current orbit



(*b*) Obliquity 41,000 years (*c*) Eccentricity 100,000 years (*d*) Precession 23,000 years

FIGURE 6 Orbital variations affect the timing of ice ages: (*a*) the current orbit; (*b*) the tilt angle variation, with period 41,000 years; (*c*) eccentricity variation, with period 100,000 years; (*d*) precession, with period 23,000 years.

with the precession, or wobble, of the Earth's spin axis (*precession*). This precession determines where in the Earth's orbit a given hemisphere's summer occurs. Figure 6 illustrates these variations.

Careful analysis of the historical record of global temperatures does show a primary cycle between glacial episodes of about 100,000 years, mixed with secondary oscillations with periods of 23,000 years and 41,000 years that match the Milankovitch theory reasonably well. Although these orbital variations only change the total annual dosage of sunlight by about 0.1 percent, the impacts on seasons and the resulting patterns of oceanic and atmospheric heat distribution around the globe are thought to be significant enough to trigger major climate changes.

Another factor that affects the amount of solar radiation reaching the top of the Earth's atmosphere is variations in the intensity of radiation emitted from the sun itself. For example, the sun spins on its axis making one complete rotation every 27 days. There are darker and brighter areas on the solar surface that cause variations of up to 0.2 percent in the amount of solar radiation reaching the Earth over that 27-day period. Those variations occur so rapidly, however, that they are not thought to be particularly important in determining average global climate.

Of greater importance is an 11-year cycle of sunspots that were first described by an amateur astronomer, Heinrich Schwabe, in 1843 (Figure 7). During peak periods of magnetic activity on the sun, the surface has large numbers of cooler, darker regions, called *sunspots,* that in essence block solar radiation, accompanied by other regions, called *faculae,* that are brighter than the surrounding surface. The

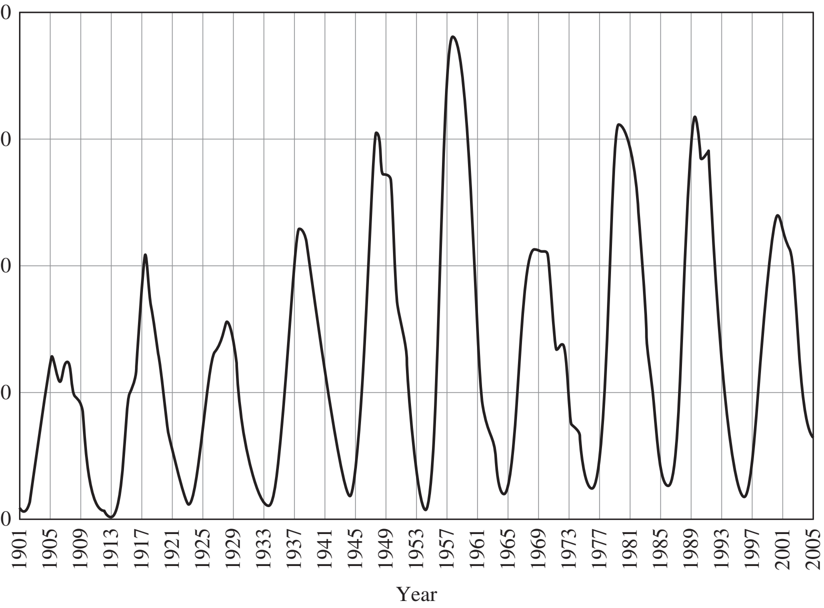


FIGURE 7 Mean annual sunspot number showing the 11-year Schwabe cycle.

net effect of sunspots that dim the sun, and faculae that brighten it, is an increase in solar intensity during periods of increased numbers of sunspots. The variation in solar radiation reaching the Earth as a result of these cycles is roughly 0.1 percent, which is thought to be enough to change the Earth's temperature by roughly 0.2°C. Sunspot activity helps provide an explanation for the continuous up and down jiggling of the Earth's temperature that complicates the problem of deciding whether a true global warming "signal" is emerging from the "noise."

A Simple Global Temperature Model

Measurements of historical global temperatures show that our planet has maintained its average temperature within a limited range, but within those bounds, it seems to be continuously changing. If we hope to predict future impacts of anthropogenic changes in our environment, we need to develop mathematical models that explain the past. Mathematically modeling of global climate and predicting the impacts of changes in key environmental parameters is an extremely important but difficult task. Such models range from very simple back-of-the-envelope calculations, to complex, three-dimensional *General Circulation Models* (GCMs, also known as *Global Climate Models*) that attempt to predict climate on a regional, seasonal, and annual basis. The most sophisticated of these models can take weeks to run on a supercomputer, yet they must still be considered primitive. In comparison, our treatment here is just the briefest of introductions. For a much more thorough introduction, see McGuffie and Henderson-Sellers (2005).

The simplest starting point for modeling climate begins with models that focus on factors influencing the single parameter, *temperature*. Obviously, other factors

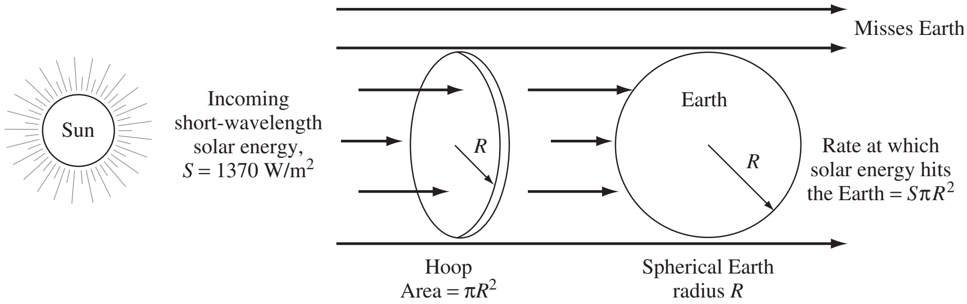


FIGURE 8 Solar energy passing through a "hoop" with the same radius as that of the Earth, hits the Earth. Radiation that misses the hoop also misses the Earth.

such as precipitation patterns, winds and storms, ocean currents, soil moisture, sea ice, glacial cover, and so forth are exceedingly important, but they are more difficult to approach with simple models. Even beginning with just temperature as the single quantity of interest, we could try to find out how temperature varies in all four dimensions (latitude, longitude, altitude, time), which is the domain of very complex GCMs. At the other end of the complexity scale is a simple "zero-dimensional" model in which a single average global temperature is obtained that is not a function of location or time. The following is such a model.

The basic zero-dimensional energy balance model equates solar energy absorbed by the Earth with the energy that the Earth radiates back to space. Radiation from the sun arrives just outside the Earth's atmosphere with an average annual intensity, called the *solar constant, S*, currently equal to about 1,370 watts/m^2. A simple way to calculate the total rate at which energy hits the Earth is to note that all of the flux passing through a "hoop" having radius equal to that of the Earth, and placed normal to the incoming radiation, strikes the Earth's surface. From Figure 8, we can write

$$\text{Rate at which solar energy strikes the Earth} = S\pi R^2 \text{ (watts)} \tag{2}$$

where

S = the solar constant, taken to be 1,370 W/m^2
R = the radius of the Earth (m)

Some of the incoming solar energy that hits the Earth is reflected back into space, as shown in Figure 9. Such reflected energy is not absorbed by the Earth or its atmosphere and does not contribute to their heating. The fraction of incoming solar radiation that is reflected is called the *albedo,* and for the Earth, the global annual mean value is now estimated to be about 31 percent. What isn't reflected is absorbed, which leads to the following expressions.

$$\text{Energy reflected by Earth} = S\pi R^2\alpha \tag{3}$$

$$\text{Energy absorbed by Earth} = S\pi R^2(1 - \alpha) \tag{4}$$

where α = the Earth's albedo, taken to be 0.31.

On the other side of the energy balance equation is the rate at which Earth sends energy back to space. Since there can be no heat transfer from the Earth to

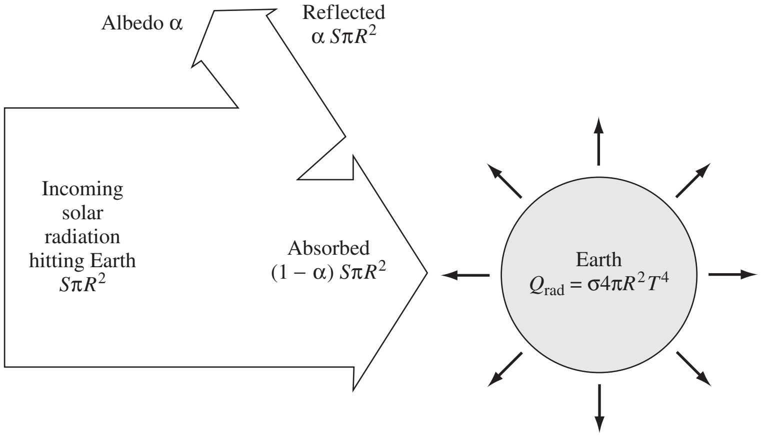


FIGURE 9 Simple global temperature model.

space by conduction or convection, the only way for the Earth to get rid of energy is by radiation. Every object radiates energy at a rate that is proportional to its surface area times its absolute temperature raised to the fourth power. For this model, we will assume the Earth is a blackbody; that is, it radiates as much as any object with the same temperature and area possibly can (emissivity = 1). We also assume it is isothermal; that is, the temperature is the same everywhere on the planet. Since the area of a spherical object is $4\pi R^2$, we can write

$$\text{Energy radiated back to space by earth} = \sigma 4\pi R^2 T_e^4 \tag{5}$$

where

σ = Stefan-Boltzmann constant = 5.67×10^{-8} W/m^2-K^4

and

T_e = earth's "effective" blackbody temperature (kelvins)

If we go on to assume steady-state conditions, that is, the Earth's temperature is not changing with time, we can equate the rate at which energy from the sun is absorbed (4), with the rate at which energy is radiated from Earth back to space (5):

$$S\pi R^2(1-\alpha) = \sigma 4\pi R^2 T_e^4 \tag{6}$$

Solving for T_e,

$$T_e = \left[\frac{S(1-\alpha)}{4\sigma}\right]^{1/4} \tag{7}$$

Substituting appropriate values into (7) yields

$$T_e = \left[\frac{1{,}370 \text{ W/m}^2(1-0.31)}{4 \times 5.67 \times 10^{-8} \text{ W/m}^2\text{K}^4}\right]^{1/4} = 254 \text{ K} = -19°\text{C} \tag{8}$$

Notice the conversion from kelvins to Celsius (°C = K − 273). The actual value of the Earth's average surface temperature is currently about 288 K (15°C). While we are off by only 12 percent, which might seem a modest error, in terms of life on Earth, the 254 K (−19°C) estimate for T_e is terribly wrong. We need to find an explanation for why the Earth is (fortunately) not that cold. The key factor that makes our model differ so much from reality is that it does not account for interactions between the atmosphere and the radiation that is emitted from the Earth's surface. That is, it does not include the greenhouse effect.

4 The Greenhouse Effect

The surface of the Earth is 34°C higher than what is predicted by (8). Wien's displacement rule gives the wavelength at which a blackbody spectrum peaks as a function of its absolute temperature:

$$\lambda_{\max}(\mu\text{m}) = \frac{2{,}898}{T(\text{K})} \tag{9}$$

The sun can be represented as a blackbody with temperature 5,800 K, so its spectrum peaks at 0.5 μm. The Earth, at 288 K, has its peak at 10.1 μm. Figure 10 shows these two spectra. The area under these curves, between any two wavelengths, is the total radiant energy in that range of frequencies. For example, the area under the curve of incoming solar radiation just outside the atmosphere is the total solar radiant flux; that is, it is the solar constant, 1,370 W/m^2. Notice nearly all the incoming solar energy as it arrives just outside the Earth's atmosphere has wavelengths less than 3 μm, while the outgoing energy radiated by the Earth has essentially all of its energy in wavelengths greater than 3 μm. With so little overlap, it is convenient to speak of solar energy as being *short-wavelength* radiation, while energy radiated from the Earth's surface is *long-wavelength,* or *thermal,* radiation. The infrared (IR) portion of the spectrum begins at about 0.7 μm and extends out to 100 μm, so some of the incoming solar radiation and all of the outgoing thermal radiation is IR.

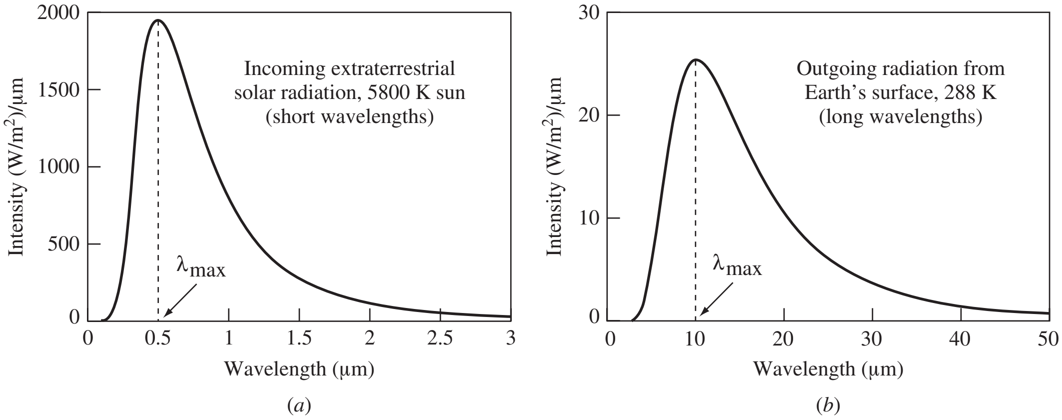
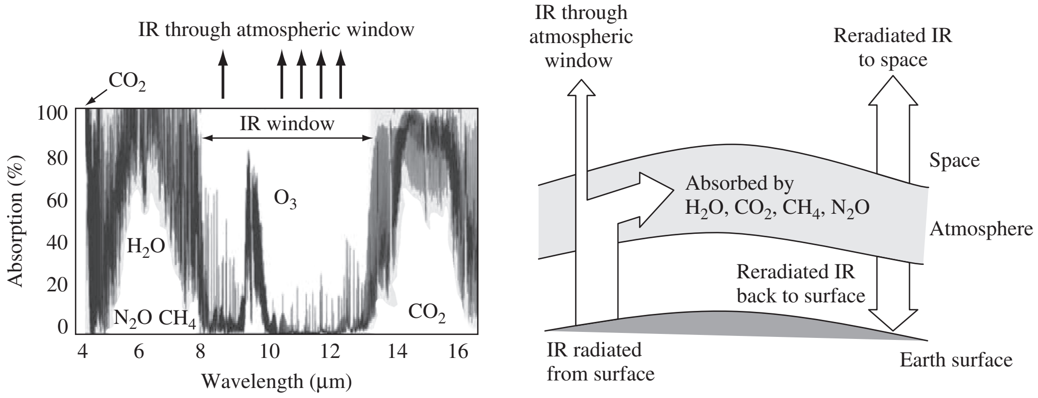


FIGURE 10 Blackbody radiation at 5,800 K and 288 K. (*a*) Incoming solar radiation just outside of the Earth's atmosphere. (*b*) Radiation from the Earth's surface at 288 K.



(*a*) (*b*)

FIGURE 11 Some IR radiated from the Earth's surface passes through a relatively clear IR window between about 7 and 13 μm (*a*). Most, however, is absorbed by greenhouse gases in the atmosphere (*b*).

As radiant energy attempts to pass through the atmosphere, it is affected by various gases and aerosols in the air. Those atmospheric constituents can let the radiant energy pass through unaffected, they can scatter the energy by reflection, or they can stop it by absorption. The key phenomenon of interest here is the ability of gases to absorb radiant energy. As the atoms in gaseous molecules vibrate toward and away from each other (vibrational energy) or rotate around each other (rotational energy), they absorb and radiate energy in specific wavelengths. When the frequency of these molecular oscillations is close to the frequency of the passing radiant energy, the molecule can absorb that energy. This absorption occurs over a rather limited range of frequencies, not just at the oscillatory frequency of the molecule, and results in an absorptivity spectrum, which is a plot of the fraction of incoming radiant energy that is absorbed as a function of wavelength.

Figure 11*a* shows the absorption spectra for the key greenhouse gases of concern in this chapter, along with the their effect on outgoing infrared (IR) radiation emitted by the Earth's surface. Most of the long-wavelength energy radiated by the Earth is absorbed by a combination of radiatively active gases, most important of which are water vapor (H_2O), carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), and ozone (O_3). Water vapor, which is by far the most important greenhouse gas, strongly absorbs thermal radiation with wavelengths less than 8 μm and greater than 18 μm. Carbon dioxide shows a strong absorption band centered at 15 μm, as well as bands centered at 2.7 μm and 4.3 μm. Between 7 μm and 13 μm there is a relatively clear sky for outgoing thermal radiation, referred to as the *atmospheric radiative window*. Radiation in those wavelengths easily passes through the atmosphere, except for a small, but important, absorption band between 9.5 μm and 10.6 μm associated with O_3.

Radiatively active gases that absorb wavelengths longer than 4 μm are called greenhouse gases. This absorption heats the atmosphere, which, in turn, radiates energy back to the Earth as well as out to space, as shown in Figure 11*b*. These greenhouse gases act as a thermal blanket around the globe, raising the Earth's surface temperature beyond the equivalent temperature calculated earlier. The importance of

TABLE 2

Application of the Simple Model (7) to Compute Effective Temperatures, Compared with Actual Surface Temperatures

Planet	Distance to Sun (10^6 km)	Atmosph. Pressure (atm)	Solar Constant S (W/m^2)	Albedo α (%)	Effective Temp. T_e (K)	Surface Temp. T_s (K)	Greenhouse Warming (°C)
Venus	108	90	2,620	76	229	750	521
Earth	150	1	1,370	31	254	288	34
Mars	228	0.006	589	25	210	218	8

Note: Mars, with little atmosphere, shows almost no greenhouse effect while on Venus it is quite pronounced.
Source: Hoffert, 1992.

water vapor as a greenhouse gas is quite evident on clear nights, when the Earth cools much more rapidly than it does on cloudy nights. It is also interesting to note that the term *greenhouse effect* is based on the concept of a conventional greenhouse with glass acting much like the aforementioned gases. Glass, which easily transmits short-wavelength solar energy into the greenhouse, absorbs almost all of the longer wavelengths radiated by the greenhouse interior. This radiation trapping is partly responsible for the elevated temperatures inside the greenhouse, although much of the effect is simply due to the reduction in convective cooling of the interior space caused by the enclosure. The elevated interior temperature of your car after it has been parked in the sun is another common example of the greenhouse effect.

If the Earth did not already have a greenhouse effect, its temperature would be 254 K as predicted by (8). That is, the planet would have an average temperature of −19°C, or about −2°F. In fact, one way to quantify the magnitude of the greenhouse effect is to compare the effective temperature, T_e, given in (7), with the actual surface temperature, T_s

$$\text{Magnitude of greenhouse effect} = T_s - T_e \tag{10}$$

Thus, since the actual temperature of the Earth is 288 K, and its effective temperature is 254 K, we can say that the greenhouse effect adds 34°C of warming to the surface of the Earth.

In Table 2, this notion is applied to Venus and Mars. Even though the atmosphere of Mars is almost entirely carbon dioxide, there is so little atmosphere that the greenhouse effect is barely apparent. The atmospheric pressure on Venus, on the other hand, is nearly 100 times that of Earth, and its atmosphere is 97 percent CO_2. The greenhouse effect on Venus is correspondingly very pronounced. It is interesting to note that without the greenhouse effect, the greater albedo of Venus would make it cooler than the Earth in spite of its closer proximity to the Sun.

5 Global Energy Balance

As suggested in Figure 11, a simple model of the greenhouse effect treats the Earth's surface, the atmosphere, and outer space as three separate regions. Figure 12 expands the diagram to include incoming solar energy as well as outgoing IR and

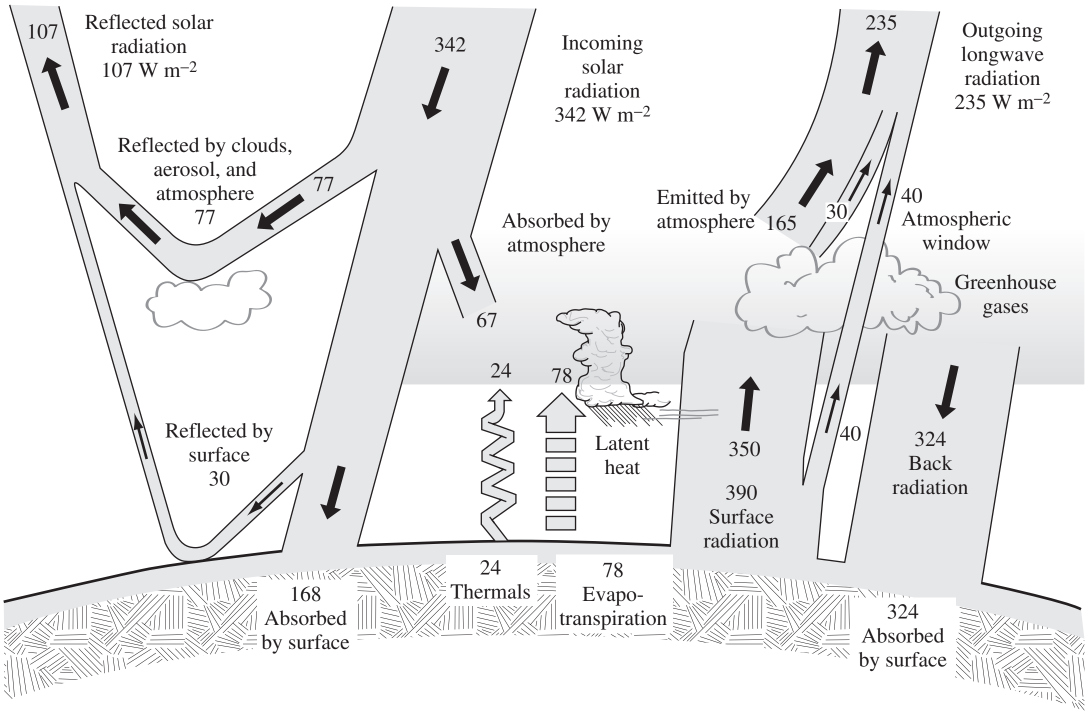


FIGURE 12 Average energy flows among space, the atmosphere, and the Earth's surface for a global equilibrium model. Units are watts per square meter of surface area. Values given are from Kiehl and Trenberth (1996) as they appear in IPCC (1996).

adds some quantitative detail to help us track energy flows. Some of this detail is relatively easy to derive. Begin by normalizing energy flows between these three regions by expressing them in terms of rates per unit of surface area of the Earth. For example, (2) indicates that the total amount of solar radiation striking the Earth is $S\pi R^2$. Distributed over the entire surface of the Earth, the average incoming solar radiation is equal to

$$\frac{\text{Incoming solar radiation}}{\text{Surface area of Earth}} = \frac{S\pi R^2}{4\pi R^2} = \frac{S}{4} = \frac{1{,}370\ \text{W/m}^2}{4} = 342\ \text{W/m}^2 \quad (11)$$

Since the albedo is 31 percent, the amount of incoming radiation reflected back into space per square meter of the Earth's surface is

$$\begin{aligned}\frac{\text{Solar energy reflected}}{\text{Surface area of Earth}} &= \frac{S\pi R^2\alpha}{4\pi R^2} = \frac{S}{4}\alpha \\ &= 342\ \text{W/m}^2 \times 0.31 = 107\ \text{W/m}^2\end{aligned} \quad (12)$$

Of this 107 W/m^2, it is estimated that 77 W/m^2 is reflected off the atmosphere itself, while the remaining 30 W/m^2 is reflected off the Earth's surface. The solar radiation that is not reflected is absorbed by the Earth and its atmosphere. Calling that absorbed energy Q_{abs} (again, with units of watts per square meter of surface) gives

$$\begin{aligned}\frac{\text{Solar radiation absorbed}}{\text{Surface area of Earth}} &= Q_{\text{abs}} = \frac{S\pi R^2(1-\alpha)}{4\pi R^2} = \frac{S}{4}(1-\alpha) \\ &= 342\ \text{W/m}^2 \times (1-0.31) = 235\ \text{W/m}^2\end{aligned} \tag{13}$$

Of that 235 W/m^2, 67 W/m^2 is absorbed by the atmosphere, and the remaining 168 W/m^2 is absorbed by the surface of the Earth.

If we assume that global temperatures are unchanging with time, then the rate at which the Earth and its atmosphere receive energy from space must equal the rate at which energy is being returned to space. The 107 W/m^2 of reflected energy is already balanced; that is, 107 W/m^2 hits the Earth/atmosphere, and 107 W/m^2 is reflected back into space so we can ignore that component for now. The Earth and its atmosphere absorb the remaining 235 W/m^2, so the same amount must be radiated back into space. If the Earth's surface were at 254 K, it would radiate 235 W/m^2, which is just enough to balance the incoming energy. We know, however, that greenhouse gases would absorb most of that outgoing 235 W/m^2, so the required energy balance would not be realized. Therefore, to force enough energy through the atmosphere to create the necessary balance, the temperature of the Earth's surface must be higher than 254 K.

If we treat the Earth as a blackbody radiator, we can use (4) to estimate the rate at which energy is radiated from the Earth's surface toward the atmosphere. With the surface of the Earth at its actual temperature of 288 K, it will radiate the following amount per unit of surface area.

$$\begin{aligned}\frac{\text{Energy radiated by surface}}{\text{Surface area of Earth}} &= \frac{\sigma 4\pi R^2 T_s^4}{4\pi R^2} = \sigma T_s^4 \\ &= 5.67 \times 10^{-8}\ \text{W/m}^2\text{K}^4 \times (288\ \text{K})^4 = 390\ \text{W/m}^2\end{aligned} \tag{14}$$

Of that 390 W/m^2, only 40 W/m^2 passes directly through the atmosphere, mostly through the atmospheric radiative window between 7 and 13 μm. The remaining 350 W/m^2 is absorbed by greenhouse gases in the atmosphere. The atmosphere then radiates 324 W/m^2 back to the surface.

There is also heat transfer from the surface to the atmosphere by convective heating, evaporation, and condensation of water. Convection transfers 24 W/m^2 to the atmosphere, while condensation of water vapor provides 78 W/m^2 of latent heat.

All of these energy flows are shown in Figure 12. If this model is an internally self-consistent equilibrium model, the rate of energy gain should equal the rate of energy loss in each of the three regions: space, the atmosphere, and the Earth's surface. Consider the following checks:

	Rate of energy gain = Rate of energy loss?	
Earth's surface:	$168 + 324 + 30 = 78 + 24 + 30 + 390$	(checks)
Atmosphere:	$67 + 78 + 24 + 350 = 165 + 30 + 324$	(checks)
Space:	$107 + 165 + 30 + 40 = 342$	(checks)

So, the model shows the necessary balances.

6 Carbon Dioxide: The Principal Greenhouse Gas

Thus far, the greenhouse effect has been described as a natural phenomenon that is responsible for Earth having an average surface temperature 34°C warmer (288 K vs 254 K) than it would have if it didn't have radiatively active gases in the atmosphere. As is now well known, anthropogenic sources of a number of gases and aerosols are enhancing the natural greenhouse effect, leading us into a future of uncertain global climate. The principal greenhouse gases are carbon dioxide, methane, nitrous oxide, and a category of carbon-based gases called *halocarbons,* which are molecules that contain atoms of carbon plus fluorine, chlorine, and/or bromine. In addition, emissions of black carbon, or soot, particles also add to atmospheric warming. We begin our exploration of these gases and particulates with the most important one, carbon dioxide. Carbon dioxide has been recognized for its importance as a greenhouse gas for over a century. The Swedish chemist Svante August Arrhenius is usually credited with the first calculations on global temperature as a function of atmospheric CO_2 content, and his results are not that far from those obtained today.

Atmospheric Concentration of CO_2

The first continuous, precise, and direct measurements of atmospheric carbon dioxide began in 1957 at the South Pole, and 1958 at Mauna Loa, Hawaii. At that time, the concentration in Mauna Loa was 315 ppm and growing at less than 1 ppm per year. In 1979, NOAA began assembling data from a network of sites to determine a global average value. The global average in 1990 was 354 ppm and growing at 1.6 ppm per year. In 2006, it reached 380 ppm, and the rate of growth had increased to about 2 ppm/yr.

Figure 13 shows NOAA data for monthly global CO_2 for the recent past. The oscillations are caused by seasonal variations in the rates of photosynthesis and respiration. During photosynthesis, carbon is transferred from the air into plant material (indicated by the carbohydrate, glucose, in the following reaction). During spring and summer, when plants are growing their fastest, atmospheric CO_2 levels drop, tending to reach their lowest point in the Northern Hemisphere in about October.

$$\text{Photosynthesis:}\quad 6\ CO_2 + 6\ H_2O \xrightarrow{\text{sunlight, chlorophyll}} C_6H_{12}O_6 + 6\ O_2 \qquad (15)$$

Reversing the preceding reaction yields the equation describing respiration, which is the process that living things use to obtain energy. During respiration, complex organic molecules are broken down, returning carbon to the atmosphere. When the rate of respiration exceeds the rate of photosynthesis, as tends to occur in the fall and winter seasons, there is a net replacement of carbon into the atmosphere, which globally results in peak concentrations around May.

$$\text{Respiration: } 6\ C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O + \text{energy} \qquad (16)$$

Carbon thus moves continually from the atmosphere into the food chain during photosynthesis and returns to the atmosphere during respiration.

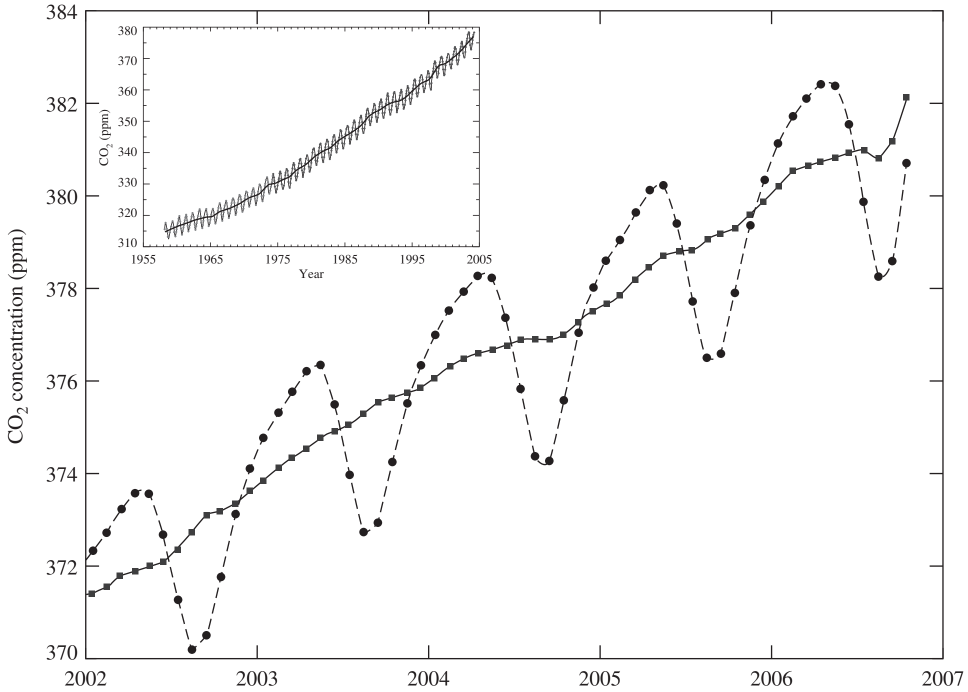


FIGURE 13 Recent global CO_2 concentrations. The oscillations are month-by-month mean values; the smoothed line is a moving average over 10 adjacent months. (*Source:* NOAA Web site http://www.cmdl.noaa.gov/ccgg/trends/index.php#mlo, 2006.)

Atmospheric carbon dioxide concentrations inferred from Antarctic ice cores and other evidence over the past 1,000 years, combined with more recent direct measurements, are shown in Figure 14. Over most of that time period, the concentration of carbon dioxide hovered at close to 280 ppm, and that is the value that is commonly used as a reference point for comparison with current readings and future projections. Carbon dioxide concentrations are now more than one-third higher than they were just before the industrial revolution.

The Carbon Cycle

The atmosphere contains about 800 GtC, where 1 GtC means 1 gigaton of carbon (10^9 metric tons or 10^{15} g). Since almost all of that carbon is stored in the form of CO_2 (less than 1 percent is in other carbon-containing compounds such as methane and carbon monoxide), in most circumstances it is reasonable to assume that atmospheric carbon is entirely CO_2. The amount of carbon locked up in terrestrial vegetation (610 GtC) is of the same order of magnitude as that in the atmosphere, but both of these amounts are dwarfed by the 39,000 GtC stored in the oceans.

Natural processes continuously transport enormous amounts of carbon back and forth among the atmosphere, biosphere, and the oceans. The carbon flux into and out of the atmosphere during photosynthesis and respiration is on the order of 60 GtC/yr. The oceans absorb around 90 Gt/C/yr and store almost all of it in the

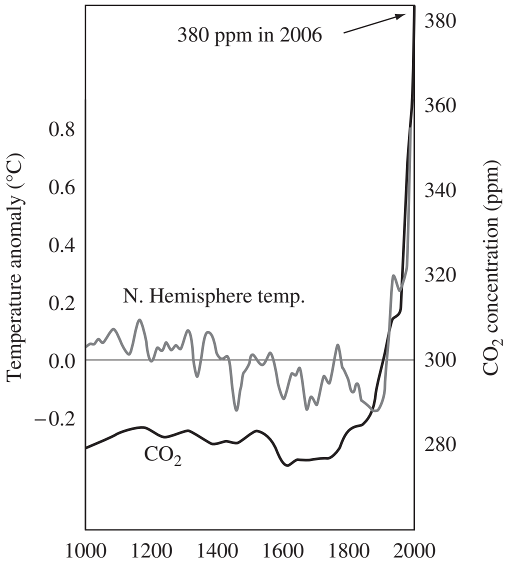


FIGURE 14 Carbon dioxide concentration and Northern Hemisphere temperature over the past 1,000 years.

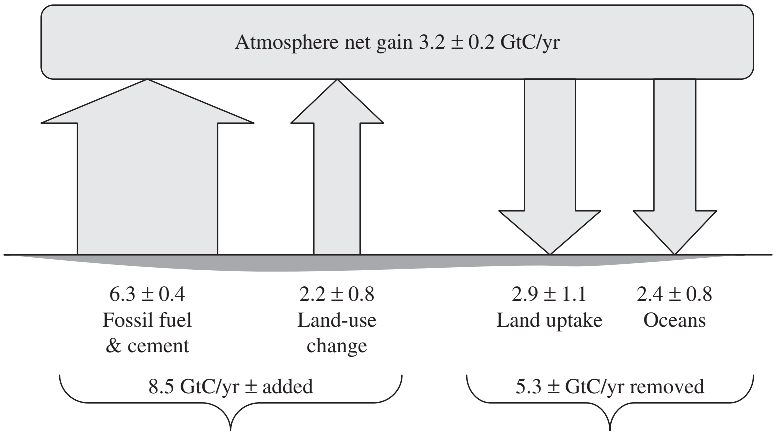


FIGURE 15 Human perturbations to the global carbon cycle during the 1990s. (*Source:* Based on data from Houghton, 2003.)

form of bicarbonate ions (HCO_3^-), but some becomes part of the marine food chain. A similar quantity is returned to the atmosphere. A very small portion of nonliving organic matter each year ends up in sediments. The slow, historical accumulation of that organic carbon is the source of our fossil fuels—oil, natural gas, and coal. When these are burned, ancient carbon is returned to the atmosphere.

By comparison with the natural fluxes of carbon, the additional amounts added to the atmosphere by combustion, cement production, and changes in land use are modest, but they are enough to cause a significant response by the climate system. Figure 15 summarizes the impact of human perturbations to the carbon fluxes into and out of the atmosphere during the 1990s. Fossil fuel combustion and cement production deliver 6.3 GtC/yr to the atmosphere, while land use changes, such as biomass burning and harvesting of forests, add another 2.2 GtC/yr. Not all of that 8.5 GtC/yr remains in the atmosphere, however. Greater carbon uptake through purposeful reforestation efforts, as well as stimulated plant growth caused by higher CO_2 levels and increased nitrogen deposition on the soils from fossil fuel combustion, removes around 2.9 GtC/yr. Finally, higher atmospheric CO_2 increases the oceans' absorption of carbon, providing an additional 2.4 GtC/yr sink.

As Figure 15 suggests, fossil fuel combustion and land-use changes in the 1990s added about 8.5 GtC/yr to the atmosphere. Of that amount, 5.3 GtC/yr was returned to the oceans or other terrestrial sinks, leaving about 3.2 GtC/yr remaining in the atmosphere. The ratio of the amount of anthropogenic carbon emitted to the amount that remains in the atmosphere is known as the *airborne fraction*. Using these data, the airborne fraction has been:

$$\text{Airborne fraction} = \frac{3.2\ \text{GtC/yr remaining in atmosphere}}{8.5\ \text{GtC/yr anthropogenic additions}} = 0.38 = 38\% \qquad (17)$$

Thus, roughly speaking, somewhat less than half of the carbon we emit stays in the atmosphere. But the airborne fraction is not necessarily a fixed quantity. For example, if large areas of land are deforested, the ability of the biosphere to absorb carbon would be reduced, and the atmospheric fraction would increase. Likewise, CO_2 fertilization of terrestrial biomass can stimulate plant growth, which increases the rate of removal of atmospheric carbon, so the airborne fraction could get smaller. The airborne fraction also depends on how fast carbon is being added to the atmosphere. For scenarios with little or no growth in emissions or even declining emissions, the oceans and plants have more time to absorb carbon, so the atmospheric fraction could be lower. On the other hand, for rapidly increasing emission rates, carbon sinks cannot keep up and the fraction remaining in the atmosphere may be higher.

The following example develops another useful relationship, this time between the concentration of CO_2 and the tons of carbon in the atmosphere. This ratio, coupled with an estimate of the airborne fraction, provides the key to predicting future CO_2 concentrations for various carbon emission scenarios.

EXAMPLE 2 Carbon Content of the Atmosphere

Find a relationship between the concentration of carbon dioxide and the total amount of carbon in the atmosphere. The total mass of the atmosphere is estimated to be 5.12×10^{21} g.

Solution We will need to know something about the density of air. That, of course, varies with altitude, but finding it under some particular conditions will work. From Table 1 we know the concentration of each gas in air. Recall from Section 1.2 that 1 mole of each gas occupies $22.414 \times 10^{-3}\ \text{m}^3$ at Standard

Temperature and Pressure (0°C and 1 atmosphere), which is 44.61 mol/m^3. The following table organizes the calculation:

Gas	$\frac{\text{m}^3\text{ gas}}{\text{m}^3\text{ air}}$	×	g/mol	×	$\frac{\text{mol}}{\text{m}^3\text{ gas}}$	=	$\frac{\text{g}}{\text{m}^3\text{ air}}$
N_2	0.7808		28		44.61		975.3
O_2	0.2095		32		44.61		299.1
Ar	0.0093		40		44.61		16.6
CO_2	0.00038		44		44.61		0.75
						Total	1291.8 g/m^3

If all of the atmosphere were at standard temperature and pressure, it would have a density of 1291.8 g/m^3, and its mass would still be 5.12×10^{21} g. Putting these together gives

$$1\text{ ppm} = \frac{1\text{ m}^3\text{ CO}_2}{10^6\text{ m}^3\text{ air}} \cdot 44.61\frac{\text{mole}}{\text{m}^3\text{ CO}_2} \cdot 12\frac{\text{g C}}{\text{mole}} \cdot \frac{5.12 \times 10^{21}\text{ g air}}{1291.8\frac{\text{g air}}{\text{m}^3\text{ air}}} \cdot 10^{-15}\frac{\text{GtC}}{\text{g C}} = 2.12\text{ GtC}$$

Notice this calculation has taken advantage of the fact that volumetric concentrations (ppm) are independent of temperature or pressure.

From Example 2, we have the following very useful relationship:

$$1\text{ ppm CO}_2 = 2.12\text{ GtC} \tag{18}$$

For example, knowing the concentration of CO_2 in 2006 was 380 ppm, the total amount of carbon in the atmosphere can be estimated to be

$$380\text{ ppm} \times 2.12\text{ GtC/ppm} = 806\text{ GtC}$$

EXAMPLE 3 Estimating the Rate of Change of CO_2

Suppose global fossil fuel combustion emits 7.4 GtC/yr and cement production adds another 0.5 GtC. Assuming the an airborne fraction of 0.38 and assuming no change in emissions associated with land use, what rate of change in CO_2 concentration would you expect?

Solution Including the 2.2 GtC/yr land-use emissions from Figure 17 gives a total emission rate of 2.2 + 7.4 + 0.5 = 10.1 GtC/yr. Using the 0.38 airborne fraction along with the 2.12 GtC/ppm ratio gives

$$\Delta\text{CO}_2 = \frac{10.1\text{ GtC/yr} \times 0.38}{2.12\text{ GtC/ppm CO}_2} = 1.8\text{ ppm CO}_2\text{/yr}$$

Carbon Emissions from Fossil Fuels

Use of energy and the resulting emissions of carbon vary considerably from country to country. The United States, with less than 5 percent of the world's population, emits 22 percent of global energy-related CO_2. The second largest emitter is China,

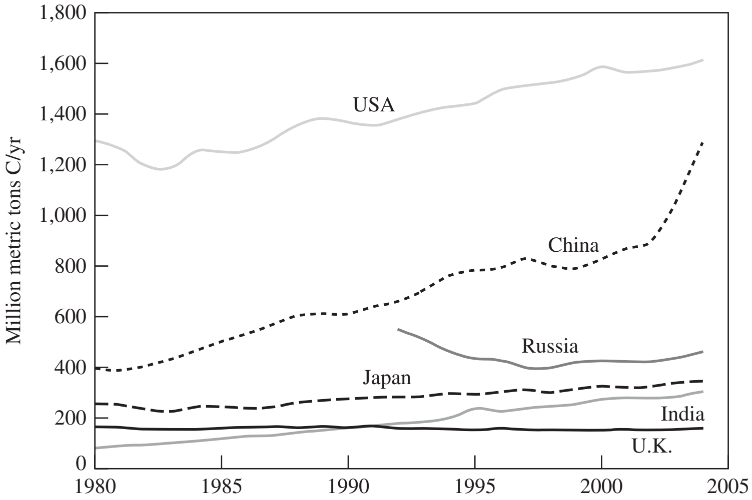


FIGURE 16 Fossil fuel carbon emissions.
(*Source:* EIA data, 2006.)

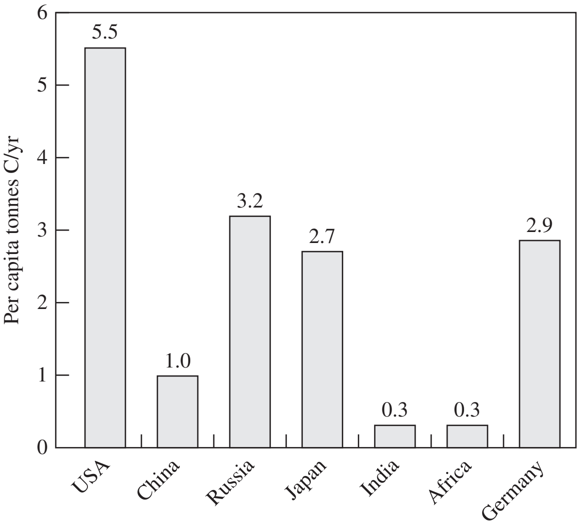


FIGURE 17 Per capita carbon emissions.
(*Source:* EIA data, 2006.)

and as Figure 16 indicates, it is very rapidly closing the gap with the United States. In fact, it is projected that China's emissions will exceed those of the United States by 2010. While China's emissions growth is certainly worrisome, two other measures provide some insight into the root cause of our global carbon problem. On a per capita basis, the United States emits far more carbon than any other country—roughly double that of most other advanced countries and more than five times as much as China (Figure 17). Moreover, it is the accumulated emissions from developed countries that are overwhelmingly responsible for the rising CO_2 concentrations in the atmosphere (Figure 18).

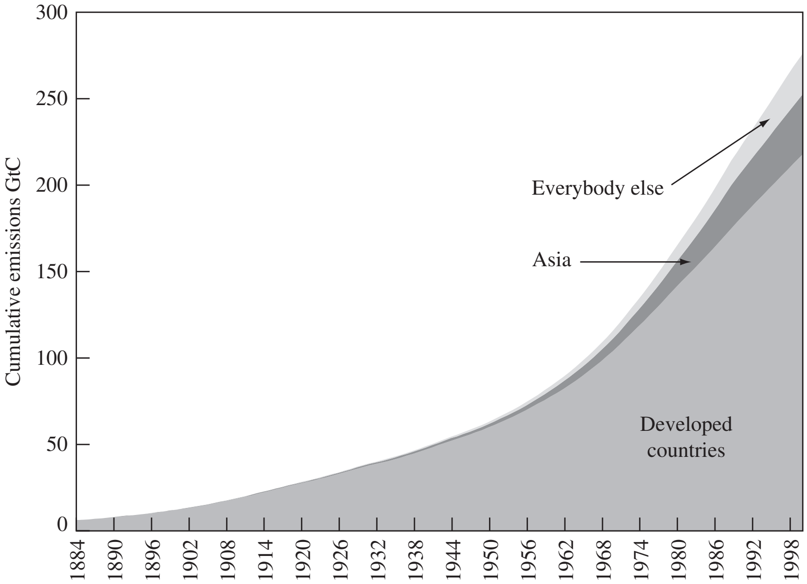


FIGURE 18 Cumulative carbon emissions by region.
(*Source:* E. Wanless with data from G. Marland, Oak Ridge National Labs.)

Carbon Emissions from Industrial Processes

A number of manufacturing processes result in carbon emissions on site that are not included in the usual accounting for fossil fuel combustion. The primary source of these industrial emissions is the calcination of limestone ($CaCO_3$) to create lime (CaO). These two compounds are basic materials in the production of cement, iron and steel, and glass. Other industrial emissions result from the production and use of soda ash (Na_2CO_3), the manufacture of carbon dioxide, and the production of aluminum.

The largest single industrial CO_2 source, however, results from cement production. Concrete, which is probably the most important building material in the world, is made up of a mixture of Portland cement, fine aggregate (sand), coarse aggregate (crushed stone), and water. Portland cement, which is the binding agent for concrete, derived its name in the early nineteenth century from its similarity to a particular type of stone found on the Isle of Portland in Dorset, England. It is typically on the order of 12 to 15 percent by weight of the concrete mix.

The major raw material needed for the manufacture of Portland cement is limestone ($CaCO_3$), along with a source of silicon such as clay or sand. Processing of those materials is done in high-temperature kilns, usually fired with fossil fuels whose combustion leads to CO_2 emissions. In addition, the calcination of $CaCO_3$ into lime (CaO) that occurs in those kilns emits its own CO_2 as the following reaction suggests:

$$\text{Limestone } (CaCO_3) + \text{Clay or Sand } (Si) + \text{Heat} \rightarrow \text{Cement } (CaO, SiO_2) + CO_2 \tag{19}$$

The average intensity of carbon emissions from cement production is about 0.222 tons of carbon per ton of cement, with about half of that being the result of calcinations and half released during combustion (Worrell et al., 2001). In addition, power plants supplying electricity for plant operations emit their own carbon, but that is usually not included in the category of industrial process emissions. All told, cement manufacturing globally contributes close to 5 percent of all anthropogenic carbon emissions. The total industrial emissions category accounts for about 0.8 GtC/yr.

As it turns out, fly ash from coal-fired power plants can be used as a replacement for some of the cement in concrete. This *fly-ash concrete* not only reduces carbon emissions by roughly one ton of CO_2 per ton of replaced cement, but it also results in a concrete that is stronger and more durable than its conventional counterpart. It also recycles a relatively useless waste product that would otherwise have to be disposed of. Concrete mixtures with more fly ash than cement are now becoming popular in the emerging green building industry.

Carbon Intensity of Fossil Fuels

The amount of carbon released per unit of energy delivered is called the *carbon intensity*. Some fuels have high carbon intensity, such as coal, and some conversion systems release no direct carbon at all, such as wind turbines or nuclear power. Interestingly, biomass fuels may also be used in ways that emit little or no carbon when new plants are grown to replace the ones that were burned. No energy system is likely to have zero carbon emissions, however, since it is hard to avoid such emissions during the mining, materials processing, and construction of any energy facility.

EXAMPLE 4 Carbon Intensity of Methane

Find the carbon intensity of methane based on its higher heating value (HHV) of 890 kJ/mol (which includes the energy of condensation of the water vapor formed. Then find the carbon intensity based on the lower heating value (LHV) of 802 kJ/mol.

Solution First, write a balanced chemical reaction for the oxidation of methane:

$$CH_4 + 2\ O_2 \rightarrow CO_2 + 2\ H_2O$$

So, burning 1 mol of CH_4 liberates 890 kJ of energy while producing 1 mol of CO_2. Since 1 mol of CO_2 has 12 g of carbon, the HHV carbon intensity of CH_4 is

$$\text{HHV carbon intensity} = \frac{12 \text{ g C}}{890 \text{ kJ}} = 0.0135 \text{ gC/kJ} = 13.5 \text{ gC/MJ}$$

Similarly, the LHV carbon intensity would be

$$\text{LHV carbon intensity} = \frac{12 \text{ gC}}{802 \text{ kJ}} = 0.015 \text{ gC/kJ} = 15.0 \text{ gC/MJ}$$

The LHV carbon intensity assumes the latent heat of the water vapor produced is not available as usable energy.

TABLE 3

LHV and HHV Carbon Intensities and Emissions for Typical Fossil Fuels

Fuel	LHV Carbon Intensity (gC/MJ)	HHV Carbon Intensity (gC/MJ)	2004 Global Carbon Emissions (GtC/yr)
Natural gas	15.3	13.8	1.58 (21.2%)
Petroleum	20.0	19.7	2.96 (39.8%)
Coal	25.8	24.2	2.89 (38.9%)

Sources: Carbon intensities from NAS, 1992; emissions from EIA, 2006.

We can get a quick estimate of the carbon intensity of other fuels by using the fact that the energy released during combustion of carbon-based fuels is approximately proportional to the amount of oxygen they consume (Baird, 1995). For example, if we consider coal to be purely carbon and oil to be approximately CH_2, we can write the following oxidation reactions:

Coal: $$C + O_2 \rightarrow CO_2 \quad (20)$$

Oil: $$CH_2 + 1.5\ O_2 \rightarrow CO_2 + H_2O \quad (21)$$

Natural gas: $$CH_4 + 2\ O_2 \rightarrow CO_2 + 2\ H_2O \quad (22)$$

The same amount of carbon is released for each of these reactions, but since energy is roughly proportional to oxygen consumption, per unit of carbon emitted we would expect about 1.5 times as much energy from oil as coal and twice as much energy from natural gas compared to coal. Turning that around, coal is more carbon-intensive than oil, and oil emits more carbon than natural gas.

The actual LHV and HHV carbon intensities for typical fossil fuels, combined with estimates of total carbon emissions, are given in Table 3. Note the slight difference between the carbon intensity of natural gas and the intensity of methane itself found in Example 4. Natural gas is mostly methane, but it also includes other hydrocarbons, which alters the calculation slightly. Since the proportions of methane and other hydrocarbons vary from one source to another, the carbon intensities given in Table 3 are "typical," and you are likely to find other, slightly different, values in the literature. The same goes for variations in types of coal. Another comment worth making is that energy consumption figures in the United States are often based on the higher heating value of a fuel, while most of the rest of the world, including the IPCC, uses LHV values. The LHV value makes more sense for power plants since they never recover the latent heat of their water vapor emissions, but HHV values are more appropriate for high-efficiency condensing furnaces and water heaters, which do capture that heat. To be consistent with the IPCC, in this chapter, all energy data and carbon intensities will be based on the LHV values.

The carbon intensity data given in Table 3 suggest that sizable reductions in carbon emissions are possible by switching from coal to natural gas. It is unfortunately the case, however, that most of the world's fossil fuel resources are in the form of coal. It is interesting to note that almost 90 percent of the world recoverable resources of coal are in just three regions: the United States, the former USSR, and China.

Table 4 presents data on the world's fossil fuel resources. These resources are shown as a resource base, which consists of already identified reserves plus an estimate at the 50-percent probability level of remaining undiscovered resources. The

TABLE 4

Energy Content of Global Fossil Fuel Resources and Occurrences, in Exajoules (EJ)

Fuel	Conventional Resources	Unconventional Resources	Total Resource Base	Additional Occurrences
Natural gas	9,200	26,900	36,100	>832,000
Petroleum	8,500	16,100	24,600	>25,000
Coal	25,200	100,300	125,500	>130,000
Totals	42,900	143,300	186,200	>987,000

Source: Nakicenovic, 1996.

resources are also described as coming from conventional sources of the type now being exploited, as well as unconventional sources that might be usable in the future. Unconventional sources of oil include oil shale, tar sands, and heavy crude; unconventional natural gas sources include gas in Devonian shales, tight sand formations, geo-pressurized aquifers, and coal seams. An additional column in Table 4 is labeled "additional occurrences." These are additional resources with unknown certainty of occurrence and/or with unknown or no economic significance in the foreseeable future. Enormous amounts of methane locked in methane hydrates ($CH_4 \cdot 6H_2O$) under the oceans (estimated at over 800,000 EJ) are the most important of these.

Example 5 suggests what might happen to the CO_2 concentration in the atmosphere if we were to burn all of the world's coal resource base.

EXAMPLE 5 **Burning the World's Coal**

Estimate the increase in atmospheric CO_2 if the 125,500 EJ of coal were to be burned. Assume a constant airborne fraction of 38 percent.

Solution We can first estimate the carbon content using the LHV value from Table 3:

$$125{,}500 \text{ EJ} \times 25.8 \text{ gC/MJ} \times 10^{12} \text{ MJ/EJ} \times 10^{-15} \text{ GtC/gC} = 3{,}238 \text{ GtC}$$

which is roughly four times as much carbon as currently exists in the atmosphere. Converting this to CO_2 and including the 0.38 airborne fraction gives

$$\Delta CO_2 = \frac{3{,}238 \text{ GtC} \times 0.38}{2.12 \text{ GtC/ppm } CO_2} = 580 \text{ ppm } CO_2$$

That would result in 2.5 times as much CO_2 in the atmosphere as we have today. In fact, it would likely be higher than that if the airborne fraction increases due to the oceans ceasing to be such a good carbon sink.

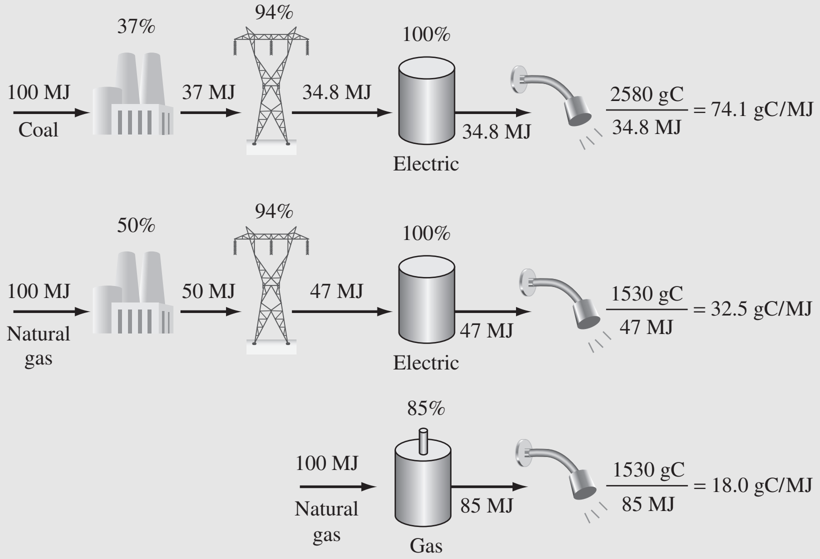
It is interesting to note that a similar calculation to that shown in Example 5 suggests that the oil and gas resource base has the potential to add only about one-third as much CO_2 to the atmosphere as would burning all of the world's coal resources. That calculation is less certain, however, because of the relatively unknown carbon emission factors that would be appropriate for the unconventional oil and gas resources. It is also complicated by the fact that the principal component of natural gas, methane, is a much more potent greenhouse gas than CO_2, which means methane leakage can amplify its potential global warming.

The carbon-intensity factors given in Table 3 suggest that switching from coal to oil or natural gas would reduce emissions significantly. For example, it would appear that switching from coal to natural gas would reduce carbon emissions by about 41 percent while delivering the same amount of energy. Example 6 shows how efficiency advantages as well as carbon advantages associated with using natural gas can substantially increase that advantage.

EXAMPLE 6 Efficiency and Carbon Intensity Combined

Compare the carbon emissions to heat household water using the following three energy systems: 1) a very good, 37 percent-efficient coal-fired power plant delivering electricity to a 100 percent-efficient electric water heater; 2) a new, 50 percent-efficient natural-gas-fired combined-cycle power plant for that same electric water heater; and 3) an 85 percent-efficient gas-fired water heater. Assume 6 percent losses in electrical transmission lines.

Solution Let us base our comparison on 100 MJ of energy provided to each system. Using the LHV values given in Table 3, burning 100 MJ of coal releases 2,580 g of carbon, while 100 MJ of natural gas releases 1,530 gC. As suggested here, the coal-plant system delivers $100 \times 0.37 \times 0.94 = 34.8$ MJ to heat water; the more efficient gas-fired power plant delivers $100 \times 0.50 \times 0.94 = 47$ MJ to heat water; and using gas in the water heater delivers 85 MJ of heat.



The carbon intensities for each system are shown in the figure. Switching from coal to gas in the power plant reduces carbon emissions for the electric water heater by 56 percent. Using gas directly in the water heater reduces carbon emissions by 75 percent (much more than the 41 percent improvement expected based just on the carbon intensity of natural gas versus coal).

Estimating Emissions: The Kaya Identity

Predicting future concentrations of carbon dioxide depends on numerous assumptions about population growth, economic factors, energy technology, and the carbon cycle itself. The usual approach involves developing a range of emission scenarios that depend on those factors and then using those scenarios to drive mathematical models of how the atmosphere and climate system will react to those inputs. At the level of treatment given in this short section, we can't begin to approach the complexity of those models; however, we can make a few simple calculations to at least give a sense of some of the important factors.

One way to build simple models of environmental problems is to start with the notion that impacts are driven by population, affluence, and technology, which is sometimes referred to as the *IPAT identity* (Ehrlich and Holdren, 1971).

$$\text{Environmental Impact} = (\text{Population}) \times (\text{Affluence}) \times (\text{Technology}) \tag{23}$$

The following application of IPAT to carbon emissions from energy sources is often referred to as the *Kaya identity* (Kaya, 1990).

$$\text{C} = \text{Population} \times \frac{\text{GDP}}{\text{Person}} \times \frac{\text{Primary Energy}}{\text{GDP}} \times \frac{\text{Carbon}}{\text{Primary Energy}} \tag{24}$$

where

$\text{C} = \text{carbon emission rate (GtC/yr)}$

$\frac{\text{GDP}}{\text{Person}} = \frac{\text{GDP}}{\text{P}} = \text{per capita gross domestic product (\$/person-yr)}$

$\frac{\text{Primary Energy}}{\text{GDP}} = \frac{\text{PE}}{\text{GDP}} = \text{primary energy intensity, (EJ/\$)}$

$\frac{\text{Carbon}}{\text{Primary Energy}} = \frac{\text{C}}{\text{PE}} = \text{carbon intensity, (GtC/EJ)}$

Equation (24) incorporates the key quantities that drive our energy-related carbon emissions. It includes economic and population scenarios plus two factors that are central to energy: energy intensity and carbon intensity. Carbon intensity has already been introduced. *Energy intensity* is the amount of energy required to create a unit of economic activity as measured by gross domestic product (GDP). It is usually thought of as a surrogate for the country's energy efficiency. For example, Japan, which only needs half the energy to produce a unit of GDP, is often considered to be roughly twice as energy efficient as the United States. While there is some truth to that assertion, it sometimes masks differences in the standard of living and the climate in each country. For example, the United States has larger houses that are kept warmer in more severe winters, so if more energy is required it may have more to do with those factors than whether or not homes are better insulated in one country or the other. Japan is also a small, densely populated country with relatively short travel distances, so transportation energy would likely be less as well, even with an equivalent level of transportation efficiency.

For example, the Kaya identity for the year 2010 looks something like this:

$$\begin{aligned}\text{C} &= 6.9 \times 10^9 \text{ people} \times \$4{,}605/\text{person-yr} \times 14.9\ \text{EJ}/\$10^{12} \times 0.016\ \text{GtC/EJ} \\ &= 7.6\ \text{GtC/yr}\end{aligned}$$

TABLE 5

1990 to 2020 Average Annual Growth Rates (%/yr) Used in the IPCC IS92a Scenario for Energy-Related Carbon Emissions

Region	Population	$\frac{\text{GDP}}{\text{Person}}$	$\frac{\text{PE}}{\text{Person}}$	$\frac{\text{Carbon}}{\text{PE}}$
China and centrally planned Asia	1.03	3.91	−1.73	−0.32
Eastern Europe and ex-USSR	0.43	1.49	−0.66	−0.24
Africa	2.63	1.25	0.26	−0.21
United States	0.57	2.33	−1.81	−0.26
World	1.40	1.53	−0.97	−0.24

Equation (24) expresses the carbon emission rate as the product of four terms: population, GDP, carbon intensity, and energy intensity. Recall from Section 3.2 that if each of the factors in a product can be expressed as a quantity that is growing (or decreasing) exponentially, then the overall rate of growth is merely the sum of the growth rates of each factor. That is, assuming each of the factors in (23) is growing exponentially, the overall growth rate of carbon emissions r is given by

$$r = r_{\mathrm{P}} + r_{\mathrm{GDP/P}} + r_{\mathrm{PE/GDP}} + r_{\mathrm{C/PE}} \tag{25}$$

By adding the individual growth rates as has been done in (25), an overall growth rate is found, which can be used in the following emission equation:

$$C = C_0 e^{rt} \tag{26}$$

where

C = carbon emission rate after t yrs (GtC/yr)
C_0 = initial emission rate (GtC/yr)
r = overall exponential rate of growth (yr^{-1})

Table 5 shows population, economic growth, carbon intensity, and energy intensity values that have been used in one of the most-cited IPCC emission scenarios (IS92a) for energy. For the world as a whole, energy intensity and carbon intensity are both improving over time, which helps offset population and economic growth.

The cumulative emissions from a quantity growing exponentially at a rate r, over a period of time T is given by

$$C_{\mathrm{tot}} = \int_0^T C_0 e^{rt}\, dt = \frac{C_0}{r}\left(e^{rT} - 1\right) \tag{27}$$

Combining (27) with an estimate of the atmospheric fraction, along with our 2.12 GtC/ppm CO_2 conversion factor, lets us make simple estimates of future CO_2 concentrations in the atmosphere, as Example 7 demonstrates.

EXAMPLE 7 Kaya Estimate of Future Carbon Emissions

Emissions from fossil-fuel combustion in 2010 are estimated to be 7.6 GtC/yr. In the same year, atmospheric CO_2 concentration is estimated to be 390 ppm. Assume the atmospheric fraction remains constant at 0.38.

a. Assuming the energy growth rates shown in Table 5 don't change, estimate the energy-related carbon-emission rate in 2050.
b. Estimate the cumulative energy-related carbon added to the atmosphere between 2010 and 2050.
c. Add into your scenario carbon emissions from industrial processes (especially cement) of 0.7 GtC/yr in 2010 and growing at 1.3%/yr. Also add a constant 0.9 GtC/yr from land-use changes. Estimate the CO_2 concentration in 2050.

Solution

a. The overall growth rate in energy-related carbon emissions is just the sum of the individual growth rates:

$$r = 1.40\% + 1.53\% - 0.97\% - 0.24\% = 1.72\% = 0.0172/\text{yr}$$

With 40 years of growth at 1.72% per year, the emission rate in 2050 would be

$$C_{2050} = C_{2010}e^{rT} = 7.6\, e^{0.0172\times 40} = 15.1 \text{ GtC/yr}$$

b. Over those 40 years, the cumulative energy emissions would be

$$C_{\text{tot}} = \frac{C_0}{r}(e^{rT} - 1) = \frac{7.6}{0.0172}(e^{0.0172\times 40} - 1) = 437 \text{ GtC}$$

c. The cumulative carbon emissions from industrial processes and land-use changes is

$$\text{Industrial } C_{\text{tot}} = \frac{C_0}{r}(e^{rT} - 1) = \frac{0.7}{0.013}(e^{0.013\times 40} - 1) = 37 \text{ GtC}$$

$$\text{Land Use } C_{\text{tot}} = 0.9 \text{ GtC/yr} \times 40 \text{ yrs} = 36 \text{ GtC}$$

Using the 2.12 GtC/ppm CO_2 ratio and a 0.38 atmospheric fraction makes our estimate of CO_2 in 2050

$$CO_2 = 390 + \frac{(437 + 37 + 36)\text{ GtC} \times 0.38}{2.12 \text{ GtC/ppm } CO_2} = 481 \text{ ppm}$$

One of the weakest aspects of the Kaya identity as expressed in (24) is its use of primary energy per GDP as a measure of energy efficiency. Primary energy, which is in essence energy as it is taken out of the ground, can be reduced by efficiency improvements on the supply side (e.g., more efficient electric power plants) as well as on the demand side (e.g., by making light bulbs more efficient). One way to address the two different efficiency improvements is to introduce another factor, call it *final energy FE,* which is the energy purchased by consumers (e.g., gasoline, natural gas, electricity). Another extension is based on prospects for carbon capture and

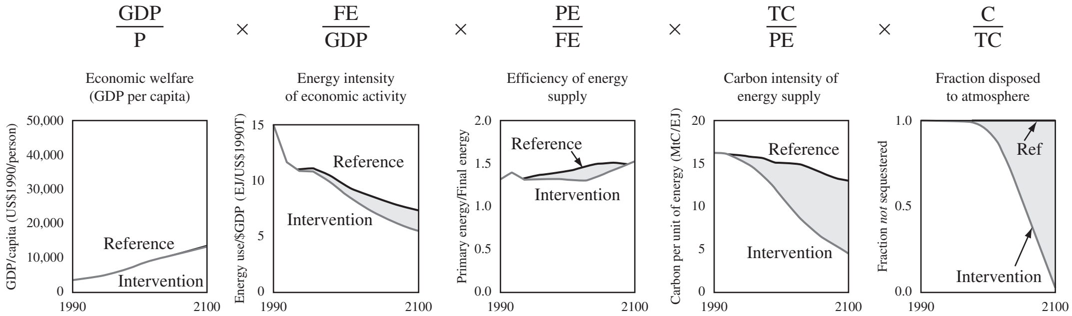


FIGURE 19 An expanded Kaya analysis showing a reference scenario and an intervention scenario designed to stabilize atmospheric CO_2.
(*Source:* Hummel, 2007 based on the IPCC A2-4.5 W/m^2 scenario.)

storage in the future. Letting *TC* be total carbon in the fuel and *C* be carbon actually emitted to the atmosphere, the Kaya identity can be expanded as follows:

$$C = P \times \frac{GDP}{P} \times \frac{FE}{GDP} \times \frac{PE}{FE} \times \frac{TC}{PE} \times \frac{C}{TC} \tag{28}$$

This much more useful disaggregation breaks the overall carbon emissions into much tighter packages for analysis. One such analysis showing a comparison between a more or less business-as-usual carbon scenario with one designed to ultimately stabilize CO_2 is shown in Figure 19.

A Climate Sensitivity Parameter

As we pump more and more CO_2 into the atmosphere, the marginal impact of each additional ton decreases as its absorption bands approach saturation. That suggests a nonlinear relationship between CO_2 and the resulting global warming that it causes. One commonly used representation of this phenomenon is given in (29).

$$\Delta T_e = \frac{\Delta T_{2X}}{\ln 2} \ln\left[\frac{(CO_2)}{(CO_2)_0}\right] \tag{29}$$

where

ΔT_e = the equilibrium, global, mean surface temperature change
ΔT_{2X} = the equilibrium temperature change for a doubling of atmospheric CO_2
$(CO_2)_0$ = the initial concentration of CO_2
(CO_2) = the concentration of CO_2 at another time

The increase in surface temperature that results from a doubling of CO_2 in the atmosphere is called the *climate sensitivity,* ΔT_{2X}. Notice what happens to (29) when CO_2 is double the initial amount. The change in surface temperature is what it should be, that is,

$$\Delta T_e = \frac{\Delta T_{2X}}{\ln 2} \ln\left[\frac{2(CO_2)_0}{(CO_2)_0}\right] = \frac{\Delta T_{2X}}{\ln 2} \ln 2 = \Delta T_{2X}$$

If the concentration of CO_2 is quadrupled,

$$\Delta T_e = \frac{\Delta T_{2X}}{\ln 2} \ln\left[\frac{4(CO_2)_0}{(CO_2)_0}\right] = \frac{\Delta T_{2X}}{\ln 2} \ln (2^2) = 2\Delta T_{2X}$$

In other words, the logarithmic function suggests that for every doubling of CO_2, the surface temperature goes up by the same amount. For example if ΔT_{2X} is 3°C, then the first doubling raises the surface temperature by 3°C, doubling it again to four times its initial value raises the temperature by another 3°C, and so on, as illustrated in Figure 20. The climate sensitivity ΔT_{2X} is a convenient benchmark that is often used to compare various general circulation computer models. It has long been estimated to be between 1.5°C and 4.5°C, with the best estimate as of 2006 being somewhere around 3.0°C.

Equilibrium Temperature and Realized Temperature

The surface temperatures dealt with thus far, T_e and ΔT_e, are known as *equilibrium* temperatures; that is, they are temperatures that would eventually be reached for a given atmospheric configuration. In reality, of course, the temperature of the Earth will not adjust instantaneously to changes in greenhouse gas concentrations. A considerable period of time is needed to warm the upper layer of the oceans and the surface of the land, which means the actual *realized* surface temperature of the Earth will lag somewhat behind the equilibrium temperature. As shown in Figure 21, at any

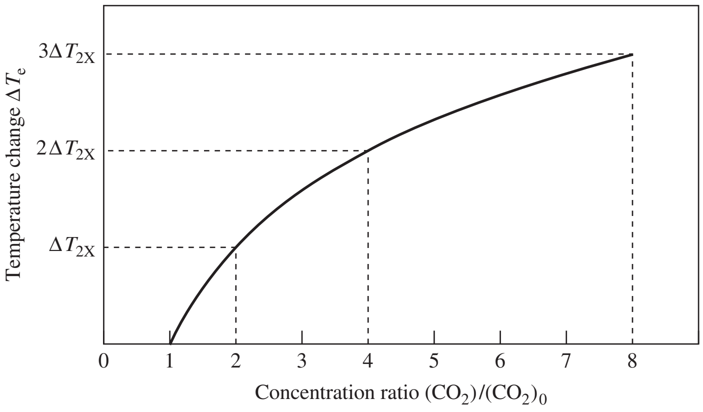


FIGURE 20 For every doubling of CO_2, the global equilibrium temperature increases by ΔT_{2X}.

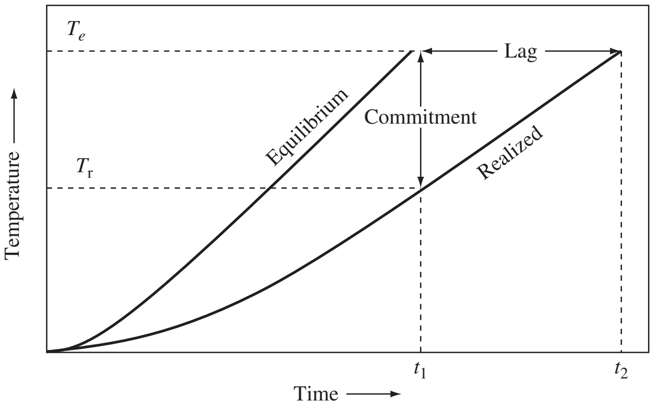


FIGURE 21 At time t_1, the realized surface temperature of the Earth is T_r, but even if there are no further increases in greenhouse gases, it will continue to rise to the equilibrium temperature T_e with a lag time of $t_2 - t_1$.

given time, the difference between realized and equilibrium temperature is referred to as the temperature *commitment,* while the time it takes to achieve equilibrium is the *lag time*. Coupled atmosphere-ocean general circulation models indicate that the realized temperature is roughly 60 to 85 percent of the equilibrium temperature.

7 The Other Greenhouse Gases and Aerosols

While most of the attention on greenhouse gases has been directed toward CO_2 emissions, a number of other gases, as well as aerosols, contribute to our global warming problem. After introducing each one, we will see how they interact with each other and with the atmosphere to affect not only climate but also ozone depletion in the stratosphere.

Methane (CH_4)

Methane is another naturally occurring greenhouse gas that is increasing in concentration as a result of human activities. As shown in Figure 22, methane concentrations in the atmosphere were approximately 750 parts per billion (ppb) for many hundreds of years before they began their rapid climb in the 1800s. After a period of very rapid rise, their concentration in the atmosphere in the early years of the twenty-first century seems to be leveling off at about 1,750 ppb. Whether this is some new steady-state level or a temporary pause, remains to be seen.

Methane is produced by bacterial fermentation under anaerobic conditions, such as occurs in swamps, marshes, rice paddies, landfills, as well as in the digestive tracks of ruminants. Total emissions have been estimated to be on the order of 600 teragrams per year, of which about 60 percent are related to human activities such as agriculture, fossil fuel use, and waste disposal. Recent estimates of methane emissions from short-lived biomass and plant litter under aerobic conditions suggest methane's contribution to the observed global warming may have been significantly underestimated and a reassessment may be called for (Keppler et al., 2006).

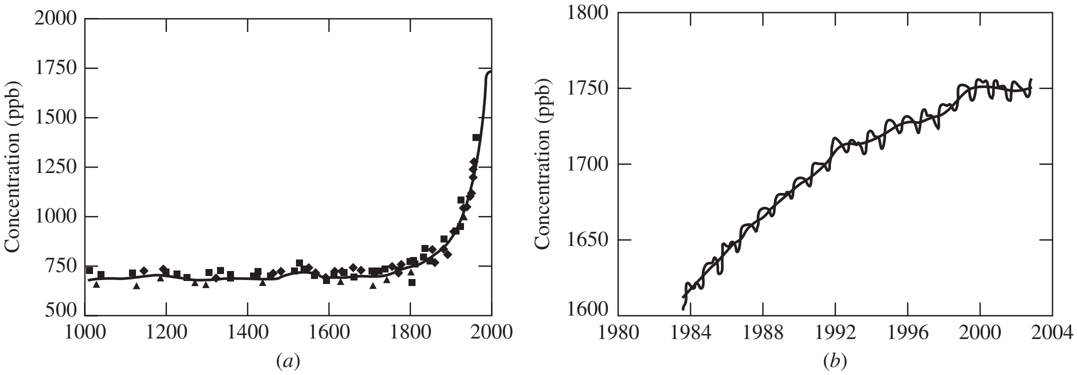


FIGURE 22 Methane concentrations from Antarctic ice cores (*a*) combined with more recent direct measurements (*b*).
(*Source:* IPCC, 2007.)

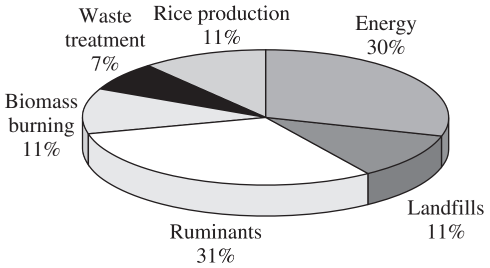


FIGURE 23 Anthropogenic sources of methane emissions, totaling some 370 Tg/yr. (*Source:* Based on Houweling et al., 1999.)

Almost half of anthropogenic methane emissions are the result of human food production, especially rice production and livestock, including cattle, sheep, and buffalo, which belch some 115 Tg of methane per year. Clearing and burning of biomass, in part to prepare land for grazing and crops, also contribute. Almost one-third of anthropogenic emissions are associated with fossil fuel use (Figure 23).

Methane has an absorption band centered at 7.7 μm, which places it right on the edge of the atmospheric window and makes it a very potent greenhouse gas. Its relatively short residence time in the atmosphere, on the order of 9 to 15 years, coupled with its high potency, suggests reducing methane emissions could be an effective means to reduce global warming on a relatively short timescale.

Methane is removed from the atmosphere primarily through reactions with the hydroxyl radical (OH), as the following reaction suggests:

$$CH_4 + OH + 9\ O_2 \rightarrow CO_2 + 0.5\ H_2 + 2\ H_2O + 5\ O_3 \qquad (30)$$

Methane is of course a greenhouse gas, so it has direct effects on the Earth's radiation balance. There are, however, a number of indirect effects that provide additional impacts, which reaction (30) helps explain. First, when methane reacts with hydroxyl (OH), the concentration of OH decreases. With less OH available, the rate of removal of CH_4 slows down, lengthening the atmospheric lifetime of the remaining methane. With a longer lifetime for CH_4 in the atmosphere, it continues to absorb infrared for a longer time, increasing its global warming potential. The second indirect effect that (30) implies is that as methane reacts with hydroxyl, it produces water vapor. When this reaction occurs in the troposphere, the increase in water vapor is insignificant, but in the stratosphere, it becomes quite important. Finally, (30) indicates the destruction of methane produces an increase in ozone, which is itself a greenhouse gas. The sum of all of these indirect effects greatly increases methane's climate impact.

There is some concern for the possibility that global warming could free large amounts of methane currently frozen in permafrost in the far northern regions of the world and could allow anaerobic decomposition of organic matter also frozen in permafrost, producing more methane. This is an important example of a positive feedback loop. Warming leading to increased releases of the greenhouse gas, methane, could reinforce the original warming.

Nitrous Oxide (N_2O)

Nitrous oxide ("laughing gas") is another naturally occurring greenhouse gas that has been increasing in concentration due to human activities (Figure 24). The current

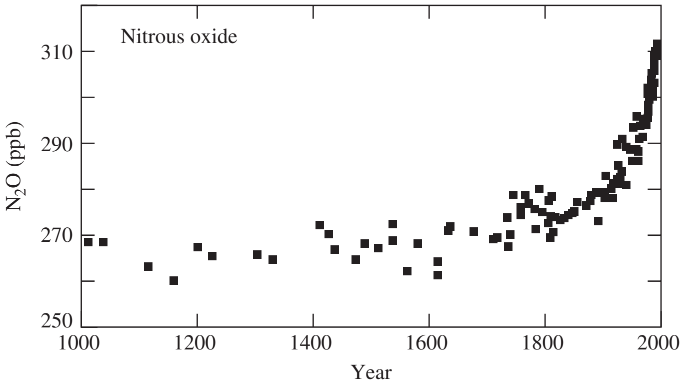


FIGURE 24 Increasing concentration of nitrous oxide and corresponding radiative forcing. (*Source:* IPCC, 2001.)

atmospheric concentration is about 320 ppb, which is a 16 percent increase over the preindustrial concentration of about 275 ppb.

Nitrous oxide is released into the atmosphere mostly during the nitrification portion of the nitrogen cycle:

NH_4^+	→	N_2	→	N_2O	→	NO_2^-	→	NO_3^-
Ammonium ion		Molecular nitrogen		Nitrous oxide		Nitrite ion		Nitrate ion

It is estimated that natural sources of N_2O deliver 9 Tg per year to the atmosphere, with most of that coming from oceans and wet forest soils. Anthropogenic sources, which add another 6 Tg/yr, are largely the result of forest clearing and nitrogen fertilizers. Other sources include three-way catalytic converters on cars, combustion of fuels containing nitrogen, and a variety of industrial processes such as the manufacture of nylon.

Apparently, there are no significant tropospheric sinks for N_2O, and it is only slowly degraded in the stratosphere by photolysis. As a result, it has a long atmospheric lifetime, estimated at about 120 years, which means perturbations in the natural cycle will have long-lasting repercussions. Nitrous oxide has an absorption band at 7.8 μm that is associated with a stretching of the bonds, and another at 8.6 μm associated with bending of the bond angle. The band at 7.8 μm is on the shoulder of the atmospheric window, and the band at 8.6 μm is right in the window, so N_2O is a very potent greenhouse gas.

Halocarbons

Halocarbons are carbon-based molecules that have chlorine, fluorine, or bromine in them. The carbon-to-fluorine bonds in halocarbons oscillate, and hence absorb, at wavelengths around 9 μm, and other bond stretching and bending in halocarbons also occur with frequencies in the atmospheric window, so these molecules are all potent greenhouse gases. They are environmentally important not only because they contribute to global warming, but also because chlorine and bromine atoms that find their way into the stratosphere have the ability to catalytically destroy ozone, as will be described later in this chapter.

Subcategories of halocarbons include *chlorofluorocarbons* (CFCs), which have only carbon, fluorine, and chlorine, but no hydrogen; *hydrochlorofluorocarbons* (HCFCs), which are like CFCs, but do contain hydrogen; *hydrofluorocarbons* (HFCs), which contain no chlorine; and *halons,* which are carbon-based molecules containing bromine along with fluorine and perhaps chlorine. These halocarbon gases differ from all of the other radiatively active gases in that they do not occur naturally, and their presence in the atmosphere is due entirely to human activities. Other important halocarbons include carbon tetrachloride (CCl_4), methyl chloroform (CH_3CCl_3), and methyl bromide (CH_3Br). With so many categories, it may help to show some examples, which Table 6 does.

Chlorofluorocarbons (CFCs) are nontoxic, nonflammable, nonreactive, and are not water soluble, which led to the belief when they were first invented that they were truly benign chemicals. But because they are inert and don't dissolve easily in water, they aren't destroyed by chemical reactions or removed from the troposphere by rain. That means they have long atmospheric lifetimes, as Table 6 suggests. The only known removal mechanism is photolysis by short-wavelength solar radiation, which occurs after the molecules drift into the stratosphere. It is the chlorine freed during this process that can go on to destroy stratospheric ozone. Similarly, halons, which contain bromine, have no tropospheric sinks, and their only removal mechanism is also photochemical decomposition in the stratosphere, which releases the ozone-depleting bromine.

TABLE 6

Examples of Halocarbons

Formula	Designation	Atmospheric Lifetime (yrs)
Chlorofluorocarbons (CFCs)		
$CFCl_3$	CFC-11	50
CF_2Cl_2	CFC-12	102
$CF_2ClCFCl_2$	CFC-113	85
Hydrochlorofluorocarbons (HCFCs)		
CHF_2Cl	HCFC-22	12.1
CH_3CFCl_2	HCFC-141b	9.4
$CF_3CF_2CHCl_2$	HCFC-225ca	2.1
Hydrofluorocarbons (HFCs)		
CF_3CH_2F	HFC-134a	14.6
CH_3CF_3	HFC-143a	48.3
Perfluorocarbons (PFCs)		
CF_4	Tetrafluoromethane	50,000
C_2F_6	Perfluoroethane	10,000
Halons		
CF_3Br	H-1301	65
CF_2ClBr	H-1211	20
Other		
CH_3CCl_3	Methyl chloroform	4.9
CH_3Br	Methyl bromide	0.7
CCl_4	Carbon tetrachloride	26

Source: IPCC, 1996.

Hydrochlorofluorocarbons (HCFCs) are being introduced as replacements for CFCs. By adding hydrogen to the molecules, they are no longer chemically inert, which means chemical reactions can destroy them in the troposphere before they have a chance to drift into the stratosphere. Notice the much shorter atmospheric lifetimes for the HCFCs in Table 6. HCFCs are only temporary replacements for CFCs, however, since they do still have some potential to deplete the ozone layer, and they are still potent greenhouse gases.

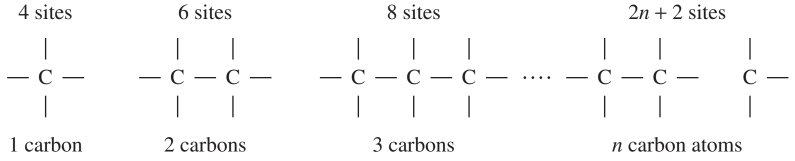
Hydrofluorocarbons (HFCs) have no chlorine at all, so they are even better than HCFCs in terms of stratospheric ozone protection. The hydrofluorocarbon CH_2FCF_3 (HFC-134a) is quickly becoming the refrigerant of choice for automobile air conditioners and refrigeration equipment. Even this chemical, however, has a sizable atmospheric lifetime (14.6 years), and, as a greenhouse gas, it is 1,400 times as potent as CO_2.

Perfluorocarbons (PFCs) are hydrocarbons containing only carbon and fluorine atoms. They are characterized by extreme stability, nonflammability, low toxicity, zero ozone depleting potential, and high global warming potential. For example, tetrafluoromethane (CF_4) has an atmospheric lifetime of 50,000 years, and its global warming potential is 6,500 times that of carbon dioxide.

Halons have the ozone-destroying element bromine in them. They are very stable molecules with no tropospheric sinks, so they only release that bromine when they eventually drift into the stratosphere and are broken apart by photolysis. The primary use of halons is in fire extinguishers. They are nontoxic and leave no residue when sprayed onto fires, so they are ideal for use in confined spaces containing critical equipment such as computers.

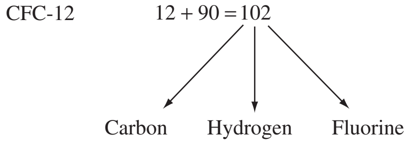
Halocarbon Numbering Systems. In Table 6, most of the halocarbons are identified by a simple numerical designation as well as a chemical formula. The CFCs, HCFCs, and HFCs are referred to using a number system developed years ago by DuPont. For example, trichlorofluoromethane, $CFCl_3$, is CFC-11, and the hydrochlorofluorocarbon CHF_2Cl is HCFC-22. To determine the chemical formula from a fluorocarbon number, begin by adding "90" to the number and then interpret the three-digit result as follows: The leftmost digit is the number of carbon atoms, the middle digit is the number of hydrogen atoms, and the right digit is the number or fluorines.

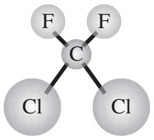
To determine the number of chlorine atoms, begin by visualizing molecules in which each carbon atom forms four single bonds to other atoms (if the other atoms are all hydrogens, this is the familiar alkane series: methane, ethane, propane, and so on):



A single carbon atom has four sites to fill, two carbon atoms have six sites, and so forth. All of the sites will be occupied by hydrogen, fluorine, or chlorine. So, to find the number of chlorines, just subtract the number of hydrogens and fluorines from the total available sites. Each vacant bonding site not taken up by fluorine or hydrogen is occupied by chlorine.

For example, to figure out what CFC-12 is, add 90 to 12 giving 102. So, there is 1 carbon, no hydrogens, and 2 fluorines. With 1 carbon there are 4 sites available, 2 of which are taken by fluorine, leaving 2 for chlorine.





CFC-12

Thus, CFC-12 is CF_2Cl_2.

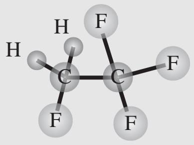
The halons also have a number system, but this one is not so complicated. Halons are given a four-digit designation, with the leftmost digit being the number of carbons, the second is fluorines, the third is chlorines, and the fourth is bromine. For example, H-1211 is CF_2ClBr.

EXAMPLE 8 Halocarbon Numbering

a. What is the chemical composition of CFC-115?
b. What is the CFC number for CH_2FCF_3?
c. What is H-2402?

Solution

a. CFC-115: Adding 90 to 115 gives 205. Thus, a molecule contains 2 carbons, no hydrogens, and 5 fluorines. Two carbons have 6 bonding sites, 5 of which are taken by fluorine. The remaining site is taken by chlorine. The chemical formula would therefore be C_2F_5Cl (more correctly written as $CClF_2CF_3$).
b. CH_2FCF_3 has 2 carbons, 2 hydrogens, and 4 fluorine atoms. Subtracting 90 from 224 gives 134. There is no chlorine or hydrogen, so this halocarbon is a hydrofluorocarbon, HFC-134. Notice that we can't tell whether this is CHF_2CHF_2 or CH_2FCF_3. To distinguish one isomer from another, a letter designation is added. For example, CHF_2CHF_2 is HFC-134 (1,1,2,2-tetrafluoroethane), and CH_2FCF_3 is HFC-134a (1,1,1,2-tetrafluoroethane).



HFC-134a

c. H-2402 is $C_2F_4Br_2$ ($CBrF_2CBrF_2$).

Refrigerants. Chlorofluorocarbons have many unusual properties that have led to their widespread use. For example, they are easily liquefied under pressure, and when that pressure is released, they evaporate and produce very cold temperatures.

In fact, they were originally developed to satisfy the need for a nontoxic, nonflammable, efficient refrigerant for home refrigerators. Before they were introduced in the early 1930s, the most common refrigerants were ammonia, carbon dioxide, isobutane, methyl chloride, methylene chloride, and sulfur dioxide. All had significant disadvantages. They were either toxic, noxious, highly flammable, or required high operating pressures that necessitated heavy, bulky equipment. From those perspectives, CFCs are far superior to any of the gases that they replaced. As a side note, fluorocarbons used in refrigeration equipment are often referred to with a "Refrigerant" number. For example, CFC-12 is often called "Refrigerant-12" or, more simply, R-12. The DuPont trade name Freon has also been used, so occasionally it is still called F-12. The "R" designation is also applied to other hydrocarbons such as methane (R-50) and ethane (R-170).

Until recently, most refrigerators, freezers, and automobile air conditioners used CFC-12 as the refrigerant, and large building air conditioning systems have tended to use CFC-11. CFC refrigerants do not wear out, so as long as they are sealed into equipment, the total emission rate can be small. Automobile air conditioners, however, tend to develop leaks that necessitate periodic recharging of their systems. Not long ago, when car air conditioners were serviced, the old refrigerant was usually vented to the atmosphere rather than being captured and recycled, which compounded the loss rate. As a result, automobile air conditioners used to contribute on the order of 20 percent of all emissions of CFCs in the United States. The Montreal Protocol and subsequent changes in the Clean Air Act have changed that picture significantly. Production and importation of CFCs ended in 1996, and air conditioners on new cars now tend to use HFC-134a and HFC-152a, which contain no chlorine. Only licensed facilities that use CFC recycling equipment can service older automobile air conditioners.

Aerosol Propellants. When CFCs were first hypothesized as a danger to the ozone layer by Molina and Rowland (1974), over half of worldwide emissions were from aerosol propellants for products such as deodorants, hair spray, and spray-paint cans. At that time, the United States alone was using over 200,000 tons per year of CFC-11 and CFC-12 in aerosols. The EPA responded rather quickly to the threat, and acting under the Toxic Substances Control Act, it banned the use of CFCs in nonessential aerosol propellant applications beginning in 1979. Norway, Sweden, and Canada adopted similar restrictions, but the rest of the world lagged far behind until recently. The Montreal Protocol has led to total bans in most of the developed world. Replacements for CFCs in aerosols include isobutane, propane, and carbon dioxide. In some applications, simple pumps or "roll on" systems have replaced the propellants.

Foamed Plastics. The second most common use for CFCs in general, and the most common use for CFC-11 in particular, has been in the manufacture of various rigid and flexible plastic foams found in everything from seat cushions to hamburger "clamshell" containers to building insulation. When liquid CFCs are allowed to vaporize in plastic, they create the tiny bubbles that make the plastic foamy.

Sheets of rigid "closed cell" urethane foams are used primarily as thermal insulation in buildings and refrigeration equipment. In such applications, the CFCs or HCFCs, which are poor thermal conductors, trapped in foam cells reduce the heat transfer capabilities of the product. CFCs have also been used to manufacture nonurethane, rigid foams such as extruded polystyrene (Dow's trade name is Styrofoam) used extensively for egg cartons and food service trays, and expanded polystyrene foam, which is used to make drinking cups. Since CFCs are trapped in the holes in closed-cell foams, they are only slowly released into the atmosphere as the material ages or is eventually crushed. It is estimated that the "bank" of CFCs and HCFCs trapped in rigid foam insulation is equivalent to about 19 Gt of CO_2. New production of HCFCs will be banned by 2020.

Replacements for these foams are possible. Polyisocyanurate foams made with hydrocarbon blowing agents such as pentane have no global warming potential or ozone-depletion potential, and they are excellent insulators. Fiberglass insulation, although thermally not as effective per unit of thickness, also contains no CFCs. Various cardboards and other paper products can be used to replace many of the polystyrene applications in the food industry.

Flexible foams, which are used in furniture, automobile seats, and packaging, have cells that are open to the atmosphere (open-cell). Hence, CFC release is almost immediate. These foams are made using carbon dioxide as the primary blowing agent, but the CO_2 is often augmented with methylene chloride.

Ozone (O_3)

Ozone has a strong absorption band at 9 μm, right in the middle of the atmospheric window (refer to Figure 11), so clearly it is a greenhouse gas of importance. It has, however, proven to be a difficult one to understand. It is not a "well-mixed" gas in the atmosphere the way all of the greenhouse gases described up to this point are. Not only does its concentration vary from place to place and time to time around the globe, but its effect on climate depends on its vertical distribution as well.

Ozone in the troposphere is formed by photochemical reactions involving relatively short-lived precursor gases, including NO_x, nonmethane-hydrocarbons, and CO. It is the principal gas in photochemical smog. And since smog is associated with the major industrialized areas of the world, it is not surprising that tropospheric ozone concentrations are higher in the Northern Hemisphere than in the Southern Hemisphere. Those concentrations also vary seasonally as sunnier summer months energize the formation of ozone. Ozone transported from the industrialized countries in the Northern Hemisphere to the Arctic has been identified as a major contributor to the current rapid warming of that region.

In the stratosphere, ozone concentrations are decreasing as a result of the attacks by chlorine and bromine released by UV-exposed CFCs and halons. These stratospheric ozone losses also vary by geographic location and by season as the annual appearance of the ozone hole over Antarctica in September and October certainly demonstrates. This ozone loss helps open the atmospheric window, which offsets some of the global warming caused by halocarbon emissions. As emissions of CFCs and halons into the atmosphere are curtailed, it is expected that ozone will recover in the early part of the twenty-first century, so this helpful impact of ozone depletion will diminish.

Aerosols

Suspensions of particles having an effective diameter of less than 10 μm are called aerosols. Some particles enter the atmosphere as solids (e.g., soil dust), and others are formed in the atmosphere when gases such as sulfur dioxide condense into liquid particles such as sulfates. Combustion of fossil fuels and biomass burning are the principal anthropogenic sources of aerosols. Natural sources of aerosols include wind-blown soil dust, evaporation of sea-spray droplets, and volcanic eruptions. While the natural emission rates from these and other sources far exceeds anthropogenic emissions, most of these particulates are so large that they quickly drop out of the troposphere. Smaller particles, especially ones that reach the stratosphere, are much more important in terms of their impact on climate as well as human health.

Aerosols have characteristics that make them considerably different from the well-mixed, long-lived greenhouse gases CO_2, CH_4, N_2O, and halocarbons. For one, they have atmospheric lifetimes that are measured in days or weeks in the troposphere, and a few years in the stratosphere, while greenhouse gases have lifetimes measured usually in decades, but for some, in thousands of years. The aerosols in the troposphere today are mostly particles that entered the atmosphere within the past few days. That means aerosol concentrations can vary considerably from week to week and from place to place. That short lifetime also implies that the radiative impacts of aerosols are likely to be regional phenomena, centered around the industrialized areas of the world, whereas greenhouse gas impacts are much more uniformly distributed around the globe. In fact, in some heavily industrialized parts of the world, the cooling caused by aerosols can be greater than the warming due to greenhouse gases. That doesn't necessarily mean, however, that such an area would experience cooling, since regional temperatures are so much affected by other factors.

Aerosols are also different from most greenhouse gases in that they cause both heating and cooling of the atmosphere. As Figure 25 suggests, they affect the Earth's energy balance in three ways: (1) they can reflect incoming solar radiation back into space, which increases the Earth's albedo; (2) they can provide cloud condensation nuclei, which increases cloud reflectivity and cloud lifetime, which also increases albedo; and (3) carbonaceous particles such as soot from fossil-fuel combustion can increase the atmospheric absorption of incoming solar energy. Sulfates and white smoke increase the albedo by enhancing atmospheric and cloud reflection,

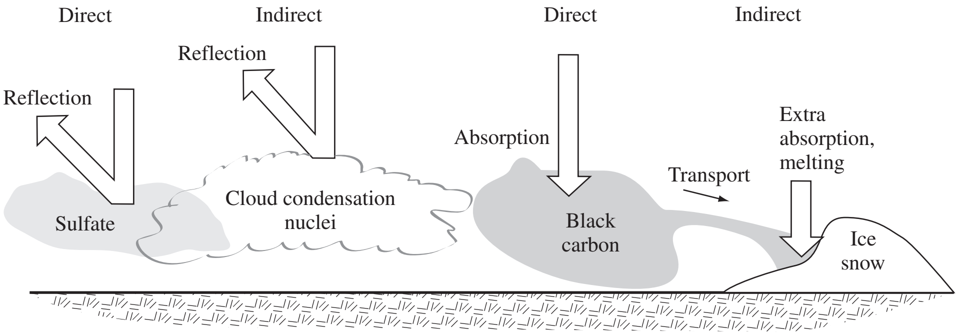


FIGURE 25 Aerosols provide cooling by increasing solar reflection off sulfates (direct effect) and white clouds (indirect). Black carbon warms the planet by absorbing sunlight (direct) and reducing albedo by helping to blacken and melt ice (indirect).

which helps cool the planet, and solar absorption by black smoke causes warming. The term *black carbon* (BC) is used to describe that black smoke, which consists of various carbonaceous products of incomplete combustion, including chars, charcoals, and soots.

The direct effect of black carbon absorption of incoming solar energy, warming the air itself, is amplified by its indirect effect as soot is transported and deposited onto snow-covered areas. By making the snow darker, its albedo is reduced causing further warming. And, as the soot-blackened snow melts, it exposes darker undersurfaces, further reducing albedo. These indirect albedo effects are estimated to more than double the warming impact of black carbon, making it a very important contributor to current global warming. Its importance, coupled with its extremely short atmospheric lifetime, suggest reducing black carbon may be the quickest way to slow the current rate of global warming.

8 Radiative Forcing of Climate Change

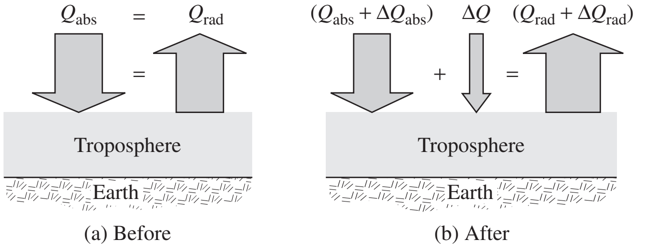
In the global equilibrium model shown earlier in Figure 12, the incoming 235 W/m^2 of solar energy absorbed by the Earth and its atmosphere is exactly balanced by 235 W/m^2 of outgoing long-wave radiation. If, for some reason, an additional amount of energy is added to the incoming energy, then the balance will be temporarily upset. Over time, however, the climate system will adjust to that change either by either increasing or decreasing the surface temperature of Earth until a balance is once again attained.

Mathematically, we can represent that process as follows. In Figure 26, the incoming energy absorbed, and outgoing energy being radiated, are shown as they exist at the top of the troposphere—that is, at the tropopause. Initially, the balanced system has energy absorbed equal to energy radiated:

$$Q_{abs} = Q_{rad} \tag{31}$$

When the system is perturbed by adding *radiative forcing*, ΔQ (W/m^2), to the incoming absorbed energy, a new equilibrium will eventually be established so that

$$(Q_{abs} + \Delta Q_{abs}) + \Delta Q = (Q_{rad} + \Delta Q_{rad}) \tag{32}$$



(a) Before (b) After

FIGURE 26 Radiative forcing, ΔQ, perturbs the balance between incoming solar energy absorbed, Q_{abs}, and outgoing radiant energy, Q_{rad}. (*a*) The balanced system before perturbation; (*b*) the balanced system after radiative forcing is added.

where the deltas refer to changes in the quantity in question. Subtracting (31) from (32) gives

$$\Delta Q = \Delta Q_{\text{rad}} - \Delta Q_{\text{abs}} \tag{33}$$

As we shall see, the value of the radiative forcing concept is that it provides a common currency for comparing the impacts of changes in individual greenhouse gases and aerosols, changes in albedo, and changes in the sun's radiation. Determining the radiative forcing associated with these has been one of the most important research areas for atmospheric scientists.

Another Climate Sensitivity Parameter

The key question, of course, is how much does the equilibrium surface temperature change ΔT_e for a given change in radiative forcing, ΔQ. The quantity that links the two is called the *mean climate sensitivity parameter* (λ) with units of °C of temperature rise per W/m^2 of radiative forcing. A linear expression for this relationship is often used:

$$\Delta T_e = \lambda \Delta Q \tag{34}$$

It has been established that a doubling of CO_2 is equivalent to a radiative forcing ΔQ of about 4.2 W/m^2. We already had another climate sensitivity parameter ΔT_{2X}, which is the equilibrium surface temperature change for a doubling of CO_2. The relationship between these two climate sensitivities is therefore

$$\lambda \approx \frac{\Delta T_{2X}}{4.2}\left(\frac{°\text{C}}{\text{W/m}^2}\right) \tag{35}$$

Following a procedure outlined by the National Research Council (NRC, 2003) as a first cut at estimating λ, we might consider the Earth as a blackbody with no greenhouse effect and no complicating feedback factors, just as we did in Figure 9. The rate at which this Earth radiates energy (on a per-unit-of-area basis) is given by the Stefan-Boltzmann equation,

$$Q = \sigma T_e^4 \tag{36}$$

from which,

$$\frac{dQ}{dT} = \frac{d}{dT}\,(\sigma T_e^4) = 4\sigma T_e^3 \tag{37}$$

Using (8), we already have derived the equilibrium surface temperature T_e for this simplified Earth and found it to be 254 K. Substituting that value, along with the Stefan-Boltamann constant, into (34) gives

$$\lambda_B = \frac{1}{dQ/dT} = \frac{1}{4 \times 5.67 \times 10^{-8}\,\text{W/m}^2\text{K}^4 \times (254\text{ K})^3} = 0.27°\text{C/(W/m}^2) \tag{38}$$

where the "B" in λ_B indicates this was derived assuming a blackbody Earth. If this sensitivity is used in (34) with a ΔT_{2X} forcing of 4.2 W/m^2, we would expect a doubling of CO_2 to cause an equilibrium surface temperature increase of 4.2 × 0.27 = 1.1°C, which is far below what most general circulation models predict.

Not surprisingly, (38) is not a very good estimate because it does not account for any of a number of complicating feedback factors.

Ice-Albedo Feedback. Ice and snow are highly reflective. If global warming decreases their surface area, they will reflect less solar radiation (decreasing the albedo) and absorb more of it, which leads to greater warming. That is an example of a *positive feedback* effect in that a perturbation in one direction or another feeds back onto itself, increasing the original perturbation.

Water-Vapor Feedback. Water vapor is the most important greenhouse gas. Global warming increases evaporation, which enhances the absorption of long-wavelength energy radiated from the surface. That means the atmosphere warms up and radiates more energy back to the surface, which increases surface temperature. It is another positive feedback loop.

Cloud Feedbacks. Since clouds can be highly reflective, surface warming could lead to more clouds, which could increase the albedo, which could lower the surface temperature. This is an example of a *negative feedback effect,* in which feedback reduces the original perturbation. On the other hand, clouds contain water vapor, which is a greenhouse gas, so increasing clouds could lead to an enhanced greenhouse effect, which leads to more surface warming—positive feedback. Modeling both phenomena has proved to be one of the most challenging problems in climate science, but the consensus seems to be that the net effect of clouds yields a positive feedback.

Lapse-Rate Feedbacks. Water vapor and clouds in the upper troposphere are colder than the surface of the Earth, which means they radiate less energy toward space than does the surface. The difference between the two is, in essence, the greenhouse effect. If the lapse rate changes, which is the rate of decrease of temperature with altitude, then the temperature difference from surface to upper troposphere also changes, which can lead to greater or lesser greenhouse effect. When surface warming causes the lapse rate to increase, the greenhouse effect will be enhanced, which is a positive feedback.

Figure 27 shows one way to model feedback loops using a summing device and a feedback loop with a gain factor of g. The input is the blackbody sensitivity λ_B. The output λ is the climate sensitivity factor, including the impacts of feedback. The feedback factor g is positive for processes having positive feedback, and negative for those with negative feedback. From the diagram, we can write the following:

$$\lambda = \lambda_B + g\,\lambda \qquad (39)$$

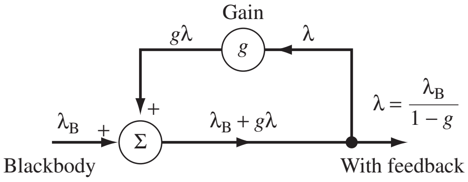


FIGURE 27 A simple model used to modify the blackbody climate sensitivity factor λ_B into a climate sensitivity factor λ that includes feedback phenomena.

Solving for λ gives

$$\lambda = \frac{\lambda_B}{1 - g} \tag{40}$$

The gain g used in the feedback model can be expressed as the sum of individual feedback factors

$$g = g_{\text{water vapor}} + g_{\text{cloud}} + g_{\text{surface ice}} + g_{\text{lapse rate}} + g_{\text{other}} \tag{41}$$

EXAMPLE 9 Climate Sensitivity with Feedback

Suppose the combined radiative forcing of greenhouse gases, aerosols, and solar intensity is 4.2 W/m^2, which is the focing that is thought to result from a doubling of CO_2. Using the blackbody sensitivity found in (38), $\lambda_B = 0.27°C/(W/m^2)$, find the global equilibrium temperature increase when the following feedback factors are included:

a. Water-vapor feedback, $g = 0.5$
b. Water-vapor feedback, $g = 0.5$, plus ice-albedo feedback, $g = 0.25$
c. Water-vapor feedback, $g = 0.5$, plus cloud feedback, $g = -0.25$
d. What feedback factor would yield an equilibrium $\Delta T_{2X} = 3.0$ K?

Solution: Including the specified feedbacks in (40) gives

a. The new sensitivity factor is $\lambda = \frac{\lambda_B}{1 - g} = \frac{0.27}{1 - 0.5} = 0.54°C/(W/m^2)$.

From (34), $\Delta T_e = \lambda \Delta Q = 0.54°C/(W/m^2) \times 4.2\ W/m^2 = 2.3°C$.

b. Including an additional $g = 0.25$ yields

$$\Delta T_e = \frac{\lambda_B}{1 - g}\Delta Q = \frac{0.27}{1 - 0.5 - 0.25} \times 4.2 = 4.5°C$$

c. With the -0.25 negative feedback provided by clouds,

$$\Delta T_e = \frac{\lambda_B}{1 - g}\Delta Q = \frac{0.27}{1 - 0.5 + 0.25} \times 4.2 = 1.5°C$$

d. From (35), for $\Delta T_{2X} = 3.0°C$,

$$\lambda = \frac{\Delta T_{2X}}{4.2} = \frac{3.0}{4.2} = 0.714°C/(W/m^2)$$

Rearranging (40) gives a feedback factor of

$$g = 1 - \frac{\lambda_B}{\lambda} = 1 - \frac{0.27}{0.714} = 0.62$$

The results shown in Example 9 span the typical range of climate sensitivity factors ΔT_{2X} used in global circulation models, ranging from 1.5°C to 4.5°C, and shows the feedback factor needs to be about 0.62 to produce the often-used $\Delta T_{2X} = 3°C$. Figure 28 points out how sensitive λ becomes as the feedback factor

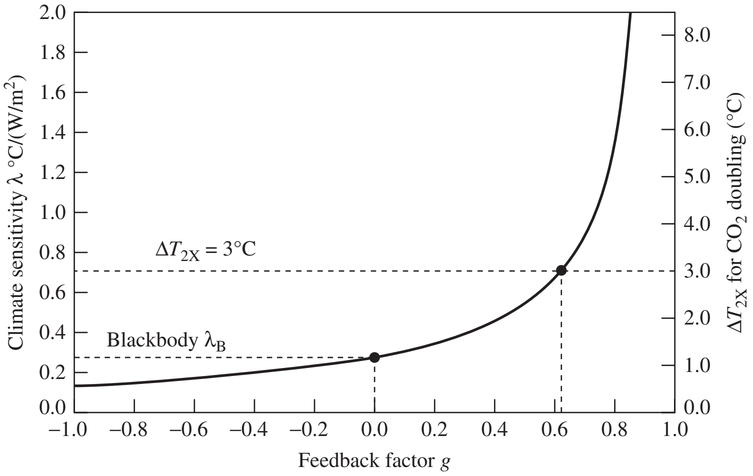


FIGURE 28 The climate sensitivity λ becomes very sensitive to the feedback factor as g approaches 1. An example with a 3°C equilibrium surface temperature change for a CO_2 doubling (ΔT_{2X}) is also shown.

increases. Once a strong positive feedback is present in the system, the impacts of other feedback processes are amplified.

The preceding analysis depends on a number of gross simplifications. It is a linear response model, which is really only valid for small perturbations around the equilibrium climate system. It assumes the various feedback processes (cloud, ice, lapse rate, and so on) are independent and additive, when in fact they do interact with each other. And it is an equilibrium analysis that ignores thermal capacitance and lag time responses, when the actual climate system is certainly not in equilibrium and won't be so for a good long time. Finally, the mathematics of (40) blows up for a feedback factor of $g = 1$, at which point the slightest radiative forcing would produce an infinite temperature change. It is introduced here as a rough guide to the relative importance of feedback processes.

Radiative Forcings Since Preindustrial Times

The concept of radiative forcing of climate change can be applied to the accumulation of greenhouse gases in the atmosphere, changes in aerosols from natural and anthropogenic sources, ozone depletion in the stratosphere, photochemically produced ozone accumulation in the troposphere, and the natural variability in solar intensity reaching Earth's outer atmosphere. Both positive and negative forcings are possible. Positive forcings contribute to global warming, while negative forcings tend to cool the Earth.

Gases and particulate matter added to the atmosphere can exert both *direct* and *indirect* radiative forcing effects. Direct forcing is caused by substances in the atmosphere that have actually been emitted from some source. Indirect forcings are those that occur when those substances go on to cause other atmospheric changes that affect radiative properties of the atmosphere. For example, aerosols have a direct effect on forcing when they absorb or reflect sunlight. Aerosols can also cause an indirect effect when they induce changes in the albedo of clouds. Halocarbons provide another example of direct and indirect effects. The direct effect of halocarbons is an increase in radiative forcing as these gases absorb long-wave radiation from the Earth. They also, however, cause an indirect effect by destroying ozone in the stratosphere. Recall from Figure 11 that ozone absorbs in the middle of the atmospheric window, so ozone destruction opens the window and allows the Earth to cool more readily. The direct effect of halocarbons therefore contributes to global warming, while the indirect effect of destroying ozone works in the opposite direction to help to cool the planet.

Figure 29 shows global average radiative forcing estimates associated with CO_2, the other greenhouse gases (CH_4, CFCs, N_2O), aerosols, land-use changes, and the estimated increase in solar intensity that has occurred in this period. Aerosol effects are broken out into the positive forcing caused by black carbon and the negative forcings caused by the indirect impact of reflective aerosols and induced cloud changes.

Several features stand out. Carbon dioxide provides the largest forcing, but the others are significant; conventional air pollution, responsible for black carbon and ozone, is also substantial; and the largest uncertainties are associated with aerosol

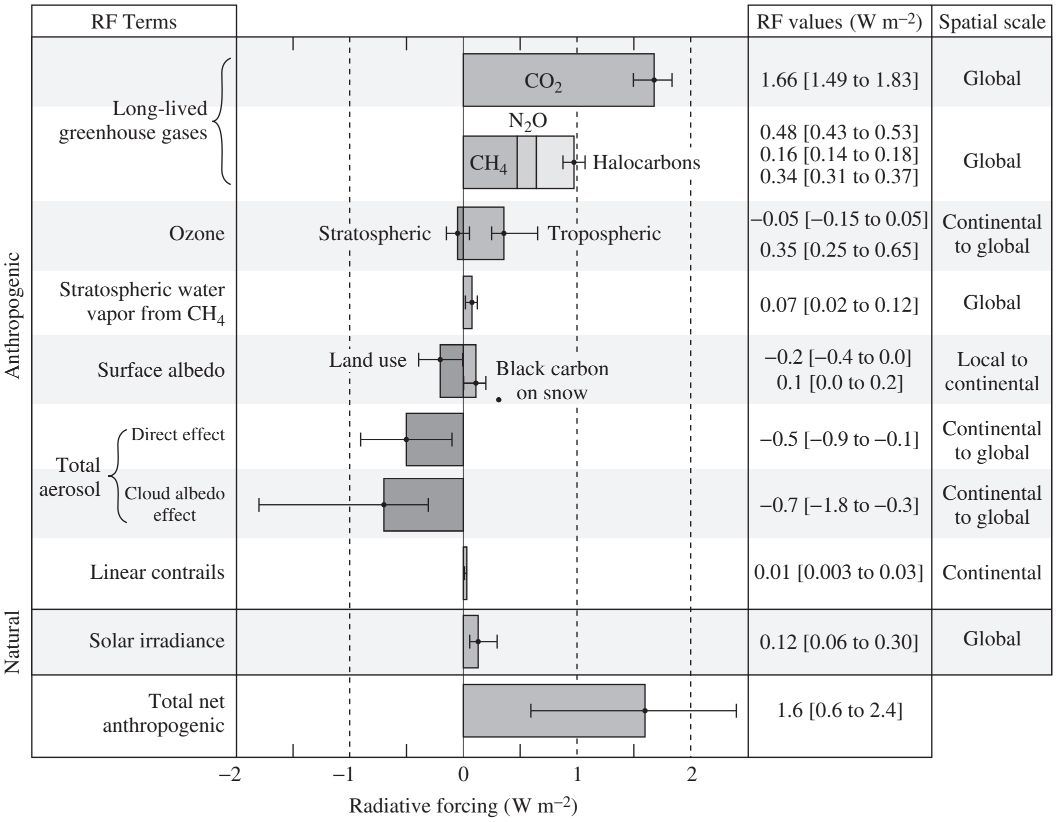


FIGURE 29 Global mean radiative forcing during the Industrial Era, 1750–2005, in W/m². Carbon dioxide is the best understood and ranks as the most important. Estimates of aerosol effects are significant and show the greatest uncertainty. Error bars are based on one standard deviation probabilities. (*Source:* IPCC, 2007.)

effects, both direct and on clouds. In spite of the significant black carbon contribution to warming, the net effect of aerosols seems to be negative, meaning global cooling, but the uncertainties shown make that not a clear assertion.

The 2007 IPCC 4th Assessment Report (AR4) asserts with very high confidence that the globally averaged net effect of human activities since 1750 has been one of warming with a radiative forcing of $+1.6\ \text{W/m}^2$ (+0.6 to +2.4). The combined forcing due to greenhouse gases themselves is $+2.3\ \text{W/m}^2$, and its rate of increase during the industrial era is very likely to have been unprecedented in more than 10,000 years.

EXAMPLE 10 Combined Forcings

Using the IPCC estimate of $1.6\ \text{W/m}^2$ net radiative forcing shown in Figure 29, along with an equilibrium temperature increase caused by those forcings of 0.8°C, find the following:

a. The climate sensitivity factor λ, $°\text{C}/(\text{W/m}^2)$

b. The feedback gain factor g

c. The temperature climate sensitivity ΔT_{2X} (°C)

Solution

a. From (34), we get the climate sensitivity:

$$\lambda = \frac{\Delta T_e}{\Delta Q} = \frac{0.8}{1.6} = 0.5°\text{C}/(\text{W/m}^2)$$

b. From (40), the feedback gain factor is

$$g = 1 - \frac{\lambda_B}{\lambda} = 1 - \frac{0.27}{0.5} = 0.54$$

c. From (35), we estimate the temperature sensitivity for a doubling of CO_2:

$$\Delta T_{2X} = 4.2\lambda = 4.2 \times 0.54 = 2.27°\text{C}$$

The simple summation of forcings suggested in the preceding example assumes that $1\ \text{W/m}^2$ of one constituent (e.g., CO_2) has the same impact as $1\ \text{W/m}^2$ of another (e.g., solar intensity). A new *radiant forcing efficacy* factor E is based on weighing each climate sensitivity in terms of its global warming impact with respect to CO_2.

$$E = \frac{\lambda}{\lambda_{CO_2}} \tag{42}$$

For example, the well-mixed greenhouse gases have an efficacy of 1.0, while the direct and indirect effect of aerosols might have an efficacy of 0.72 (National Research Council, 2005). If, for example, the forcing of greenhouse gases is $3.0\ \text{W/m}^2$ while aerosols have a net forcing of $-1.8\ \text{W/m}^2$, the combined impact of the two with and without considering efficacy would be

$$(CO_2 + \text{aerosol})\text{-forcing without efficacy} = 3.0 + (-1.8) = 1.2\ \text{W/m}^2$$

$$(CO_2 + \text{aerosol})\text{-forcing with efficacy} = 3.0 \times 1.0 + (-1.8) \times 0.72 = 1.7\ \text{W/m}^2$$

Clearly, the efficacy factor concept could have a significant impact on models based on radiant forcing.

9 Global Warming Potential

The *Global Warming Potential* (GWP) is a weighting factor that allows comparisons to be made between the cumulative global warming impact over a specified period of time of some greenhouse gas and a simultaneous emission of an equal mass of CO_2. There are three primary factors that affect GWPs. The first is the radiative forcing associated with the addition to the atmosphere of a unit mass of each greenhouse gas. The second is based on estimates of the rate at which that unit mass injected decays over time. The third is related to the cumulative radiative forcing that the unit addition to the atmosphere will have over some period of time into the future.

Some examples of how the GWP index can be used include the following:

1. By combining GWP with estimates of the cost of curtailing emissions of each greenhouse gas, a country-by-country, least-cost approach to prevention of climate change could be identified.
2. GWPs could facilitate trading of emissions reductions among countries. For example, one country might decide the least expensive way to offset emissions of CO_2 might be to reduce emissions of CH_4 in another. Buying and selling international carbon emission offsets could lead to technology and economic transfer from the developed to the developing countries (Swisher and Masters, 1991).
3. Country by country rankings of their individual contributions to climate change are made possible. Quantifiable goals can then be established for future reductions.
4. Combined with other indices, they could be part of an overall environmental impact assessment for products and industrial process audits. Included in an environmental labeling system, consumer choices could be affected.

Calculating the GWP

Mathematically, we imagine an impulse function in which 1 kg of the greenhouse gas in question and 1 kg of CO_2 are emitted into the atmosphere at the same time. As the concentration of each gas decreases with time, so does the radiative forcing associated with the remaining amount of gas. Mathematically, the GWP of a greenhouse gas is given by

$$\mathrm{GWP}_g = \frac{\int_0^T F_g \cdot R_g(t)\, dt}{\int_0^T F_{\mathrm{CO}_2} \cdot R_{\mathrm{CO}_2}(t)\, dt} \tag{43}$$

where

F_g = radiative forcing efficiency of the gas in question, $(\mathrm{W/m^2})/\mathrm{kg}$
F_{CO_2} = radiative forcing efficiency of CO_2, $(\mathrm{W/m^2})/\mathrm{kg}$
$R_g(t)$ = fraction of the 1 kg of gas remaining in the atmosphere at time t
R_{CO_2} = fraction of the 1 kg of CO_2 remaining at time t
T = the time period for cumulative effects (years)

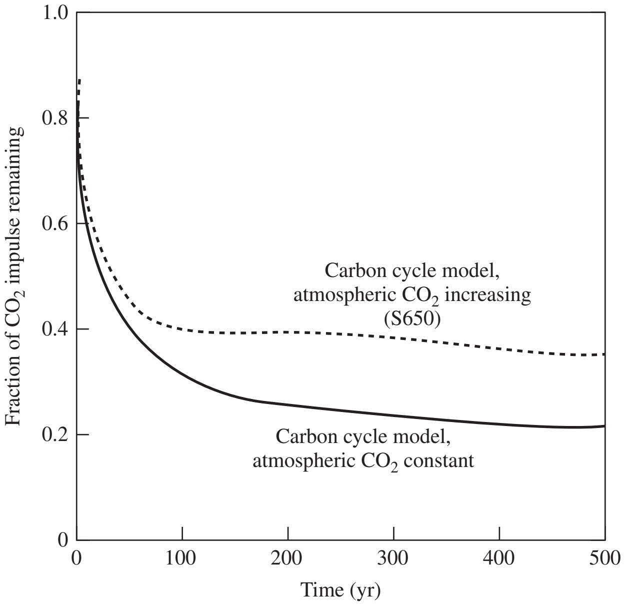


FIGURE 30 The impulse response for an injection of 1 kg of CO_2 into the atmosphere; that is, $R_{CO_2}(t)$ in (43). The solid line is for an atmosphere in which the CO_2 concentration is constant; the dashed line is for an atmosphere that stabilizes at 650 ppm of CO_2 by the year 2200. (*Source:* IPCC, 1995.)

For some greenhouse gases, R_g can be modeled using a simple exponential decay function. And for some, the radiative efficiency is a constant. Neither is the case for the comparison gas, CO_2. As Figure 30 suggests, CO_2 decay is rapid during the first few decades as the biosphere absorbs the carbon, and then for next few hundred years, it decays at a much slower rate corresponding to the slow uptake of the oceans. To further complicate matters, the decay also depends on the assumptions made for the background concentration of CO_2. Two curves for R_{CO_2} are shown, one corresponding to an atmosphere in which CO_2 concentration is unchanging, while the other has been drawn for a more realistic atmosphere in which CO_2 reaches 650 ppm by the year 2200.

The choice of the GWP time horizon that policy makers might use depends on the issues that are being addressed. For example, shorter time horizons might be used when it is the rate of change of global temperature, rather than the ultimate temperature increase, that is of concern. Short time horizons would also seem appropriate when the most effective greenhouse abatement strategy is needed to head off possibly abrupt climate changes that might be triggered when warming reaches some threshold level. On the other hand, long time horizons might be used to evaluate strategies to avoid long, slow, irreversible impacts such as sea level changes.

Carbon Dioxide Equivalents of Greenhouse Gases

Table 7 shows GWPs for a number of widely used chemicals. As can be seen, all of these species have very high GWPs relative to CO_2 because they all absorb in the atmospheric window. Notice the three standard time horizons for GWPs are 20 years, 100 years, and 500 years. By including a time horizon, the GWP accounts for the

TABLE 7

Global Warming Potentials (GWPs) Relative to Carbon Dioxide (kilograms of CO_2 per kilogram of gas)

			Global Warming Potential (Time Horizon in Years)		
Chemical Species	Chemical Formula	Lifetime (yrs)	20-yr	100-yr	500-yr
Carbon dioxide	CO_2	50–200	1	1	1
Methane	CH_4	12	62	23	7
Nitrous oxide	N_2O	114	275	296	156
CFC-11	$CFCl_3$	45	6,300	4,600	1,600
CFC-12	CF_2Cl_2	100	10,200	10,600	5,200
CFC-113	$CF_2ClCFCl_2$	85	6,100	6,000	2,700
CFC-115	CF_3CClF_2	1,700	4,900	7,200	9,900
HCFC-22	CHF_2Cl	12	4,800	1,700	540
HCFC-141b	CH_3CFCl_2	9.3	2,100	700	220
HCFC-142b	CH_3CF_2Cl	19	5,200	2,400	740
HFC-23	CHF_3	270	9,400	11,700	10,000
HFC-134a	CH_2FCF_3	14	3,300	1,300	400
HFC-152a	CH_3CHF_2	1.4	410	120	37
Tetrafluoromethane	CF_4	50,000	3,900	5,700	8,900
Hexafluoroethane	C_2F_6	10,000	8,000	11,900	18,000
Sulfur hexafluoride	SF_6	3,200	15,100	22,200	32,400
Carbon tetrachloride	CCl_4	35	2,700	1,800	580
Methyl bromide	CH_3Br	1	16	5	1
Halon-1301	CF_3Br	65	7,900	6,900	2,700

Source: IPCC, 3rd Assessment, 2001.

greater impact that a gas with a long atmospheric lifetime will have compared with one that quickly disappears after it has been emitted. For example, CFC-115 and a replacement HCFC-22 have almost identical 20-year GWPs, but the longer atmospheric lifetime of CFC-115 (1,700 vs. 12 years) means its 100-year GWP is more than five times higher (7,200 vs. 1,700), and its 500-year GWP is almost 20 times higher.

Notice the extremely high GWPs for several gases in this table. Some of the HFCs, which were developed as alternatives to ozone-depleting substances, stand out—especially trifluoromethane (HFC-23). HFC-23 has a lifetime of 270 years and a 100-yr GWP of 11,700. Reducing HFC-23 in developing countries has become one of the most sought-after (least expensive) approaches to meeting Kyoto Protocol obligations. The perfluorocarbons CF_4 (tetrafluoromethane) and C_2F_6 (hexafluoroethane), used in semiconductor manufacturing and primary aluminum production, have extremely long atmospheric lifetimes and GWPs thousands of times higher than CO_2. One gas, sulfur hexafluoride (SF_6) has a GWP of 22,200, which gives it the distinction of being the most potent greenhouse gas the IPCC has ever evaluated. SF_6 has excellent dielectric properties and was widely used in electric power and distribution equipment.

The purpose of the GWPs is to assess the relative importance of various emissions, which means the GWP for each gas should be multiplied by the emission rate for the gas. The 100-year GWPs are often used to describe non-CO_2 emissions in terms of *carbon dioxide equivalents* (CDE, or CO_2-eq), as Example 11 demonstrates.

EXAMPLE 11 Carbon Dioxide Equivalent of Greenhouse Gases

Annual anthropogenic emissions of CO_2, CH_4, and N_2O are estimated to be 27,000 $MtCO_2$/yr, 370 $MtCH_4$/yr, and 6 MtN_2O/yr (where "Mt" means million metric tons). Compare the impacts of these three gases over a 100-year time horizon, and find the total equivalent CO_2 emissions.

Solution The comparison will be based on the products of emission rates and GWPs. GWPs are found in Table 7, and the emission rates are given. Carbon dioxide has a GWP of 1 (by definition) over all of the time horizons, so its product doesn't change:

$$\begin{aligned} CO_2&: \quad GWP_{100} \times \text{emissions} = 1 \times 27{,}000 \text{ Mt/yr} = 27{,}000 \text{ MtCO}_2 \\ CH_4&: \quad GWP_{100} \times \text{emissions} = 23 \times 370 \text{ Mt/yr} = 8{,}510 \text{ MtCO}_2\text{-eq} \\ N_2O&: \quad GWP_{100} \times \text{emissions} = 296 \times 6 \text{ MtN}_2\text{O/yr} = 1{,}776 \text{ MtCO}_2\text{-eq} \end{aligned}$$

The total carbon dioxide equivalence of the combined emissions is

$$27{,}000 + 8{,}510 + 1{,}776 = 37{,}286 \text{ MtCO}_2\text{-eq}$$

If we multiply this by the C/CO_2 ratio of 12/44 and switch the units, we can get an equivalence expressed in gigatons of carbon:

$$\text{Total } CO_2, CH_4, N_2O \text{ emissions} = 37.3 \text{ GtCO}_2\text{-eq} \times 12/44 = 10.1 \text{ GtC-eq}$$

Of this 10.1 GtC total, 7.4 GtC is actual carbon dioxide, and 2.8 GtC is other gases.

Example 11 suggests that almost three-fourths of the impact for the next century from current emissions of the three key greenhouse gases will be caused by CO_2. If we redo the calculations using 20-year GWPs, the near-term importance of controlling methane emissions stands out. Over the next 20 years, today's CH_4 emissions will have nearly as much impact on climate (85 percent as much) as today's CO_2 emissions.

Applying the above carbon dioxide equivalence approach to emissions puts the combination of greenhouse gases into perspective. Figure 31 shows the GWP-weighted emissions of U.S. greenhouse gases, and Figure 32 shows a "spaghetti chart" representation of their sources and the end uses responsible for those emissions.

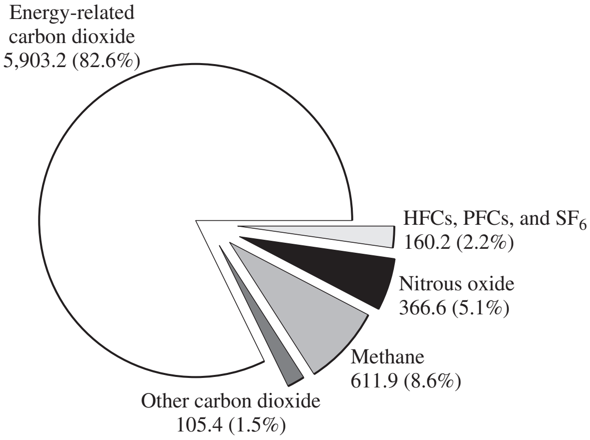


FIGURE 31 U.S. greenhouse gas emissions using 100-year GWPs (millions of metric tons of CO_2 equivalents and percentages of the total.
(*Source:* EIA, 2006.)

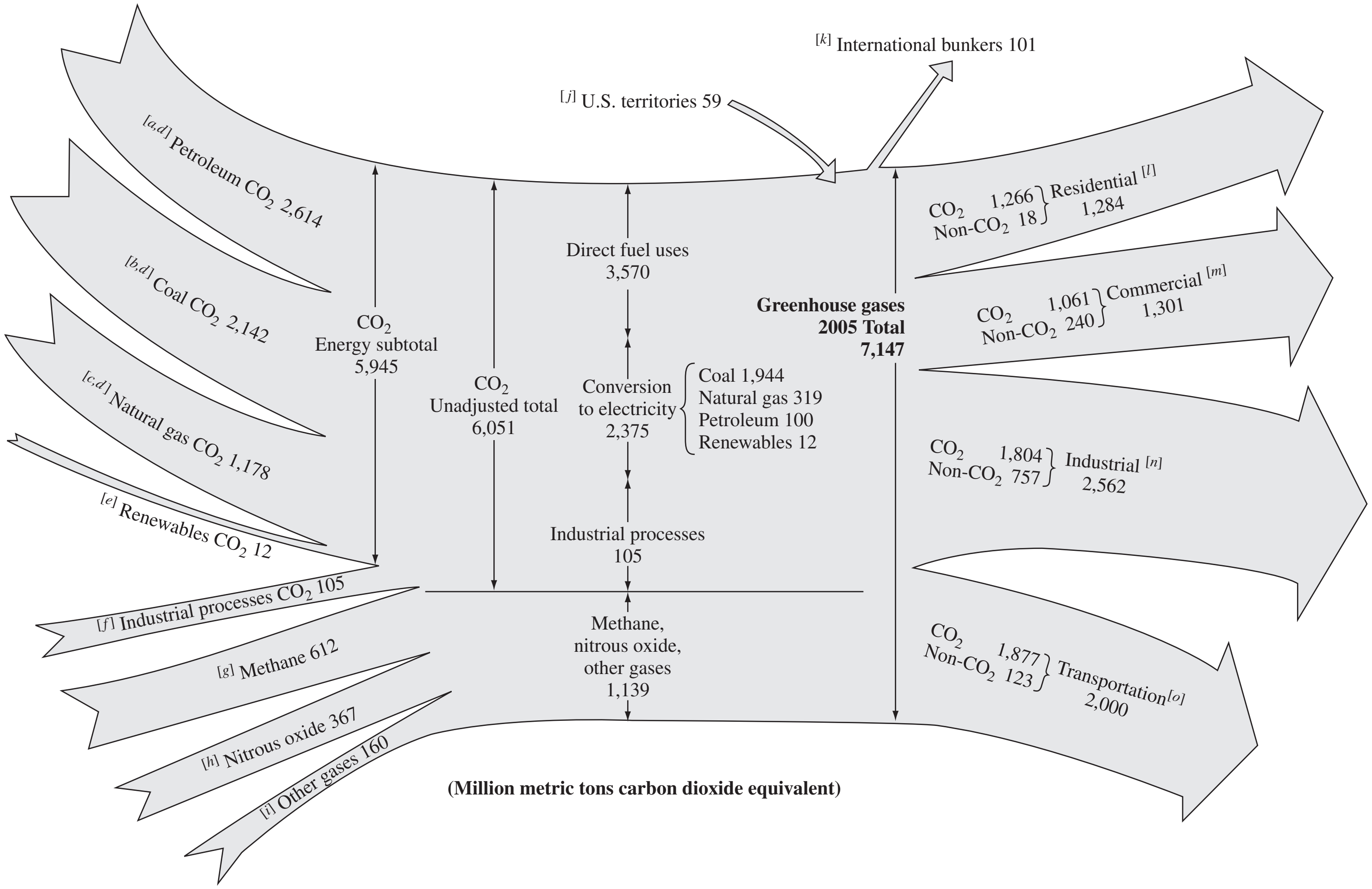
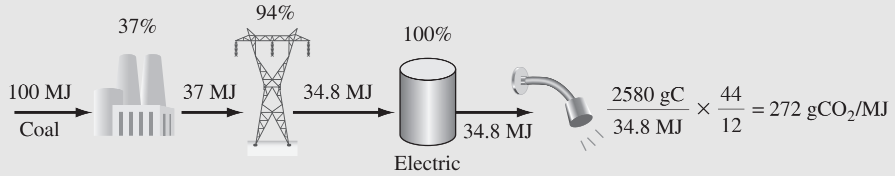


FIGURE 32 Greenhouse gas emissions in the U.S. economy, 2005.
(*Source:* EIA, 2006.)

EXAMPLE 12 Water Heater Emissions, Including Leakage

In Example 6, CO_2 emissions from a coal-fired power plant delivering power to an electric water heater were compared with emissions for a natural-gas-fired water heater.



That analysis, however, didn't include the global warming impacts of the roughly 1.5 percent methane leakage to the atmosphere during the mining, transportation, and storage of natural gas before the gas makes it to the house. Including that leakage, what is the equivalent carbon emission rate for the 85 percent-efficient gas water heater? Use the 20-year GWP for natural gas.

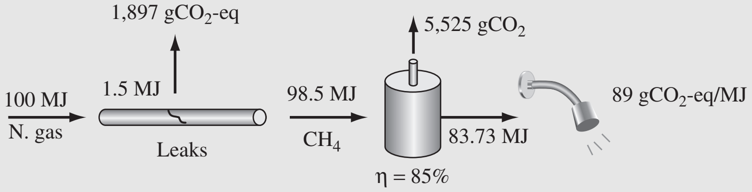
Solution Starting with 100 MJ of natural gas, we'll lose 1.5 MJ in leakage and deliver 98.5 MJ to the gas-fired water heater.

From Table 3, the LHV carbon intensity of natural gas is 15.3 gC/MJ. From Table 7, the 20-year GWP of methane is 62 gCO_2/gCH_4. So, the CO_2 equivalent of that 1.5 MJ of methane leakage is

$$1.5\ \text{MJ} \times 15.3\ \text{gC/MJ} \times \frac{16\ \text{gCH}_4}{12\ \text{gC}} \times \frac{62\ \text{gCO}_2}{\text{g CH}_4} = 1{,}897\ \text{gCO}_2\text{-eq}$$

The actual CO_2 emissions from the 85 percent-efficient water heater are

$$\text{CO}_2\ \text{emissions} = 15.3\ \text{gC/MJ} \times 98.5\ \text{MJ} \times \frac{44\ \text{gCO}_2}{12\ \text{gC}} = 5{,}525\ \text{gCO}_2$$



Total emissions from the gas-fired water heater per MJ actually heating water are

$$\frac{1{,}897\ \text{gCO}_2\text{-eq} + 5{,}525\ \text{gCO}_2}{83.73\ \text{MJ heat to water}} = 89\ \text{gCO}_2\text{-eq/MJ}$$

which is still a 68 percent savings compared to the 272 gCO_2 for the electric water heater (without including the leakage, the savings in Example 6 was 75 percent).

10 IPCC Assessment Reports

The Intergovernmental Panel on Climate Change (IPCC) is an organization founded in 1988 by the United Nations Environment Program (UNEP) and the World Meteorological Organization (WMO). Its role is to assess scientific, technical, and socioeconomic information related to the risks, potential impacts, and options for adaptation and mitigation of human-induced climate change. The IPCC itself does not carry out research but instead relies on literally thousands of climate and policy experts from around the globe who contribute their expertise as authors and reviewers of IPCC reports. Its principal output has been a series of assessment reports, which are based on peer-reviewed and published scientific and technical literature. The first IPCC Assessment Report, completed in 1990, provided the overall scientific and policy basis for addressing climate change adopted by the 1992 UN Framework Convention on Climate Change (UNFCCC), which entered into force in 1994. The Second Assessment Report (SAR), *Climate Change 1995,* was instrumental in negotiations that led to the adoption of the Kyoto Protocol by the UNFCCC in 1997.

The Third Assessment Report (TAR), *Climate Change 2001,* and the Fourth Assessment Report (AR4), which was released in 2007, continue to be the most authoritative summary documents on the status of climate research. The IPCC also prepares special reports and technical papers to fill the gaps between those major, comprehensive assessment reports.

The IPCC is organized around three focus areas: Science (Working Group I), Impacts and Adaptation (Working Group II), and Mitigation (Working Group III). The mitigation group develops scenarios that describe possible combinations of population, economic growth, and technology drivers that lead to future greenhouse gas emissions. These emissions scenarios are then handed off to the climate modeling community, which converts them into regional and global temperatures and climate impacts. Those, in turn, can inform decision makers about the potential to mitigate these impacts (by modifying the emissions driving forces) and/or adapting to those climate changes (e.g., raise the levees). Figure 33 summarizes these interactions and helps frame the climate problem in terms of both sustainable development and environment.

The Special Report on Emissions Scenarios (SRES)

The IPCC *Special Report on Emissions Scenarios* (SRES) published in 2000 is based on a set of four storylines that weave demographic, technological, and economic driving forces into four distinct scenarios that can be translated into future emissions. Scenarios provide alternative images about how the future might unfold. They are not forecasts, nor do they carry with them any probabilities of their likelihood. No preferences for one scenario over another are offered by IPCC, and they are not meant to infer policy recommendations. They are narratives, around which emission scenarios can be built.

The SRES report describes four narrative storylines, designated as A1, A2, B1, and B2. The A storylines focus more on economic growth, while the B storylines have more of an environmental emphasis. The A1 and B1 lines are based on more of a homogeneous, globalized future, while the A2 and B2 imagine a more fractured

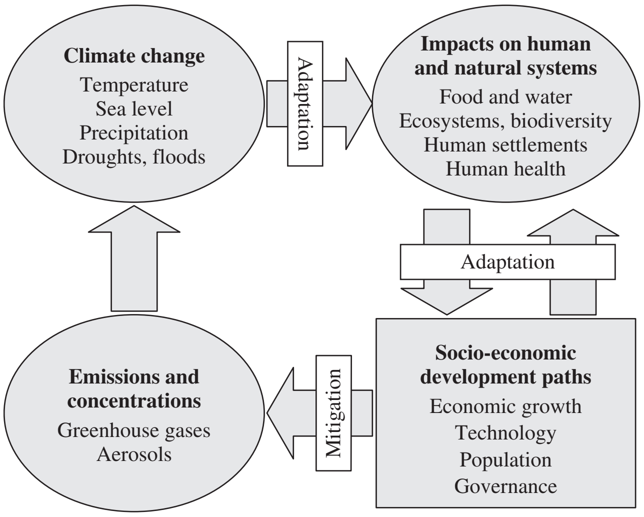


FIGURE 33 Sustainable development, adaptation, and mitigation interactions. (*Source:* IPCC, TAR, 2006.)

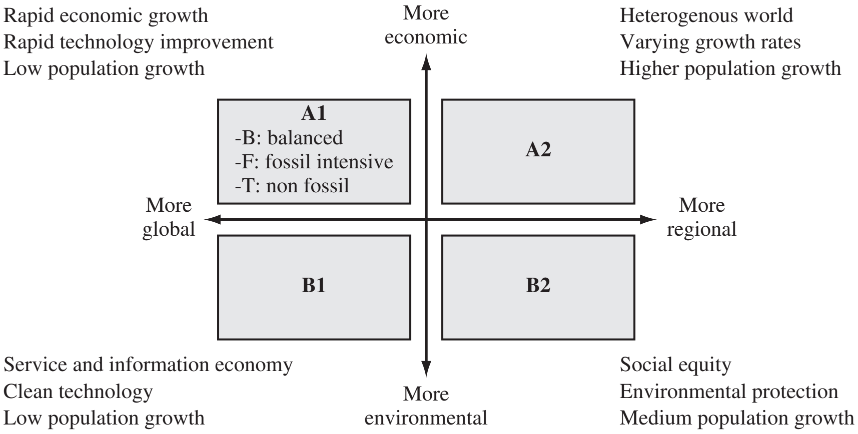


FIGURE 34 The four families of storylines in the SRES emission scenarios.

world with different regional priorities and growth rates. The A1 line is further designated to be either fossil-fuel intensive (A1-F), nonfossil fuel, technology emphasis (A1-T), or a more balanced combination of the two (A1-B). These characterizations are summarized in Figure 34.

Within each of the four storylines, multiple scenarios have been developed. Table 8 summarizes population, world GDP, and primary energy that result from a single illustrative scenario for each of the six categories (A1F, A1B, A1T, A2, B1, B2), but in reality there are many more within each category. One clue to the emphasis placed on technology in each scenario is the relative dependence on coal versus zero-carbon energy sources for each scenario, also shown in the table. Carbon dioxide emissions that result from these scenarios are shown in Figure 35.

TABLE 8

Some Main Driving Forces for Illustrative Scenarios from the Six SRES Storylines (each of the six categories has multiple scenarios not shown)

	A1F1	A1B	A1T	A2	B1	B2
Population (billion)						
1990	5.3	5.3	5.3	5.3	5.3	5.3
2020	7.6	7.5	7.6	8.2	7.6	7.6
2050	8.7	8.7	8.7	11.3	8.7	9.3
2100	7.1	7.1	7.0	15.1	7.9	10.4
World GDP (10^{12} $\$1990$/yr)						
1990	21	21	21	21	21	21
2020	53	56	57	41	53	51
2050	164	181	187	82	136	110
2100	525	529	550	243	328	235
Primary Energy (10^{18}J/yr)						
1990	351	351	351	351	351	351
2020	669	711	649	595	606	566
2050	1431	1347	1213	971	813	869
2100	2073	2226	2021	1717	514	1357
Share of Coal in Primary Energy (%)						
1990	24	24	24	24	24	24
2020	29	23	23	22	22	17
2050	33	14	10	30	21	10
2100	29	4	1	53	8	22
Share of Zero Carbon in Primary Energy (%)						
1990	18	18	18	18	18	18
2020	15	16	21	8	21	18
2050	19	36	43	18	30	30
2100	31	65	85	28	52	49

Source: IPPC, SRES, 2000.

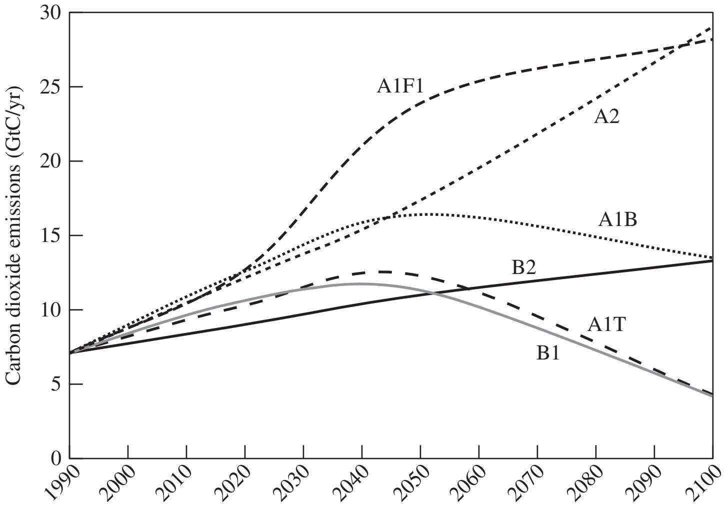


FIGURE 35 Illustrative SRES scenarios for fossil-fuel plus land-use carbon dioxide emissions corresponding to the scenarios shown in Table 8.
(*Source:* Based on IPCC, SRES, 2000.)

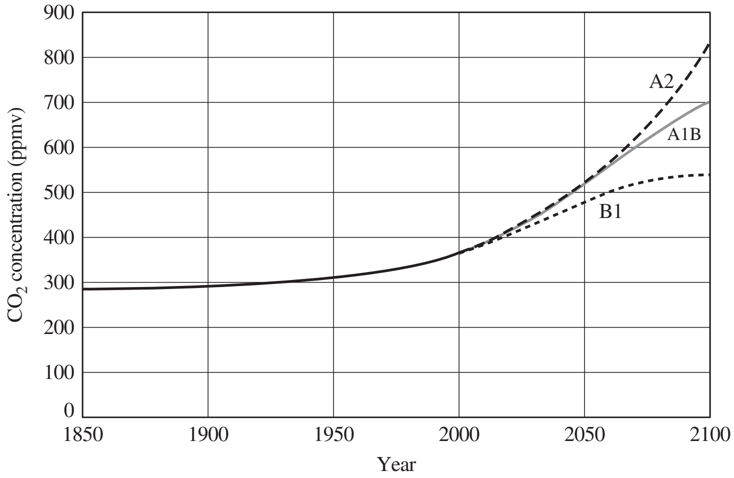


FIGURE 36 Simulations of CO_2 concentrations for three SRES scenarios done for the Fourth Assessment Report (AR4).
(*Source:* Max Planck Institute for Meteorology, 2006.)

The SRES scenarios provide estimates for greenhouse gas emissions over the next 100 years or so. Those emission rates can then be used as inputs for climate models that estimate future atmospheric concentrations of CO_2, CH_4, N_2O, and other greenhouse gases, from which GWP-weighted equivalent CO_2 concentrations can be estimated. Figure 36 shows one set of simulations done at the Max Planck Institute for Meteorology (MPI-M) in Hamburg for three of the SRES scenarios, A1B, A2, and B1, which cover the range of scenarios shown in Figure 35. Under the most optimistic scenario B1, CO_2 stabilizes at close to double the preindustrial concentration.

General Circulation Models

Scenarios for future greenhouse gas emissions are used as inputs to computer models that attempt to predict likely impacts of these perturbations on the Earth's climate systems. At this point in time, even the largest supercomputer cannot handle all of the mathematical complexity that could be included if scientists tried to model everything they know about how the climate system works. So, a hierarchy of models has been created, with the most sophisticated of those being *general circulation models* (GCMs). There are both atmospheric-GCMs and ocean-GCMs, which can be coupled together to form an atmospheric-ocean coupled general circulation model (AOGCM). As GCMs have evolved in complexity, they have gotten better and better at matching the historical record, which adds confidence in their predictions for the future. A complete global climate model would couple an AOGCM with a land model, a sea-ice model, and models for aerosols, carbon cycle, vegetation dynamics, and atmospheric chemistry.

A 3D (actually 4D if you include time) atmospheric general circulation model is based on a cartesian grid in which the atmosphere is divided into cells, such as are shown in Figure 37. For the ECHAM5/MPI-OM coupled atmosphere-ocean

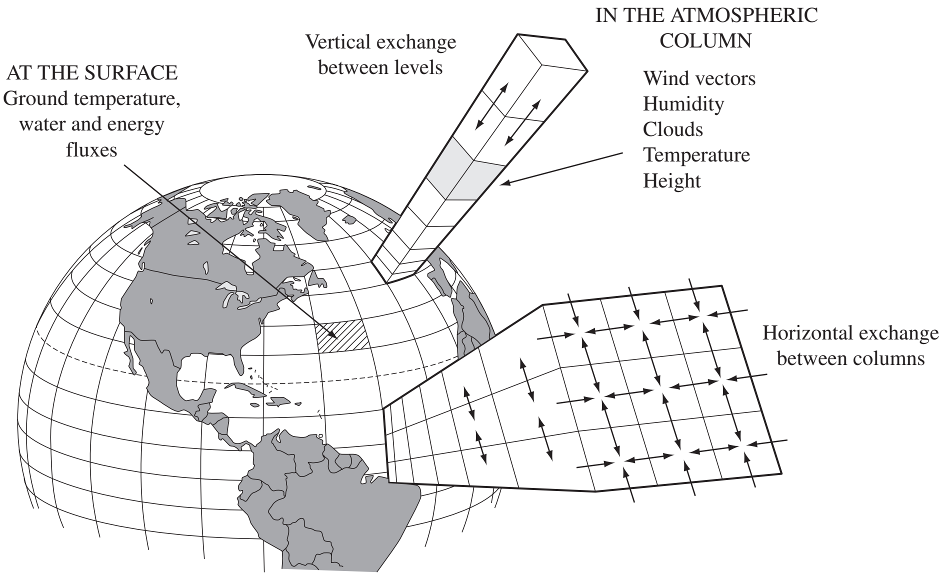


FIGURE 37 A cartesian (or rectangular) grid GCM in which horizontal and vertical exchanges are handled between adjacent columns and layers.
(*Source:* Based on Henderson-Sellers and McGuffie, 1987.)

model used at the Max Planck Institute for Meteorology in Germany, the cells have a horizontal resolution of approximately 200 km with 31 vertical levels. Their ocean GCM has regionally varying horizontal resolution between approximately 10 km and 150 km. Each cell is characterized by a number of variables such as temperature, humidity, pressure, cloudiness, and so forth. Those cells then communicate (mathematically) with adjacent atmospheric cells and are updated with some specified time increment. A grid coupling of atmosphere, ocean, and land models is shown in Figure 38.

Global Mean Temperature Simulations

After having run simulations for accumulating concentrations of carbon dioxide and other greenhouse gases, GCMS attempt to predict future global climate responses. A major uncertainty introduced at this stage of modeling is properly estimating the appropriate climate sensitivity factor, ΔT_{2X} (the change in mean global temperature resulting from a doubling of the equivalent CO_2 concentration above its preindustrial level).

The IPCC has often used three possible values of climate sensitivity, 1.5°C, 2.5°C, and 4.5°C, but a more recent approach has been based on trying to use a probability density function (pdf) and its accompanying cumulative probability distribution to help quantify the uncertainty in this key parameter. Figure 39 shows two such probability estimates. One (labeled WR, after its author's initials), is based on a log-normal fit to the long-stated IPCC 90 percent probability range for ΔT_{2X} of 1.5°C to 4.5°C. The WR pdf has a median ΔT_{2X} probability of 2.6°C and indicates a 10 percent chance that it is greater than 4°C. The other (labeled AS) has a median

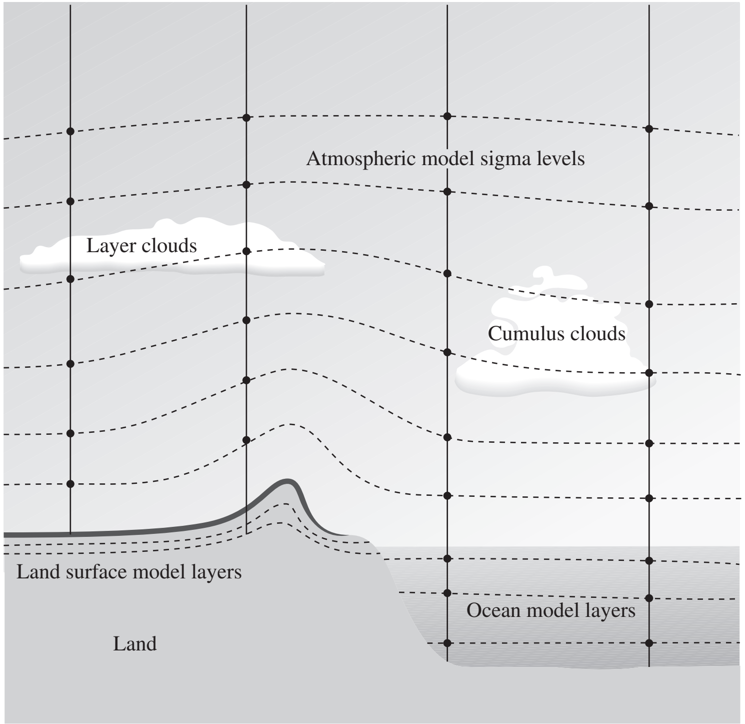


FIGURE 38 Coupled atmosphere, ocean, and land GCM grid.
(*Source:* Bureau of Meteorology, Australia.)

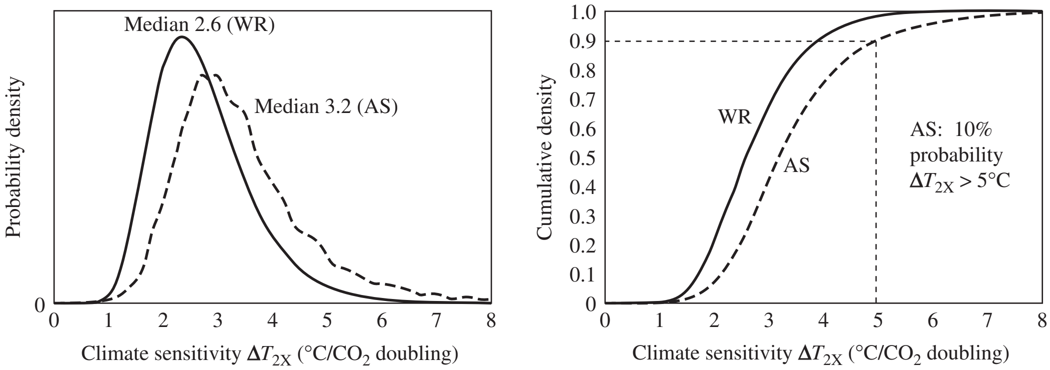


FIGURE 39 The probability density function and cumulative probability function for two estimates of the climate sensitivity factor, ΔT_{2X}. The dotted lines for the AS pdf indicate a 10 percent probability that the sensitivity factor is over 5°C.
(*Source:* WR = Wigley and Raper, 2001; AS = Andronova and Schlesinger, 2001.)

value of 3.2°C, a 90 percent probability that ΔT_{2X} is over 2.2°C, and a 10 percent probability that ΔT_{2X} is greater than 5.0°C. The difference at the high end of climate sensitivity for these two pdfs has a huge impact on the likelihood that we will be able to avoid dangerous climate changes in the future.

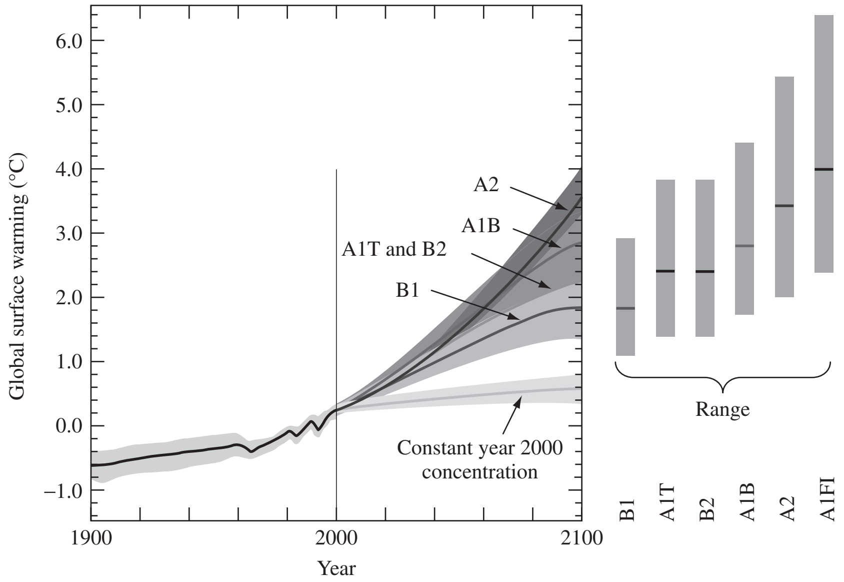


FIGURE 40 IPCC estimates of future global average surface warming relative to 1980 to 1999 for the six SRES scenarios. The gray bars to the right show best estimates and likely ranges for each scenario.
(*Source:* MTI-M, 2006.)

Figure 40 shows simulations of global average surface temperature for the six SRES scenarios. According to these 2007 simulations, a mean global warming between 2.4° and 4.4°C relative to mid-nineteenth-century levels is likely. Anywhere within that range would make the planet warmer than it has been for hundreds of thousands of years.

Dangerous Anthropogenic Interference with Climate

The United Nations Framework Convention on Climate Change calls for the "stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent *dangerous anthropogenic interference* (DAI) with the climate system" (UNFCC, 1992). Note the difference between stabilizing the rate at which carbon is emitted and stabilizing its concentration in the atmosphere. In fact, with a constant carbon emission rate at the current level, the equivalent carbon dioxide level would likely be double the preindustrial concentration of 280 ppm by 2100. Moreover, even if we stabilize the concentration of greenhouse gases in the atmosphere, it will take decades for the climate system to adjust to its new equilibrium temperature.

One way to begin thinking about dangerous climate change is to outline the sorts of impacts that global warming could have. The IPCC has done that, and their list of likely impacts in the relatively short term includes more frequent heat waves such as the one experienced in Europe in 2003, responsible for 35,000 deaths; more frequent intense storms like Katrina in 2005, leading to more extreme flooding and a surge in property damage; drier summers and increased risk of severe droughts with accompanying loss of farm productivity, especially in mid-latitude areas such as the

Mediterranean, Central America, and Southern Africa; less snow and reduced glacial extent, impacting areas that rely on snowmelt for water supply such as China, India, and Peru; sea-level rise and increased risk of coastal flooding that would displace millions of people in areas such as China, Bangladesh, and Egypt; and increasing ocean acidity as carbon dioxide is absorbed and converted to carbonic acid, leading to widespread impacts on marine ecosystems and biodiversity.

In the longer term, future warming could cause large-scale, potentially rapid and nonlinear climate responses, including the deglaciation of polar ice sheets, such as those in Greenland and West Antarctica, which could lead to several meters of sea level rise above that already expected due to thermal expansion of the oceans. Some climate changes could be nonlinear, producing responses that cause the climate system to jump from one stable state to another, as would occur if the North Atlantic thermohaline circulation system were to collapse (more on this later).

The question then arises as to how likely those impacts will become as a function of global temperatures increases. We have already seen the introduction of probability density functions in the context of the climate sensitivity factor. Some have tried to link probabilities to avoiding dangerous anthropogenic interference with climate. One such study shown in Figure 41 combines a number of assessments of the likelihood of exceeding the 2°C temperature rise thought to be capable of everything from widespread coral bleaching to disintegration of the Greenland ice sheet and perhaps a collapse of the West Antarctic ice sheet. As shown, a number of studies suggest that the stabilized CO_2-eq concentration needs to remain below about 450–500 ppm for the odds to be in favor of not exceeding a 2°C temperature rise.

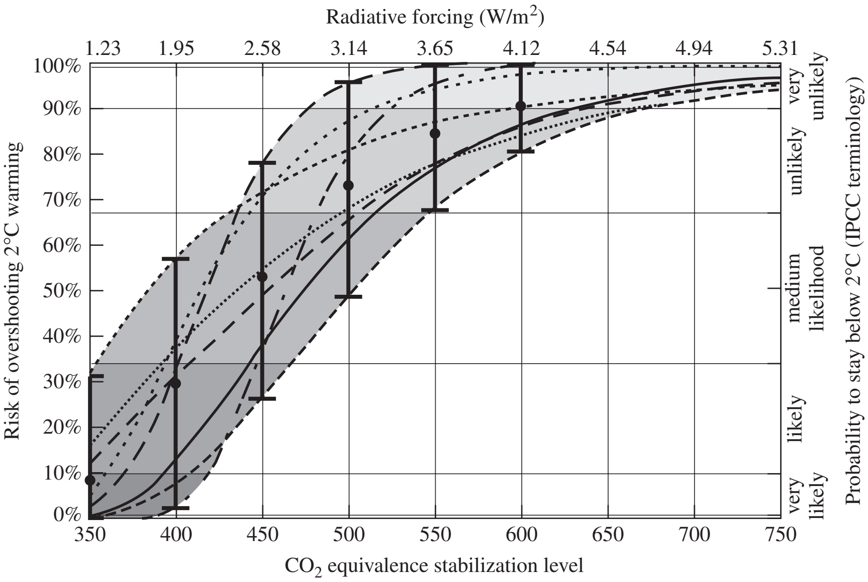


FIGURE 41 Estimated risks of exceeding a 2°C global mean temperature rise depending on the stabilized equivalent-CO_2 concentration. For example, if CO_2-eq stabilizes at 550 ppm (doubling of preindustrialized levels), the risk of overshooting a 2°C warming appears to be above 70 percent. Dotted and dashed lines correspond to different studies.
(*Source:* Meinhausen and den Elzen, 2004.)

11 Stabilizing Greenhouse Gases

The warning provided by studies of the likelihood of dangerous anthropogenic interference with climate increases our awareness of the need for mitigation strategies to help control greenhouse gas emissions. Since greenhouse gases accumulate in the atmosphere, to stabilize their concentration, we must do more than simply hold their emission rates constant. They need to be reduced to the point where their rate of addition to the atmosphere equals the rate at which they are removed. Figure 42 shows how emissions would have to decrease for the atmospheric concentration of CO_2 to stabilize at various levels. Notice avoiding dangerous anthropogenic interference would require emissions to start declining within the next 30 to 40 years.

Figure 43 suggests mitigation approaches that can contribute to reducing fossil-fuel carbon dioxide emissions. It consists of a reference emission scenario as a

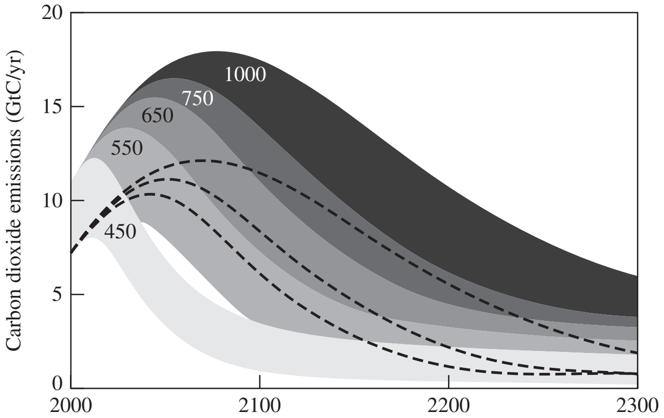


FIGURE 42 Pathways to climate stabilization. The numerical values refer to the eventual CO_2 (ppm) stabilization level achieved.
(*Source:* IPCC, 2001.)

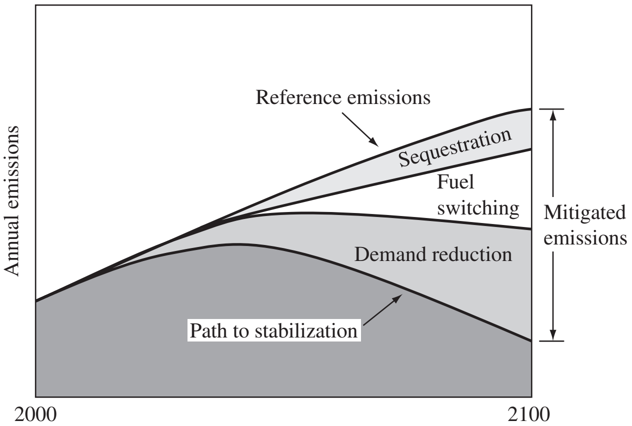


FIGURE 43 An example of a mitigation strategy for reducing fossil-fuel carbon dioxide emissions from a reference scenario to a stabilization path.

starting point, a target emission rate that would lead to the goal of climate stabilization, and an array of mitigation techniques that could be used to achieve that stabilization goal. Approaches to mitigation of carbon dioxide include *demand reduction* by increasing the efficiency with which we use energy and/or structural changes in the economy (e.g., a shift from manufacturing to a more service-oriented economy); *fuel switching,* which includes shifting from one fossil fuel to another (e.g., coal to natural gas) as well as increasing generation from nonfossil-fuel sources (e.g., nuclear, wind, solar, biomass, geothermal, and hydro); and *sequestration,* which includes carbon capture and storage (CCS) at power plants as well as increased sequestration in biomass.

An intuitively persuasive way to frame the problem of controlling future carbon emissions was introduced by Pacala and Socolow (2004). The concept is simple. Their reference scenario is in essence a 50-year projection of current carbon emission growth rates from 7 GtC/yr in 2005 to 14 GtC/yr in 2055, which makes it quite similar to the SRES B2 scenario. Their mitigation goal is to at least hold emissions constant at 7 Gt/yr over that 50-year period. To do so would require the elimination of 175 GtC of emissions over the next 50 years, which is the area of the emissions stabilization triangle shown in Figure 44.

As shown in Figure 45, they then proceed to break that emissions triangle into seven *stabilization wedges,* with each wedge having the ability to eliminate 1 GtC/yr in 2055. The Princeton Carbon Mitigation Initiative is in the process of identifying a significant number of strategies already available that could reduce emissions equivalent to one wedge of carbon emissions if deployed on a large scale. Their list includes the following:

1. Double the fuel efficiency of every automobile by 2050.
2. Drive half as many miles.
3. Utilize the best efficiency practices for all residential and commercial buildings.
4. Double the efficiency of coal-fired power plants.
5. Using natural gas instead of coal at 1,400 power plants.

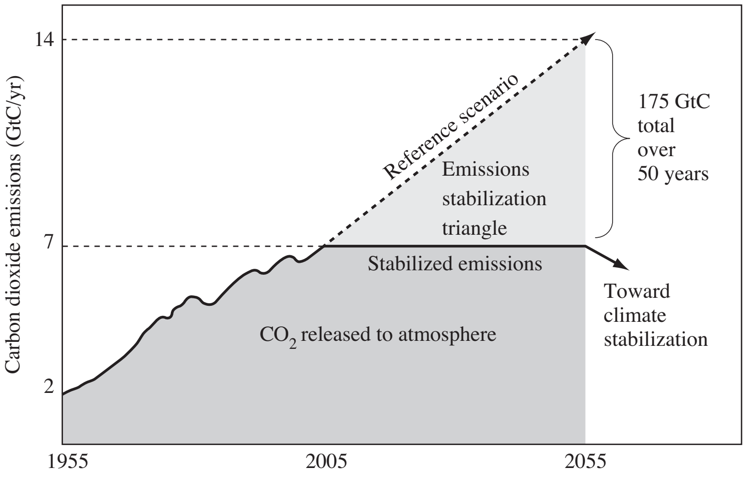


FIGURE 44 A 50-year emissions stabilization scenario. (*Source:* Pacala and Socolow, 2004.)

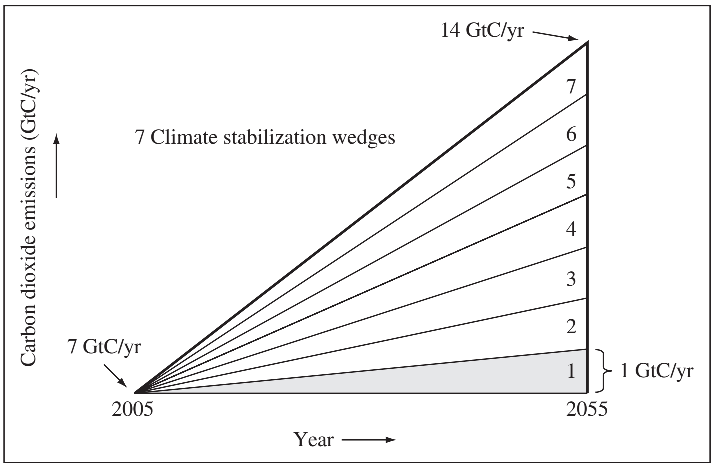


FIGURE 45 Each stabilization wedge eliminates 1 GtC/yr by 2055.
(*Source:* Pacala and Socolow.)

6. Capture and store carbon emissions from 800 coal-fired power plants.
7. Triple the current nuclear power generation capacity.
8. Increase existing wind turbine capacity by a factor of 50.
9. Use 4 million wind turbines to produce hydrogen fuel.
10. Install 700 times the existing solar power capacity.
11. Increase ethanol production by a factor of 50 using cellulosic sources.
12. Eliminate tropical deforestation and double the rate of new forest planting.
13. Increase soil conservation tillage in all agricultural soils.
14. Capture and store carbon from 180 coal-based synfuels plants.
15. Increase hydrogen fuel production from fossil-fuel sources by a factor of 10.

Every one of these options is already available on a commercial scale. Full deployment of seven of them offers the enticing potential to at least stabilize emissions over the next half century. Realize, of course, that stabilizing emissions does not stabilize the concentrations of greenhouse gases in the atmosphere, and to avoid dangerous anthropogenic interference with our climate system, we must do much better than that.

12 The Oceans and Climate Change

The oceans cover 70 percent of the Earth's surface and contain over 97 percent of the Earth's water. They have high heat capacity, so they warm slowly, and their currents distribute heat around the globe, greatly affecting local and regional temperatures on the land. They store 50 times as much carbon as the atmosphere, and they exchange carbon dioxide back and forth across the atmosphere-ocean interface at a rate that dwarfs the contribution by humankind. In other words, the oceans significantly affect and are affected by climate change. Greenhouse enhancement can affect the ocean in many ways, but three of the most important changes have to do with ocean acidification, rising sea levels, and changes in the ocean circulation. In addition, biological processes that result in the removal of carbon from the atmosphere can be affected by climate change, and in turn, those changes can affect climate.

Ocean Acidification

The oceans are performing a great service to the climate system by absorbing so much of the carbon dioxide we continuously pump into the atmosphere. That absorption comes at a price, however, since it is causing subtle chemical changes in the oceans that may have enormous implications for marine ecosystems.

Dissolved inorganic carbon exists in three forms: dissolved CO_2 gas, bicarbonate (HCO_3^-), and carbonate (CO_3^{2-}) ions. When CO_2 dissolves into sea water, some of it creates carbonic acid (H_2CO_3), which in turn ionizes to form hydrogen and bicarbonate:

$$CO_2 + H_2O \Leftrightarrow H_2CO_3 \Leftrightarrow H^+ + HCO_3^- \qquad (44)$$

The bicarbonate ions (HCO_3^-), in turn, ionize to some extent into hydrogen and carbonate ions:

$$HCO_3^- \Leftrightarrow H^+ + CO_3^{2-} \qquad (45)$$

The natural pH of the oceans is between 8.0 and 8.3, which means it is somewhat alkaline. The preceding reactions release hydrogen ions, which make the water more acidic (less alkaline) as CO_2 absorption increases. Studies have shown that the oceans have already experienced a drop in pH of about 0.1 since preindustrial times, and it seems likely ocean pH will fall by another 0.3 by the year 2100 (Caldeira and Wickett, 2003). If that happens, the oceans will be more acidic than they have been for hundreds of millions of years.

The preceding reactions proceed in both directions, and the resulting balance of aqueous CO_2, bicarbonate, and carbonate ions is a function of the pH (Figure 46). As Figure 46 shows, one consequence of a decrease in pH is that carbonate decreases as well. And carbonate is needed to build the calcium carbonate ($CaCO_3$)

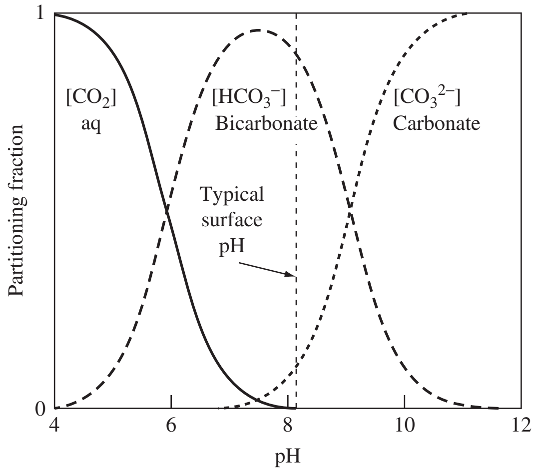


FIGURE 46 The relative proportions of dissolved CO_2, bicarbonate (HCO_3^-), and carbonate (CO_3^{2-}) depends on pH. When pH drops, less carbonate is available for calcifying organisms to grow and maintain calcium carbonate shells.

in shells and other hard parts of marine organisms. Some of the most important organisms likely to be affected by a lack of carbonate are certain types of plankton and small marine snails that are a major source of food for fish and marine mammals, including some species of whales. Coral reefs would also be affected since they are built from secretions of calcium carbonate by small sea anemone-like animals and coralline algae, which colonize and literally create the reefs.

Another impact of the drop in carbonate ions in seawater is the increased potential for shells to literally dissolve. The balance between the rate at which solid calcium carbonate dissociates and the rate at which it is formed is related to the concentration of carbonate ions:

$$CaCO_3 \Leftrightarrow Ca^{2+} + CO_3^{2-} \tag{46}$$

With less carbonate, the reaction shifts to the right, increasing the tendency for calcium carbonate to dissolve. The dissociation constant that goes with (46) is temperature and pressure dependent, and in cold, deep waters, the shift is sufficient to dissolve calcium carbonate shells. There is a depth, called the *saturation horizon*, below which shells dissolve and above which they survive. With more atmospheric CO_2, the shift in pH and carbonate has caused the saturation horizons to shift closer to the surface, in some areas by as much as 50 to 200 meters compared to where they were in the 1800s. Thinning of the shell-friendly, upper layer of the oceans, with unknown consequences, is a rising concern for climate and marine scientists trying to predict the impacts of global warming.

Rising Sea Level

Water expands as it warms, decreasing its density and increasing its volume, which causes sea level to rise. In addition, melting glaciers and ice caps contribute to increasing ocean volume and rising sea levels. Thermal expansion is the primary cause of the current 3 mm per year rise in sea level, but melting glaciers have the potential to accelerate that rise substantially in the future.

Almost all of the world's nonoceanic water is stored in ice caps and glaciers, and almost all of that is contained in just two great ice sheets—the Antarctic ice sheet and the Greenland ice sheet. Table 9 shows the volume of ice in these ice sheets plus the remaining ice stored in glaciers and ice caps along with the equivalent rise in sea level that complete melting could incur. As shown there, the Antarctic ice sheet stores enough water to raise the level of the oceans by 73 meters.

The Antarctic ice sheet is projected to grow slightly in the future due to increased precipitation coupled with extremely cold temperatures, so it may not

TABLE 9

Some Characteristics of the World's Ice

	Area (10^6 km^2)	Volume (10^6 km^3)	Sea Level Equivalent (m)
Antarctic ice sheet	12.1	29	73
Greenland ice sheet	1.71	2.95	7.4
Glaciers and ice caps	0.64	0.1 ± 0.02	0.3 ± 0.05

Source: IPCC, 1995.

contribute to rising sea levels. A portion of Antarctic ice, however, rests on the ocean floor, and there has been some concern that if that West Antarctic Ice Sheet (WAIC) were to break loose and slide into the ocean, it could cause a rapid rise in sea level of perhaps 5 or 6 meters. Indeed, a 200-m thick, 3,200 km^2 chunk of the Larsen Ice Shelf did collapse into the sea in 2002, amplifying concern that ocean warming and the acceleration of ice flows may be destabilizing the ice sheet that could cause a runaway discharge into the oceans. Some studies indicate a critical threshold for WAIC collapse might be in the range of a 2°C to 5°C global temperature rise, but too many uncertainties remain to be confident that this is an accurate estimate.

Inland portions of the Greenland ice sheet have shown a slight increase in volume, but significant melting is occurring along the coasts. Meltwater seeping down through the crevices of the melting ice seems to be acting as a lubricant that is accelerating glacial movement toward the ocean. There is great concern for a potential tipping point, beyond which the surface temperature of the ice sheet will become too warm to allow winter glacial growth to offset summertime melting. At that point, which is estimated by some to be a 2 to 3°C mean global temperature increase above preindustrial levels, the Greenland ice sheet will begin to melt irreversibly.

Atmospheric-ocean coupled general circulation models have been used to estimate future sea-level change for a great many SRES scenarios. For the illustrative A1B and B2 scenarios shown in Figure 47, a sea level rise on the order of one quarter of a meter by 2100 is projected. Multiple variations on all of the SRES scenarios show much wider variation, with most lying somewhere between 0.2 to 0.7 meters in 2100. The enormous thermal capacitance of the oceans means that sea level will continue to rise well beyond the time when carbon dioxide concentrations stabilize in the atmosphere as the oceans slowly adjust to the increased radiative forcing.

Rising sea levels coupled with the projected increase in the fraction of hurricanes and storms that are predicted to be highly dangerous may result in large areas of densely populated coastline having to be abandoned. With approximately 20 percent

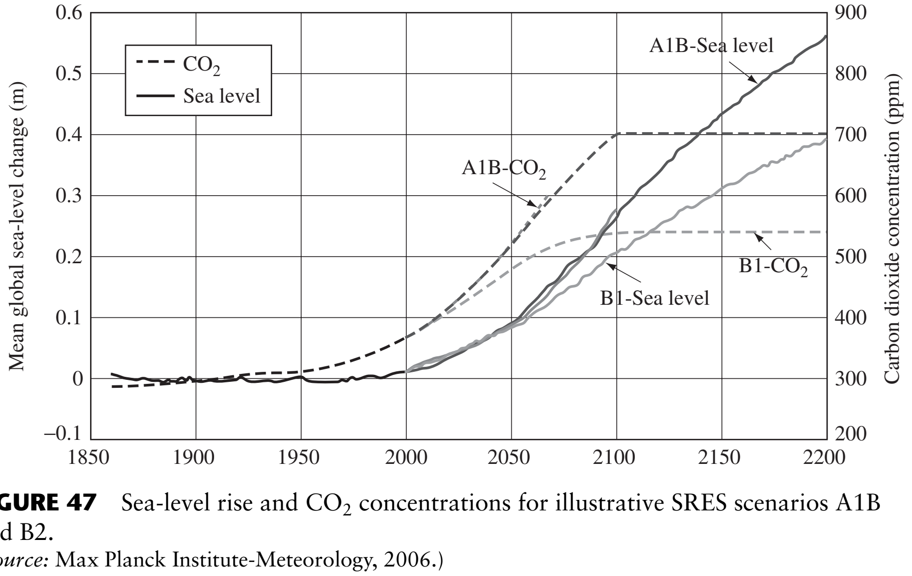


FIGURE 47 Sea-level rise and CO_2 concentrations for illustrative SRES scenarios A1B and B2.
(*Source:* Max Planck Institute-Meteorology, 2006.)

FIGURE 48 Flooding in New Orleans after Hurricane Katrina in 2005 caused $81 billion in damages and 1,800 deaths.

of the world's population living within 30 km of a coastline, tens of millions of people may become climate change refugees. Damage due to storm surges could be devastating to coastal areas that have not been adequately fortified, which places an added burden on countries that don't have the resources to adapt to rising water levels. Hurricane Katrina (Figure 48), which slammed into the Gulf Coast in 2005, provided a stark example of how costly, both in terms of human lives (over 1,800 deaths) and property damage (over $80 billion), storm events can be. Other potential impacts of rising sea levels include increased shoreline erosion, exacerbated coastal flooding, inundation of coastal wetlands, and increased salinity of estuaries and aquifers.

Thermohaline Circulation and the Biological Carbon Pump

Atmospheric carbon dioxide is absorbed by the oceans and forms inorganic dissolved bicarbonate and carbonate ions. A small fraction of that carbon is taken up by phytoplankton during photosynthesis and becomes part of the food chain in the upper layer of the oceans (the euphotic zone). Photosynthesis, respiration, and decomposition taking place in the surface layers of the ocean remove and replace some 40 to 50 GtC per year. There is a net removal, however, of about 10 GtC per year (nearly twice the carbon emission rate from the combustion of fossil fuels) from the surface waters as dissolved organic carbon, particulate carbon, and $CaCO_3$ in the "hard" parts of marine algae and animals sink into the intermediate and deep ocean. That removal process is known as the *biological carbon pump*.

Related to the biological pump is another process called the *thermohaline circulation* system, or sometimes the *Atlantic conveyor,* that transports carbon-rich surface waters into the deep oceans. As shown in Figure 49, there is an enormous

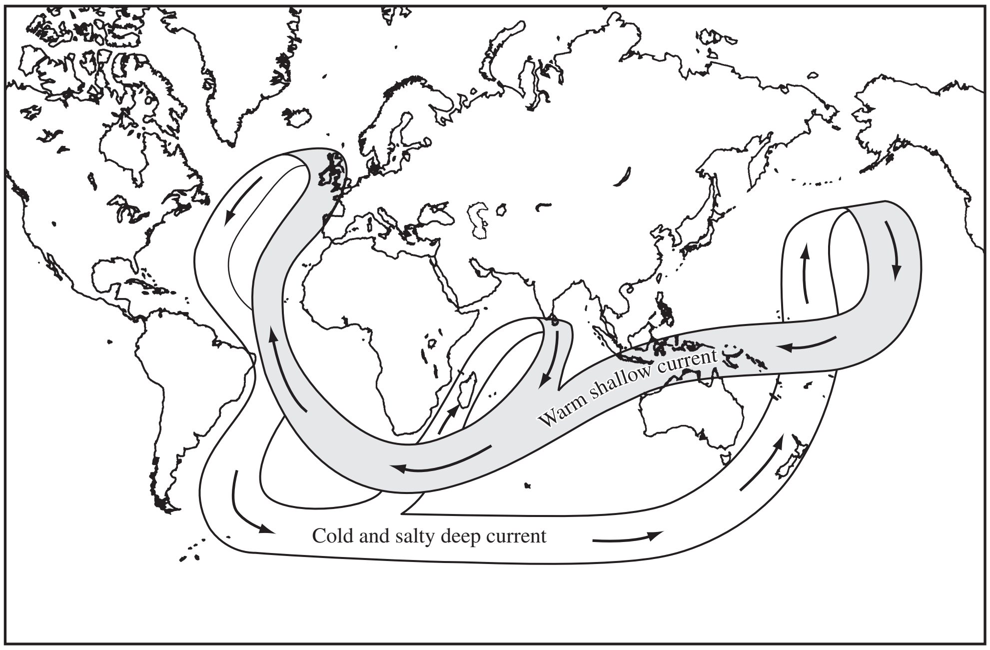


FIGURE 49 The thermohaline circulation system is driven by cold, salty water sinking into the North Atlantic abyss and upwelling near India and in the Pacific.

flow of ocean water around the globe, equal to the combined flow of all the world's rivers, that transports heat and nutrients from one place to another. Relatively warm seawater near the surface flows into the North Atlantic, where winds encourage evaporation and help sweep the surface waters aside, allowing the warm subsurface water to rise. That rising warm water gives off an enormous amount of heat as it reaches the surface. It is this source of heat that accounts for Western Europe's relatively mild winters. As that warm surface water evaporates and cools, it becomes more salty and dense, eventually becoming so dense that it sinks down to the ocean bottom. That North Atlantic deep water then moves southward, around the southern tip of Africa. Some emerges off the coast of India, but most upwells in the Pacific ocean where it starts its path back to the North Atlantic.

What makes the thermohaline circulation especially important in studies of climate change is the fact that it appears to have several stable states. It runs at a fast rate, an extremely slow rate, or at its present rate, which is somewhere in the middle. The transition between those stable states can occur in a relatively few years. Shifting from one stable state to another is thought to be responsible for a rather dramatic climate shift that occurred in northern Europe just after the last glacial period ended. Some 11,000 years ago, while glaciers were retreating and temperatures were back to interglacial levels, northern Europe and northeastern North America suddenly plunged back into glacial conditions. The temperature in Greenland dropped by 6°C in less than 100 years, stayed that way for 1,000 years, then jumped back to warm conditions in about 20 years (Broeker and Denton, 1990). The explanation for this event, known as the *Younger Dryas* (named for an Arctic flower that grew in Europe during this period), is based on the sudden stopping, then restarting

1,000 years later, of the Atlantic conveyor. Apparently, glacial meltwater draining from the North American ice sheet into the North Atlantic diluted the salty surface water enough to halt the density-driven dive into the abyss, which stopped the entire conveyor. Europe lost the warming influence of the current, and glacial conditions suddenly returned.

Most coupled ocean-atmosphere models show a future decrease in the strength of the Atlantic conveyor, which on its own would lead to some cooling of Europe and the east coast of the United States. That cooling would offset some but not all of the projected regional warming that would otherwise be expected due to an enhanced greenhouse effect. The possibility of a complete collapse of thermohaline circulation that could result if precipitation and runoff patterns change sufficiently is another of the potential "surprises" that worry some climatologists.

13 Changes in Stratospheric Ozone

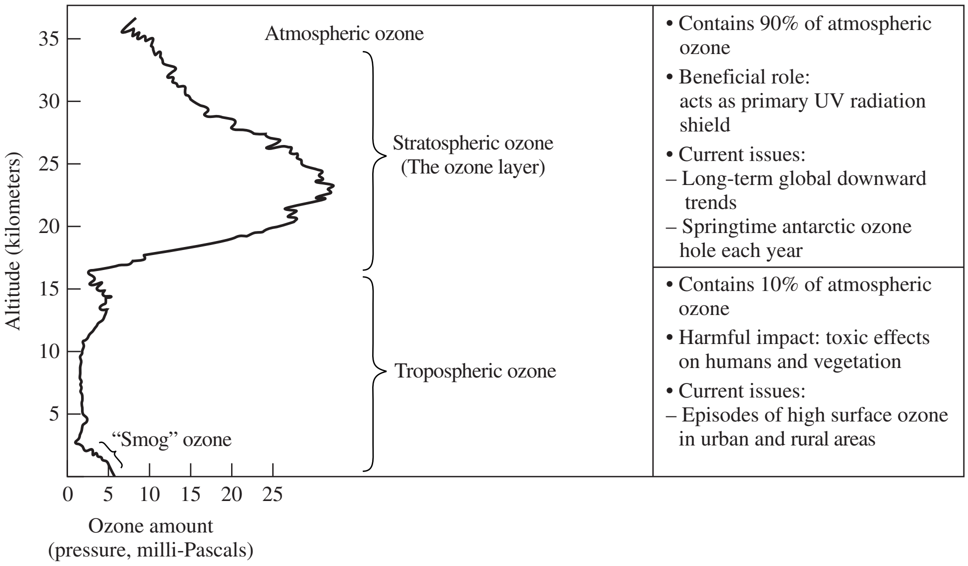
The changes occurring in the stratosphere's protective layer of ozone are closely linked to the greenhouse problem just discussed. Many of the same gases are involved, including CFCs, halons, methane, and nitrous oxide, as well as aerosols and ozone itself. Some of the gases that enhance the greenhouse effect, such as methane, actually reduce stratospheric ozone depletion, so the two problems really must be considered together.

Ozone has the unusual characteristic of being beneficial when it is in the stratosphere protecting us from exposure to ultraviolet radiation, while ozone formed near the ground in photochemical smog reactions is harmful to humans and other living things. About 90 percent of the atmosphere's ozone is contained in the stratosphere between roughly 10 and 50 km in what is commonly referred to as the *ozone layer*. The remaining 10 percent or so is contained in the troposphere, near the ground, above cities and industrialized areas. Figure 50 shows this distribution of ozone in the atmosphere along with some comments on its beneficial and harmful attributes.

If all of the atmospheric ozone overhead at any given spot on the Earth were to be brought down to ground level so that it would be subjected to 1 atm of pressure, it would form a layer only a few millimeters thick. In fact, one of the common methods of expressing the total amount of ozone overhead is in terms of *Dobson units* (DU), where 1 DU is equivalent to a layer of ozone 0.01 mm thick at 1 atm of pressure at 0°C. At midlatitudes, the ozone overhead is typically about 350 DU (3.5 mm at 1 atm); near the equator, it is closer to 250 DU. During the annual appearance of the ozone hole over Antarctica, the ozone column drops below 100 DU.

The Ultraviolet (UV) Portion of the Solar Spectrum

Ozone (O_3) is continuously being created in the stratosphere by photochemical reactions powered by short-wavelength ultraviolet (UV) radiation, while at the same time it is continuously being removed by other photochemical reactions that convert it back to diatomic oxygen molecules (O_2). The rates of creation and removal at any given time and location dictate the concentrations of ozone present.



- Contains 90% of atmospheric ozone
- Beneficial role: acts as primary UV radiation shield
- Current issues:
  - Long-term global downward trends
  - Springtime antarctic ozone hole each year

- Contains 10% of atmospheric ozone
- Harmful impact: toxic effects on humans and vegetation
- Current issues:
  - Episodes of high surface ozone in urban and rural areas

FIGURE 50 Amounts of ozone in the troposphere and stratosphere, including comments on the beneficial and harmful roles it plays in each region.
(*Source:* UNEP, 1994.)

The energy required to drive some of these photochemical reactions comes from sunlight penetrating the atmosphere. As you may recall, it is convenient to consider such electromagnetic radiation in some contexts as a wave phenomenon and in others as a packet of energy called a photon. The energy content of a single photon is the proportional to its frequency, and inversely proportional to its wavelength:

$$E = h\nu = \frac{hc}{\lambda} \tag{47}$$

where

E = the energy of one photon (J)
h = Planck's constant (6.626218×10^{-34} J s)
c = the speed of light (2.997925×10^{8} m/s)
ν = frequency (s^{-1})
λ = wavelength (m)

Thus, photons with shorter wavelengths have higher energy content.

The entire solar spectrum was introduced in Figure 10, where it was used to help derive the effective temperature of the Earth. In the context of stratospheric ozone depletion, however, it is only the UV portion of the spectrum that is of interest. Photons associated with those shorter wavelengths can have sufficient energy to break apart molecules, initiating *photochemical reactions*. Molecules that are dissociated by UV light are said to have undergone *photolysis*.

The energy required to drive photochemical reactions is often expressed as an amount of energy per mole of gas under standard conditions. If we assume the energy required under stratospheric conditions is not much different, we can easily derive the maximum wavelength that a photon must have to cause photolysis. A single photon (with sufficient energy) can cause photolysis in one molecule. And since there are Avogadro's number (6.02×10^{23}) of molecules in one mole of gas, we can convert energy needed per mole into the energy that one photon must have, as follows:

$$E(\text{J/photon}) = \frac{E(\text{J/mol})}{1 \text{ photon/molecule} \times 6.02 \times 10^{23} \text{ molecule/mol}} \tag{48}$$

Now we can use (47) to solve for the wavelength.

EXAMPLE 13 Photon Energy for Photolysis

What is the maximum wavelength that a photon can have to photodissociate diatomic oxygen if 495 kJ/mol are required?

$$O_2 + h\nu \rightarrow O + O$$

Solution The energy that a photon must have is given by (48):

$$E = \frac{495 \text{ kJ/mol} \times 10^3 \text{ J/kJ}}{1 \text{ photon/molecule} \times 6.02 \times 10^{23} \text{ molecule/mol}} = 8.22 \times 10^{-19} \text{ J/photon}$$

Rearranging (46) and substituting the energy that the photon must have gives us

$$\lambda_{\max} = \frac{hc}{E} = \frac{6.626 \times 10^{-34}(\text{Js}) \cdot 2.998 \times 10^{8}(\text{m/s})}{8.22 \times 10^{-19} \text{ J}} = 241.6 \times 10^{-9}\text{m}$$
$$= 241.6 \text{ nm}$$

Thus, photons with wavelengths longer than 241.6 nm will not have sufficient energy to break apart the oxygen molecule.

The wavelengths that are most important in the context of stratospheric ozone depletion are in the 100 to 400 nm portion of the ultraviolet (UV) spectrum. It is called ultraviolet since wavelengths just above 400 nm are violet in color and visible to the human eye. For analogous reasons, wavelengths above 700 are called infrared, since 700 nm is the edge of the red portion of the visible spectrum. The most important portion of the UV spectrum is divided into three regions, designated as UV-A, UV-B, and UV-C, which have wavelengths as shown in Figure 51.

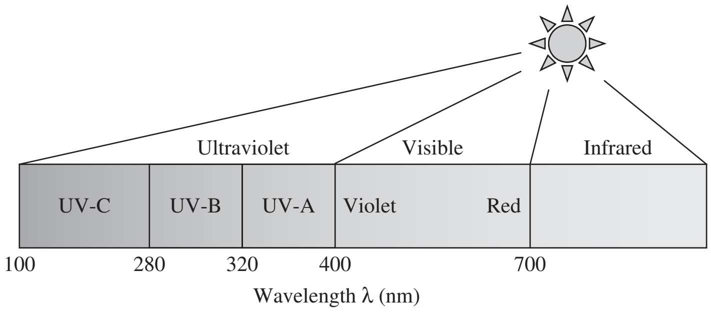


FIGURE 51 The ultraviolet (UV), visible, and infrared (IR) portions of the solar spectrum, including the biologically important UV-A, UV-B, and UV-C regions.

About 99 percent of the incoming solar UV that makes it through the atmosphere is UV-A (320 to 400 nm). It is sometimes referred to as "black light" since it is invisible to the human eye, though some animals can use it for vision. UV-A penetrates deeply into human skin, causing tanning (but little skin reddening), DNA damage, photoaging (toughening of the skin, wrinkles), and possibly skin cancer. Some exposure to UV-A benefits human health by helping to induce the formation of vitamin D. UV-B radiation (280–320 nm) poses the greatest danger to human health. It causes reddening of the skin (erythema) and reduction of vitamin-D synthesis in the short term; in the long term, it is a cause of skin cancer, cataracts, and suppression of the immune system. The "SPF rating" of sunscreen products refers to their ability to protect us from UV-B. Some sunscreens now include ingredients that also help block UV-A. Photons within the UV-C portion of the spectrum (100 to 280 nm) readily destroy DNA, which makes them extremely dangerous but also lets artificially created UV-C be used as a very effective disinfectant for water treatment and other germicidal applications. Fortunately, the atmosphere effectively shields us from these very potent photons.

The Ozone Layer as a Protective Shield

Ozone formation in the stratosphere can be described by the following pair of reactions. In the first, atomic oxygen (O) is formed by the photolytic decomposition of diatomic oxygen (O_2).

$$O_2 + h\nu \rightarrow O + O \tag{49}$$

where $h\nu$ represents a photon. As shown in Example 13, for this dissociation to take place, the photon must have a wavelength of no more than 242 nanometers. Photons with less energy (longer wavelengths) cannot cause the reaction to take place; all they can do is heat up the molecule that absorbs them. Photons with higher energy than the minimum required use some of their energy to cause photolysis, and what is left over is dissipated as heat. Diatomic oxygen (O_2) has its maximum absorption at about 140 nm, and it is a very effective absorber of UV-C radiation between 130 and 180 nm.

The atomic oxygen formed by (49) reacts rapidly with diatomic oxygen to form ozone,

$$O + O_2 + M \rightarrow O_3 + M \tag{50}$$

where M represents a third body (usually a nearby N_2 molecule) needed to carry away the heat generated in the reaction.

Opposing the preceding ozone formation process is ozone removal by photodissociation:

$$O_3 + h\nu \rightarrow O_2 + O \tag{51}$$

The absorptance of ozone extends from about 200 to 320 nm and reaches its peak at 255 nm. The reaction shown in (51) is very effective in removing UV-C radiation and some of the UV-B before it reaches the Earth's surface.

The preceding combination of reactions (49 to 51) forms a chain in which oxygen atoms are constantly being shuttled back and forth between the various molecular forms. A principal effect is the absorption of most of the short-wavelength, potentially damaging, UV radiation as it tries to pass through the stratosphere.

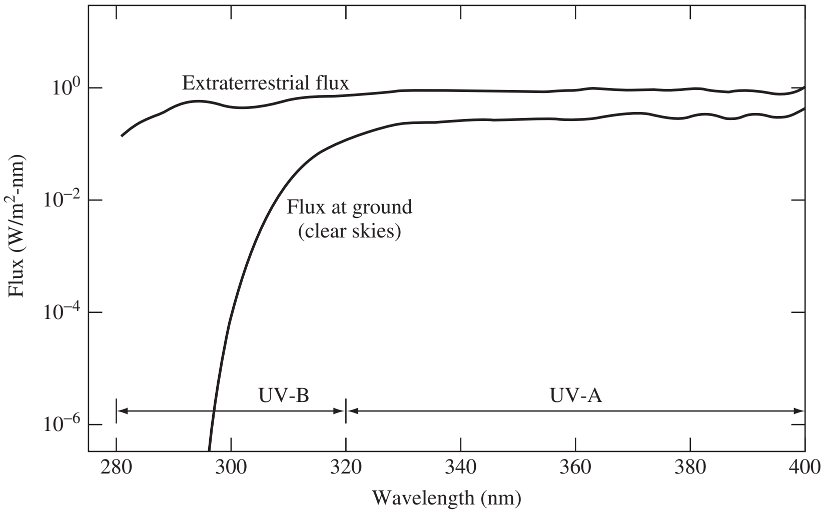


FIGURE 52 Extraterrestrial solar UV flux and expected flux at the Earth's surface on a clear day with the sun 60° from the zenith.
(*Source:* Frederick, 1986.)

In addition, that absorption also heats the stratosphere, causing the temperature inversion shown earlier in Figure 1. That temperature inversion, which is what defines the stratosphere, produces stable atmospheric conditions that lead to long residence times for stratospheric pollutants.

The effectiveness of these reactions in removing short-wavelength UV is demonstrated in Figure 52, in which the extraterrestrial solar flux and the flux actually reaching the Earth's surface are shown. The radiation reaching the surface has been drawn for a clear day with the sun assumed to be 60° from the zenith (overhead), corresponding roughly to a typical mid-latitude site in the afternoon. As can be seen, the radiation reaching the Earth's surface is rapidly reduced for wavelengths less than about 320 nm. In fact, an intact ozone layer shields us from almost all of the UV having wavelengths shorter than about 290 nm.

Catalytic Destruction of Stratospheric Ozone

As is now common knowledge, the natural formation and destruction of ozone in the stratosphere is being affected by gases that we release into the atmosphere. In particular, ozone destruction is enhanced by catalytic reactions that can be represented as follows:

$$\begin{array}{ll} & X + O_3 \rightarrow XO + O_2 \\ \text{and} & \underline{XO + O \rightarrow X + O_2} \\ \text{for a net} & O + O_3 \rightarrow 2\,O_2 \end{array} \qquad (52)$$

where X is a free radical such as Cl, Br, H, OH, or NO (recall that free radicals are highly reactive atoms or molecules that have an odd number of electrons, which means one electron is not paired with another atom). Notice in (52) how the X radical that enters the first reaction is released in the second, freeing it to go on and participate in another catalytic cycle. The net result of the preceding pair of reactions is the destruction of one ozone molecule. The original catalyst that started the reactions, however, may go on to destroy thousands more ozone molecules before it eventually leaves the stratosphere.

Chlorofluorocarbons (CFCs) are a major source of ozone-destroying chlorine. Recall that CFCs have long atmospheric lifetimes since they are inert and nonwater-soluble, so they aren't removed in the troposphere by chemical reactions or rainfall. Eventually, CFCs drift into the stratosphere, where they are exposed to UV radiation, which breaks apart the molecules, freeing the chlorine. CFCs typically require photons in the UV-C range with wavelengths shorter than 220 nm for photolysis. For example, CFC-12, which has an atmospheric lifetime of 100 years (Table 7), undergoes photolysis as follows:

$$CF_2Cl_2 + \text{UV-C} \rightarrow CF_2Cl + Cl \tag{53}$$

Eventually the second chlorine atom is also released. The freed chlorine then acts as a catalyst as described by (52):

$$\begin{aligned} & Cl + O_3 \rightarrow ClO + O_2 \\ & \underline{ClO + O \rightarrow Cl + O_2} \\ \text{Net:}\quad & O + O_3 \rightarrow 2\,O_2 \end{aligned} \tag{54}$$

The chlorine atom makes this loop tens of thousands of times, but eventually it reacts with methane to become water-soluble HCl that can diffuse into the troposphere, where it can be washed out of the atmosphere by rainfall.

$$Cl + CH_4 \rightarrow HCl + CH_3 \tag{55}$$

Chlorine that has been incorporated into HCl, as in (55), is *inactive;* that is, it does not participate in the catalytic destruction of ozone represented by (54). Notice that methane, which is a potent greenhouse gas, helps remove the ozone-destroying, active form of chlorine from the stratosphere. In the context of climate change, methane is part of the problem; in the context of ozone depletion, it is part of the cure.

Another way that chlorine is removed from the catalytic chain reaction is by its incorporation into chlorine nitrate, $ClONO_2$:

$$ClO + NO_2 \rightarrow ClONO_2 \tag{56}$$

At any given time, as much as 99 percent of the chlorine in the stratosphere is tied up in these inactive molecules, HCl and $ClONO_2$. Unfortunately, those inactive molecules are subject to chemical reactions and photolysis that can restore the chlorine to its active forms, Cl and ClO, especially when there are nearby particles available that provide platforms upon which the reactions can take place. Sulfate aerosols, especially from volcanic eruptions, and clouds of nitric acid ice over the Antarctic provide such surfaces. When gases and particles are both involved, the chemistry is referred to as *heterogeneous* reactions, as opposed to *homogeneous* reactions involving gases alone.

Bromine in the stratosphere is another potent ozone-depleting gas, as the following reactions suggest:

$$\begin{aligned} Br + O_3 &\rightarrow BrO + O_2 \\ BrO + ClO &\rightarrow Br + ClO_2 \end{aligned} \tag{57}$$

As is the case with chlorine, bromine acts as a catalyst destroying ozone until it eventually combines with methane to form hydrogen bromide (HBr). Hydrogen bromide deactivates the bromine, but not for long. Photolysis rather quickly decomposes HBr, sending bromine back into its catalytic loop with ozone. In this regard, bromine is much more potent than chlorine since stratospheric chlorine exists mostly in its inactive forms, HCl and $ClONO_2$.

The main source of bromine in the stratosphere is methyl bromide (CH_3Br), about half of which comes from the oceans, and about half is due to anthropogenic sources. Methyl bromide is used extensively in agriculture to sterilize soil and fumigate some crops after they are harvested. Other anthropogenic sources include biomass burning and automobile exhaust emissions when leaded fuels are used. Bromine has also been used as a flame retardent in fire extinguishers.

The Antarctic Ozone Hole

Chlorine that has been incorporated into hydrogen chloride (HCl) or chlorine nitrate ($ClONO_2$) is inactive, so it doesn't contribute to the destruction of ozone. In the Antarctic winter, however, a unique atmospheric condition known as the *polar vortex* traps air above the pole and creates conditions that eventually allow the chlorine to be activated. The polar vortex consists of a whirling mass of extremely cold air that forms over the South Pole during the period of total darkness in the Antarctic winter. The vortex effectively isolates the air above the pole from the rest of the atmosphere until the Antarctic spring arrives in September. Stratospheric temperatures in the vortex may drop to below −90°C, which is cold enough to form polar stratospheric clouds (PSCs) even though the air is very dry. The ice crystals that make up polar clouds play a key role in the Antarctic phenomenon by providing reaction surfaces that allow chemical species to stay together long enough to react with each other.

A number of reactions take place on the surfaces of polar stratospheric cloud particles that result in the formation of chlorine gas (Cl_2). For example,

$$ClONO_2 + H_2O \rightarrow HOCl + HNO_3 \tag{58}$$

$$HOCl + HCl \rightarrow Cl_2 + H_2O \tag{59}$$

$$ClONO_2 + HCl \rightarrow Cl_2 + HNO_3 \tag{60}$$

Once the sun rises in the Antarctic spring in September, the chlorine gas (Cl_2) formed during the darkness of winter photolytically decomposes into atomic chlorine:

$$Cl_2 + h\nu \rightarrow 2\ Cl \tag{61}$$

which then destroys ozone by reinitiating the catalytic destruction process described in (54):

$$(Cl + O_3 \rightarrow ClO + O_2) + (ClO + O \rightarrow Cl + O_2) = O + O_3 \rightarrow 2\ O_2 \tag{62}$$

The destruction of ozone as the sun first appears in the Antarctic spring proceeds as described until the nitric acid (HNO_3) formed in (58) and (60) photolyzes and forms inactive chlorine nitrate ($ClONO_2$), which stops the ozone destruction process. As the vortex breaks down in the spring, ozone from nearby areas rushes in

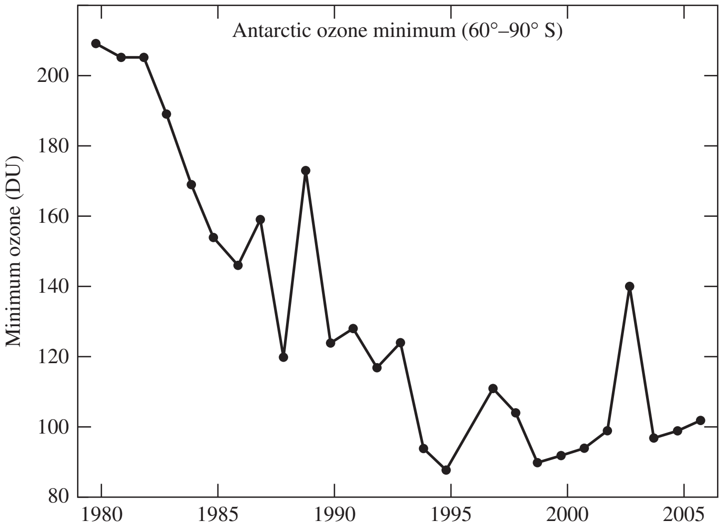


FIGURE 53 Antarctic ozone minimun (60°–90° S), suggesting the beginning of recovery in about the year 2000.
(*Source:* NASA Website.)

and replenishes the ozone above Antarctica. Conversely, the transport of ozone-depleted air from polar regions is thought to be contributing to decreases in stratospheric ozone at middle latitudes. The entire ozone-hole cycle typically lasts from mid-August until the end of November, reaching its greatest depth and aerial extent near the first of October.

Concern over possible destruction of stratospheric ozone was first expressed by two scientists, F. Sherwood Rowland and Mario J. Molina in 1974; however, it wasn't until 1985, with the dramatic announcement of the discovery of a "hole" in the ozone over Antarctica, that the world began to recognize the seriousness of the problem. In 1995, Rowland, Molina, and Paul Crutzen were awarded the Nobel Prize in Chemistry for their pioneering work in explaining the chemical processes that lead to the destruction of stratospheric ozone.

Figure 53 shows how deep the ozone hole has been over the past couple of decades. The values shown (in Dobson units) correspond to the minimum values measured anywhere within the hole each year. When measurements first began, in the early 1980s, those minimums were a little above 200 DU. They reached their lowest level in 1995 (88 DU) and now seem to be somewhat on the rise as efforts to control emissions of ozone-depleting substances begin to have effect. The area of the ozone hole has shown a similar pattern to the depth (Figure 54). The hole grows in area to about 25 million km^2 in late September, making it roughly equivalent in size to the entire North American continent.

The very low stratospheric temperatures over Antarctica, which form those polar stratospheric clouds, are not duplicated in the Arctic. The combination of land and ocean areas in the Arctic results in warmer temperatures and much less of a polar vortex, factors that do not encourage the formation of polar clouds. The result is a much less dramatic thinning of the ozone layer over the Arctic in the spring, although it is still significant. There is concern, however, that greenhouse effect

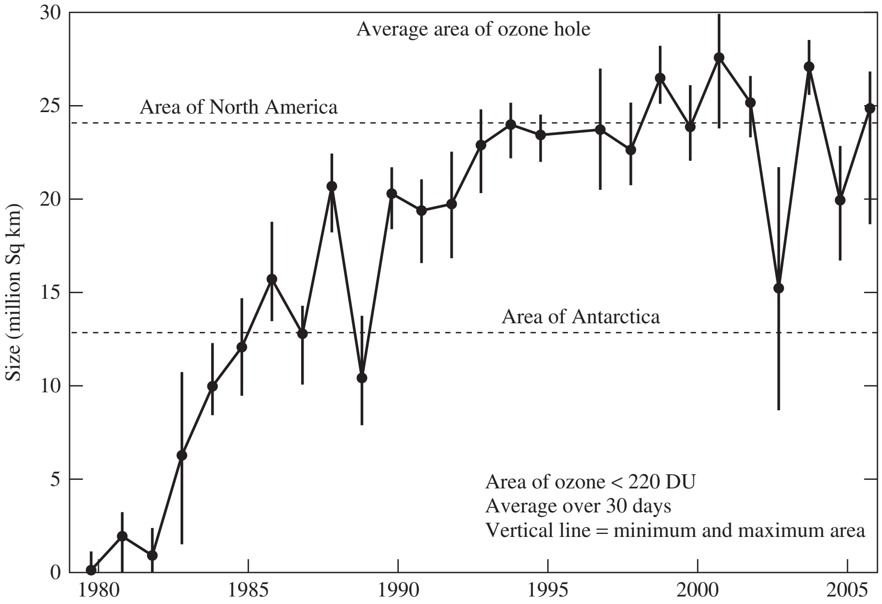


FIGURE 54 The average area of the ozone hole, defined as the region within which ozone is less than 220 DU, is about the size of North America.
(*Source:* NASA.)

enhancement could lead to stratospheric cooling, which would make an Arctic ozone hole more likely. Should an extensive ozone hole develop in the Northern Hemisphere, far more people would be exposed to elevated UV radiation.

Ozone Depletion Potential (ODPs)

The concept of Ozone Depletion Potentials (ODPs) for ozone-depleting substances (ODSs) is analogous to the Global Warming Potentials (GWPs) for greenhouse gases already described. In the case of ODPs, the reference gas is CFC-11 ($CFCl_3$). The ODP of a gas is defined as the change in total ozone per unit mass emission of the gas, relative to the change in total ozone per unit mass emission of CFC-11. Notice the definition of ODPs is independent of the length of time required for the depletion to take place, so, in that sense, they are somewhat different from time-dependent GWPs.

ODPs have become an integral part of the regulatory approach to controlling emissions of ozone-depleting substances. The Montreal Protocol on Substances that Deplete the Ozone Layer and its subsequent Amendments, which governs international production and use of halocarbons, and the U.S. Clean Air Act Amendments of 1990, which controls domestic uses, both dictate the phase-out of ozone-depleting substances based on their Ozone Depletion Potentials.

Table 10 lists ODPs for a number of CFCs, halons, and possible replacements. By definition, CFC-11 has an ODP of 1.0, which is the highest ODP of any of the CFCs. Halons-1211 and 1301 have much higher ODPs than CFC-11 due to the higher reactivity of bromine and the reduced effectiveness of deactivation traps

TABLE 10

Ozone Depletion Potentials (ODPs) and 100-Year Global Warming Potentials (GWPs) Measured Relative to CFC-11

Species	Chemical Formula	Ozone Depletion Potential Relative to CFC-11	Global Warming Potential Relative to CFC-11
CFC-11	$CFCl_3$	1.0	1.0
CFC-12	CF_2Cl_2	1.0	2.3
CFC-113	$C_2F_3Cl_3$	1.0	1.3
HCFC-22	CF_2HCl	0.05	0.4
HCFC-123	$C_2F_3HCl_2$	0.02	0.02
HCFC-124	C_2F_4HCl	0.02	0.12
HCFC-141b	$C_2FH_3Cl_2$	0.12	0.16
HFC-134a	$C_2H_2F_4$	0.0005	0.28
Halon-1211	CF_2ClBr	6	0.28
Halon-1301	CF_3Br	12	1.5
Methyl chloroform	CH_3CCl_3	0.1	0.03
Methyl bromide	CH_3Br	0.4	0.001

Source: WMO, *The Scientific Assessment of Ozone Depletion,* 2002.

compared with chlorine. Notice the replacement HCFCs have low ODPs, typically less than 0.1, which is a reflection of their short atmospheric lifetimes.

For comparison, 100-year GWPs relative to CFC-11 rather than the usual CO_2 reference molecule have been added to Table 10. These GWPs tell us that the HCFC and HFC replacements for CFCs are not only much better in terms of ozone depletion, but they also cause considerably less global warming. The halons, on the other hand, tend to be worse than CFCs in both categories.

Impacts of Increased Exposure to UV

The biological response to increased ultraviolet radiation is often represented by an *action spectrum*. The action spectrum provides a quantitative way to express the relative ability of various wavelengths to cause biological harm. For example, the action spectra for plant damage, DNA damage, and erythema (skin reddening) are shown in Figure *55a*. All of these spectra show increasing damage as wavelength decreases, and they all show significant damage in the UV-B (280 to 320 nm) region. There are some differences, however. Plants, for example, show no damage from wavelengths above about 320 nm, but erythema damage occurs well into the UV-A (320 to 400 nm) portion of the solar spectrum. Sunscreens that block UV-B, but not UV-A, can give a false sense of security by allowing people to spend more time exposed to the sun without burning, but skin damage that may lead to skin cancer is occurring nonetheless.

By combining wavelength-dependent action spectra, such as are shown in Figure *55a*, with ground-level solar irradiances, a biological weighted effectiveness of radiation is obtained. Figure *55b* shows a representative action spectrum superimposed on two solar irradiance curves—one showing irradiance with an intact ozone layer (dotted line), and the other with an ozone layer that has been thinned somewhat (solid line). As the ozone column is depleted, the irradiance in those

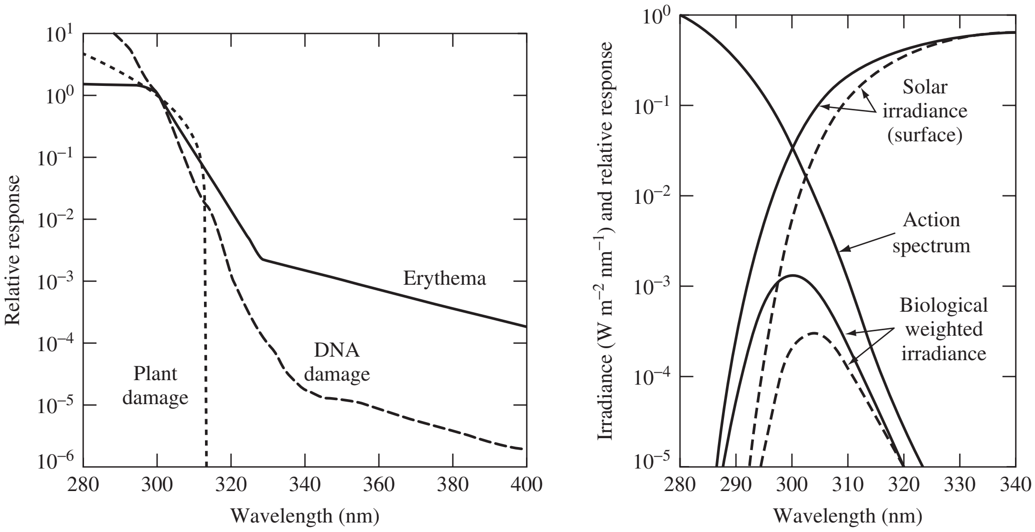


FIGURE 55 (*a*) Action spectra for damage to plants, DNA, and human skin; (*b*) by combining the action spectra with solar irradiance a biological weighted irradiance is obtained. The dotted lines represent irradiance under normal ozone conditions, while the solid lines correspond to a reduction in ozone.
(*Source:* Simon, 1993; Madronich, 1993.)

critical short UV-B wavelengths, where organisms are most sensitive, rises rapidly, as does the potential biological damage. The biologically weighted irradiance shows the critical wavelengths in terms of biological damage between about 290 and 310 nm (UV-B). Shorter wavelengths are effectively blocked by the remaining ozone layer, while longer wavelengths do little biological damage.

Most of our concern for increasing exposure to UV radiation has focused on the increased probability of developing skin cancer. Numerous studies have shown that skin cancer is readily induced in laboratory animals exposed to broad spectrum UV radiation, and the International Agency for Research on Cancer (1992) states unequivocally that "there is sufficient evidence in humans for the carcinogenicity of solar radiation. Solar radiation causes cutaneous malignant melanoma and non-melanocytic skin cancer." In fact, exposure to solar radiation is probably the main cause of all forms of skin cancer and an important, if not the main, cause of cancer of the lip (Armstrong, 1993).

Skin cancer is usually designated as being either the more common non-melanoma (basal cell or squamous cell carcinomas), or the much more life-threatening malignant melanoma. Epidemiological studies directly link skin cancer rates with exposure to UV-B. At lower latitudes, there is more surface-level UV-B, and skin cancer rates increase proportionately. Using such data, it has been estimated that a 1 percent increase in UV-B is estimated to cause a 0.5 percent increase in the incidence rate of melanoma and a 2.5 percent increase in nonmelanoma (Armstrong, 1993). Both types of skin cancer are more prevalent among people with fair complexion than among individuals who have more of the protective pigment, melanin.

Nonmelanoma is most likely to occur on areas of the skin habitually receiving greatest exposure to the sun, such as the head and neck. While it is rarely fatal, it often causes disfigurement. Melanomas, on the other hand, are as common on intermittently exposed parts of the body, such as the trunk and legs, as they are on the head and neck. Melanomas seem to occur more frequently on individuals who have had severe and repeated sunburns in childhood. This observation has led to a hypothesis that melanoma is less related to the accumulated dose of radiation, as seems to be the case for nonmelanoma, than it is to the pattern of exposure.

The incidence rate of melanoma among fair skinned individuals in the United States is increasing at 2 to 3 percent annually, which is faster than any other form of cancer. In 1935, the lifetime risk of an American developing melanoma was 1 in 1,500 individuals, while in 2002 the risk was 1 in 68. Approximately 54,000 melanomas are diagnosed each year, and 7,600 die of the disease. While some are tempted to relate that increase to the thinning of the ozone layer, it is much more likely to be the result of lifestyle changes rather than increased surface-level UV irradiation. More people are simply spending more time outdoors in the sun than in the past. Moreover, the decrease in stratospheric ozone is a relatively recent phenomenon, and evidence suggests that skin cancer may have a long latency period. There very well may be a 20-year lag time between changes in solar irradiance and measurable increases in skin cancer caused by that exposure.

To help alert the public to the dangers of UV exposure, various versions of a simple index have been created. A UV-index is a nondimensional, single number, in the range of 0 to 12, that represents an erythemal-weighted irradiance expected at solar noon (when the sun is at its highest) for the following day. The UV-index is now a routine part of Web and newspaper weather forecasts. To make the index more appealing and understandable for the public, it is linked to descriptive words ranging from "Low" to "Extreme," and it also provides an estimate of the length of time required for a fair-skinned person to get sunburned. The World Health Organization index is shown in Table 11.

Other human health problems associated with UV exposure include ocular damage (cataracts and retinal degeneration) and immune system suppression. The effects of increasing UV-B radiation would not be felt by humans alone, of course. There is considerable evidence that UV-B reduces photosynthesis and other plant processes in terrestrial and aquatic plants. UV-B radiation is known to penetrate to ecologically significant depths in the oceans, and besides impairing photosynthesis, it threatens marine organisms during their critical developmental stages. For example, amphibian eggs have little protection when exposed to UV-B irradiation, and the worldwide decline in frogs and toads, especially at high altitudes where solar exposure is greatest, may be in part caused by our thinning ozone layer.

TABLE 11

WHO Exposure Levels and Sunburn Warnings

UV Index	Category Descriptor	Sunburn Time
1–2	Low	More than 1 hour
3–5	Moderate	45 minutes
6–7	High	30 minutes
8–10	Very High	15 minutes
11+	Extreme	Less than 10 minutes

Political Response to the Ozone Depletion Problem

The political response to warnings of ozone depletion by the scientific community has been unusually effective and timely. When CFCs were first created in the 1930s they were hailed as wonder chemicals. They replaced toxic, explosive, and unstable refrigerants and made possible the refrigeration and air conditioning luxuries that we now take for granted.

A British scientist, James E. Lovelock (later famous for the "Gaia"), began experimenting with CFCs in the 1970s and concluded in 1973 that essentially all of the CFCs ever produced were still in the atmosphere. The first hint of potential problems came with the publication of Molina and Rowland's paper in 1974, which warned of the danger to the stratospheric ozone layer that CFCs presented. The EPA responded in a remarkably short period of time and banned nonessential spray-can uses of CFCs by 1979. A few countries followed the U.S. lead, including Canada and some of the Scandinavian countries, but most continued to manufacture and use them without concern for the consequences. It wasn't until the discovery of the ozone hole over Antarctica in 1985 that the rest of the world awakened to the lurking danger. At that time, worldwide CFC use as an aerosol propellant was the largest single source of CFC emissions.

In 1987 an international meeting convened in Montreal that led to the signing of the *Montreal Protocol on Substances that Deplete the Ozone Layer*. That Protocol called for a 50 percent reduction in use of CFCs by 1999. It was soon realized, however, that the called-for reductions were inadequate, and in 1990, 93 nations convened in London and created a new timetable calling for a complete phase-out of CFCs, most halons, and carbon tetrachloride by the year 2000. The plight of the developing countries was recognized and their use of CFCs was extended to 2010. A fund was established to help pay for technology transfer to enable the less-developed countries to take advantage of proposed replacement chemicals.

The next year, 1991, Mt. Pinatubo erupted, and the ozone hole grew to record proportions, which led to the London Amendments to the Protocol in 1992. The phase-out of CFCs was moved up to 1996.

Meanwhile, in the United States, Title VI of the Clean Air Act Amendments of 1990 brought us into compliance with the London Amendment. Major provisions include the following:

- Two categories of ozone-depleting substances (ODS) were created: Class I chemicals are those with high ozone depletion potentials, which includes CFCs, halons, and carbon tetrachloride, and Class II substances have lower ODPs and include the HCFCs. Production of Class I chemicals was scheduled to cease by 2000, while Class II chemicals are subject to an accelerated phase-out schedule that will eliminate all production and use by 2030. Exceptions for necessary medical, aviation, safety and national security purposes, and for export to developing countries are allowed.
- Motor vehicle air conditioners received special attention since they had been the largest single source of CFCs. CFC recycling equipment for air conditioner servicing is required, and only certified repair persons are allowed to do the work. CFCs must be removed from car air conditioners before they are crushed for disposal.

- Nonessential products that use ozone-depleting substances such as noise horns, "silly string," and some commercial photographic equipment, are banned.
- Warning labels are required for products that contain ODS such as refrigerators and foam insulation.
- Modifications to the Montreal Protocol that accelerate the phasing out of ODS must be adhered to.

Two years later, in 1992, the Copenhagen Amendments to the Montreal Protocol called for participating nations to move the ban of many ODSs to 1996. The EPA responded in 1993 with an accelerated phase-out schedule. Production and imports of halons ceased in 1994, and CFCs, methyl chloroform (1,1,1-trichloroethane), and carbon tetrachloride ceased in 1996. Most HCFCs were eliminated by 2003, but some will still be allowed until 2030.

Has this flurry of regulatory activity been effective? Stratospheric ozone depletion is one of those problems that cannot be reversed in a short period of time. Most of the CFCs ever produced are still drifting around in the atmosphere and they will be with us for decades into the future. The production cuts dictated by the Montreal Protocol and subsequent Amendments have already stopped the steady, historical climb in CFC atmospheric concentration. Figure 56 sums up the good news. CFC production has stopped, though HCFCs will still be produced until 2030; the major ozone-destroying chlorine and bromine concentrations in the stratosphere have begun to decline; and the overall global stratospheric ozone concentration is beginning to recover. Not only have countless lives been saved, but the Montreal Protocol has provided a model for effective international cooperation on a major global environmental problem.

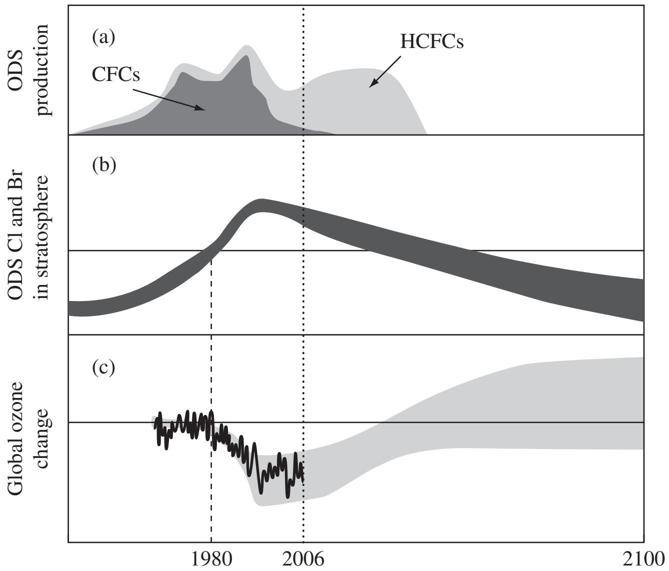


FIGURE 56 A global environmental success story. (a) Production of CFCs has ended, and HCFCs are scheduled to be phased out; (b) stratospheric chlorine and bromine are beginning to decline; (c) global ozone shows signs of recovery.
(*Source:* WHO/UNEP, *Scientific Assessment of Ozone Depletion*, 2006.)

PROBLEMS

1 Suppose an ocean sediment sample shows ($^{18}O/^{16}O$) = 0.0020150. Using the VSMOW for seawater, what is the value of $\delta^{18}O$? Is this a sign of more or less glaciation at the time the sediment was deposited?

2 The following data are from a Vostok ice core. Plot the data, and from the slope, estimate the temperature change in δD per °C.

Depth (m)	Years Before Present	Delta D ($^0/_{00}$)	Temperature (°C)
280	12087	−448.5	−3.09
327	14904	−458.1	−4.68
346	16201	−468.5	−6.34
355	16889	−480.7	−8.33
365	17706	−478.0	−7.84

3 Suppose an ice core measurement taken from the last glacial period resulted in the isotope ratio ($^2H/^1H$) = 8.100×10^{-5}.

(a) Find the corresponding $\delta D(^0/_{00})$ using the 0.00015575 VSMOW standard for deuterium.

(b) If an interglacial sample had $\delta D(^0/_{00}) = -435$ per mil, how much warmer was the interglacial period assuming a 6 per mil change in $\delta D(^0/_{00})$ corresponds to a 1°C temperature change?

4 A relationship between the mean annual surface temperature of Greenland and the value of $\delta^{18}O$ of the snow pack on the Greenland ice sheet is given by

$$T(°C) = 1.5\delta^{18}O(^0/_{00}) + 20.4$$

An ice core sample dating back to the last glacation has a value of $\delta^{18}O$ equal to −35. What would the estimated surface temperature have been at that time?

5 Use the following ice core data to derive a relationship between T(°C) and δD similar to the one shown in Problem 4 for ^{18}O.

Depth (m)	Years Before Present	Delta D ($^0/_{00}$)	Temperature (°C)
0	0	−438.0	0
12	234	−449.1	−1.84
20	420	−435.9	0.35
47	1247	−446.9	−1.48
68	2049	−440.3	−0.38
143	5397	−435.0	0.5

6 Suppose the Earth is really flat (Figure P6). Imagine an Earth that is shaped like a penny, with one side that faces the sun at all times. Also suppose that this flat Earth is the same temperature everywhere (including the side that faces away from the sun). Neglect any

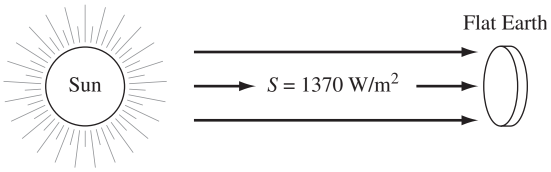


FIGURE P6

radiation losses off of the rim of the Earth, and assume there is no albedo or greenhouse effect. Treating it as a perfect blackbody, estimate the temperature of this new, flat Earth.

7 The solar flux S striking a planet will be inversely proportional to the square of the distance from the planet to the sun. That is, $S = k/d^2$, where k is some constant and d is the distance. Using data from Table 2 to find k,

(a) Estimate the solar constant for the planet Mercury.

(b) If Mercury is 58×10^6 km from the sun and has an albedo of 0.06, find its effective temperature.

(c) At what wavelength would the radiation spectrum from Mercury reach its peak?

8 The solar flux S arriving at the outer edge of the atmosphere varies by ±3.3 percent as the Earth moves in its orbit (reaching its greatest value in early January). By how many degrees would the effective temperature of the Earth vary as a result?

9 In the article "The Climatic Effects of Nuclear War" (*(Scientific American,* August 1984), the authors calculate a global energy balance corresponding to the first few months following a 5,000-megaton nuclear exchange. The resulting smoke and dust in the atmosphere absorb 75 percent of the incoming sunlight (257 W/m^2), while the albedo is reduced to 20 percent. Convective and evaporative heating of the atmosphere from the Earth's surface is negligible, as is the energy reflected from the Earth's surface. The Earth's surface radiates 240 W/m^2, all of which is absorbed by the atmosphere. Assuming that the Earth can be modeled as a blackbody emitter as shown in Figure P9, find the following (equilibrium) quantities:

(a) The temperature of the surface of the Earth (this is the infamous "nuclear winter")

(b) X, the rate at which radiation is emitted from the atmosphere to space

(c) Y, the rate of absorption of short-wavelength solar radiation at the Earth's surface

(d) Z, the rate at which the atmosphere radiates energy to the Earth's surface

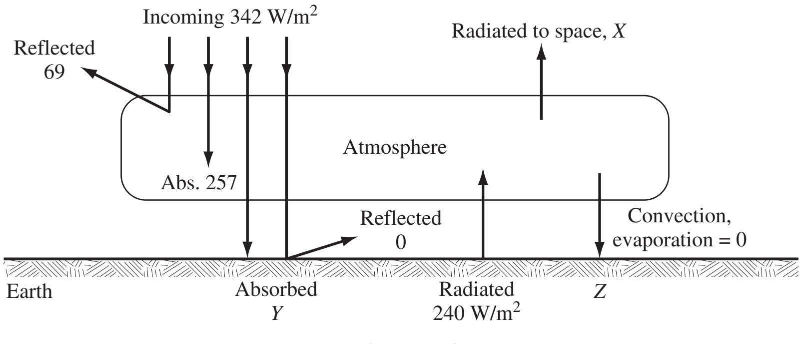


FIGURE P9

10 Consider the two-layer atmospheric model shown in Figure P10. An advantage of this model is it allows us to model radiation so that each layer radiates the same amount off its top and bottom. Find the unknown quantities W, X, Y, and Z to make this model balance. What values of T_1 and T_2 would radiate W and Z?

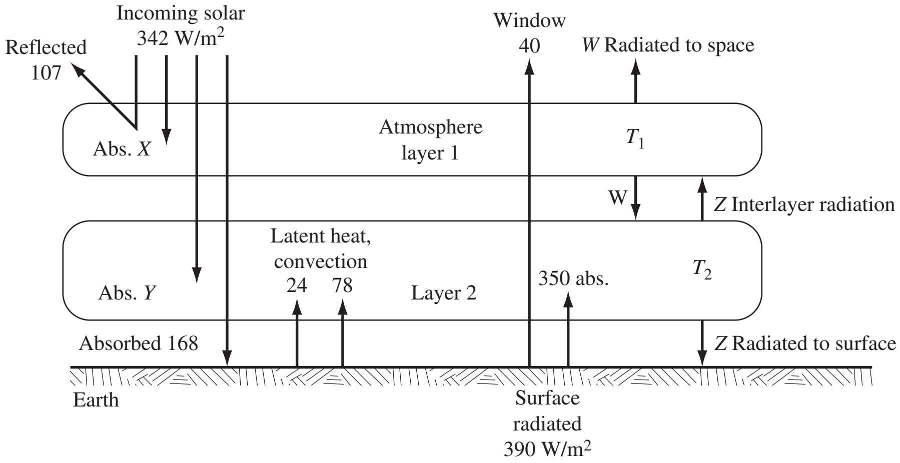


FIGURE P10

11 In Figure 12, the average rate at which energy is used to evaporate water is given as 78 W/m^2. Using 2,465 kJ/kg as the latent heat of vaporization of water, along with the surface area of the Earth, which is about 5.1×10^{14} m^2, estimate the total world annual precipitation in m^3/yr (which is equal to the total water evaporated). Averaged over the globe, what is the average annual precipitation in meters of water?

12 Suppose a greenhouse effect enhancement raises the surface temperature of the Earth to 291 K, melting enough snow and ice to reduce the albedo to the point where only 100 W/m^2 are reflected (see Figure P12). The atmospheric window is closed somewhat so that only 30 W/m^2 now pass directly from the Earth's surface to space. If the latent and sensible heat transfer to the atmosphere do not change, and if incoming solar energy absorbed by the atmosphere does not change, find the solar radiation absorbed at the

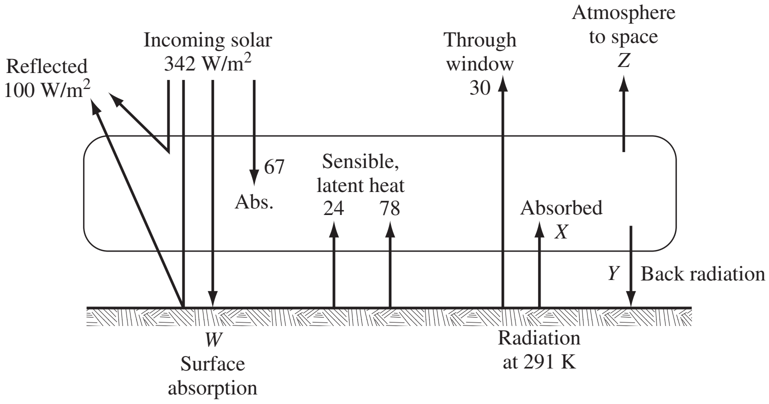


FIGURE P12

surface *W*, the surface radiation *X* absorbed by the atmosphere, the back radiation from atmosphere to the surface *Y*, and the outgoing radiation from the atmosphere *Z* so that energy balance is maintained in each of the three regions (space, atmosphere, surface).

13 What would the atmospheric concentration of CO_2 be 50 years from now if carbon emissions rise linearly from 10 GtC/yr to 16 GtC/yr over that period of time. Assume the initial CO_2 concentration is 380 ppm, and use a 40 percent airborne fraction.

14 Suppose atmospheric CO_2 is growing at 2 ppm/yr, fossil fuel and cement emissions are 9 GtC/yr, and the airborne fraction is 38 percent. If the only other carbon emissions are associated with land use changes, estimate their net carbon emission rate.

15 The United States derives about 40 percent of its energy from oil, 23 percent from coal, and 23 percent from natural gas, with the remaining 14 percent from essentially carbon-free sources.

(a) Using LHV values for carbon intensity for each fuel, estimate the overall carbon intensity.

(b) Suppose all of that coal is replaced with carbon-free nuclear and renewable energy systems; what would be the new carbon intensity?

(c) Suppose that transition takes 100 years. Modeled as an exponential change, what rate of growth in carbon intensity does that correspond to?

16 Using estimates of the total resource base of coal, petroleum, and natural gas given in Table 4, along with LHV carbon intensities given in Table 3, and with an assumed airborne fraction of 50 percent, calculate the total increase in atmospheric CO_2 that would be caused by burning all of the

(a) Natural gas

(b) Petroleum

(c) Coal with 50 percent of the CO_2 captured and stored

(d) Burning all three as stated previously, what would be the equilibrium global temperature increase with a climate sensitivity factor $\Delta T_{2X} = 2.8°\text{C}$ and a current CO_2 concentration of 380 ppm?

17 Suppose that in 100 years the world is virtually out of natural deposits of oil and gas and energy demand is double the current rate of 330 EJ/yr. At that time, imagine that 28 percent of world energy comes from coal, 60 percent is synthetic gas and oil from coal with a LHV carbon intensity of 44 gC/MJ, and the remainder is derived from noncarbon-emitting sources. Using LHV values of carbon intensity,

(a) What would be the carbon emission rate then (Gt/yr)?

(b) If carbon emissions grow exponentially from the current rate of 6.0 GtC/yr to the rate found in part a, what rate of growth *r* would have prevailed during that 100 years?

(c) If the airborne fraction is 50 percent, how much of the carbon emitted in the next 100 years would remain in the atmosphere?

(d) What would be the atmospheric concentration of CO_2 in 100 years if there is no net contribution from biomass (and there are 750 Gt of carbon in the atmosphere now)?

(e) If the equilibrium temperature increase for a doubling of CO_2 is 3°C, what would be the equilibrium temperature increase in 100 years if the initial concentration is 356 ppm?

18 Suppose in 100 years energy consumption is still 330 EJ/yr, but coal supplies only 20 percent of total demand, natural gas supplies 30 percent, and oil 10 percent. The remaining

40 percent of demand is met by noncarbon-emitting sources such as nuclear and solar energy. Using this conservation scenario, repeat parts a through e of Problem 17.

19 One reason for sometimes using HHVs rather than LHVs is related to the strange efficiencies that may emerge when a conversion device actually takes advantage of the latent heat normally lost in flue gases. Suppose a methane-fired condensing furnace has an HHV efficiency of 95 percent. What would its LHV efficiency be?

20 For each of the following fuels, find the carbon intensity based on the higher heating values (HHV) given:

(a) Ethane, C_2H_6, HHV = 1,542 kJ/mol

(b) Propane, C_3H_8, HHV = 2,220 kJ/mol

(c) *n*-Butane, C_4H_{10}, HHV = 2,878 kJ/mol

21 Consider the following four ways to heat a house. The first uses a new high-efficiency pulse-combustion gas furnace; the second uses a conventional gas furnace; the third uses an electric heat pump that delivers 3 units of energy to the house for each unit of electrical energy that it consumes (the other 2 units are heat taken from the ambient air); and the fourth uses conventional electric heating. Using HHV values of carbon intensity (since these efficiencies are based on American definitions) and power plants fueled with coal, compute the carbon emissions per unit of heat delivered to the house (gC/MJ).

Option	Description	Furnace Efficiency (%)	Power Plant Efficiency (%)
1	Pulse-gas	95	—
2	Conventional gas	70	—
3	Heat pump	300	35
4	Electric	100	35

22 Supplement Example 6 with a water heater that burns propane (C_3H_8) with 85 percent efficiency.

(a) Propane releases 2,200 kJ of energy per g-mol; find its carbon intensity (gC/MJ).

(b) Find the ratio of carbon released to energy that actually heats the water (gC/MJ).

(c) What percentage of carbon emissions are saved using propane compared with electric heating from a natural-gas-fired power plant as computed in Example 6?

23 Disaggregated growth rates for three scenarios are given in the following table:

	Population (%/yr)	GDP/person (%/yr)	Energy/GDP (%/yr)	C/Energy (%/yr)	Airborne Fraction	Sensitivity ΔT_{2X}(°C)
(A)	1.0	0.3	−2.0	−0.7	0.4	3
(B)	1.5	1.5	−0.2	0.4	0.5	2
(C)	1.4	1.0	−1.0	−0.2	0.5	3

For each scenario,

(a) Predict CO_2-induced global equilibrium temperature increase in 70 years, if the initial CO_2 concentration is 356 ppm, the initial emission rate is 6.0 Gt/Cyr, and the initial atmospheric carbon content is 750 GtC.

(b) If those growth rates continue, in what year would CO_2 concentrations be equal to double the initial 356 ppm?

24 The expanded Kaya analysis (28) applied to SRES scenario A2 averaged from 1990 to 2100 has the following energy-component growth rates:

Population growth rate $dP/dt = 0.8\%$
Per capita GDP growth rate $d(\text{GDP/P})/dt = 1.3\%$
Final Energy per GDP growth rate $= d(\text{FE/GDP})/dt = -0.7\%$
Primary Energy to Final Energy growth rate $d(\text{PE/FE})/dt = 0.1\%$
Carbon per unit of Primary Energy growth rate $d(\text{TC/PE})/dt = -0.2\%$
Carbon Sequestration growth rate $d(\text{C/TC})/dt = 0.0\%$

With 1990 energy-related carbon emissions equal to 6.0 GtC/yr and growing at the above rates, and land-use and industrial emissions a constant 2.5 GtC/yr:

(a) What would the carbon emission rate be in 2100?

(b) What total carbon would have been emitted?

(c) With an Airborne Fraction of 0.5, what would be the increase in CO_2 concentration due to these energy-related carbon emissions?

(d) With 360 ppm CO_2 in 1990, what would be the concentration in 2100?

(e) With $\Delta T_{2X} = 2.8°\text{C}$, estimate the increase in global equilibrium temperature in 2100 relative to 1990.

25 Indicate whether the following halocarbons are CFCs, HCFCs, HFCs, or halons, and give their designation numbers:

(a) C_3HF_7

(b) $C_2FH_3Cl_2$

(c) $C_2F_4Cl_2$

(d) CF_3Br

26 Write chemical formulas for the following:

(a) HCFC–225

(b) HFC–32

(c) H–1301

(d) CFC–114

27 What feedback climate sensitivity λ and factor g would correspond to the following values of ΔT_{2X}? For each, estimate the change in ΔT_{2X} that would result if the feedback factor g increases by 0.1.

(a) $\Delta T_{2X} = 2.5°\text{C}$

(b) $\Delta T_{2X} = 3.5°\text{C}$

28 Figure 39 shows two probability density functions for the climate sensitivity factor ΔT_{2X} labeled WR and AS.

(a) Estimate the AS probability that ΔT_{2X} is less than 2.5°C.

(b) Estimate the WR probability that ΔT_{2X} is greater than 3°C.

(c) Estimate the AS probability that ΔT_{2X} is between 3°C and 4°C.

(d) Estimate the WR probability that ΔT_{2X} is between 3°C and 4°C.

29 The radiative forcing as a function of concentration for N_2O is sometimes modeled as follows, where C and C_0 are final and initial ppb concentrations:

$$\Delta F = k_2(\sqrt{C} - \sqrt{C_0})$$

Assuming it has been in that region since preindustrial times when the concentration was 275 ppb, find an appropriate k_2 if the current concentration is 311 ppb and the forcing is estimated to be 0.14 W/m^2. Estimate the added radiative forcing in 2100 if it reaches a concentration of 417 ppb.

30 The following is an estimate for radiative forcing caused by the principal greenhouse gases:

$$\Delta F = 6.3\ \ln\frac{[(\mathrm{CO_2})]}{[(\mathrm{CO_2})_0]} + 0.031(\sqrt{\mathrm{CH_4}} - \sqrt{(\mathrm{CH_4})_0}) + 0.133(\sqrt{(\mathrm{N_2O})} - \sqrt{(\mathrm{N_2O})_0}) + 0.22[(\mathrm{CFC{-}11}) - (\mathrm{CFC{-}11})_0] + 0.28[(\mathrm{CFC{-}12}) - (\mathrm{CFC{-}12})_0]$$

where concentrations are expressed in ppb and ΔF is in W/m^2. Using the following data for atmospheric concentrations:

Year	CO_2 (ppm)	CH_4 (ppb)	N_2O (ppb)	CFC–11 (ppb)	CFC–12(ppb)
1850	278	700	275	0	0
1992	356	1714	311	0.268	0.503
2100	710	3616	417	0.040	0.207

(a) What would be the combined radiative forcing caused by these gases from 1850 to 1992?

(b) What would be the forcing from 1992 to 2100?

(c) What would be the forcing from 1850 to 2100?

31 What would be the equilibrium temperature change from 1850 to 2100 for the combination of greenhouse gases described in Problem 30 if the climate sensitivity parameter λ is 0.57°C per W/m^2?

32 For most greenhouse gases, injection of 1 kg into the atmosphere decays exponentially with time (but not CO_2), so their GWP equation (43) can be simplified to the following:

$$\mathrm{GWP_g} = \frac{\int_0^T F_g \cdot R_g(t)dt}{\int_0^T F_{CO_2} \cdot R_{CO_2}(t)dt} = \left(\frac{F_g}{F_{CO_2}}\right) \cdot \frac{\int_0^T e^{-t/\tau}dt}{\int_0^T R_{CO_2}(t)dt}$$

where τ is their residence time (Table 7). For CO_2, the integral for 20-yr, 100-yr, and 500-yr GWPs has been estimated to be roughly the following

$$\int_0^{20} R_{CO_2}(t)\,dt \approx 13.2 \text{ yrs}; \quad \int_0^{100} R_{CO_2}(t)\,dt \approx 43.1 \text{ yrs}; \quad \int_0^{500} R_{CO_2}(t)\,dt \approx 138 \text{ yrs}$$

For HFC-134a, $(F_g/F_{CO_2}) = 4{,}129$ and $\tau = 14$ yrs. Compute its GWP for the following time horizons and compare your results with Table 7.

(a) 20 years

(b) 100 years

(c) 500 years

33 Using the procedure described in Problem 32, compute the global warming potential for a greenhouse gas having an atmospheric lifetime $\tau = 42$ years and a relative forcing per unit mass that is 1,630 times that of CO_2. Do this for a time period of:

(a) 20 years
(b) 100 years
(c) 500 years

34 The IS92a emission scenario suggests anthropogenic emissions in 2025 equal to the following: $44{,}700 \times 10^9$ kg CO_2; 320×10^9 kg CH_4; and 22×10^9 kg N_2O. Weighting these by their global warming potentials, what fraction of the total impact of that year's emissions of these three gases would be caused by each gas over the following time periods?

(a) 20 years
(b) 100 years
(c) 500 years

35 In 2006, the United States emitted 6,000 million metric tons (Mt) of CO_2, 26.6 MtCH_4, and 1.2 MtN_2O. Using 100-year global warming potentials, what is the equivalent CO_2 emission rate of the total for these three gases, in GtC-eq/yr.

36 Direct greenhouse gas radiative forcing in the mid-1990s is estimated to be about 2.45 W/m^2. If realized temperature is 75 percent of equilibrium temperature, what negative forcing by aerosols, and so on, would be needed to match the 0.6°C realized global temperature increase actually observed? Assume that an appropriate climate sensitivity factor λ is 0.57.

37 Repeat Example 12, but this time use a 100-yr instead of the 20-yr GWP for methane leakage.

38 Based on a 1990 world fossil-fuel carbon emission rate of 6.0 Gt/yr and total fossil fuel carbon ever emitted equivalent to the consumption of 200,000 EJ of coal (LHV) for each of the following values of maximum emission rate, plot a graph of carbon emissions vs. time using a Gaussian emission function. In what year would maximum emissions occur?

(a) Maximum emission rate of 22 GtC/yr
(b) Maximum emission rate of 34 GtC/yr
(c) Maximum emission rate of 58 GtC/yr

39 Many economists favor a carbon tax as a way to discourage CO_2 emissions. Suppose such a tax were to be set at \$20 per metric ton of carbon emissions as (CO_2). Consider a small, 50-MW, 35 percent-efficient coal-fired power plant. Using a carbon intensity for coal of 24 gC/MJ,

(a) What would the annual carbon tax be for this power plant assuming capacity factor of 100%?
(b) Suppose a tree plantation sequesters (removes from the atmosphere and stores in biomass) on the order of 5,000 kg of carbon per year per acre over the 40 years that the trees are growing (after which time the forest is mature, and no further accumulation of carbon occurs). What area of forest would have to be planted to "offset" the power plant's emissions over the next 40 years (roughly the lifetime of the power plant)?
(c) How much could the owners of the plant pay for the forestry project (\$/acre per year) and still have it be cheaper than paying the \$20/metric ton carbon tax?

40 Suppose a landfill leaks 10 tonnes (1 tonne = 1 metric ton = 1000 kg) of methane CH_4 into the atmosphere each year.

(a) Using methane's 20-year GWP, what would the warming (radiative forcing) be equivalent to in terms of tonne/yr of CO_2 emissions?

(b) Suppose a soil-vapor extraction system is installed at the landfill to suck up the methane before it leaks into the atmosphere. If that methane is burned, the methane is converted to CO_2. What would be the CO_2 emissions per year now?

(c) What is the equivalent CO_2 savings by burning the methane? How many tonnes per year of C emissions (as CO_2) would be saved?

(d) If a carbon tax of \$20/tonne of C (as CO_2) is enacted, how much tax could be saved per year by burning the methane instead of letting it leak out of the landfill?

(e) A carbon tax of \$20/tonne of C is the same as a tax of \$5.45 per tonne of CO_2 (carbon is 12/44ths of the mass of CO_2). Using the tons of CO_2 equivalents found in a, how much tax could be saved per year if the methane is burned instead of letting it leak out of the landfill?

41 Gasoline is approximately C_7H_{15}, and 1 gallon of it weighs about 6.15 pounds. Assume that all of the carbon in gasoline is emitted as CO_2 when it is burned.

(a) Suppose an old car that only gets 12 miles per gallon (mpg) will be driven 40,000 more miles before it ends up in the junk yard. How much carbon will it emit during that time?

(b) If the car weighs 4,000 pounds, and it is driven 10,000 miles per year, what is the ratio of the weight of carbon emitted per year to the weight of the car?

(c) Suppose there is a carbon tax of \$15 per ton of carbon (\$4.08 per ton of CO_2). What would the carbon tax be on 1 gallon of gasoline?

(d) How much would carbon emissions be reduced over those 40,000 miles if that old car is replaced with a new one that gets 40 mpg?

(e) If an electric utility is trying to reduce its carbon tax by offering incentives to get older cars off the road, how much should the utility be willing to pay to encourage the owner of the old clunker to trade it in on a 40-mpg car?

42 Consider the potential carbon emissions from a gasoline-powered car compared with the carbon emissions to make the electricity for an electric car.

(a) Suppose the conventional car gets 40 miles per gallon (mpg). What are the carbon emissions (gC/mi) assuming that gasoline contains 5.22 pounds of carbon per gallon?

(b) An efficient combined-cycle, natural-gas-fired power plant needs about 8,000 kJ of heat to generate 1 kWh of electricity. At 13.8 gC/MJ for natural gas, what would be the carbon emissions (gC/mi) for an electric vehicle that gets 5 miles/kWh?

(c) Suppose an older, 30 percent-efficient coal-fired power plant generates the electricity that powers that electric car. At a carbon intensity of 24 gC/MJ for coal, what would the carbon emissions be (gC/mi) for that 5 mi/kWh electric car?

43 The photon energy required to cause the following reaction to occur is 306 kJ/mol. What is the maximum wavelength that the photon can have?

$$NO_2 + h\nu \rightarrow NO + O$$

44 Photodissociation of ozone requires 104.6 kJ/mol. What maximum wavelength can a photon have to drive this reaction?

$$O_2 + h\nu \rightarrow O_2 + O$$

REFERENCES

Armstrong, B. K., 1993, Implications of increased solar UVB for cancer incidence, *The Role of the Stratosphere in Global Change,* M. Chanin (ed.), Springer-Verlag, Berlin.

Arrhenius, S., 1896, On the influence of carbonic acid in the air upon the temperature of the ground, *Philosophical Magazine and Journal of Science,* S. 5, 41(251), 237–276.

Andronova, N. G., and M. E. Schlesinger, 2001, Objective estimation of the probability density function for climate sensitivity, *J. of Geophysical Research D: Atmospheres,* 106(D19), 22605–22611.

Baird, C., 1995, *Environmental Chemistry,* W. H. Freeman, New York.

Broecker, W. S., and G. H. Denton, 1990, What drives glacial cycles?, *Scientific American,* January, 49–56.

Caldeira, K., and M. E. Wickett, 2003, Anthropogenic carbon and ocean pH, *Nature,* 435, 365, September 25.

Dansgaard, W., and H. Oeschger, 1989, Past environmental long-term records from the Arctic, *The Environmental Record in Glaciers and Ice Sheets,* H. Oeschger and C. C. Langway (eds.), John Wiley & Sons, New York.

Edmonds, J., and J. Reilly, 1983, Global energy production and use to the year 2050, *Energy,* 8, 419.

Ehrlich, P., and J. Holdren, 1971, Impact of population growth, *Science*, 171, 1212–1217.

Energy Information Administration (EIA), 2006, *Emissions of Greenhouse Gases in the United States 2005,* U.S. Department of Energy, Washington, DC.

Frederick, J. E., 1986, The ultraviolet radiation environment of the biosphere, *Effects of Changes in Stratospheric Ozone and Global Climate,* Vol. 1, USEPA and UNEP, Washington, DC.

Henderson-Sellers, A., and K. McGuffie, 1987, *A Climate Modelling Primer,* John Wiley & Sons, New York.

Hansen, J. E., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl, 2001, A closer look at United States and global surface temperature change, *J. Geophys. Res.,* 106, 23947–23963.

Hansen, J., 2004, The role of soot in global climate change, *Black Carbon Emissions and Climate Change: A Technical Workshop,* Oct 13–15, National Renewable Energy Laboratory, San Diego, CA.

Houghton, R. A., 2003, Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000, *Tellus,* 55B, 378–390.

Hummel, H., 2006, *Interpreting Energy and Emission Scenarios: Methods for Understanding and Communicating Policy Insights,* Ph.D. Thesis, Stanford University, Stanford, CA.

Houweling, S., T. Kaminski, F. Dentener, and J. Lelieveld, 2006, Inverse modeling of methane sources and sinks using the adjoint of a global transport model, *Nature*, 439, 187–191, January, 12.

Imbrie, J., J. D. Hays, D. G. Martinson, A. McIntyre, A. C. Mix, J. J. Morley, N. G. Pisias, W. I. Pell, and N. J. Shackleton, 1984, Orbital theory of the Pleistocene climate: Support from a revised chronology of the marine $\delta^{18}O$ record, in *Milankovitch and Climate,* Part I, A. Berger, J. Imbrie, J. Hays, G. Kukla, and B. Saltzman (eds.), D. Reidel Publishing Co., Dordrecht. pp 269–305.

Intergovernmental Panel on Climate Change (IPCC), 1990, *Climate Change: The IPCC Scientific Assessment,* Cambridge University Press, Cambridge, UK.

Intergovernmental Panel on Climate Change (IPCC), 1992, *Climate Change 1992,* Cambridge University Press, Cambridge, UK.

Intergovernmental Panel on Climate Change, 1995, *Climate Change 1994: Radiative Forcing of Climate Change,* Cambridge University Press, Cambridge, UK.

Intergovernmental Panel on Climate Change, 1996, *Climate Change 1995: The Science of Climate Change,* Cambridge University Press, Cambridge, UK.

Intergovernmental Panel on Climate Change (IPCC), 2000, *IPCC Special Report: Emissions Scenarios,* Cambridge University Press, Cambridge, UK.

Intergovernmental Panel on Climate Change (IPCC), 2007, *Climate Change 2007: The Physical Science Basis,* Cambridge University Press, Cambridge, UK.

International Agency for Research on Cancer, 1992, *Solar and Ultraviolet Radiation,* IARC monographs on the evaluation of carcinogenic risks to humans, Vol. 55, IARC, Lyon.

Jouzel, J., N. I. Barkov, J. M. Barnola, M. Bender, J. Chappellaz, C. Genthon, V. M. Kotlyakov, V. Lipenkov, C. Lorius, J. R. Petit, D. Raynaud, G. Raisbeck, C. Ritz, T. Sowers, M. Stievenard, F. Yiou, and P. Yiou, 1993, Extending the Vostok ice-core record of paleoclimate to the penultimate glacial period, *Nature,* 36, 407–412.

Kiehl, J. T., and K. E. Trenberth, 1996, Earth's annual global mean energy budget, *Bulletin of the American Meteorological Society* (submitted).

Kaya, Y., 1990, Impact of carbon dioxide emission control on GNP growth: Interpretation of proposed scenarios, *IPCC Energy and Industry Subgroup, Response Strategies Working Group*, Paris, Cambridge University Press, Cambridge, UK.

Keppler, J., J. T. G. Hamilton, M. J. Bra, and T. Röckmann, 2006, Methane emissions from terrestrial plants under aerobic conditions, *Nature*, 439: 187–191.

Lean, J., and D. Rind, 1996, The Sun and climate, *Consequences,* 2(1), 27–36.

Lisiecki, L. E., and M. E. Raymo, 2005, A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18}O records, *Paleoceanography,* 20, PA1003.

Madronich, S., 1993, Trends in surface radiation, *The Role of the Stratosphere in Global Change,* M. Chanin (ed.), Springer-Verlag, Berlin.

Molina, M. J., and F. S. Rowland, 1974, Stratospheric sink for chlorofluormethanes: Chlorine atom catalyzed destruction of ozone, *Nature,* 249, 810–812.

McGuffie, K., and A. Henderson-Sellers, 2005, *A Climate Modelling Primer,* 3rd ed, John Wiley & Sons, West Sussex, England.

Nakicenovic, N., 1996, Energy primer, in *Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses, Intergovernmental Panel on Climate Change,* Cambridge University Press, Cambridge, UK.

National Research Council, Committee on Radiative Forcing Effects on Climate, 2005, *Radiative Forcing of Climate Change: Expanding the Concept and Addressing Uncertainties,* National Academy Press, Washington, DC.

National Research Council, Panel on Climate Change Feedbacks, 2003. *Understanding Climate Change Feedbacks,* National Academy Press, Washington, DC.

Pacala, S., and R. Socolow, 2004, Stabilization wedges: Solving the climate problem for the next 50 years with current technologies, *Science,* 305(5686), 968–972.

Petit, J. R., J. Jouzel, D. Raynaud, N. I. Barkov, J. M. Barnola, I. Basile, M. Bender, J. Chappellaz, J. Davis, G. Delaygue, M. Delmotte, V. M. Kotlyakov, M. Legrand, V. Lipenkov, C. Lorius, L. Pépin, C. Ritz, E. Saltzman, and M. Stievenard, 1999, Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, *Nature*, 399, 429–436.

Reddy, A. K. N., and J. Goldemberg, 1990, Energy for the developing world, *Scientific American,* September, 263(3), 110–118.

Simon, P. C., 1993, Atmospheric changes and UV-B monitoring, *The Role of the Stratosphere in Global Change,* M. Chanin (ed.), Springer-Verlag, Berlin.

Swisher, J. N., and G. M. Masters, 1991, Buying environmental insurance: Prospects for trading of global climate-protection services, *Climatic Change*, 19, 233–240.

Schneider, S. H., and J. Lane, 2006, *Avoiding Dangerous Climate Change*, H. J. Schellnhuber (ed.), Cambridge University Press, Cambridge.

Siegenthaler, U., T. F. Stocker, E. Monnin, D. Lüthi, J. Schwander, B. Stauffer, D. Raynaud, J. Barnola, H. Fischer, V. Masson-Delmotte, and J. Jouzel, 2005, Stable carbon cycle–climate relationship during the late Pleistocene, *Science,* 25, 310, 5752, 1313–1317.

Spahni, R., J. Chappellaz, T. K. Stocker, L. Loulergue, G. Hausammann, K. Kawamura, J. Flückiger, J. Schwander, D. Raynaud, V. Masson-Delmotte, and J. Jouzel, 2005, Atmospheric methane and nitrous oxide of the late Pleistocene from Antarctic ice cores, *Science,* 310(5752), 1317–1321.

United Nations Environment Programme (UNEP), 1994, *Montreal Protocol on Substances that Deplete the Ozone Layer, Scientific Assessment of Ozone Depletion*, World Meteorological Organization Global Ozone Research and Monitoring Project—Report No. 37, Geneva, Switzerland.

Warren, S. G., and S. H. Schneider, 1979, Seasonal simulation as a test for uncertainties in the parameterization of a Budkyo-Zellers zonal climate model, *Journal of the Atmospheric Sciences*, 36, 1377–1391.

Wuebbles, D. J., 1995, Weighing functions for ozone depletion and greenhouse gas effects on climate, *Annual Review of Energy and Environment*, 20, 45–70.

Wigley, T., and S. Raper, 2001, Interpretation of high projections for global-mean warming, *Science,* 293(5529), 451–454.

World Meteorological Organization (WMO), 2002, *The Scientific Assessment of Ozone Depletion, 2002*.

Worrell, E., L. Price, N. Martin, C. Hendriks, and L. Meida, 2001, Carbon dioxide emissions from the global cement industry, *Annual Review of Energy and the Environment*, 26, 303–329, November.

Solid Waste Management and Resource Recovery

The Congress hereby declares it to be the national policy of the United States that pollution should be prevented or reduced at the source whenever feasible; pollution that cannot be prevented should be recycled in an environmentally safe manner whenever feasible; pollution that cannot be prevented or recycled should be treated in an environmentally safe manner whenever feasible; and disposal or other release into the environment should be employed only as a last resort and should be conducted in an environmentally safe manner.

—*Pollution Prevention Act of 1990*

From *Introduction to Environmental Engineering and Science*. Third Edition. Gilbert M. Masters, Wendell P. Ela.

1 Introduction

On March 22, 1987, the garbage barge *Mobro 4000* and its tug, the *Break of Dawn,* left the port of New York loaded with 3,200 tons of solid waste on its way to a port in North Carolina where its cargo was meant to become fuel for an ill-conceived waste-to-methane project. Along the way, questions emerged about medical wastes possibly being part of its load, which led to the revocation of its permit to dock. Thus began the *Mobro*'s famous six-month, 6,000 mile odyssey as it searched for a friendly port willing to take its load of trash. It was refused entry in six states and three foreign countries before returning to New York Harbor and, after a heated legal battle, was finally allowed to dock and deliver its refuse to an incinerator in Brooklyn. The press distorted the story, and the *Mobro* inadvertently became a powerful symbol of a not-in-my-backyard (NIMBY), "throw-away" society that had run out of a convenient "away."

Perhaps stimulated by the *Mobro* episode, much has changed in the field of solid waste management in the past two decades. The per capita rate at which the United States generates municipal solid waste (MSW) grew from 2.7 pounds per day in the 1960s to 4.5 pounds by 1990, at which point it came to a virtual standstill, with the 2005 rate still being 4.5 pounds per day per person. Meanwhile, the recycling rate has grown substantially, from 10 percent of generated MSW in 1980 to 32 percent in 2005, which means less has to go to landfills. In fact, in the United States, we now have 80 percent fewer landfills than existed at the time of the *Mobro* in 1987. Most of the sites that closed down were smaller, older sites, which were not designed with today's level of environmental controls. Those that remain are bigger and better.

Historically (certainly well into the twentieth century), solid waste management was of the most rudimentary sort even in the most developed countries of the world. Garbage, if it was collected at all, usually ended up at the local dump where open burning was commonly used to control both the volume and public health dangers associated with the waste. It wasn't until roughly 1950 that most urban areas in the United States ceased operating unsightly, rat-infested open dumps. More recently, and for a number of reasons, the technologies used to manage our wastes have become increasingly sophisticated. As the quote at the beginning of this chapter illustrates, present policies emphasize waste reduction and recycling to minimize the volume and toxicity of the materials that must be disposed of. The "away" that the remaining wastes are heading toward may include incineration, perhaps with energy recovery, and/or disposal in carefully engineered and managed sanitary landfills.

2 RCRA Wastes

The Resource Conservation and Recovery Act (RCRA), which is the major federal statute that governs solid waste, delineates two categories of waste: *hazardous* and *nonhazardous* waste. Most of the focus in RCRA is on Subtitle C, which defines and regulates hazardous waste. In this chapter, however, we want to address Subtitle D,

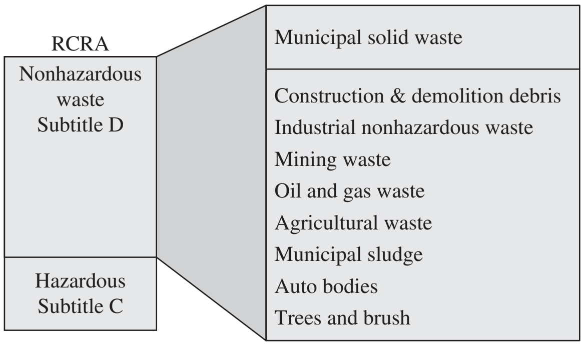


FIGURE 1 About 95 percent of RCRA regulated solid wastes are categorized as nonhazardous, with about 2 percent of that being municipal solid waste (MSW).

TABLE 1

Ore Required to Yield 100 kg of Product

Mineral	Average Grade (%)	Ore (kg)	Product (kg)	Residue (kg)
Aluminum	23	435	100	335
Copper	0.91	10,990	100	10,890
Iron	40	250	100	150
Lead	2.5	4,000	100	3,900
Nickel	2.5	4,000	100	3,900
Others (avg.)	8.1	1,234	100	1,134

Source: Based on data from Young, 1992.

nonhazardous wastes (Figure 1). Of the roughly 12 billion tons per year of nonhazardous wastes, most is generated by industry during raw material extraction, material processing, and product manufacturing. Most of this chapter, however, will be looking at the roughly 2 percent of total wastes that make up the category referred to as municipal solid waste (MSW).

The magnitude of the earth- and ore-moving task needed to provide the minerals society demands is truly monumental. Digging holes, removing the ores, and piling up the leftover residues creates enormous aesthetic, environmental, economic, and energy problems. Table 1 provides an estimate of typical amounts of materials that must be moved and processed to produce some of our most important minerals. Producing the metals and other basic materials used in our society requires considerable amounts of energy, which adds its own costs and environmental impacts. In some cases, such as aluminum, the cost of that energy is a major fraction of the total cost of production, which provides a major incentive to recycle it rather than producing it from virgin materials.

Even though MSW is dwarfed in size (and environmental impact) by the other sectors, it should be remembered that all of that industrial waste is created in the process of providing us with the material things that ultimately may end up in our trash. Consuming less therefore saves not only the wastes that would have ended up in the municipal wastestream, but it also reduces the energy, materials, and waste associated with providing those items that we just might be able live without. Clearly, attention needs to be directed not just at the management of consumer wastes but at the complete set of processes that result in the products our society seems to think it needs.

3 Municipal Solid Waste

Before we can more carefully consider the material flows and environmental impacts of wastes, some definitions need to be provided. Figure 2 may help keep some of these terms straight.

Solid wastes are wastes that aren't liquid or gaseous, such as durable goods, nondurable goods, containers and packaging, food scraps, yard trimmings, and miscellaneous inorganic wastes. Solid waste is more or less synonymous with the term *refuse,* but *solid waste* is the preferred term.

Municipal solid waste (MSW) is solid waste from residential, commercial, institutional, and industrial sources, but it does not include such things as construction waste, automobile bodies, municipal sludges, combustion ash, and industrial process wastes even though those wastes might also be disposed of in municipal waste landfills or incinerators.

Garbage, or *food waste,* is the animal and vegetable residue resulting from the preparation, cooking, and serving of food. This waste is largely putrescible organic matter and moisture. Home kitchens, restaurants, and markets are sources of garbage, but the term usually does not include wastes from large food-processing facilities such as canneries, packing plants, and slaughterhouses.

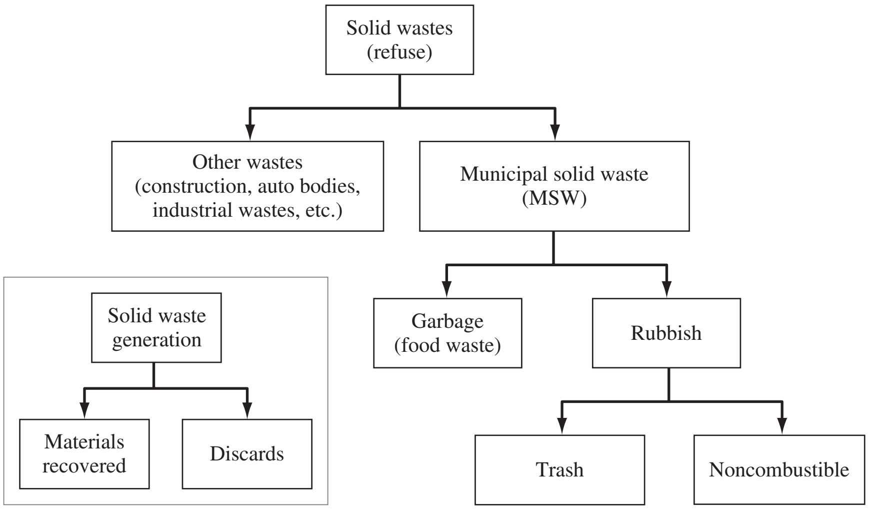


FIGURE 2 Illustrating some definitions used in solid waste management.

Rubbish consists of old tin cans, newspaper, tires, packaging materials, bottles, yard trimmings, plastics, and so forth. Both combustible and noncombustible solid wastes are included, but rubbish does not include garbage.

Trash is the combustible portion of rubbish.

Generation refers to the amounts of materials and products that enter the waste stream. Activities that reduce the amount or toxicity of wastes before they enter the municipal waste system, such as reusing refillable glass bottles or reusing plastic bags, are not included.

Materials recovery is the term used to cover the removal of materials from the wastestream for purposes of recycling or composting.

Discards are the solid waste remaining after materials are removed for recycling or composting. These are the materials that are burned or buried. In other words,

$$\text{Waste generation} = \text{Materials recovered} + \text{Discards} \tag{1}$$

MSW Generation in the United States

The EPA estimates that the United States generated 246 million tons of municipal solid waste in 2005, which is about 4.5 pounds per person per day. As shown in Figure 3, paper and paperboard products are the largest single component of MSW (34.2 percent by weight; another 37 percent is roughly equally divided among yard trimmings, food scraps and plastics). Figure 4 shows the historical growth in generation of MSW, by source. Over the years, the single largest component—paper and paperboard products—grew steadily until 2000, but it seems, finally, to be leveling off. In recent years, the second largest component, yard trimmings, has been declining in importance as backyard composting and mulching lawnmowers have become more common. One category that continues to grow steadily is plastics.

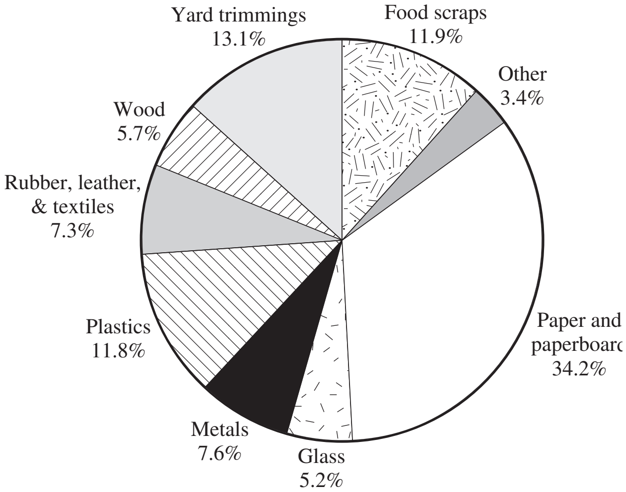


FIGURE 3 Materials generated in municipal solid waste by weight, 2005. Total generation was 246 million tons.
(*Source:* U.S. EPA, 2006a.)

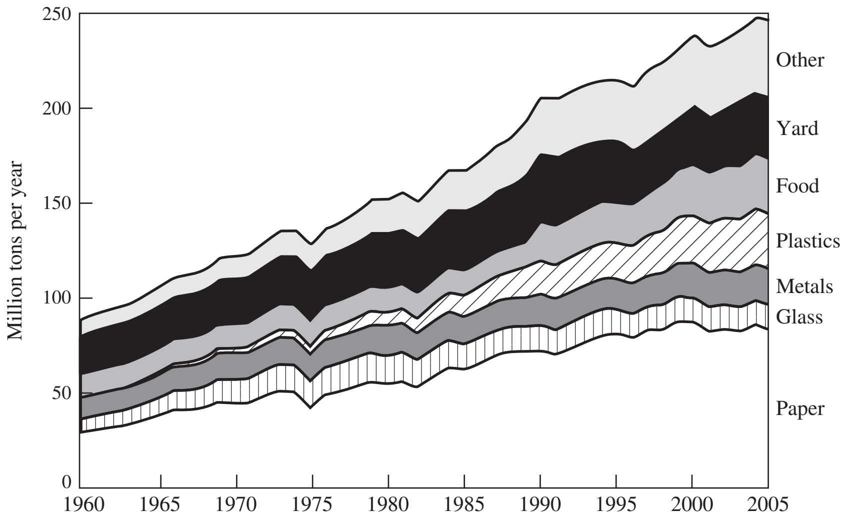


FIGURE 4 Generation of materials in MSW, 1960 to 2005.
(*Source:* U.S. EPA, 2006a.)

4 Solid Waste Management

Management of municipal solid wastes includes recovery of materials for recycling and composting, combustion with or without energy recovery, and finally disposal in landfills or other locations (Figure 5).

Integrated Solid Waste Management

An approach called *integrated solid waste management* is intended to help guide decisions about the generation of wastes, recycling of materials, and ultimate disposal of waste residues. An outline of some of the highlights of integrated solid waste management is suggested in Table 2. Source reduction and recycling in almost all circumstances are given the highest priority, and it is easy to understand why:

- To reduce the amount of solid waste that has to be burned or buried
- To reduce pollution associated with mining, use, and disposal of resources
- To reduce the rate of consumption of scarce resources.

While the order in which the entries in Table 2 have been listed is purposely intended to suggest a hierarchy of priorities, that hierarchy is not rigid. For example, is lightweight packaging that uses less material always better than more substantial packaging? If the heavier packaging can more easily be recycled, it might be better than the lightweight materials that go straight to the dump. Perhaps the heavyweight materials would burn cleanly in a waste-to-energy incinerator, while toxics in the lightweight packaging might make it inappropriate for combustion.

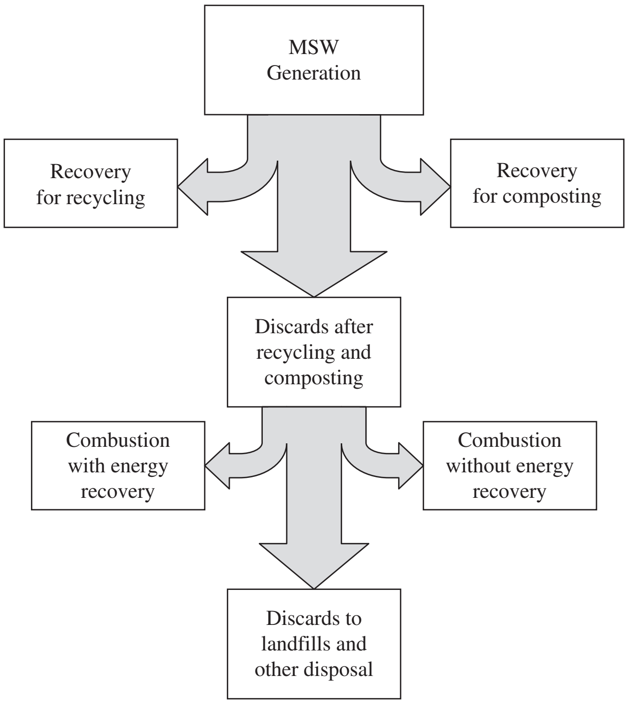


FIGURE 5 Management of generated MSW includes recovery for recycling and composting, combustion with and without energy capture, and finally disposal of discards to landfills.

TABLE 2

Integrated Solid Waste Management
Source Reduction • Reduce toxicity • Less packaging • Product reuse • More durable products • On-site mulching and composting
Recycling • Collecting • Processing • Using recycled materials in products • Composting
Disposal • Combustion with energy recovery • Landfill • Incineration without energy recovery

The disposal priorities suggested in Table 2, which would appear to rank combustion with energy recovery above landfilling, are controversial. Not only would combustion characteristics of the waste affect the decision, but also the relative suitability of local topography and meteorology for landfills versus incinerators needs to be evaluated. The history of MSW combustion has, in fact, been a checkered one. In 1960, about 30 percent of MSW was burned in incinerators, typically with no air pollution controls and no energy recovery. As concern for air quality grew in the 1960s and 1970s, those old incinerators were closed, and the fraction of our wastes that was burned dropped to a low of 10 percent in the early 1980s. As incinerators with emission controls and energy recovery began to come on line, the quantity of MSW burned rose quickly in the 1980s, rising to about 16 percent in the 1990s, but their performance has been somewhat problematic, and there has been little growth since then.

The Status of Integrated Waste Management in the United States

The impact of the integrated waste management perspective can be demonstrated with some statistics. Of the 246 million tons of MSW generated in the United States in 2005, just under one-third was recovered for recycling or composting, up from 7 percent in 1960 and 17 percent in 1990. Of the wastestream that was not recycled or composted, 80 percent went to landfills, and the remainder was combusted with energy recovery. As can be seen in Figure 6, the rate at which MSW is disposed of

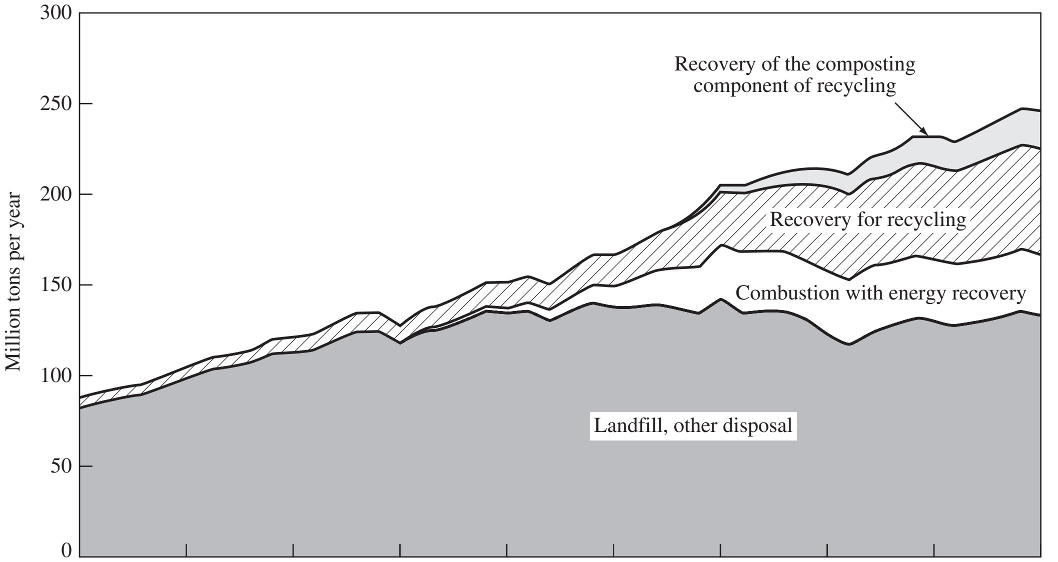


FIGURE 6 Trends in municipal solid waste generation, recovery, and disposal, from 1960 to 2005. Disposal in landfills peaked in 1990.
(*Source:* U.S. EPA, 2006a.)

TABLE 3

Municipal Solid Waste in Selected Countries[a]

Country	Year of Estimate	Per Capita Generation (kg/yr)	Percent Landfilled (%)	Percent Incinerated (%)	Relative Population Density
Australia	1980	681	98	2	0.1
Austria	1988	355	68	8	3.4
Canada	1989	625	84	9	0.1
France	1989	303	45	41	3.8
Germany	1987	318	66	30	9.2
Italy	1989	301	31	16	7.1
Japan	1988	394	33	64	12.1
Spain	1988	322	77	5	2.9
United Kingdom	1989	357	78	14	8.7
United States	1993	730	62	16	1.0

[a] Some data may not be directly comparable given differences in reporting methods.
Source: World Resources Institute, 1992, and U.S. EPA, 1994.

in landfills has remained relatively steady and has even declined some over the past 25 years, in spite of a nearly 50 percent increase in population.

The per capita generation of municipal solid waste in the United States is typically at least twice as high as it is for most European countries, as shown in Table 3. For comparison, the percentages of waste generated that is landfilled and the population densities are also shown in the table. While there is some correlation between population density and landfilling, there are notable exceptions. Sparsely populated countries, such as Australia and Canada, tend to landfill very high fractions of their wastes, as might be expected, while densely populated countries such as Japan and Italy landfill less and incinerate more. Japan, for example, is 12 times as densely populated as the United States, and it landfills only 33 percent of its municipal solid waste, while incinerating 64 percent. The United Kingdom, on the other hand, is also much more densely populated than the United States, but it sends a higher fraction of its waste to landfills than we do.

5 Source Reduction

"Garbage that is not produced does not have to be collected," is a simple enough concept, yet, in the United States, it has taken a long time to receive the attention that it deserves. In this regard, we have been lagging behind many other advanced countries in the world—especially those with modest domestic resources and limited land space for disposal, such as Germany and Japan.

Green Product Design Strategies

Design that concerns itself with reducing the environmental impacts associated with the manufacture, use, and disposal of products is an important part of any pollution prevention strategy. Fortunately, companies that design green products and businesses that use them are finding bottom-line, market benefits that go beyond simply feeling good about their endeavors.

A number of strategies have been identified that contribute to good green design practices, including the following (Keoleian et al., 1994).

Material Selection. A critical stage in product development is selection of appropriate materials to be used. In green design, attempts are made to evaluate the environmental impacts associated with the acquisition, processing, use, and retirement of the materials under consideration. In some cases, substituting one material for another can have a modest impact on the quality and price of the resulting product but can have considerable impact on the environmental consequences.

Designers today have a much greater range of materials to choose from in developing their products. Ceramics and composites offer superior strength with lighter weight than the traditional materials such as steel and aluminum. High-strength alloys and plastics are quickly displacing the metals used in the past. For example, the telecommunications cables that AT&T used in the 1950s consisted mostly of steel, lead, and copper, with small percentages of aluminum and about 1 percent plastics. By the mid-1980s, polyethylene had replaced virtually all of the lead that had been used in the cables' sheathing so that the fraction of plastics is now more than 35 percent. Such progress continues with fiber-optic cables, which are now replacing traditional copper cables, weighing only 3 percent as much as copper and using only 5 percent of the energy to produce (OTA, 1992).

An especially important aspect of materials selection is the need to reduce the toxicity of materials whenever possible. Toxic substances used in products can create serious environmental risks when those products are in use as well as when they end up in our wastestream. Toxic substances emitted from everyday products we use in our homes; volatile organic compounds released from building materials and glues; and insecticides, paints, and solvents that we store in our garage have the potential to cause us direct harm when we come in contact with them, and that harm can continue when they end up in landfills or incinerators. Along with these products, heavy metals, such as lead, cadmium, chromium, mercury, arsenic, copper, tin, and zinc, are quite commonly used in consumer products and an especially dangerous component in MSW. Examples of sources of heavy metals in the municipal wastestream are shown in Table 4.

TABLE 4

Examples of Toxic Metal Sources in MSW

Metal	Applications
Arsenic	Wood preservatives; household pesticides
Cadmium	Rechargeable Ni-Cad batteries; pigments in plastics (kitchenware, etc.), colored printing inks (magazines, catalogs), and enamels (pots, pans, mugs, colored glassware, porcelain, ceramics)
Lead	Automobile batteries, solder (plumbing, gutters, electronics, food cans); pigments or stabilizers in plastics, glass, ceramics
Mercury	Batteries; fluorescent lamps; fungicides; paints
Zinc	Flashlight batteries; rubber products (including tires)

Product System Life Extension. Products that don't wear out as quickly don't have to be replaced as often, which usually means resources are saved, and less waste is generated. Sometimes products are discarded for reasons that have nothing to do with their potential lifetime, such as computers that become obsolete and clothing fashions that change, but many products can continue to remain in service for extended periods of time if they are designed to be durable, reliable, reusable, remanufacturable, and repairable.

Extending product life, of course, mean consumers replace their products less often, which translates into decreased sales volume for manufacturers. The temptation toward planned obsolescence can hopefully be replaced by the realization that market share in the future may be driven by consumer demand for greener products.

Material Life Extension. Once a product has reached the end of its useful life, the materials from which it was made may still have economic value and additional savings can result from avoiding disposal. The key design parameter for extending the life of materials is the ease with which products can be recycled when they reach the end of their useful life. Manufacturers are beginning to face legislative mandates and other pressures to either physically take back their products when consumers are finished with them, or assure in some way that significant fractions of their products will be recycled.

Europe has led the way in this regard, and since that is such a major market, many major manufacturers are including design for disassembly and recyclability no matter where their products are likely to be sold. Recycling mandates for certain products have been enacted, including the *End of Life Vehicle* (ELV) Directive, which currently requires 80 percent of an ELV to be reused and recycled, and by 2015, that rate escalates to 90 percent. ELVs can only be scrapped at authorized treatment facilities, which are required to meet tight environmental standards. The cost of these facilities is paid for in large part by the vehicle manufacturers.

A similar measure called the *Waste Electrical and Electronic Equipment* (WEEE) Directive focuses on keeping hazardous materials out of landfills. The United States is beginning to follow Europe's lead, and in 2006, a number of computer manufacturers and some "big-box" retailers began *takeback* programs for computers and other electronic products. Some of these charge a nominal fee, some are free, and some offer consumers a cash rebate for returned equipment.

Reduced Material Intensiveness. Green design strategies include reducing the amount, and/or toxicity, of material required to make a given product, while maintaining the usefulness and value of the product. Changes in battery technology provide a good example. In 1974, a typical car battery contained about 30 pounds of lead, but modern batteries use less than 20 pounds. Common household batteries now use much less mercury than they did in the past, and many new battery technologies incorporate materials that are far less toxic.

The history of beverage containers provides another example of sizable reductions in materials intensiveness. Nonrefillable (one-way) 0.47 L (16 oz) glass bottles, for example, are 37 percent lighter than they were in the 1970s. Similar lightweighting for steel cans, aluminum cans, and polyethylene terephthalate (PET) one-piece plastic bottles is shown in Table 5.

TABLE 5

Mass Reductions in Soft Drink Container

Type of Container	1970s (g)	1992 (g)	Percent change (%)
One-way glass bottle (0.47 L, 16 oz)	344	218	−37
Steel can (0.36 L, 12 oz)	48	33	−31
Aluminum can (0.36 L, 12 oz)	20	16	−22
PET bottle (2 L, 68 oz, one-piece)	66	54	−18

Source: Franklin Associates, 1994.

Process Management. Manufacturing products requires raw materials and energy inputs, both of which often can be managed more efficiently.

The energy required to manufacture a product is an important component of a lifecycle assessment. Process improvements such as waste-heat recovery, use of more efficient motors and motor controls, and high-efficiency lighting are almost always cost-effective. Since electric motors account for two-thirds of the electricity used by industry (and almost half of the electricity used in the United States for all purposes), a good place to look for efficiency improvements is there. It is not uncommon for motors to be oversized and to run at constant speed, with both factors contributing to low energy efficiency. When motors are used to pump fluids, the pumping rate is usually controlled by adjusting valves or dampers rather than by adjusting the speed of the motor itself, which wastes electricity. New electronic adjustable-speed drives, coupled with more efficient motors, can often save half of the energy normally used.

Better materials management can also lead to lower environmental impacts. Wastes can be minimized by more carefully estimating and ordering needed inputs, especially when they are hazardous materials, and by more careful inventory control.

Efficient Distribution. Methods of packaging and transporting products greatly affect the overall energy and environmental impacts associated with those products.

Transportation costs are affected by the quantity and type of material shipped, which is in turn affected by packaging, trip distance, and type of carrier. The type of carrier is constrained by the terrain to be covered as well as the speed required for timely arrival. In general, shipping by boat is the least energy-intensive option, followed by rail, truck, and then air, which is fastest but requires the most energy per ton-mile transported. If pipelines are an option (as for example, transport of petroleum), they can be even less energy intensive per ton-mile than transport by ship.

Packaging can be considered a component of distribution. Reducing the amount of packaging can reduce the environmental costs of distribution, but may increase deterioration of the product as it is damaged or loses freshness. Using recycled materials for packaging along with strategies that reduce the amount of packaging can contribute to reduced environmental lifecycle costs. Once again, Europe leads the way in helping to reduce packaging waste. Germany, for example, long ago shifted the burden of packaging disposal from the consumer back onto manufacturers and distributors. Their Packaging Waste Law requires manufacturers and distributors to recover and recycle their own packaging wastes.

Eco-Labels

Surveys have consistently shown that consumers, if given a choice, are willing to purchase products that are environmentally superior to competing products, even if they cost a bit more. Clearly, from the pair of lifecycle assessments just described, a detailed LCA list of emissions, energy inputs, and disposal impacts would overwhelm even the most conscientious consumer. At the opposite extreme are a variety of relatively ambiguous terms such as “recyclable,” “recycled,” “eco-safe,” “ozone-friendly,” and “biodegradable” that often appear on product labels. Without a uniform and consistent standard for such terms, these labels are all too often meaningless or even misleading. For example, all soaps and detergents have been “biodegradable” since the 1960s, and “CFC-free, ozone-friendly” aerosols have been the norm in the United States since the banning of CFCs for such applications in 1978. Labels claiming a product is “eco-safe” or “environmentally safe” usually have little meaning since the terms are largely undefined.

Clearly, a credible labeling system certifying that the products and packaging bearing such labels have been independently certified to meet certain environmental standards can be a powerful motivator in the marketplace. It is interesting to note that the EPA once proposed the establishment of a national eco-label but has since abandoned the idea, leaving such labels in the hands of the private sector. To that end, there are now several competing eco-labels being promulgated by private organizations in the United States. *Green Seal* labels provide a simple, overall stamp of approval for consumer products, analogous to the Underwriter Laboratories (UL) label on electrical appliances or the Good Housekeeping Seal of Approval. Scientific Certification Systems, on the other hand, develops standards and certification programs for manufacturers, including its *SCS Sustainable Choice* certification for building products. The SCS certification is based on a multifaceted rating system that includes such components as public health and environment, renewable energy and efficiency, recycled materials, corporate environmental and social responsibility, and end-of-life management for sustainable reuse of materials.

In contrast to the United States, in other parts of the world, national labels are prevalent. Examples of such labels shown in Figure 7 include the “Blue Angel” in Germany, the Japanese and Indian versions of an “EcoMark,” and the “Environmental Choice” label used in Canada.

6 Lifecycle Assessment

Solid waste management all too often focuses almost entirely on what to do with a given wastestream, with the key decision being whether to incinerate the waste or bury it. As landfills filled up and closed, as incinerators were shut down due to poor performance, and as communities became more agitated by the environmental impacts of living near either type of facility, a new, broader, approach to the problem has emerged. Using an energy and materials balance approach at every stage in the lifecycle of a product can yield new insights into not only the solid waste problem but also problems of air and water pollution.

A conceptual diagram for the *lifecycle assessment* (LCA) of a product is presented in Figure 8. Inputs include energy and raw materials utilized in each stage

Canada (Environmental Choice)

Japan (EcoMark)

United States (Green Seal)

United States (SCS Sustainable Choice)

Germany (Blue Angel)

India (EcoMark)

FIGURE 7 Eco-labels from around the world.

of production, use, and disposal of the product. The central box in the figure suggests the various stages in the product lifecycle, which includes the acquisition and processing of materials (mining, smelting, and so on), the actual manufacturing of the product itself, the packaging and distribution of the product, the use of the product, and, finally, the ultimate disposal of the product. Outputs are the air, water, and solid waste effluents associated with each stage, along with the waste heat dissipated in the environment plus energy that may be recovered during disposal.

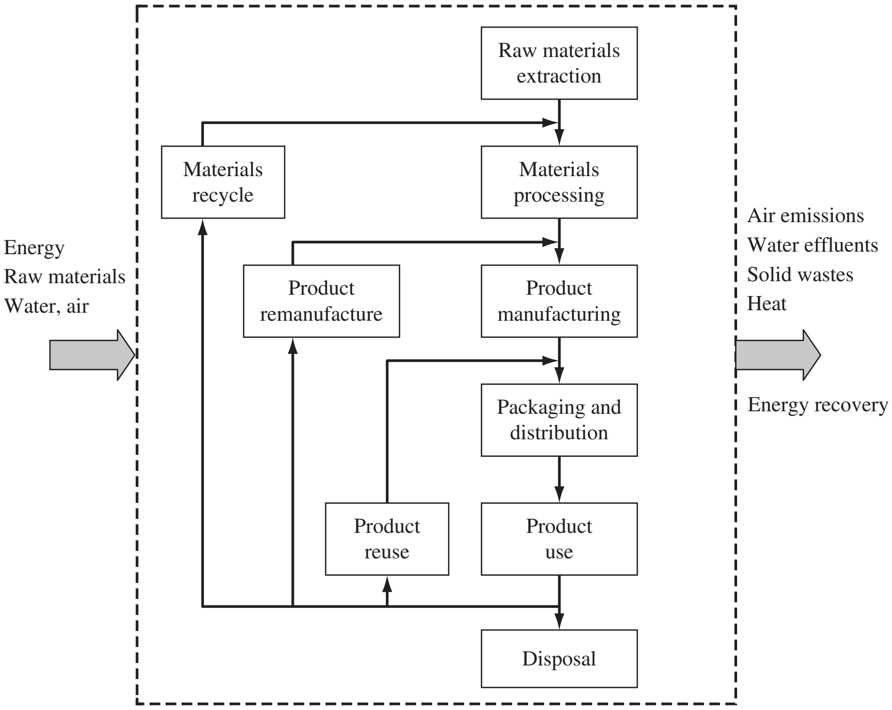


FIGURE 8 The lifecycle of a product, including clarification of the terms reuse, remanufacturing, and recycling.

Figure 8 also helps define some of the terms that are used to describe the potential recovery of materials in the product lifecycle.

> *Reusing* a product in the same application for which it was originally intended saves energy and resources. For example, a plastic bag can carry groceries home from the market over and over again and a polystyrene cup might be used several times before disposal. Returnable glass bottles for soft drinks are another example. A product can also be reused for some other purpose, such as occurs when glass jars are reused in a workshop to hold small objects such as screws or nails.
>
> *Remanufacturing* refers to the process of restoring a product to like-new condition. The restoration begins by completely disassembling the product, cleaning and refurbishing the reusable parts, and then stocking those parts in inventory. That inventory, along with new parts, is used to remanufacture products that are equal in quality to new units. Some distinguish between remanufacturing and repair. Repair means only those parts that have failed are replaced. For example, malfunctioning components on a faulty electronic circuit board might be replaced rather than throwing out the entire product.
>
> *Recycling* is the term used to describe the act of recovering materials from the wastestream and reprocessing them so they become raw materials for new applications.

An Example Lifecycle Assessment: Polystyrene Cups

As an example of the value (and difficulty) of performing a complete lifecycle assessment, consider a comparison of the environmental impacts of single-use, 8-ounce, hot-drink containers made from polystyrene foam with similar cups made from uncoated paper (Hocking, 1991). As shown in Table 6, the raw materials inputs for the two types of cups are very different. A paper cup requires about 21 g of wood and bark plus 1.2 g of chemicals to produce an 8.2-g cup, while a 1.9-g polystyrene

TABLE 6

Life-Cycle Assessment for Single-Use, Hot-Drink Cups (per 1,000 Cups)

	8.3 g Paper Cup	1.9 g Polyfoam Cup
Raw Materials		
Wood and bark (kg)	21	0
Petroleum feedstock (kg oil)	0	2.4
Other chemicals (kg)	1.2	0.08
Purchased Energy		
Process heat (kg oil)	1.8	1.9
Electricity (kg oil)[a]	2	0.15
Water Effluent		
Volume (m^3)	1	0.05
Suspended solids (g)	80	1
BOD (g)	90	0.4
Organochlorines (g)	20	0
Inorganic salts (g)	500	30
Fiber (g)	10	0
Air Emissions		
Chlorine (g)	2	0
Chlorine dioxide (g)	2	0
Reduced sulfides (g)	10	0
Particulates (g)	20	0.8
Carbon monoxide (g)	30	0.2
Nitrogen oxides (g)	50	0.8
Sulfur dioxide (g)	100	7
Pentane (g)	0	80
Ethylbenzene, styrene (g)	0	5
Recycle/Reuse Potential		
Reuse	Possible	Easy
Recycle	Acceptable	Good
Ultimate Disposal		
Proper incineration	Clean	Clean
Heat recovery (MJ)	170	80
Mass to landfill (kg)	8.3	1.9
Volume in landfill (m^3)	0.0175	0.0178
Biodegradability (landfill)	Yes	No

[a] Calculated using 33-percent efficient power plant burning residual fuel oil.

Source: Based on data in Hocking, 1991.

cup requires 2.4 g of petroleum feedstock plus 0.08 g of other chemicals. From the materials-input perspective, it is very difficult to compare the environmental impacts associated with logging with the impacts associated with drilling, transporting, and processing petroleum.

In terms of energy requirements, processing of raw wood, to pulp, to fully bleached kraft paper for the paper cup uses large amounts of process heat (steam) and electricity. If we assume that the steam and electricity are produced by burning residual fuel oil, a total of about 3.8 g of fuel per cup are required. Converting petroleum feedstock into polystyrene also uses significant quantities of heat and electricity. Crude oil or natural gas are first cracked in the presence of steam to yield ethylene or benzene, which are in turn catalyzed to produce ethylbenzene. Ethylbenzene is then thermally decomposed to styrene, which is converted to polystyrene. Hocking's analysis indicates that roughly the same amount of process heat per cup is required to make polystyrene or bleached kraft paper, but about 13 times as much electricity is required for the paper cup. Assuming that fuel oil is burned in a 33 percent-efficient power plant to produce the electricity needed, the paper cup uses roughly 80 percent more energy than the Styrofoam cup (3.8 g of oil equivalents per cup vs. 2.1 g). However, if we add in the petrochemical feedstock needed for the polystyrene cup, the paper cup uses roughly 15 percent less petroleum (3.8 g of oil vs. 4.5 g).

Most of the air and water effluents are considerably higher for the paper cup, although the comparison is complicated by the fact that the manufacturing processes for paper and polystyrene are so different. Production of polystyrene, for example, emits significant amounts of pentane (the foam blowing agent that has replaced chlorofluorocarbons), while paper production does not emit any. While pentane is not an ozone-depleting substance, it can contribute to photochemical smog. Again, it would be difficult to decide whether, for example, pentane emissions from the manufacture of polystyrene are more important than the higher criteria air pollutants emitted during paper production.

Consider the use, reuse, and recyclability of the two cups. The polystyrene cups are stiffer and stronger, especially when holding hot liquids, and their natural insulation helps keep the drinks hotter and the outside surface cooler. It is easy to imagine reusing the same polyfoam cup several times while at a party or conference, but the poor structural integrity of a paper cup—expecially when hot—makes it less likely to be reused. Both types of cups can be recycled, but the hot melt adhesive used in paper cups makes them somewhat less attractive.

At the disposal end of the lifecycle, if the cups are incinerated properly, they both burn cleanly and they both produce about the same amount of ash. If the incineration includes energy recovery, the paper cups yield roughly twice as much heat recovery per cup. However, since paper cups have four times the mass, paper cups yield about half as much energy per kilogram. If they are buried in a landfill rather than being burned in an incinerator, the polystyrene cup is compressed to roughly the same volume as the paper cup. The paper cups, however, are biodegradable (but very slowly), which means they can eventually take up less space, but their degradation produces a potent greenhouse gas, methane, and the liquids that filter through the landfill will contribute to the BOD of the leachate. If the landfill is well managed, the methane and leachate are easily controlled, but if it is poorly managed, it might be better to have the nonbiodegradable polystyrene to deal with.

As should be obvious by now, the paper versus polystyrene cup analysis illustrates how difficult it is to reach solid conclusions when performing a lifecycle assessment. Besides the many assumptions that must be made to generate any of the numerical comparisons, what conclusions can be drawn when the two approaches produce sorts of environmental impacts? Perhaps Hocking's conclusion that "it would appear that polystyrene foam cups should be given a much more even-handed assessment as regards their environmental impact than they have received during the past few years" (Hocking, 1991, p. 745) is about as strong a statement as can be made in this controversial debate.

Soft-Drink Containers

As another example of lifecycle analysis, consider the comparison of environmental impacts associated with a number of common soft-drink containers. The energy requirements, air and water emissions, and volume of solid waste for each container system are presented in Table 7. For almost every measure, refillable glass bottles cause the least environmental impact. Unfortunately, although they were common in the past, supermarkets and convenience stores have been switching to throwaways to avoid having to accept and store returnables (although they are still popular in Europe). By comparison, nonrefillable glass bottles use several times as much energy, and their emissions and solid waste requirements are correspondingly higher as well.

In the United States, plastic bottles and aluminum cans have become the containers of choice. Although not as benign as returnable bottles, from an energy standpoint both would seem to be better than one-way glass bottles. Unfortunately, making comparisons for air and water emissions for glass, polyethylene terephthalate (PET), and aluminum containers is more difficult than the table would suggest since the types of pollution for each will be considerably different.

TABLE 7

Environmental Impacts Associated with the Delivery of 1 Liter of Soft Drink in Various Containers

Type of Container	Energy (kJ/L)	Air Emissions (g/L)	Water Emissions (g/L)	Solid Waste (10^{-4} m^3)
0.47 L (16 oz) refillable glass bottle used 8 times	4,290	6.5	1.0	2.2
0.47 L (16 oz) **nonrefillable** glass bottle	9,700	18.2	2.0	6.7
1 L (34 oz) **nonrefillable** glass bottle	10,200	19.8	2.1	7.5
0.29 L (10 oz) **nonrefillable** glass bottle	11,600	22.0	2.4	8.1
0.47 L (16 oz) PET bottle	8,810	11.1	1.9	3.4
1 L (34 oz) PET bottle	7,110	8.9	1.6	2.6
2 L (68 oz) PET bottle	5,270	6.7	1.2	1.8
3 L (102 oz) PET bottle	5,180	6.5	1.2	1.7
0.35 L (12 oz) aluminum can	9,170	11.0	3.2	1.6

Source: Rhyner et al., 1995, based on data from Sellers, 1989.

7 Waste Management, Energy, and Greenhouse Gases

A lifecycle analysis of greenhouse gas (GHG) emissions must cover a long series of steps from extraction and processing of raw materials to manufacture of products, transportation of products to markets, consumer choices, waste collection, recycling or composting, and finally how we manage the discards that end up in landfills or incinerators. At every step, from extraction of ores to waste disposal, there are greenhouse gas (GHG) emissions that can be affected by solid waste management decisions. Most processes lead to GHG emissions, but some can help reduce them by increasing carbon stored in landfills or from increased biomass growth from composted organic materials. Figure 9 summarizes these possibilities, while the following list helps describe the factors:

- *Fossil fuel combustion emissions* when making, transporting, using, and disposing products that enter the wastestream.
- *Nonenergy-related manufacturing emissions,* such as the CO_2 released when limestone is converted to lime for concrete and steel manufacturing.
- *Methane* (CH_4) emissions from landfills.
- *Methane, carbon dioxide, and nitrous oxide* (N_2O) emissions from waste combustion.
- *Carbon sequestration* in materials that are stored for a long period of time in landfills.

Solid Waste Management and Greenhouse Gases

Translating emissions as well as carbon sinks into equivalent greenhouse gases requires a careful analysis of the energy inputs (for example, coal, oil, natural gas, electricity) as well as the types of greenhouse gases that are emitted (for example, CO_2, CH_4, N_2O). Recall that non-CO_2 greenhouse gases are weighted according to their Global Warming Potentials (GWPs). The 100-year GWP of methane, for example, is 23, which means 1 ton of methane is equivalent to 23 tons of carbon dioxide. The usual unit for a mixture of greenhouse gases is metric tons of carbon equivalents (MTCE). Notice MTCE is expressed as carbon and not carbon dioxide. Since the molecular weight of CO_2 is 44, and the atomic weight of carbon is 12, 1 metric ton of CO_2 is actually $12/44 = 0.273$ MTCE.

Table 8 shows some examples of potential greenhouse gas emission reductions associated with source reduction, recycling versus landfilling, and combustion versus landfilling. Source reduction values depend on whether the reduction offsets virgin materials or the current mix of virgin and recycled inputs to the manufacturing process. Negative values mean greenhouse emissions are greater than what would occur if the material were landfilled.

EXAMPLE 1 Recycling at the Los Angeles International Airport (LAX)

A recent study of recycling at the Los Angeles International Airport found that LAX generates about 19,000 tons of solid waste per year (1.3 pounds per passenger). On an annual basis, LAX recycles 12 tons of aluminum, 2,021 tons of

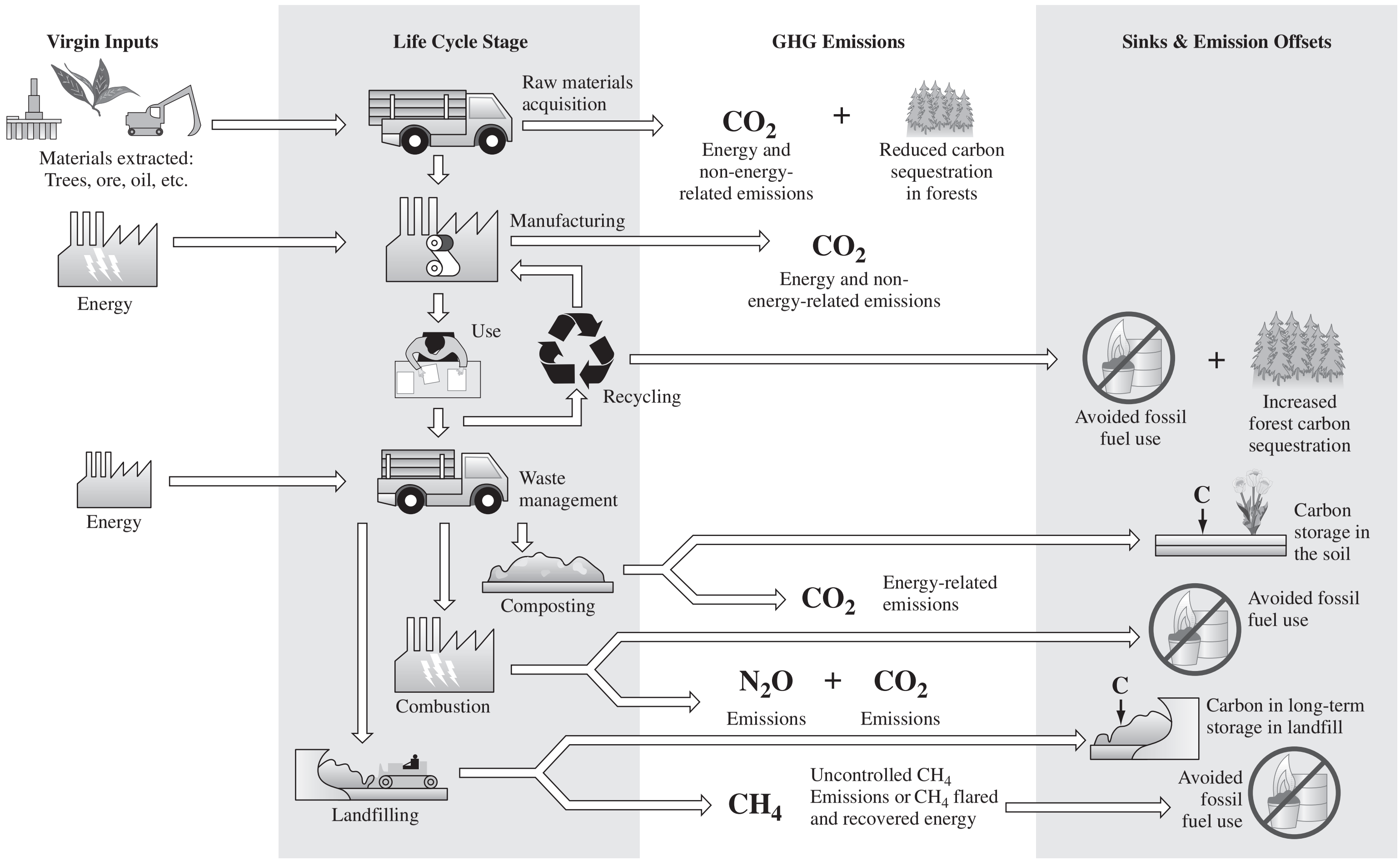


FIGURE 9 Greenhouse gas sources and sinks associated with material lifecycle.
(*Source:* U.S. EPA, 2006b.)

TABLE 8

Net Greenhouse Gas Emission Reductions Compared to Landfilling for Source Reduction, Recycling, Composting, or Combustion in Metric Tons of Carbon Equivalents Per Ton of Material (MTCE/ton)

	Source Reduction			
Materials	Current Mix of Inputs	100% Virgin Inputs	Recycling or Composting versus Landfilling	Combustion versus Landfilling
Aluminum cans	2.28	4.28	3.71	−0.01
Corrugated cardboard	1.63	2.32	0.96	0.29
Fly ash	0.01	0.01	0.25	0.01
Food waste composted	NA	NA	0.25	0.25
Glass	0.88	1.02	0.50	0.43
HDPE	0.50	0.55	0.39	−0.24
Magazines	2.28	2.36	0.76	0.05
Mixed metals	NA	NA	1.44	0.30
Mixed MSW	NA	NA	NA	0.15
Mixed paper	NA	NA	1.06	0.27
Mixed plastics	NA	NA	0.42	−0.26
Mixed recyclables	NA	NA	0.83	0.20
Newspapers	1.09	1.39	0.52	−0.03
Office paper	2.71	2.79	1.31	0.70
Personal computers	15.14	15.14	0.63	0.06
PET	0.58	0.60	0.43	−0.28
Steel cans	0.88	1.02	0.50	0.43
Textbooks	3.03	3.11	1.38	0.70
Tires	1.10	1.10	0.51	−0.04
Yard trimmings	NA	NA	−0.01	0.00

Source: U.S. EPA, 2006b.

cardboard, 527 tons of office paper, 89 tons of newspaper, 17 tons of glass, and 921 tons of plastic. In addition, they compost 271 tons of food waste (Atkins, 2006).

a. Find the equivalent greenhouse gas savings associated with these recycled and composted materials assuming they would have all gone to a landfill if they hadn't been recycled.

b. If the landfill charges $80/ton (called the *tipping fee*), how much money is saved in tipping fees by recycling and composting?

c. If, in the future, there is a carbon tax of $10 per metric ton of carbon dioxide, how much would LAX save in carbon taxes at the current recycling rate?

Solution

a. Using values from Table 8, the greenhouse gas savings are

12 tons aluminum × 3.71 MTCE/ton = 44.5 MTCE

2,021 tons cardboard × 0.96 = 1,940.2 MTCE

527 tons office paper × 1.31 = 690.3 MTCE
89 tons of newspaper × 0.52 = 46.3 MTCE
17 tons of glass × 0.50 = 8.5 MTCE
921 tons of plastic × 0.42 = 386.8 MTCE
217 tons of food waste × 0.25 = 54.3 MTCE
Total = 3,170 metric tons of equivalents

b. The total reduction in waste to be landfilled is

Saved tons = 12 + 2,021 + 527 + 89 + 17 + 921 + 217 = 3,804 tons
Saved landfill fees = 3,804 tons/yr × $80/ton = $304,320/yr

c. Total equivalent carbon reductions as $CO_2 = 44/12 \times 3{,}170 = 11{,}623$ metric tons of carbon dioxide per year. At $10 per metric ton of CO_2, the carbon tax savings would be

$$\text{Tax savings} = 11{,}623 \text{ metric tons } CO_2 \times \$10/\text{ton} = \$110{,}623/\text{yr}$$

Only about 20 percent of LAX generated waste is recycled or composted. For most airport waste, about two-thirds could be recycled or composted so the above savings of $304,320 + $110,623 = $414,943 per year could potentially triple with a more intensive recycling program.

Energy Implications of Solid Waste Management

Clearly, source reduction and recycling translate to less garbage going to incinerators and landfills, which means fewer air emissions and less water contamination. A complete lifecycle analysis, however, yields a complex array of environmental advantages, including energy savings, reduced water use, changes in the amounts and constituents of air emissions and water pollution, solid waste disposal reductions, habitat conservation achieved by avoiding raw materials extraction, and so forth. In this section, we want to focus on just the energy savings associated with solid waste management options.

Table 9 shows the net energy savings by source reduction, recycling, composting or incinerating various materials in municipal solid waste instead of sending them to a landfill. The *source reduction* values show energy savings for products that would have been manufactured from virgin materials instead of the current mix of recycled and virgin inputs. From the table, for example, by reading the news on the Web and avoiding 1 ton of newspaper, 36.87 million Btus (MMBtu) of energy would be saved if that source reduction is based on the current mix of virgin and recycled materials, and 41.1 MMBtus if only virgin materials are avoided. If old newspapers are *recycled* rather than being sent to a landfill, the savings is 16.91 MMBtu per ton of newspapers collected. That savings includes all of the energy required to collect, sort, and reprocess the materials versus deriving them from virgin sources. If old newspapers are *combusted* (with the average fraction of MSW energy recovery), the savings is 2.44 MMBtu/ton.

A graphical presentation of energy savings when products are recycled is shown in Figure 10. Aluminum recycling saves the most energy per ton, which is a reflection of the fact that manufacturing aluminum from virgin inputs is very energy

TABLE 9

Net Energy Reductions Compared to Landfilling for Source Reduction, Recycling, Composting, or Combustion (Units Are Million Btu Per Ton of Material [MMBtu/ton])

Materials	Source Reduction: Current Mix of Inputs	Source Reduction: 100% Virgin Inputs	Recycling or Composting versus Landfilling	Combustion versus Landfilling
Aluminum cans	126.71	239.41	206.95	0.11
Corrugated cardboard	22.13	26.99	15.65	2.44
Fly ash	NA	NA	5.29	NA
Food waste composted	NA	NA	−0.25	0.88
Glass	8.06	8.62	2.65	0.15
Magazines	33.62	33.66	1.09	1.98
Mixed metals	NA	NA	75.33	12.57
Mixed MSW	NA	NA	0.28	1.76
Mixed paper	NA	NA	23.19	2.47
Mixed plastics	NA	NA	52.94	5.62
Mixed recyclables	NA	NA	17.21	2.97
Newspapers	36.87	41.1	16.91	2.96
Office paper	36.59	37.28	10.09	2.14
Personal computers	957.27	957.27	43.96	5.22
PET	71.2	73.24	53.36	3.69
Textbooks	35.31	35.34	0.54	2.14
Tires	88.70	88.70	52.49	27.23
Yard trimmings	NA	NA	−0.17	1.11

Source: U.S. EPA, 2006b.

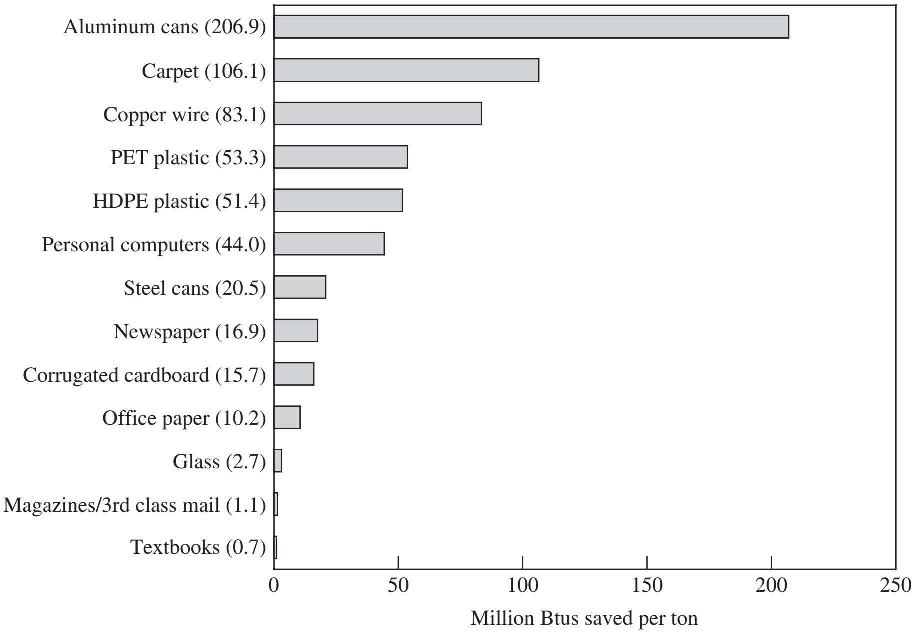


FIGURE 10 Energy savings per ton of recycled materials.
(*Source:* U.S. EPA, 2006b.)

intensive, whereas relatively little energy is required to manufacture cans from recycled aluminum. Recycling carpet also results in significant energy savings since the recycled material is turned into secondary products whose manufacture would have required energy-intensive processes.

EXAMPLE 2 **Energy Savings with LAX Recycling**

Calculate the energy savings associated with the recycling operations at Los Angeles International Airport using recycling data provided in Example 1. Compare that energy savings with the cost of natural gas at \$10 per million Btu.

Solution Combining the recycling rates with the energy data in Table 9 gives

12 tons aluminum × 206.95 MMBtu/ton = 2,483 MMBtu
2,021 tons/yr carboard × 15.65 = 31,628 million Btu/yr
527 tons office paper × 10.09 = 5,317
89 tons of newspaper × 16.91 = 1,504
17 tons of glass × 2.65 = 45
921 tons of plastic × 52.94 = 48,757
217 tons of food waste × (−0.25) = −54
Total = 89,680 million Btu/yr

If that energy costs \$10 per million Btu, this simple recycling program at one airport saves society almost \$900,000 per year.

8 Recycling

After source reduction, which is given the highest priority in the solid waste management hierarchy, the recovery of materials for recycling and composting is generally thought to be the next most important component of integrated solid waste management programs. Resource recovery means that the materials have not only been removed from the municipal wastestream, but they also must be, in essence, purchased by an end user. That distinction, for example, means yard trimmings composted at home are considered to be source reduction, not recycling. On the other hand, yard trimmings that are delivered to an offsite composting facility and then sold are considered to be recovered or recycled materials.

The term *recycling* is often misconstrued to include activities such as refilling bottles for reuse and remanufacturing products for resale to consumers, but it is better to use the term only when materials are collected and used as raw materials for new products. The process of recycling includes collecting recyclables, separating them by type, processing them into new forms that are sold to manufacturers, and, finally, purchasing and using goods made with reprocessed materials.

Another distinction that can be made is between *preconsumer* and *postconsumer* recyclable materials. Preconsumer materials consist of scrap that is recycled by manufacturers back into the original manufacturing process without ever having been turned into a useful product. Postconsumer recyclables are products that have already been used by consumers for their originally intended purposes, such as newspaper or plastic bottles, that are now part of the wastestream. Preconsumer recycled materials are not included in EPA data on generation, recovery, and discards, and they generally are not counted in procurement laws that mandate government purchases of products with significant recycled-materials content.

Fine distinctions between terms such as *resource recovery, reuse, remanufacturing, recyclable, preconsumer* and *postconsumer recycled materials,* and so forth may seem annoying or even unnecessary, but they are important when systematic collection, reporting, and unambiguous interpretation of recycling data are needed.

The rate of recovery of materials from the wastestream, both for recycling and composting, as reported by the EPA, has increased dramatically in the past few decades. Resource recovery grew from less than 10 percent of all municipal solid waste in 1980 to 16 percent in 1990 to 32 percent in 2005. Most of that increase is attributable to greater rates of recovery of paper and paperboard and increased composting of yard trimmings. Table 10 shows U.S. recovery rates for a number of categories of MSW. Notice how dominant the paper and paperboard category is in terms of total tons of material recovered (53 percent of total recovery), which tends to mask the importance of recovery rates for other materials. For example, recovered aluminum is less than 1 percent of the total mass of recovered materials, but in terms of its total economic value, it far exceeds the paper products category.

TABLE 10

Generation and Recovery of Materials in MSW in the United States, 2005 (Millions of Tons Per Year)

Material	Weight Generated	Weight Recovered	Recovery as a Percent of Generation (%)
Paper and paperboard	84.0	42.0	50.0
Glass	12.8	2.76	21.6
Metals			
Steel	13.8	4.93	35.8
Aluminum	3.21	0.69	21.5
Other nonferrous metals	1.74	1.26	72.4
Total Metals	18.7	6.88	36.8
Plastics	28.9	1.65	5.7
Rubber and leather	6.70	0.96	14.3
Textiles	11.1	1.70	15.3
Wood	13.9	1.31	9.4
Other materials	4.57	1.17	25.6
Total Materials in Products	180.7	58.4	32.3
Other wastes			
Food, other	29.2	0.69	2.4
Yard trimmings	32.1	19.9	61.9
Miscellancous inorganic wastes	3.69	Neg.	Neg.
Total Other Wastes	65.0	20.6	31.6
TOTAL MUNICIPAL SOLID WASTE	245.7	79.0	32.1

Source: U.S. EPA, 2006a.

EXAMPLE 3 Carbon and Energy Savings in Recycled Aluminum

Find the carbon and energy savings associated with the 0.69 million tons of aluminum recycled each year. For perspective, compare the energy savings with the energy needs for a typical 25 mpg car driven 12,000 miles per year burning 0.125 million Btu per gallon gasoline. Also compare the carbon emissions if the car emits 4.2 metric tons of CO_2 per year.

Solution From Table 8 we find that recycled aluminum saves 3.71 metric tons of carbon equivalents per ton:

$$\begin{aligned}\text{Carbon savings} &= 0.69 \times 10^6 \text{ tons/yr} \times 3.71 \text{ MTCE/ton}\\ &= 2.56 \times 10^6 \text{ MTCE/yr}\\ \text{CO}_2 \text{ savings} &= 2.56 \times 10^6 \text{ MTCE/yr} \times \frac{44 \text{ tons CO}_2}{12 \text{ tons C}}\\ &= 9.38 \times 10^6 \text{ mt CO}_2\text{/yr}\end{aligned}$$

At 4.2 metric tons of CO_2 per year per car, this is equivalent to the emissions from

$$\text{Car equivalents (carbon)} = \frac{9.38 \times 10^6 \text{ mt CO}_2\text{/yr}}{4.2 \text{ mt CO}_2\text{/car}} = 2.2 \text{ million cars}$$

For energy savings, Table 9 indicates aluminum recycling saves 206.95 million Btu/ton.

$$\begin{aligned}\text{Energy savings} &= 0.69 \times 10^6 \text{ tons/yr} \times 206.95 \times 10^6 \text{ Btu/ton}\\ &= 143 \times 10^{12} \text{ Btu/yr}\end{aligned}$$

$$\text{One car} = \frac{12{,}000 \text{ miles/yr}}{25 \text{ miles/gallon}} \times 0.125 \times 10^6 \text{ Btu/gallon} = 60 \text{ million Btu/yr}$$

$$\text{Car equivalents (energy)} = \frac{143 \times 10^{12} \text{ Btu/yr}}{60 \times 10^6 \text{ Btu/yr/car}} = 2.4 \text{ million cars}$$

So, the simple recycling of only 21.5 percent of our aluminum is equivalent to taking over 2 million cars off the road in terms of both carbon and energy savings.

Paper and Paperboard Recycling

The largest single category of waste generation and waste recovery is paper and paperboard products. In 2005, 84 million tons of paper products were generated and 50 percent of that was recovered. Two types of paper products, corrugated cardboard boxes and newspapers, together account for half of the paper products generated and three-fourths of paper products recovered.

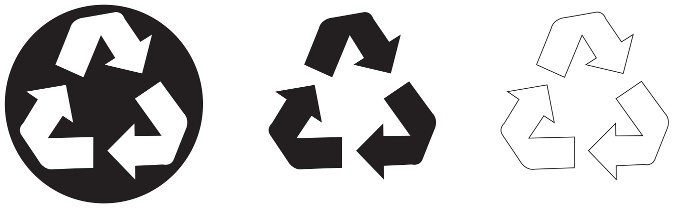
The Chinese invented paper 2,000 years ago when they discovered that vegetable fibers, such as bamboo or mulberry, when soaked in water formed mats of fibrous material that bonded itself together and when dried could be used as a writing medium. Soon after, their process shifted to use of wood as the principal fiber source. The U.S. papermaking industry began in 1860 using recycled fiber from cotton, linen rags, and waste paper, but now virgin paper production is based almost entirely on pulps made from wood chips. Other sources of fiber can include straw, jute, flax, bagasse from sugar cane, and hemp—in fact, there is quite a surge of interest these days in legalizing hemp as a source of fiber to reduce the need for forest products.

Papermaking begins with pulping. For quality paper, the cellulose fibers in wood chips must be separated from the lignins that hold the fibers together. The process involves cooking, to soften the wood and chemical or mechanical treatment to help separate the fibers. Most wood pulp is prepared using a chemical process. For low-quality paper, such as newsprint, whole chips of wood can be pulped mechanically, without removing the lignins, creating a product called *groundwood*. Pulp is naturally tan in color, so chemical bleaching may be required if the end product needs to be white. If the finished pulp needs to be transported some distance to the papermill, it will be dried and baled; otherwise, it is pumped as a thick slurry to the mill.

Recycled wastepaper has long been used as a source of fiber to augment pulp made from virgin materials. Depending on the application, various fractions of recycled and virgin materials can be incorporated—all the way up to 100 percent recycled. Each recycling of wastepaper, however, results in losses so it is not possible to use the same fiber over and over again. Each time paper is soaked in water, heated, and processed, some of the cellulose breaks down into starches and is lost; also, in each processing, some of the smaller fibers don't bond together, and they are lost as well. Not only is less fiber made available after each recycling, but the fibers are shorter and the resulting product is not as strong as product made from virgin materials. Nonetheless, there are abundant uses for recycled wastepaper.

Roughly half of U.S. recycled paper products are used to make paperboard (cereal boxes, tablet backs, shoe boxes, and so on) and containerboard (corrugated cardboard). For such applications, product whiteness is not important. When paper products are to be recycled into newsprint or other papers, an additional de-inking stage must be added to the processing, which increases costs. Recycled wastepaper is also used to make various construction products such as cellulose insulation, fiberboard, and some flooring materials. Other applications include animal bedding, nursery pots, garden mulch, and trays.

The standard recycling symbol, used worldwide, is the familiar triangle of arrows that form a Möbius strip (Figure 11). For paper products, when the arrows



Recycled content   Recyclable   Recyclable

FIGURE 11 Labels on paper products with the Möbius arrows contained within a circle means the product is made from recycled materials. When the arrows are freestanding, the product has been designed to be recyclable after use.

stand alone, whether colored or not, it means the product is recyclable. When the arrows are contained within a circle, it means the product is made from recycled materials.

The economic value of recycled paper very much depends on the characteristics of the paper collected. In fact, the Paper Stock Institute has defined 51 grades of recycled paper and paper products, and each category has its own pricing structure when sold. Prices for recycled paper products have traditionally been volatile, depending on ever-changing market conditions. For example, the price of old corrugated containers (OCC) jumped from $30/ton in 1993, to $200/ton in 1996, but fell back again to $80/ton in 1998. In 2005, OCC was selling for about $90/ton.

Recycling Plastics

The rate at which plastics enter the municipal solid wastestream in the United States has grown phenomenally from 0.4 million tons per year in 1960 to 18 million tons in 1993, and 29 million tons in 2005. Plastics now comprise almost 12 percent of the weight of all municipal solid waste generated, but since they are so bulky, they account for one-fourth of the total volume in landfills. Although all plastics are theoretically recyclable, less than 6 percent are actually recovered.

The term *plastics* encompasses a wide variety of resins or polymers with different characteristics and uses. Plastics are produced by converting basic hydrocarbon building blocks such as methane and ethane into long chains of repeating molecules called *polymers*. While there are natural polymers such as cellulose, starch, and natural rubbers, it is only the synthetic materials that are referred to as plastics. One way to categorize plastics is based on their behavior when heated. Roughly 80 percent are *thermoplastics,* which can be remelted and remolded into new products. The other 20 percent are *thermoset* plastics, which decompose when they are heated, so they are more difficult to recycle.

The Plastic Bottle Institute of the Society of the Plastics Industry has developed a voluntary coding system for plastic bottles and other rigid containers, which uses the familiar three arrows in a triangle, this time with a numerical designation that indicates the type of resin used in the plastic. Table 11 describes the labeling system and highlights the main uses of the designated plastics.

Although the three-arrow symbol is usually thought to mean the product is recyclable, in the case of plastics it is really only polyethylene terepthalate (PET or PETE) and high-density polyethylene (HDPE) that are recycled to any significant degree. In fact, while soft-drink bottles (PET) and milk and water bottles (HDPE) make up less than 6 percent of the total plastic products generated in MSW, they account for almost two-thirds of all recycled plastic. PET provides a better barrier to transmission of gases than HDPE, which makes it a better container for carbonated beverages and other foods whose aromas need to be contained. PET also fetches a much higher price in the recycling market.

When plastics are recycled, it is often important to carefully separate them by resin type and color. Different resins have different melting points, so if a batch of mixed plastics is being heated and reformed into new products, some resins may not melt at all, and some may end up burning. A single PVC bottle in a batch of 20,000 PET bottles, for example, can ruin the whole batch and potentially damage the manufacturing equipment (Moore, 1994). Similarly, if polypropylene, which is difficult

TABLE 11

Types of Resins and 1993 Generation and Recycling Rates

Designation	Polymer	Applications	Generation (1,000 tons/yr)	Recovery
1 PETE	Polyethylene terepthalate (PET)	Soft drink bottles, peanut butter jars; frequently recycled	2,860	19%
2 HDPE	High-density polyethylene	Milk, water, juice, detergent, motor oil bottles; often recycled	5,890	9%
3 V	Vinyl/polyvinyl chloride (PVC)	Cooking oil bottles, credit cards, household food wrap, building materials; rarely recycled, toxic if burned	1,640	0%
4 LDPE	Low-density polyethylene	Shrink wrap, garbage and shopping bags, lids, cups; sometimes recycled	6,450	3%
5 PP	Polypropylene	Snack food wrap, margarine tubs, straws, car batteries; occasionally recycled	4,000	0.3%
6 PS	Polystyrene	Pharmaceutical bottles, styrofoam cups, clear salad bowl lids, packing peanuts; rarely recycled	2,590	0%
7 Other resins		Multilayered, mixed material containers, squeezable bottles; rarely recycled	5,480	0.01%
TOTALS			28,910	5.7%

Source: Generation and recycling data from U.S. EPA, 2006a.

to distinguish from polyethylene, contaminates a batch of recycled polyethylene, the resulting blend would be useless. Plastic recycling is also complicated by the potential for contamination by the products that they once contained. Plastics do absorb, to varying degrees, materials that they come in contact with. Substances such as pesticides and various oils will migrate slowly into their plastic containers, and some will survive the melting and reforming processes when recycled. For that reason, the Food and Drug Administration keeps a careful eye on the use of recycled plastics for direct-food-contact applications.

Recycled plastics find application in a number of products. Recycled PET bottles, for example, are being transformed into quality fibers for jackets and other outdoor clothing, as well as fiberfill material for cushions. They are also used to make carpeting, surfboards, sailboat hulls, as well as new PET soft drink bottles. To avoid the contamination problem mentioned previously, recovered PET used in new soft drink bottles is usually sandwiched between an inner and outer layer of virgin plastic. HDPE and LDPE are often recycled into new detergent bottles, trash cans, and drainage pipes. For some of the most popular uses for recycled HDPE and LDPE, minor color variations in the final product are not particularly important. For other products, such as colored detergent bottles, variations in color would affect consumer perceptions of the product, so presorting of recyclables by color can be important.

While it is usually important for recovered plastics to be sorted carefully, some applications are emerging that can use commingled plastics. Mixtures of plastics can be shredded, melted, and extruded into useful forms. Plastic lumber, which can be used for fence posts, siding, timbers, park benches, docks, and so forth, is formed in this way. Plastic lumber is more expensive than wood lumber, but it resists insect and weather damage better than wood. In countries such as Japan, which have little timber, plastic lumber is enjoying a growing niche in the construction market.

PVC in the wastestream poses particular problems. As already mentioned, small amounts can cause major contamination problems for recycled PET. It is also a source of potentially serious air pollution problems if it ends up being burned. At elevated temperatures, the chlorine in PVC combines with hydrogen to form toxic hydrogen chloride, which can be emitted into the atmosphere. To help avoid these problems, a new mechanical separation system has been devised that automatically separates PVC from other plastics at the recycling facility. Whole or crushed plastic bottles are moved on a conveyor belt past a sensor that detects the chlorine atoms in PVC. When PVC is detected, air jets are triggered that kick the PVC bottles away from the remaining containers.

Glass Container Recycling

The amount of glass in the municipal wastestream is one of the few materials that has shown a decrease over time. In 1980, 15 million tons of glass entered the wastestream, but by 1990, it had dropped to less than 13 million tons, which is where it has remained through 2005. The impact of much lighter-weight aluminum cans and plastic drink bottles is evident. Even at these lower current rates, that still amounts to about 100 pounds of glass generated per person per year. The glass recovery rate in the United State is about 22 percent, most of which is made into new glass containers.

Of the glass used in the Unitd State, 85 percent is for bottles and jars of one sort or another. The remaining 15 percent, referred to as *durable goods,* consists of items such as window glass, fiberglass, mirrors, ceramic dishes, porcelain, glassware, ovenware, and light bulbs. The recycling rate for those durable goods is essentially zero, and, in fact, they are considered to be contaminants in container glass recycling. Ceramics and heat-resistant cookware melt at much higher temperatures than container glass, so they form solid or slightly melted chunks of impurities when the two are mixed together in a glass furnace. Metal caps and rings left on bottles also cause problems when they are mixed in with glass. Metals melt at lower temperatures than glass, and that molten metal can cause expensive corrosion problems in the glass furnace itself.

Recovered glass that will be remade into new container glass needs to be separated by color: clear (flint), green (emerald), or brown (amber). Those colors have traditionally been created with permanent dyes, but there are now colored polymer coatings that can be applied to the outside of clear glass containers. If these coatings would allow all glass used in containers to be clear, the need to separate by color as well as the need to use care to prevent bottle breakage during collection (which mixes bits and pieces of different colored glass together in the residue) would be eliminated. An additional advantage of the polymer coatings is they make the container considerably stronger, which could allow bottles to be made even lighter than they are now.

Once recovered glass has been separated by color, and the metal rings and caps removed, it needs to be crushed into smaller pieces, called *cullet*. When significant quantities of color-separated cullet are accumulated, it is usually shipped to container manufacturers to be remade into new glass bottles and jars. Glass has the unusual property of being 100 percent recyclable; that is, the same bottle can be melted down and remade over and over again without any product degradation. Moreover, cullet melts at lower temperature than the raw materials from which new glass is made (silica sand, soda ash, and limestone) which helps save energy.

Not all cullet is remade into containers. Some is used in a mix with asphalt to form a new road paving material called *glasphalt*. Cullet can also be used as a road base material in areas where suitable supplies of gravel may be limited. Other products that can be made using cullet include fiberglass, abrasives, reflective paint for road signs (made from small glass beads), lightweight aggregate for concrete, glass polymer composites, and glass wool insulation.

Aluminum Recycling

By some measures, the most valuable collectable in a municipal recycling program is aluminum cans. A ton of aluminum is typically worth 10 times as much as a ton of PET or HDPE, and it generates on the order of 20 times as much revenue per ton as glass, steel cans, or newspapers. Even though it may comprise only 2 or 3 percent of the total tonnage of collected recyclables, it often generates on the order of 20 to 40 percent of the total revenues. This economic incentive is directly attributable to the fact that recycled aluminum uses only 2 to 3 percent of the energy required to make new aluminum from bauxite ore.

In 2005, aluminum beer and soft drink cans made up 60 percent of the 3.2 million tons of aluminum in municipal solid waste. Almost half of those cans, roughly 60 billion per year, are recovered from the wastestream and recycled back into the production of aluminum. It is estimated that 95 percent of recovered aluminum goes back into the production of the next generation of cans, with a turnaround time of as little as 90 days (Bukholz, 1993). The next biggest source of aluminum in MSW is from durable and nondurable goods such as home appliances and furniture, with almost none of that being recovered for recycling.

Recycling aluminum not only helps reduce the demands placed on landfills, but doing so saves quite large amounts of energy as well. The production of new aluminum begins with the mining of bauxite ore. Then, using a procedure known as the Bayer process, impurities are removed, and the aluminum hydrates and hydroxides in bauxite are converted to aluminum oxide, or alumina. Alumina is then electrolytically reduced to aluminum metal. That electrolysis, known as the Hall-Heroult process, consumes enormous amounts of electricity. On the average, the energy required by the two-step conversion of bauxite to aluminum metal requires 220,600 kJ/kg, while only 5,060 kJ/kg are needed to make it from recycled aluminum—an energy savings of 98 percent (Table 12).

TABLE 12

Energy Requirements for Aluminum Produced from Bauxite and from Recycled Aluminum

Source	Electricity (kWh/kg)	Fossil Fuel (kJ/kg)	Primary Energy[a](kJ/kg)
Bauxite	15	60,000	220,600
Recycled aluminum	0.08	4,200	5,060
Recycling savings	14.92	55,800	215,540
Percent savings	99%	93%	98%

[a] *Source:* Based on data in Atkins et al., 1991, adjusted for 10,150 Btu/kWh heat rate.

EXAMPLE 4 **Energy for the Aluminum in a Can**

a. Using the current 50 percent recycling rate for aluminum cans, find the primary energy required to produce the aluminum in a 0.36 L (12 oz) aluminum can having a mass of 16 g.

b. How much energy is saved when a can is recycled instead of being thrown away?

c. What is the equivalent amount of gasoline wasted when one can is thrown away (gasoline has an energy content of about 35,000 kJ/L).

Solution

a. The primary energy needed for the 50 percent of the can made from bauxite is

$$\text{Primary energy (virgin)} = \frac{16\text{ g/can} \times 0.50\text{ virgin} \times 220{,}600\text{ kJ/kg}}{1{,}000\text{ g/kg}}$$
$$= 1{,}765\text{ kJ/can}$$

and the primary energy needed to produce the rest from recycled aluminum will be

$$\text{Primary energy (recycled)} = \frac{16\text{ g/can} \times 0.50\text{ recycled} \times 5{,}060\text{ kJ/kg}}{1{,}000\text{ g/kg}}$$
$$= 40\text{ kJ/can}$$

So the total for an average can with 50-percent recycled aluminum is

$$\text{Primary energy} = 1{,}765 + 40 = 1{,}805\text{ kJ/can}$$

b. Every time an aluminum can is not recycled, a new can must be made from bauxite to replace it. The energy to produce the aluminum for a new can from ore is

$$\text{Primary energy (all virgin)} = \frac{16\text{ g/can} \times 220{,}600\text{ kJ/kg}}{1{,}000\text{ g/kg}} = 3{,}530\text{ kJ/can}$$

If the can is recycled, it can be melted down and made into the next new can, which requires

$$\text{Primary energy (all recycled)} = \frac{16 \text{ g/can} \times 5{,}060 \text{ kJ/kg}}{1{,}000 \text{ g/kg}} = 81 \text{ kJ/can}$$

The energy saved by recycling is therefore

$$\text{Energy saved} = 3{,}530 - 81 = 3{,}449 \text{ kJ/can}$$

c. The equivalent amount of gasoline thrown away when a can is not recycled is

$$\text{Gasoline equivalent} = \frac{3{,}449 \text{ kJ/can}}{35{,}000 \text{ kJ/L of gasoline}} = 0.099\text{L} = 98 \text{ mL}$$

Since the can holds 355 mL, that is like throwing away almost one-third of a can of gasoline every time a can ends up at the dump.

The analysis in Example 4 tells us the amount of energy needed to produce the aluminum for one can, given the current recycling rate. It does not tell us the total amount of energy required to make an aluminum can, however. As shown in Figure 12, once an ingot of aluminum is cast, more energy is required to roll it into sheets, to transport those sheets to the can manufacturing plant, to manufacture the cans, and to transport used cans back to the plant that remelts the aluminum. Table 13 provides an estimate that shows about 3,188 kJ of energy are invested in a single 12 oz can, which is 75 percent more than the energy required to produce the aluminum itself.

Other Metals

Metals can be conveniently classified into two categories: ferrous and nonferrous. Nonferrous metals (aluminum, copper, lead, zinc) have little or no iron in them, while ferrous metals do. In keeping with EPA definitions, the only metals we will consider here are those that end up in the municipal solid wastestream as discarded refrigerators, stoves, water heaters, and so forth. That excludes, for example, preconsumer scrap that is recovered during metal manufacturing, and it also excludes old automobile and truck bodies that are crushed and recycled.

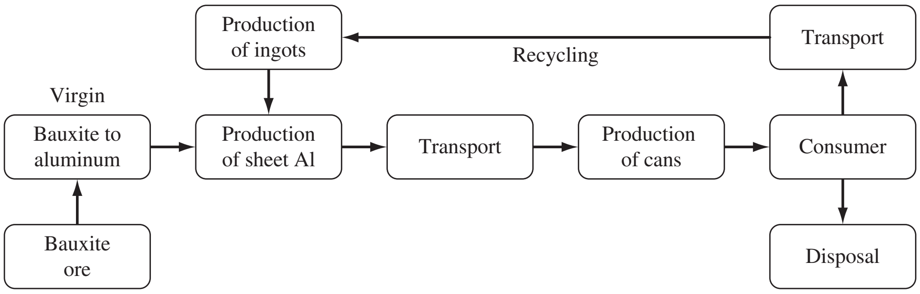


FIGURE 12 A flow diagram for aluminum cans produced from primary materials and recycled materials.

TABLE 13

Energy Required to Make a Single 16-g (12 oz) Aluminum Can[a]

Process	Energy (kJ)
Bauxite to make 8 g of aluminum metal	1,765
Recycled cans to make 8 g of aluminum	40
Production of 16 g sheet from ingot	421
Transport of 16 g sheet to can manufacturer	4
Production of a single 16 g can	955
Transport 8 g of recycled cans to foundry	3
TOTAL	3,188

[a] Assumes 50 percent recycling rate.

Source: Based on Example 4 and International Petroleum Industry Environmental Conservation Association, 1992.

In 2005, 18.7 million tons of metals entered the municipal solid waste system. Ferrous metals account for the largest chunk of that total, 74 percent. Aluminum is the second largest at 17 percent, and the remaining 9 percent are other nonferrous metals. All told, 37 percent of metals generated were recovered from the wastestream. Renewed emphasis on ferrous metal recovery has led to a rapid rise in white goods (appliances) recycling, which is now about 90 percent. About half of the steel and aluminum in packaging is now recovered. The recovery rate of "other nonferrous" metals is quite high, averaging 72 percent. That category is dominated by lead, most of which is in car batteries, with 98 percent of that battery lead being recovered.

Construction and Demolition Debris

Construction and demolition (C&D) debris is generated when new structures are built and when existing structures are renovated or demolished. While the exact definition of C&D debris varies from state to state, it usually consists of the residues from residential and nonresidential building sites, including wood, metals, gypsum wallboard, roofing, concrete, asphalt. It often includes trees, stumps, earth, and rock from the clearing of all sorts of construction sites, including public works projects, such as streets and highways.

C&D debris is a wastestream that is considered to be separate from MSW even though a significant fraction of this material ends up in MSW landfills. The EPA does not track these wastes nearly as carefully as MSW, so consistent data are hard to come by. One report, written in 1998, suggests an annual generation rate of about 2.8 pounds per capita, which can be compared with the current 4.5 pounds of conventional MSW (U.S. EPA, 1998). Table 14 shows a distribution of generation sources, with residential debris accounting for 43 percent of the total, and nonresidential debris the other 57 percent. Most of the debris from residential sources is generated during renovation; for nonresidential buildings, it is mostly demolition.

Typical construction waste for a new, detached, single-family, 2,000-square-foot house is shown in Table 15. Most of building construction waste is relatively clean, unmixed, nonhazardous material that is easily disposed of without harm to

TABLE 14

Estimated Building-Related C&D Debris Generation, 1996[a]

	Residential		Nonresidential		Totals	
Source	Million tons/yr	Rate lbs/sq.ft.	Million tons/yr	Rate lbs/sq.ft.	Million tons/yr	Percent
New construction	6.6	4.38	4.3	3.9	10.8	8
Renovation	31.9	NA	28.0	NA	59.9	44
Demolition	19.7	115	45.1	155	64.8	48
Totals	58.2		77.4		135.5	100
Percent	43%		57%		100%	

[a] Excludes roadway, bridge, and land clearing.
Source: U.S. EPA, 1998; U.S. DOE, 2006.

TABLE 15

Typical Construction Waste for a New 2,000-Square-Foot Home

	Weight (Pounds)	Percent
Solid sawn wood	1,600	20
Engineered wood	1,400	18
Drywall	2,000	25
Cardboard (OCC)	600	8
Metals	150	2
Vinyl (PVC)	150	2
Masonry	1,000	13
Hazardous materials	50	1
Other	1,050	13
Total	8,000	100

Source: U.S. DOE, 2006.

TABLE 16

Estimated Management of Building-Related Construction and Demolition (C&D) Debris in the United States, 1996

Management Option	Million tons/yr	Percent of Total
Recovered for recycling	25–40	20–30
C&D landfills	45–60	34–45
MSW landfills and other	40–45	30–40
Totals	136	100

Source: U.S. EPA, 1998.

soil or groundwater. As much as 95 percent of it can be recycled, but historically only about one-fourth of it actually does get reused or reprocessed into new materials (Table 16). That low rate is starting to change, in large part because of the green building industry. The U.S. Green Buildings Council created its own building rating system called *Leadership for Energy and Environmental Design* (better known as *LEED*), which recognizes buildings that have outstanding environmental attributes. One portion of their rating system awards points for management of construction waste (1 point for 50 percent recycling, 2 points for 75 percent, and 3 points for 90 percent).

9 Collection and Transfer Operations

After source reduction, and before recycling, comes the process of actually collecting solid wastes and transporting them to resource recovery facilities, landfills, or incinerators. This collection and transportation of solid wastes accounts for roughly three-fourths of the total cost of refuse service. In the past, collection and transport decisions by professionals involved in solid waste management were focused on selecting the proper number and size of trucks, choosing the most efficient collection routes and schedules, locating transfer stations if they were to be used, and administering the whole system. With the growing importance of recycling and composting, those basic operations have become more complicated. Now, a municipality may have separate trucks, routes, schedules, and destinations for recyclables and compostable materials—all of which need to be coordinated with already existing refuse collection systems.

A number of collection options are available for communities with recycling programs. In some areas drop-off centers, typically located in convenient places such as supermarket parking lots, accept recyclables that the consumer delivers in person. These may provide the only recycling service in a community, which is often the case for rural areas, or they may supplement curbside collection. Curbside collection systems vary considerably, and the collection vehicles that are used depend on the degree of sorting that residents are expected to perform. It has been common for a separate truck for recyclables to make its own collection run, but cocollection systems in which a single truck picks up both recyclables and residential solid waste are beginning to appear more often.

Collection of MSW

Conventional collection vehicles for municipal solid waste are distinguished by their size, how they are loaded, and how much compaction of refuse they can accomplish. Trucks designed for residential refuse service are usually rear-loaded or side-loaded. Rear-loaded trucks more readily accommodate large, bulky items, and they tend to be able to provide greater refuse compaction, while side-loaded trucks tend to cost less and are often more appropriate in densely populated areas where collection takes place on only one side of the street at a time. A range of sizes for either type of collection truck is available, from roughly 6 to 40 cubic yards, but the most common capacity for municipal service is typically 16 to 28 yd^3. Obviously, larger trucks cost more, but they don't have to make as many trips back and forth to the disposal site, which can more than offset the higher capital costs. Larger trucks, however, are also less maneuverable in crowded urban areas, and their weight may exceed allowable limits for residential streets.

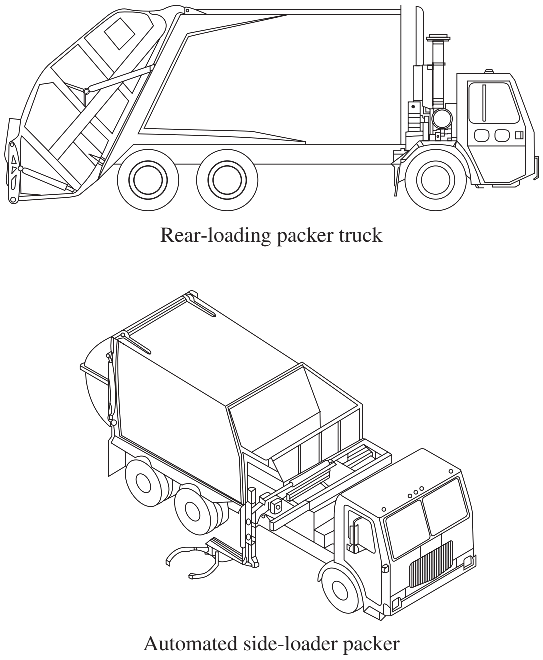


FIGURE 13 Examples of a rear-loading packer truck and a side-loading truck with an automated arm to pick up trash containers.

Municipal solid waste at the curbside has a density of around 100 to 200 lb/yd^3. At those low densities, collection vehicles fill too fast, which means multiple, time-wasting trips to the disposal site would be needed. Modern trucks, called *packers,* have hydraulic compactors that can compress that waste to as much as 1,500 lb/yd^3. The ratio of the density of waste in the truck to the density at curbside is known as the compaction ratio. Ratios of 2.5 to 3.5 were the norm in the recent past, but newer generations of trucks are being offered with 8-to-1 compaction. Packers come with a variety of loading options, including front-load, rear-load, and fully automated side-loading packers capable of picking up cans at the curb (Figure 13). Those side-loading packers may cost more, but the reduced rate of injuries to crews who no longer have to lift heavy containers all day long can more than offset the extra capital cost of the trucks.

Micro-Route Planning

Designing a route for collection vehicles to travel in order for every house to be served with a minimum of deadheading (traveling without picking up refuse) is a challenging task. While it is possible to analyze this problem mathematically for an idealized neighborhood, the realities of one-way streets, dead ends, hills, traffic, location of transfer stations, and other complications can make this approach too difficult for practical implementation. Instead, a common-sense, heuristic approach is usually taken. In a 1974 report, *Heuristic Routing for Solid Waste Collection Vehicles,* the EPA outlined a set of simple guidelines that included some obvious notions such as avoiding heavily traveled streets during rush hour, making routes

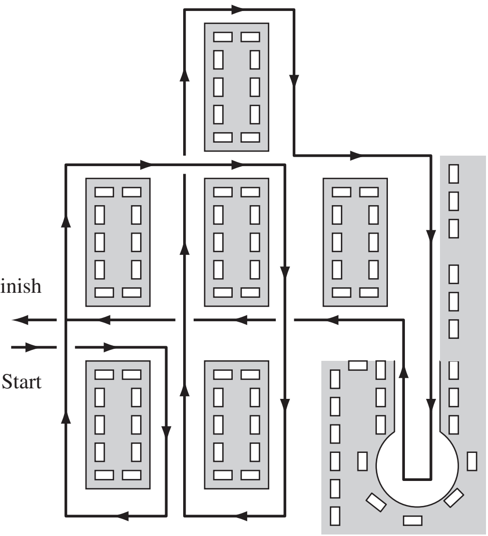


FIGURE 14 A heuristic route emphasizing right turns and a minimum amount of deadheading.

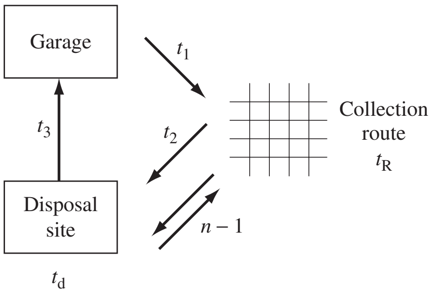


FIGURE 15 Timing diagram for collection of refuse.

compact with streets clustered in the same geographical area, and routing long, straight runs before looping. Other ideas are less obvious such as starting the run at the top of a hill rather than the bottom so the load is increasing as the vehicle travels downhill. Recognizing the hazards and delays associated with making left turns, the guidelines emphasize clockwise loops made up mostly of right-hand turns. Dead-end streets therefore are better accessed when the vehicle encounters the street on the right hand side rather than the left. One-way streets are best entered at the upper end of the street. Figure 14 shows an example of a routing that tries to follow these guidelines.

Direct Haul Systems

We can use some simple estimates to help gain a sense of how truck sizes and collection patterns might be planned. Consider Figure 15, which shows a basic collection system consisting of the garage, where collection vehicles are parked overnight, the collection route, and the disposal site where refuse is deposited. Later, the possibility of incorporating a transfer station between the collection route and the disposal site to reduce lost time by the collection vehicles will be introduced.

To work up a basic plan for refuse collection, consider the following notation:

t_1 = time to drive from garage to beginning of collection route
t_2 = time to drive between collection route and disposal site
t_3 = time to drive from disposal site back to garage at end of day
t_R = total time on route collecting refuse
t_d = time spent at the disposal site dropping one truck load
t_b = time spent on breaks, and so on
T_t = total time for one day of refuse collection
n = number of runs from the collection route to the disposal site

No matter what, each day's collection requires at least a drive from the garage to the collection route, time spent on the route, a run to the disposal site, time spent there unloading, and a drive back from the disposal site to the garage. More than one run may be made between the collection route and the disposal site, depending on how big the truck is, how many pickups are to be made, and how much refuse is picked up at each stop. The following describes a single day of collection:

$$T_t = t_1 + t_R + t_2 + t_d + (n - 1)(2t_2 + t_d) + t_3 + t_b$$
$$T_t = t_1 + (2n - 1)t_2 + t_3 + t_b + t_R + nt_d \tag{2}$$

EXAMPLE 5 Time Spent Collecting Refuse

Suppose it takes 0.4 hours to drive from the garage to the beginning of the route, 0.4 hours to drive between the route and disposal site, and 0.25 hours to return from the disposal site to the garage. It takes 0.2 hours to offload a truck at the disposal site. The crew is given two 15-minute breaks per day and another 30 minutes is allowed for unexpected delays (total 1 hour). If two runs are made to the disposal site each day, how much time is left in an 8-hour shift for actual refuse collection?

Solution Rearranging (2) and substituting appropriate values,

$$\begin{aligned} t_R &= T_t - t_1 - (2n - 1)t_2 - t_3 - t_b - nt_d \\ &= 8 - 0.4 - (2 \times 2 - 1) \times 0.4 - 0.25 - 1 - 2 \times 0.2 \\ &= 4.75 \text{ hours left for actual collection} \end{aligned}$$

To decide how many customers can be served per day by one collection truck, we need the average length of time needed to service each stop. A simple approach suggests that the time per stop is equal to the average time that it takes to drive from one stop to another, plus the time taken at each stop to empty the containers, from which we can write

$$t_s = \frac{v}{d} + nt_c \tag{3}$$

where

t_s = average time for one stop
v = average truck speed between stops
d = average distance between stops
n = number of containers to empty per stop
t_c = time to empty one container

To find the volume of an appropriate collection vehicle, we need to know the number of stops, the volume of refuse per stop, and the vehicle compaction ratio:

$$V = \frac{v_s N}{r} \tag{4}$$

where

V = volume of the collection vehicle
v_s = average volume of refuse per stop
r = compaction ratio
N = number of stops per truckload

EXAMPLE 6 Vehicle Sizing

Continuing with Example 5 (two truckloads and 4.75 hours of collection per day), suppose the pickup vehicle averages 5 mph between stops that are on the average separated by 150 feet. Each stop collects from two households that each put out two full 30-gallon (4 cubic feet) containers. It takes an average of 20 seconds to empty one container. How many households could be serviced per day, and how large should the truck be if it has a compaction ratio of 3.5:1?

Solution The time required to service each stop would be

$$t_s = \frac{d}{v} + nt_c = \frac{150 \text{ ft/stop} \times 3{,}600 \text{ s/hr}}{5 \text{ mile/hr} \times 5{,}280 \text{ ft/mile}} + 4 \text{ cans/stop} \times 20 \text{ s/can} = 100 \text{ s/stop}$$

At 100 seconds per stop, and 2 truckloads per day, the number of stops that could be made per truckload is

$$N = \frac{4.75 \text{ h/d} \times 3{,}600 \text{ s/h}}{100 \text{ s/stop} \times 2 \text{ truckloads/day}} = 85 \text{ stops/truckload}$$

The truck volume needed would be

$$V = \frac{v_s N}{r} = \frac{4 \text{ cans/stop} \times 4 \text{ ft}^3\text{/can} \times 85 \text{ stops/load}}{3.5 \times 27 \text{ ft}^3\text{/yd}^3} = 14.4 \text{ yd}^3$$

So, you would choose a truck size that provides at least 14.4 yd^3 of hauling capacity. With 2 households per stop and 2 truckloads per day, the truck would service 2 × 2 × 85 = 340 households each day.

The collection vehicle sized in Example 6 was based on an assumption that each truck would make 2 runs to the disposal site per day. If the truck is large enough, perhaps only 1 run per day to the disposal site would be needed. In fact, if we rework Examples 1 and 2 using only 1 run, we find that a 35-yd^3 truck could serve 412 customers per day instead of the 340 found in Example 2. With 3 runs per day, there is so much time taken up making trips back-and-forth to the disposal site that the time remaining to fill a truck is so low that a 7.6-yd^3 truck is all that would be needed. This little truck would only be able to serve 270 customers per day. So, the natural question to ask is, "which choice is best?" Should we design the collection system around large, expensive trucks, which make fewer trips, or go with smaller trucks, which are cheaper, but since they make more back-and-forth runs to the disposal site each day they serve fewer customers per truck?

Economics of Collection

To decide on the truck size that would provide the cheapest waste transport, we need to know the annual cost of owning and operating trucks, including the cost of the crew that makes the pickups. Finding the annualized cost of each truck involves using an engineering economy calculation in which the capital cost, amortized over the lifetime of the vehicle, is added to the estimated annual maintenance and fuel costs. The relationship between the purchase price of capital equipment, such as trucks, and the amortized yearly cost is given by

$$A = P\left[\frac{i(1+i)^n}{(1+i)^n - 1}\right] \tag{5}$$

where

A = annual cost (\$/yr)
P = purchase price (\$)
i = interest rate, discount rate (yr^{-1})
n = amortization period (yr)

The quantity in brackets in (5) is known as the *capital recovery factor,* CRF(i, n). One way to interpret (5) is to think of A as being annual payments that would pay off a loan of P dollars in n years if the loan interest rate is i (the decimal fraction representation of i should be used; for example, 10 percent should be entered as $i = 0.10/\text{yr}$). For example, a mid-size collection vehicle costing \$180,000, if amortized over a five-year period using a 10-percent discount rate, would have an annualized cost of

$$A = \$180{,}000\left[\frac{0.10(1+0.10)^5}{(1+0.10)^5 - 1}\right] = \$180{,}000 \times 0.2638/\text{yr} = \$47{,}484/\text{yr}$$

These vehicles are heavy, and their start-and-stop driving pattern leads to very low fuel efficiency—less than 2 miles per gallon is typical. An average truck driven 50 miles per day, 260 days per year, at 2 mpg and \$3.00 per gallon would cost almost \$20,000 per year in fuel alone. If we add maintenance, insurance, and other costs, this typical packer truck can easily cost on the order of \$80,000 per year—not counting the labor cost of the crew.

The annualized cost of a vehicle will depend on its size and usage, and one approach to analyzing its economics is to use a simple linear model such as the following:

$$\text{Annualized cost (\$/yr)} = \alpha + \beta V \tag{6}$$

where α and β are empirically determined estimates based on a survey of available vehicles, and V is the volume of the truck. The following example illustrates how it could be used to help make economic decisions about what size truck makes sense for a given job.

EXAMPLE 7 An Economic Analysis of Refuse Collection

Suppose the annualized cost of purchasing, fueling, and maintaining a compactor truck is given by the following expression:

$$\text{Annualized cost}(\$/\text{yr}) = 25{,}000 + 4{,}000V$$

where V is the truck volume in cubic yards. Suppose these trucks require two-person crews, with labor charged at \$24 per hour each (including benefits). Do an economic analysis of the collection system in Example 6 in which a 14.4-yd^3 truck collects refuse from 340 households each day. Each household generates 60 lbs of refuse per week. The trucks and crew work 5 days per week, and curbside pickup is provided once a week for each house. What is the cost per ton of refuse collected and what is the cost per household?

Solution If we assume 8-hour days, 5 days/week, 52 weeks/yr, the annualized cost of labor per truck would be

$$\begin{aligned}\text{Labor cost} &= 2 \text{ person} \times \$24/\text{hr} \times 8 \text{ hr/day} \times 5 \text{ day/wk} \times 52 \text{ week/yr}\\ &= \$99{,}840/\text{yr}\end{aligned}$$

The annualized cost of the truck would be

$$\begin{aligned}\text{Annualized truck cost } (\$/\text{yr}) &= 25{,}000 + 4{,}000 \times 14.4 = \$82{,}600/\text{yr}\\ \text{Total annual cost of truck and crew} &= \$99{,}840 + \$82{,}600 = \$182{,}440/\text{yr}\end{aligned}$$

Over a 5-day week, 1,700 households (5 × 340 = 1,700) are served by each truck. The total amount of refuse collected by 1 truck in 1 year is

$$\text{Annual refuse} = \frac{1{,}700 \text{ homes} \times 60 \text{ lb/week} \times 52 \text{ week/yr}}{2{,}000 \text{ lb/ton}} = 2{,}652 \text{ ton/yr}$$

The annual cost per ton of waste collected is therefore

$$\text{Annual cost per ton} = \frac{\$182{,}440/\text{yr}}{2{,}652 \text{ ton/yr}} = \$68.80/\text{ton}$$

The annual cost per household for the truck and crew

$$\text{Annual cost per household} = \frac{\$68.80/\text{ton} \times 2{,}652 \text{ ton/yr}}{1{,}700 \text{ household}} = \$107.32/\text{yr}$$

The total amount billed to each customer will be considerably higher after including transfer station fees, administrative costs, overhead, profits, and so on.

Table 17 extends the analysis of Examples 5, 6, and 7 to trucks making one run or three runs per day to the disposal site. Under the assumptions made in this analysis, mid-size trucks are optimum. Larger trucks make fewer runs to the disposal site, but their capital cost is too high. Use of smaller trucks holds their capital cost down, but the extra time spent driving back and forth reduces the number of customers served, which more than offsets that advantage.

The solution represented by Table 16 is quite sensitive to the distance between collection route and the disposal site. As that distance increases, having fewer trucks driving that long distance only once per day begins to become the most cost effective design, as is shown in Figure 16.

TABLE 17

Comparison of Costs for Trucks Making One, Two, or Three Trips Per Day to a Disposal Site

Number of Trips Per Day	Houses Served Per Truck	Minimum Truck Size (yd^3)	Annualized Costs Truck ($/yr)	Annualized Costs Labor ($/yr)	Cost Per Ton ($/ton)	Cost Per Household ($/yr)
1	2,061	34.9	164,556	99,840	$82.25	$128.31
2	1,702	14.4	82,643	99,840	$68.80	$107.32
3	1,344	7.6	55,338	99,840	$74.02	$115.47

Assumptions are as given in Examples 5, 6, and 7.

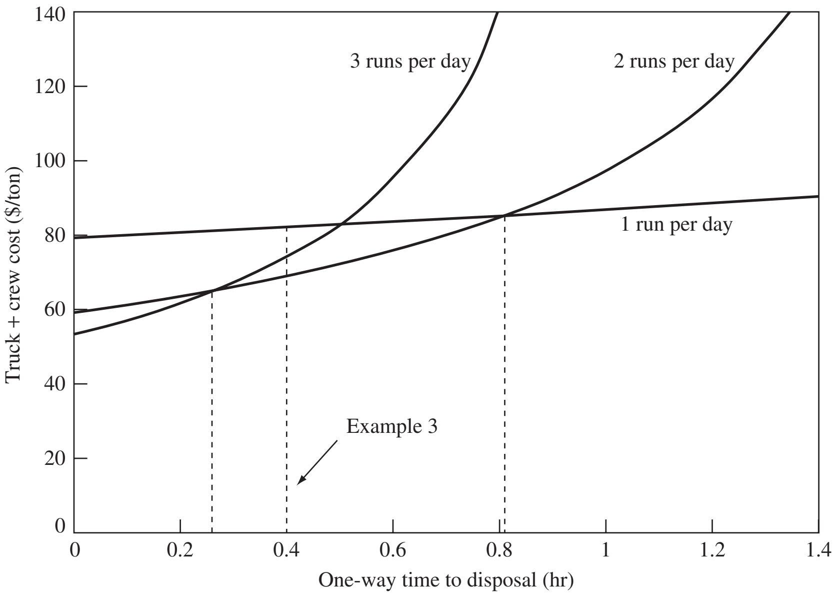


FIGURE 16 The optimum number of trips to make per day to the disposal site is sensitive to the time needed to make the drive. For short distances, smaller trucks making more trips are optimum. For long distances, large trucks that can collect all day long before disposal become more cost effective. This figure is based on assumptions in Examples 6 and 7.

Transfer Stations

As convenient, local landfills close it is often the case that the replacement site is located many miles away from the community that is being served. As that distance from the collection area increases, it takes more and more time to haul the refuse to the disposal site, which leaves less and less time for each truck to actually collect the wastes. At some point, it is better to construct a facility close to town, called a *transfer station,* that acts as a temporary repository for wastes dropped off by local garbage trucks. Larger, long-haul trucks are then used to transport wastes from the transfer station to the disposal site.

Again, a design decision must be made. At what point is the disposal site so far away that the added costs of a transfer station are justified? The problem we need to analyze is shown in Figure 17. A procedure for deriving the cost-versus-distance graph for direct haul from collection route to disposal site has already been demonstrated. What needs to be done now is an analysis of the transfer station and its associated fleet of long-haul trucks.

The cost of a transfer station itself depends on many factors, including its size, local construction costs, and the price of land. In addition, transfer stations can employ a variety of technological features that also contribute to costs. Simple transfer stations may consist of just a lightweight building shell with a thick concrete slab called a tipping floor. Collection vehicles drop their refuse onto the tipping floor while a front-loader scoops it up and loads the transfer vehicles. More complex facilities may include hoppers for direct deposit of refuse from collection vehicles into transfer vehicles. They may also have compaction equipment to compress the wastes before loading. Some transfer stations compress wastes into dense bales (1,000 lb/yd^3) that can be loaded with forklifts onto flatbed trucks. Baled refuse is especially convenient for rail or barge transport.

Trucks that haul wastes from transfer stations to the disposal site are usually large tractor-trailer vehicles. The trailers may be open-topped or closed, but closed

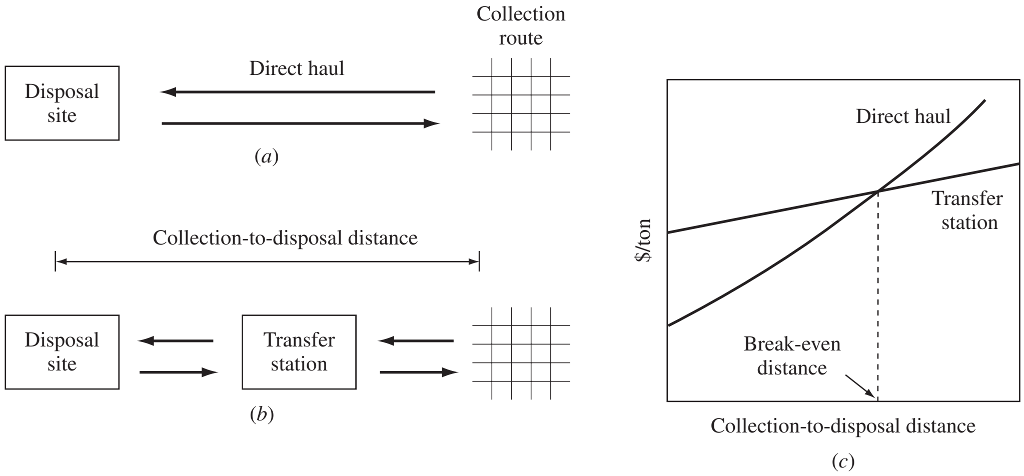


FIGURE 17 The cost curve for a direct-haul system (*a*) starts lower but rises more steeply than the cost curve when a transfer station is included (*b*). At some distance, where the curves cross, the transfer station option becomes more cost effective (*c*).

trailers are preferred since they are less likely to spread debris along the highway and the wastes can be more easily compacted in the trailer. Open trailers carry on the order of 40 to 120 cubic yards of refuse with densities of 200 to 300 lb/yd^3. Since the density of wastes in closed trailers may be higher (500 to 800 lb/yd^3), closed trailers are usually smaller (40 to 80 yd^3) in order to keep the total vehicle weight below highway load limits.

EXAMPLE 8 Costs of a Transfer Station and Its Vehicles

A transfer station handling 300 tons/day, 5 days per week, costs $5 million to build and $150,000 per year to operate. An individual tractor-trailer costs $140,000 and carries 15 tons per trip. Operation and maintenance costs (including fuel) of the truck are $50,000/yr; the driver makes $40,000 per year (including benefits). The capital costs of the building and transfer trucks are to be amortized over a 10-year period using a 12 percent discount factor.

Suppose it takes 30 minutes to make a one-way trip from the transfer station to the disposal site, and 7 round trips per day are made. Find the transfer-station and hauling cost in dollars per ton.

Solution Using (5), the capital recovery factor for a 10-year amortization period and 12 percent discount factor is

$$\text{CRF} = \frac{i(1+i)^n}{(1+i)^n - 1} = \frac{0.12(1+0.12)^{10}}{(1+0.12)^{10} - 1} = 0.1770/\text{yr}$$

The annualized cost of the transfer station is therefore

$$\$5{,}000{,}000 \times 0.1770/\text{yr} + \$150{,}000/\text{yr} = \$1{,}035{,}000/\text{yr}$$

Per ton of waste, that works out to be

$$\text{Transfer station cost} = \frac{\$1{,}035{,}000/\text{yr}}{300\ \text{ton/day} \times 5\ \text{days/week} \times 52\ \text{weeks/yr}} = \$13.27/\text{ton}$$

The annualized cost of a single truck and its driver is

$$\$140{,}000 \times 0.177/\text{yr} + \$50{,}000/\text{yr} + \$40{,}000/\text{yr} = \$114{,}780/\text{yr}$$

Per ton of waste hauled, the truck and driver cost

$$\text{Truck and driver cost} = \frac{\$114{,}780/\text{yr}}{15\ \text{ton/trip} \times 7\ \text{trip/day} \times 260\ \text{day/yr}} = \$4.20/\text{ton}$$

Total cost of transfer station, trucks and driver is therefore

$$\text{Total cost} = \$13.27 + \$4.20 = \$17.47/\text{ton}$$

In Example 8, the $4.20/ton trucking cost was worked out for a particular trip time from the transfer station to the disposal site of .5 hours each way. That is, it is $4.20/ton per .5 hours of distance between transfer station and disposal site, or $8.40 per hour of distance. The transfer station itself has a fixed cost of $13.27/ton.

By varying that trip time, we will get a variable trucking cost that we can add to the fixed annual cost of the transfer station itself, resulting in the graph shown in Figure 18.

We can now compare the economics associated with garbage trucks making their daily pick-ups and delivering their loads all the way to a disposal site with the

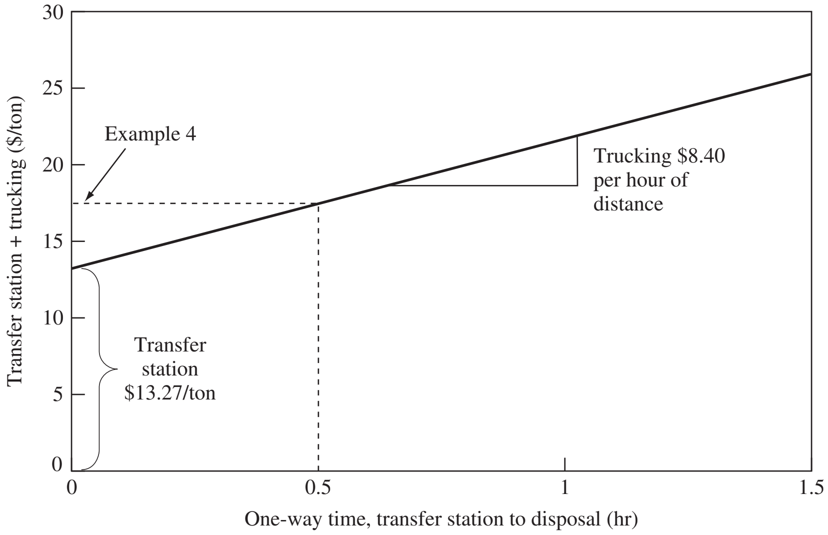


FIGURE 18 Transfer station and long-haul trucking costs depend on transfer-station-to-disposal distance. Drawn for the data given in Example 8.

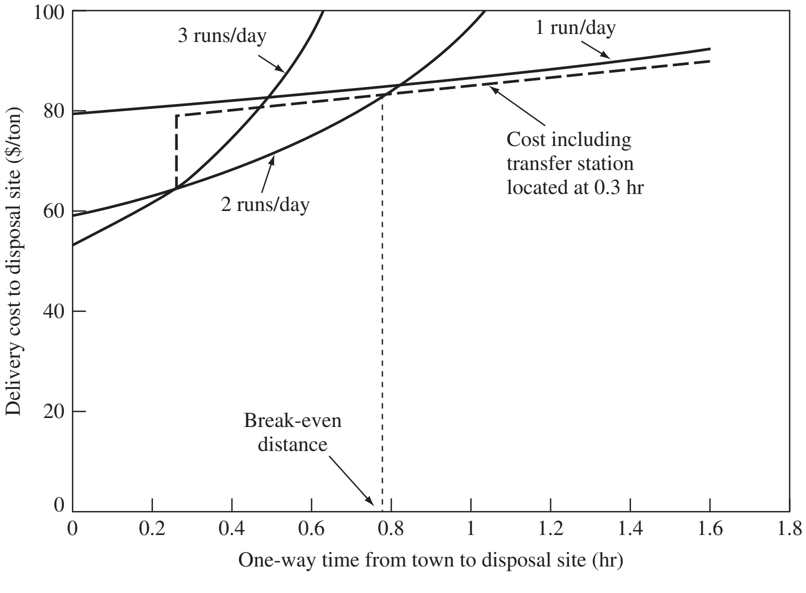


FIGURE 19 A transfer station located 0.3 hours away from the collection route becomes cost effective if the one-way drive time from route to disposal site exceeds about 0.75 hours. Based on data from Examples 7 and 8.

economics of a delivery to a much closer, but costly, local transfer station. To continue our example, let's combine the example for the costs of collection shown in Figure 16 with the costs of a transfer station shown in Figure 18. If we assume an ideal site for a local transfer station is 0.3 hours from town, the decision graph for whether or not to add a transfer station is shown in Figure 19. For one-way run times from route to ultimate disposal site above about 0.75 hours, including a transfer station is the cheapest option. These results, of course, are specific to the example presented. They should, however, give a sense of the trade-offs that need to be evaluated in actual decision making.

10 Materials Recovery Facilities and Economics

Recyclable materials need to be sorted to separate the glass, plastics, newspapers, cans, and so forth. If a community has a recycling system in place, some of this separation is usually done by consumers, some may be done by the crew of the recycling truck as they make their pick-ups, and some is typically done at a *materials recovery facility* (MRF). If there is no recycling system geared toward consumers (that is, materials that could be recycled are mixed in with all of the other refuse collected), there still may be some attempt at recovery. A facility that tries to recover recyclable materials that are mixed in with all the usual municipal solid waste is referred to as a *waste processing facility* (WPF) or sometimes as a *front-end processing system*. These definitions are not precise; for example, a not uncommon collection system has consumers put their commingled recyclables into separate (usually blue) bags, which are tossed into the truck with the rest of the mixed waste. The blue color of those bags helps them visually stand out from dark green, plastic garbage bags, which facilitates their removal at a processing facility.

The primary function of a materials recovery facility is to separate bottles by color, plastics by resin, cans by their metal content, as well as old newspapers (ONP), old corrugated containers (OCC), and compostable organics. The second function is to densify those separated materials so that they can be more easily shipped to end users. Densification includes crushing bottles, flattening metal cans, granulating and baling plastics, and baling waste paper.

Since different communities have differing degrees of preseparation by consumers and truck crews, materials recovery facilities also have a range of equipment and processes. Almost all will include a significant amount of hand separation supplemented by varying amounts of automated equipment. Smaller facilities tend to use more labor since they may not process enough waste to justify the cost of automated equipment, but even large automated facilities use workers for some amount of handpicking and quality control. An example MRF is illustrated in Figure 20.

The usual MRF is designed to receive old newspapers and corrugated cardboard that have already been separated from the bottles, cans, plastics, and glass. This not only simplifies processing, but it helps reduce contamination of paper products from food and chemical wastes that may accompany the other recyclables. A separate area of the tipping floor, with its own conveyor belt, receives newspapers and cardboard. Sorters stand by the conveyor belt, removing cardboard, which will be baled later, along with paper and plastic bags, string, and other foreign materials, which are discarded. Newspapers remain on the belt and are fed into a bailer that

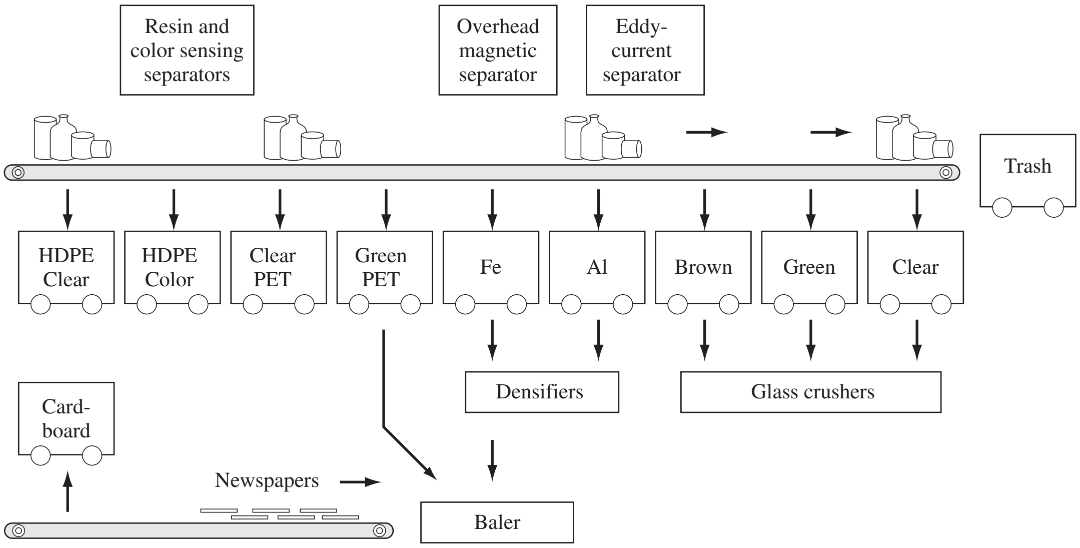


FIGURE 20 A materials recovery facility sorts commingled recyclables using some combination of hand and machine separation.

compacts them and wraps them with wire to form bales. A typical bale is approximately 0.75 m by 1.2 m by 1.5 m (30 × 48 × 60 inches). The same baler can be used to make bales of corrugated cardboard, PET, HDPE, aluminum, or ferrous metals. This removal of specific items, such as cardboard, from the conveyor is referred to as "positive sorting," to distinguish it from "negative sorting," in which the product in question, say newspaper, is left on the conveyor belt.

Another conveyor belt has sorters who push bottles, cans, and other objects into individual hoppers. Since plastics are voluminous and lightweight, they tend to stay on top so they are usually pushed into the appropriate hoppers first. To avoid mixing resins, at least three hoppers are needed: HDPE, PET, and mixed plastics. More careful separation would have hoppers for HDPE, green PET, clear PET, PVC, and mixed residue. The plastics may be fed into a granulator, or knife shredder, which cuts them into pieces roughly 1 cm in size. Or, they may be crushed and baled. As was mentioned earlier, PVC is a particularly critical contaminant of PET, and it can also be a source of chlorine emissions if burned, so sorters must keep a careful watch for it. Manual separation is difficult since PVC and PET are visually very similar. Mechanical sorters which remove PVC are now available, however. In one such sorter, plastic bottles pass over a detector array that can sense the presence of chlorine atoms in PVC bottles. When chlorine is detected, air jets blow the PVC bottles over a partition away from the remaining materials. Another sorting technology analyzes light that has been transmitted through plastic bottles to identify the type of resin and color.

Ferrous metals can be removed from the conveyor belt by hand, or magnets can pull them off. One version of a magnetic separator uses a stationary magnet that is suspended above the main conveyor belt. The magnet draws the ferrous metals to another belt running perpendicular to the main conveyor and then drops them into their hopper. Some aluminum, either from bimetal cans having ferrous metal sides and aluminum lids, or aluminum cans that get wedged in steel cans, will end up in the ferrous metal hopper. This aluminum contamination reduces the value the recycled metal when the product is sold to detinning plants, but the aluminum can actually be somewhat beneficial in steelmaking.

Aluminum cans and aluminum food trays can be picked off the conveyor by hand, as is usually the case in a small MRF, but it is also possible to do this using an *eddy-current separator*. When good electrical conductors, such as aluminum, are placed in a changing magnetic field, small eddy currents are created in the conductor. Those eddy currents, in turn, create their own magnetic field, which interacts with the field created by the magnets. That interaction generates a force that can be used to nudge aluminum out of the wastestream. Other metals, such as copper, will also experience forces induced by eddy currents, but they are much heavier than aluminum and aren't moved as easily. One version of an eddy current separator has the materials to be separated sliding down a ramp under a series of alternating polarity permanent magnets. As the aluminum slides under these magnets it experiences a changing magnetic field, which results in eddy-current forces that push the aluminum to one side of the ramp. Nonconductors slide straight down the ramp, so separation is achieved.

Bottles need to be separated by color, into clear, green, and brown. Prior to color separation, it is important to avoid breakage, but afterwards, the glass is purposely crushed. Vibrating screens, which allow the crushed glass to fall through the openings, can be used to help remove larger, unwanted contaminants such as bottle caps, labels, and other debris.

Processing Equipment for Commingled Wastes

Many communities collect commingled recyclables and mixed wastes. Such wastes could be landfilled as collected, but they contain valuable resources that should be recovered. Not only can recyclables be removed, but combustible products can be prepared for waste-to-energy incineration, and materials to be landfilled can be densified to take up less space. Processing facilities that handle such mixtures may utilize a combination of large, heavy-duty machines to screen, shred, separate, and classify the waste as is suggested in Figure 21.

Raw solid waste consists of a mix of boxes, bags, newspapers, food waste, chunks of wood, bottles, cans, garden clippings, tires, and so forth, with varying sizes and shapes. To facilitate recovery of various components of the waste it must be broken down into smaller pieces. A combination of machines that screen and shred the waste accomplish this task.

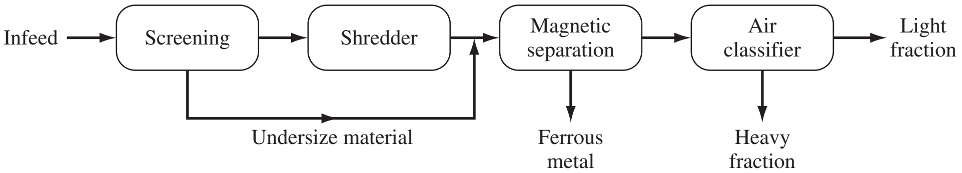
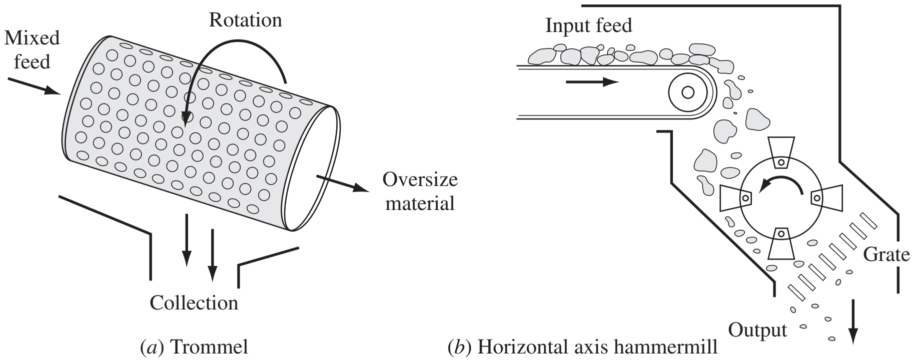


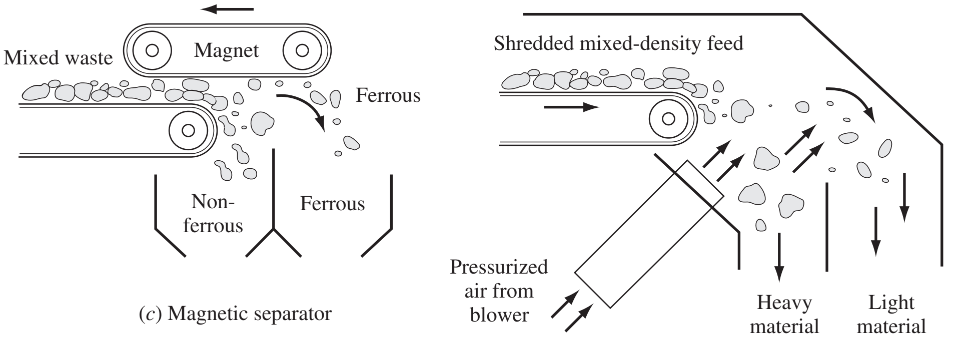
FIGURE 21 An example resource recovery system for mixed solid waste.

Screening. Figure 21 suggests a simple screening of the waste as a first step, though this is not always the case. The most common screen is a large-diameter, tilted, rotating drum with appropriately sized holes, called a *trommel*. These drums have a diameter of roughly 1 meter to 3 meters, and the holes are often about 10 cm across—big enough to let most of the metal and glass fall through but too small to pass much of the paper and larger objects. The rate of rotation is such that particles move at least halfway up the side of the drum before falling back. If the speed is too high, centrifugal forces can hold the wastes against the drum so they never fall back to the bottom. Typical rotation speeds are on the order of 20 to 30 rpm. A sketch of a trommel is shown in Figure 22*a*.

Shredding. Shredders are brute-force machines that pound, crush, pulverize, and shred the wastes. A number of types of shredders are available, but the one most commonly used in recycling operations is the horizontal axis hammermill shown in Figure 22*b*. Wastes dropped into the top are pounded by hammers until they are reduced to a small enough size to fall through the bottom grate. As a way to help prolong the relatively short life of the hammers, wastes are often prescreened before the hammermill to remove objects that are already sufficiently small. Other types of shredders include flail mills, which are similar to hammermills except the hammers are more like knives, and there are no grates, so materials make only one pass by the



(*a*) Trommel (*b*) Horizontal axis hammermill



(*c*) Magnetic separator

(*d*) Air knife

FIGURE 22 Examples of typical equipment used to process commingled recyclables and mixed wastes.

hammers. There are also knife shredders, which slice up objects like rags and bottles; shear shredders, with counterrotating blades that cut materials into strips; and pulverizers that crush glass. Even if there is no materials recovery, shredding has several advantages when used as a precursor to landfilling. Shredded refuse compacts better within a landfill, reducing the storage volume needed. Moreover, shredded waste in a landfill may not need daily covering with earth when spread in thin enough layers to maintain aerobic conditions, which controls odors. Since it is relatively dry, and there are no large food particles, rat and insect problems are also minimized without daily covering.

Magnetic Separators. After materials have been broken into small enough particles, they can be passed by magnets to remove ferrous metals. A number of design variations are possible, but they are all similar in nature. Ferrous metals are picked up by the magnet and moved away from the remaining materials on the main conveyor belt. Figure 22*c* shows one version of a magnetic separator.

Air Classification. Once the wastestream has been reduced to relatively small particle sizes, a blast of air, along with gravity, can be used to separate heavy particles from light ones. In one version of an air classifier, called an air knife, light particles such as plastic and paper are blown over barriers that separate a number of collection bins. Heavy particles are not affected much by the air stream, so they fall into the first hopper. The air knife shown in Figure 22*d* has two receiving bins, but more are possible if more sensitive separation is needed.

Economics of Recycling

While the environmental and resource values of recycling are clear, the economic viability under constantly fluctuating market and regulatory conditions is less certain. To make an economic assessment we need to explore the collection costs, the cost of materials processing, the market value of recycled materials, and the tipping fees avoided by reducing the amount of waste that ends up at the local landfill or incinerator.

To a significant extent, there is a trade-off between the cost of collection and the cost of processing. For example, separate collection of source-separated recyclables is expensive, but the processing that follows collection is relatively cheap. Moreover, overall resource recovery rates may be lower if consumers balk at being asked to do most of the source separation. At the other extreme, using the same packer to pick up recyclables and mixed waste cuts collection costs, but raises the cost of processing. In this case, everyone participates whether they like it or not, so potential recovery rates are high. Also complicating the economics is the trade-off between extensive hand sorting at the materials recovery facility, common when volumes are low, and mechanized sorting, which predominates larger operations. Finally, data on costs is hard to come by, difficult to interpret, and often based on widely varying assumptions and accounting practices.

The big economic question is, of course, how does the cost of collection and processing recyclables, which seems to be very roughly around \$180 per ton (\$120 for collection, \$60 for MRF), compare with the revenues received when the recovered materials are sold? Unfortunately, that comparison is not a simple one for the

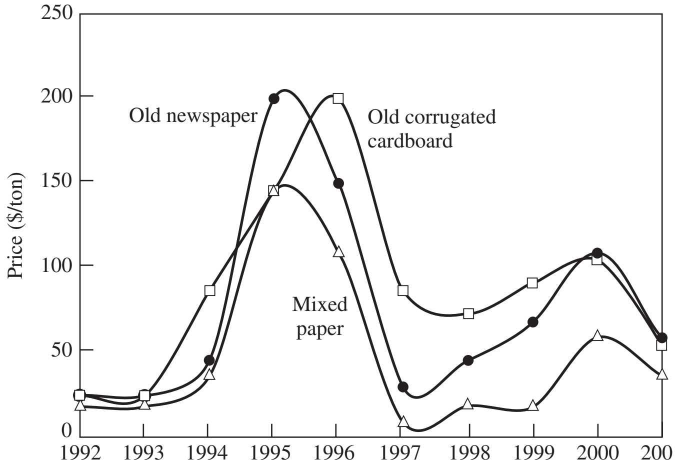


FIGURE 23 Prices for recycled newspaper, corrugated cardboard, and mixed paper. (*Source*: Drawn from Corporations Supporting Recycling data.)

general case. Prices vary considerably over time, both within a given year, and from year to year. For example, prices for recycled aluminum cans tend to rise in the spring as can-makers prepare for the heavy summer demand for beverages. In the fall, prices drop as the supply of returned cans increases and demand drops. Prices also vary from year to year, sometimes by large amounts. For example, there was little demand for recycled paper in the early 1990s, and paper processors often had to pay mills to take it off their hands. Stimulated in part by President Clinton's 1993 executive order directing every federal agency to purchase printing and writing paper containing at least 20 percent postconsumer material by the end of 1994, and 30 percent by the end of 1998, the recycled paper market suddenly exploded (Figure 23). In 1995, it was not uncommon for mills to be paying $140 per ton of mixed paper that had in previous years brought only $25 per ton. The market quickly saturated, and the price collapsed a few years later. After various oscillations in price, in 2005, mixed paper was selling for about $75 per ton.

Given the uncertainties in prices paid for recycled materials, it is difficult to do an economic analysis of recycling that will be valid for any length of time or for any location in general. Nonetheless, it is worth sketching out a simple example to suggest the approach. Table 18 combines representative 2005 prices for recovered materials with an estimate of the relative amounts of various materials typically recovered from MSW. The product of the two gives the revenue derived from the sale of each recovered material. For example, revenue from sale of recovered corrugated cardboard is $40.50 per ton of recovered MSW. Aluminum cans, with much less mass, provide $24 per ton of recovered MSW.

The total revenue from the sale of all recovered materials in the example is $107 per ton. If we use an estimate of $180/ton to collect and process recyclables, the net cost of recycling would be $73 per ton. Is this good or bad? To answer that question, we must estimate the cost of collecting MSW and sending it to the local landfill. As a rough estimate, let's say ordinary MSW collection without recycling costs $100/ton. Add to that the cost of disposing waste in a landfill (called the tipping charge). With a tipping charge of $50/ton, the total cost to dispose of ordinary

TABLE 18

An Example Estimate of Recovered Materials from MSW and Revenues Derived from Their Sale

Material	Recycle Mass (tons)	Recycle Price ($/ton)	Recycle Revenue ($)	Recycle Revenue (%)
Corrugated cardboard	0.45	90	40.50	38
Aluminum cans	0.03	800	24.00	22
Newspaper	0.22	65	14.30	13
Soft mixed paper	0.09	75	6.75	6
HDPE	0.01	600	6.00	6
Steel cans	0.05	90	4.50	4
Magazines	0.04	95	3.80	4
PET	0.01	500	5.00	5
Glass	0.10	25	2.50	2
TOTAL	1 Ton		$107.35	100

MSW is $150/ton. Compared to our estimate of $73 per ton to recycle, it would appear that recycling saves us $77 per ton of recovered materials. This is not a robust conclusion, however, since there are simply too many uncertainties and variations in the numbers used here to justify a strong general statement, but it is encouraging.

Recycling and Carbon Credits

It seems likely that sometime in the near future pouring greenhouse gases into the atmosphere will no longer be cost free. It is easy to imagine a carbon trading scheme, such as the one established with the Kyoto Protocol, that would give value to carbon-saving technologies. By avoiding carbon emissions, recovering and recycling resources from MSW could very well earn carbon credits that could add to the economic value of recycling.

EXAMPLE 9 Carbon Credits for Recycled Corrugated Cardboard

Suppose 1 ton of recovered materials from MSW includes 0.45 tons of corrugated cardboard. By recycling that cardboard instead of landfilling it, greenhouse gas emissions will be reduced. If carbon is traded (or taxed) at $50 per metric ton of carbon equivalents, what would be the extra revenue associated with selling that saved carbon?

Solution From Table 8, recycling one ton of cardboard reduces greenhouse gas emissions by the equivalent of 0.96 metric tons of carbon. The greenhouse gas reduction is therefore

$$0.45 \text{ tons} \times 0.96 \text{ metric-tons-carbon/ton} = 0.432 \text{ MTCE}$$

If carbon is trading at $50/MTCE, the value of this carbon savings would be

$$0.432 \text{ MTCE} \times \$50/\text{MTCE} = \$21.60$$

Table 19 extends Example 9 to the entire suite of recovered materials shown in Table 18. For each of the recovered products, the carbon equivalent from Table 8 is applied to determine the carbon credit earned by recycling. At $50/MTCE the total carbon-savings revenue that would be realized is $43.34, which, when added to the $107.35 received by selling the recovered materials, gives a total of $150.69. The economic value of recycling would certainly be significantly improved by including this greenhouse gas adder. Figure 24 extends this example to show the impact of varying the value of the carbon credits.

TABLE 19

Extension of Table 16 to Include Revenue Earned by Reducing Greenhouse Gas Emissions when Materials Are Recovered and Recycled

Material	Recycle Mass (tons)	Recycle Price ($/ton)	Recycle Revenue ($)	Carbon Equivalent (MTCE/ton)	Carbon Credit (MTCE)	Carbon Revenue ($)	Total Revenue ($)	Revenue (%)
Corrugated cardboard	0.45	90	40.50	0.96	0.432	21.60	62.10	41
Aluminum cans	0.03	800	24.00	3.71	0.111	5.57	29.57	20
Newspaper	0.22	65	14.30	0.52	0.114	5.72	20.02	13
Soft mixed paper	0.09	75	6.75	1.06	0.095	4.77	11.52	8
HDPE	0.01	600	6.00	0.39	0.004	0.20	6.20	4
Steel cans	0.05	90	4.50	0.50	0.025	1.25	5.75	4
Magazines	0.04	95	3.80	0.76	0.030	1.52	5.32	4
PET	0.01	500	5.00	0.43	0.004	0.22	5.22	3
Glass	0.10	25	2.50	0.50	0.050	2.50	5.00	3
TOTAL	1 Ton		$107.35			$43.34	$150.69	100

Carbon credits are assumed to earn $50 per metric ton of carbon equivalents (MTCE) saved.

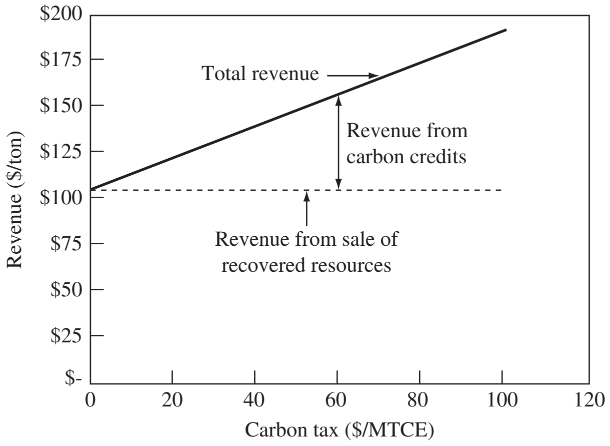


FIGURE 24 Extending the results from Table 17 to show the impact of varying carbon credits on revenue derived from recycling. Revenue is from selling recovered materials and avoiding carbon emissions. Units $ per ton of recovered material.

11 Composting

Yard trimmings and food waste account for one-fourth of the mass of all municipal solid waste generated in the United States. Prior to the 1990s, essentially all of that ended up as discards sent to the local landfill or incinerator. As the remaining life of landfills becomes more limited, and the difficulty in permitting new sites grows, it is clear that source reduction and recycling programs need to be implemented for these wastes as well as all the others. In fact, by the mid-1990s, over half of the population in the United States lived in states having regulations restricting disposal of yard trimmings. The impact of those programs is beginning to be felt. The generation rate of yard trimmings, for example, is dropping as more and more households implement backyard composting, and the recovery rate for yard trimmings is increasing rapidly as municipalities implement their own composting programs. Notice the distinction between backyard composting, which is a form of source reduction since it reduces the amount of waste that has to be collected, and municipal composting, which is considered a form of recycling since it creates a marketable product out of collected wastes.

Composting is the term used to describe the aerobic degradation of organic materials under controlled conditions, yielding a marketable soil amendment or mulch. It is a natural process that can be carried out with modest human intervention, or it can be carefully controlled to shorten the composting time and space required, and to minimize offensive odors. The stabilized end product of composting is rich in organic matter, which makes it a fine soil conditioner, but the concentrations of key nutrients such as nitrogen, phosphorus, and potassium are typically too low for it to compete with commercial fertilizers.

In the simplest composting systems, yard wastes (sometimes mixed with sewage sludge) are stacked in long, outdoor piles called windows (Figure 25). A typical windrow might be 2 meters high, 3 or 4 meters wide, and tens of meters long. Their length is determined by the rate of input of new materials, the length of time that materials need for decomposition, and the cross-sectional area of the pile. More complex systems use biological reactor vessels. Some systems involve both technologies: reactor vessels are used to start the decomposition process, and windrows are used to for the finishing stages.

The composting process is affected by temperature, moisture, pH, nutrient supply, and the availability of oxygen. Bacteria and fungi are the principal players in

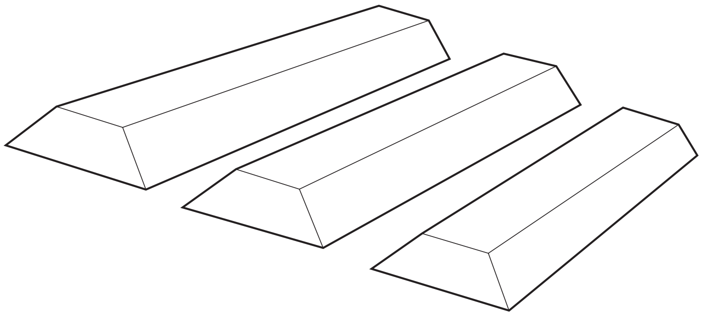


FIGURE 25 Composting piles, called windrows.

the decomposition process, but macroorganisms such as rotifers, nematodes, mites, sow-bugs, earthworms, and beetles also play a role by physically breaking down the materials into smaller bits that are easier for microoganisms to attack.

Temperature

In the early stages of decomposition, mesophilic microorganisms (bacteria and fungi that grow best at temperatures between 25 and 45°C) generate heat while they metabolize the waste, which raises the temperature of the pile. When temperatures reach 45°C, the activity of mesophiles stops, and thermophilic microorganisms, which prefer temperatures between 45 and 70°C, take over the decomposition. Inside the compost pile, temperatures continue to increase as long as nutrient and energy sources are plentiful for the thermophiles. If the pile temperature stays above 55°C for more than 72 hours, most pathogens and weed seeds will be destroyed, making a more marketable end product. Eventually, the nutrient supply drops, thermophiles die off, the temperature falls, and mesophilic microorganism once again dominate the decomposition process until stable end products are formed. Figure 26 shows an example of this rise and fall of pile temperature.

pH

In the early stages of decomposition, organic acids are formed, which cause the pH of the pile to drop to about 5. When the pH is this low, acid-tolerating fungi dominate. Eventually microorganisms break down the organic acids and the pile pH rises, as suggested in Figure 26.

Nutrient Levels

A number of nutrients must be available to the microorganisms that are attacking and degrading the compost pile. The most important nutrients needed are carbon

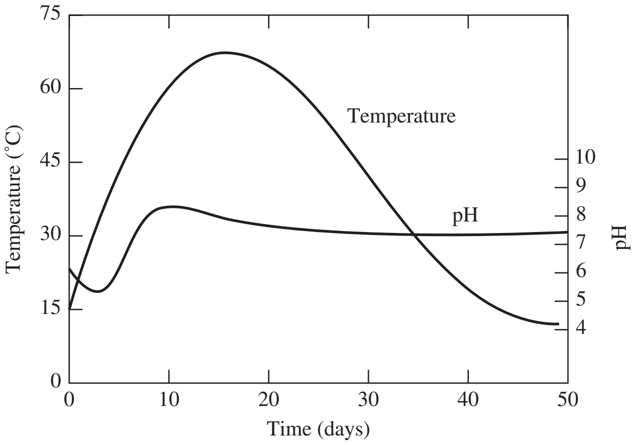


FIGURE 26 Example of temperature and pH variation with time in a compost pile. (*Source:* U.S. EPA, 1994b.)

TABLE 20

Carbon-to-Nitrogen Ratios of Various Materials

Type of Feedstock	C:N Ratio
High Carbon Content	
Corn stalks	60:1
Foliage	40–80:1
Dry leaves and weeds	90:1
Mixed MSW	50–60:1
Sawdust	500:1
High Nitrogen Content	
Cow manure	18:1
Food scraps	15:1
Grass clippings	12–20:1
Humus	10:1
Fresh leaves	30–40:1
Nonlegume vegetable scraps	12:1

Source: U.S. EPA, 1994b.

for energy, nitrogen for protein synthesis, and phosphorus and potassium for cell reproduction and metabolism. In addition, a number of nutrients are needed in trace amounts, including calcium, cobalt, copper, magnesium, and manganese.

One of the best indicators of the likely health of microorganisms is the ratio of carbon to nitrogen available to them. An ideal C:N ratio is roughly 25 to 35 parts of available carbon to available nitrogen. High C:N ratios inhibit the growth of microorganisms, slowing the decomposition. Low C:N ratios accelerate the rate of decomposition but may cause loss of nitrogen as ammonia gas and rapid depletion of the available oxygen supply leading to foul-smelling anaerobic conditions. Leaves, cornstalks, rice hulls, and paper are examples of materials that have high C:N ratios, while grass clippings and sewage sludge have low C:N ratios. Examples of high carbon content and high nitrogen content feedstocks are given in Table 20.

By properly combining different kinds of solid wastes, the C:N ratio can be brought into the desired range. Mixing dry leaves with grass clippings, for example, helps balance the ratio, leading to rapid decomposition without foul odors. Sewage sludge, which is high in water as well as nitrogen, nicely complements municipal solid waste, which tends to be low in moisture and nitrogen. When composted properly, high compost temperatures not only kill pathogens, but they also promote the drying of sewage sludges, which helps reduce the cost of sludge dewatering.

Oxygen

As aerobic microbes degrade waste, they remove oxygen from the pile. If the supply of oxygen is insufficient to meet their needs, anaerobic microorganisms will take their place, slowing the degradation process and producing undesirable odors. Oxygen can be supplied by simply mixing or turning the pile every so often, or piles can be aerated with forced ventilation.

Compost piles that are aerated by mixing need to be turned anywhere from once or twice a week, to once per year. More frequent turning speeds the composting process and helps prevent anaerobic conditions, but may cause the pile to dry out or cool down too much. Such turned windrows can complete the composting of yard trimmings in roughly three months to one year. Composting can be speeded up (which also reduces the land area required) by incorporating a forced aeration system. Wastes are stacked on top of a grid of perforated pipes and a blower forces air through the pipes and composting materials. Composting can be completed in as little as three to six months using this method.

12 Discarded Materials

Recall the definitions used to describe the flow of solid waste materials through our society. Source reduction (for example, lightening containers) and reuse (for example, refilling glass bottles) are activities that reduce the amount of materials that enter the municipal wastestream. What remains are referred to as MSW materials generated. Some of those materials are recovered from the wastestream and recycled or composted, leaving materials that are *discarded*. Those remaining discards are burned or buried.

Table 21 shows historical quantities of various categories of materials discarded in the United States. The impacts of recycling and composting are evident. Paper and paperboard discards are declining as are yard trimmings and ferrous metals. The ever-increasing use of plastics in our society, along with only modest efforts to recycle, are reflected in the continued rapid rise in plastics discards.

TABLE 21

Materials Discarded in the Municipal Wastestream, 1960 to 2005 (Millions of Tons Per Year)

Materials	1960	1970	1980	1990	2000	2005	2005 (%)
Paper and paperboard	24.9	37.5	43.4	52.5	50.2	42.0	25.2
Glass	6.6	12.6	14.4	10.5	9.7	10.0	6.0
Ferrous metals	10.3	12.2	12.3	10.4	8.9	8.8	5.3
Aluminum	0.3	0.8	1.4	1.8	2.3	2.5	1.5
Other nonferrous metals	0.2	0.4	0.6	0.4	0.5	0.5	0.3
Plastics	0.4	2.9	6.8	16.8	24.0	27.3	16.4
Rubber and leather	1.5	2.7	4.1	5.4	5.7	5.7	3.4
Textiles	1.7	2.0	2.4	5.2	8.2	9.4	5.6
Wood	3.0	3.7	7.0	12.1	11.8	12.6	7.6
Other	0.1	0.5	2.0	2.5	3.2	3.4	2.0
Food scraps	12.2	12.8	13.0	20.8	25.8	28.5	17.1
Yard trimmings	20.0	23.2	27.5	30.8	14.8	12.2	7.3
Miscellaneous organic wastes	1.3	1.8	2.2	2.9	3.5	3.7	2.2
TOTAL (million tons)	82.5	113.1	137.1	172.1	168.6	166.6	100

Source: U.S. EPA, 2006a.

13 Waste-to-Energy Combustion

The history of combustion of solid wastes in the United States clearly demonstrates the controversial nature of this method of waste disposal. In 1960, a little over 30 percent of MSW discards were burned in incinerators solely as a volume reduction practice. That percentage dropped rapidly during the 1960s and 1970s, reaching a low of less than 10 percent in 1980. With increased emphasis on avoiding landfilling and with better incinerators that were designed to control emissions while generating useful steam or electricity, combustion of MSW has returned to about 15 percent of MSW generation. As Table 3 illustrated, a number of developed countries around the world rely much more heavily on incineration than does the United States—especially if they have high population densities. Japan, for instance, burns almost two-thirds of its wastes, while the incineration rate in western Europe is roughly 30 to 40 percent.

In the early 1990s, the United States had 190 MSW-combustion facilities in operation that burned about 30 million tons of waste per year. By 2000, that number had dropped to 102, and by 2005, it was down to 88. Those 88 facilities, however, are bigger and cleaner and they burn slightly more waste than those in operation in the early 1990s. Most of these are located in the Northeast and South, while the West has only half a dozen.

Incineration of MSW as a method of waste disposal has a number of favorable attributes, including volume reduction, immediate disposal without waiting for slow biological processes to do the job, much less land area requirements, destruction of hazardous materials, and the possibility of recovering useful energy. There are tradeoffs. For example, it is reasonable to question whether burning paper for its energy value, or recycling it to reduce the environmental impacts of virgin pulping, is the higher use for the material. On the negtive side of combustion, poorly operated incinerators release toxic substances such as dioxins into the air, the ash recovered may be classified as hazardous materials that requires special handling, and the public has generally been reluctant to accept the technology—especially if a facility is being proposed in their own area. There is also a concern for whether incineration competes with recycling. Once an incinerator has been built, a reliable stream of refuse to burn must be maintained to pay it off. There can be reluctance then to expand a recycling program if it may reduce the fuel supply, and hence cash flow, for the incineration system. In the general heirarchy of solid waste management, recycling is considered a greater good than incineration, so there are advocates who feel it is important to write into an incineration contract the freedom to reduce the amount of garbage provided to the facility at any time, without penalty or other economic risk, when that reduction is due to an expansion in recycling.

Energy Content of MSW

The energy content of MSW depends on the mix of materials that it contains as well as its moisture content. The standard test used to determine the heating value of a material involves completely burning a sample in a bomb calorimeter and then

TABLE 22

Examples of "As Received" HHV of Various Components of MSW as Well as Fossil Fuels

Material	kJ/kg	Btu/lb
Mixed paper	15,800	6,800
Mixed food waste	5,500	2,400
Mixed green yard waste	6,300	2,700
Mixed plastics	32,800	14,000
Rubber	26,100	11,200
Leather	18,500	8,000
Textiles	18,700	8,100
Demolition softwood	17,000	7,300
Waste hardwood	15,000	6,500
Coal, bituminous	28,500	12,300
Fuel oil, no. 6	42,500	18,300
Natural gas	55,000	23,700

Source: Solid waste data from Niessen, 1977.

measuring the rise in temperature of a surrounding water bath. The result obtained is known as the higher heat value (HHV), or gross heat value. Included in the HHV is energy that was used to vaporize moisture, which is usually not recovered in normal combustion unless the flue gases are condensed. A more realistic estimate of the energy that can be recovered during combustion is known as the lower heat value (LHV), or net energy. Table 22 gives example higher heating values of various components of municipal solid waste "as received"—that is, without any particular effort to remove the moisture. For comparison, the energy content of fossil fuels is also included in the table.

EXAMPLE 10 **Energy Density of MSW Discards**

Estimate the energy density in Btu/lb of "as received" discarded solid waste in the United States.

Solution Using data given in Table 21 for discarded materials and data from Table 22 for energy density, we can set up a table based on 1 pound of discards. We need to make a few assumptions since some data are missing. For wood, we will use an average HHV of hardwood and softwood $(7{,}300 + 6{,}500)/2 = 6{,}900$ Btu/lb. Since the mix of rubber and leather is not given in Table 21, we will use an average value of $(11{,}200 + 8{,}000)/2 = 9{,}600$ Btu/lb. For "miscellaneous other" wastes, we will assign an energy value of 0. Also, metals and glass will be assumed to have 0 heating value even though coatings on cans do provide modest amounts of energy.

Materials	Pounds	Btu/lb	Btu	% of Total
Paper and paperboard	0.252	6,800	1,714	28.7
Glass	0.060	—	—	0.0
Ferrous metals	0.053	—	—	0.0
Aluminum	0.015	—	—	0.0
Other nonferrous metals	0.003	—	—	0.0
Plastics	0.164	14,000	2,294	38.4
Rubber and leather	0.034	9,600	328	5.5
Textiles	0.056	8,100	457	7.6
Wood	0.076	6,900	522	8.7
Other	0.020	—	—	0.0
Food scraps	0.171	2,400	411	6.9
Yard trimmings	0.073	2,700	198	3.3
Miscellaneous organic wastes	0.022	2,400	53	0.9
TOTAL (lbs and Btu/lb)	1.000		5,977	100.0

Thus, typical discarded solid waste in the United States has an energy content of approximately 5,977 Btu/lb (13,902 kJ/kg), which is about half the energy density of coal and roughly one-third that of fuel oil. Notice the importance of plastics, especially, and paper products as well.

The composition of MSW varies considerably around the world, and so does its energy content. In developing countries, a relatively small fraction of the waste consists of manufactured materials such as paper, metals, plastic, and glass, which means that the percentage that is food waste is high—usually over 40 percent. Since food waste has a relatively low energy density, the heating value of MSW in the developing world tends to be lower than that of more industrialized countries. Khan and Abu-Gharrarah (1991) have developed an equation that attempts to predict the heating value of MSW based on the paper and food fractions, plus a term that accounts for plastic, leather, and rubber.

$$\text{HHV (kJ/kg)} = 53.5(\text{F} + 3.6\ \text{CP}) + 372\ \text{PLR} \tag{7}$$

where F is food, CP is cardboard and paper, and PLR is plastic, leather, and rubber in the waste mixture, all expressed as mass percentages. For example, Colombia, with 56 percent food waste, 22 percent paper, and 5 percent plastic, leather, and rubber, (7) would estimate the fuel value of its MSW to be

$$\text{HHV} = 53.5(56 + 3.6 \times 22) + 372 \times 5 = 9{,}090\ \text{kJ/kg}$$

This result agrees with that given in Table 23, which shows estimates of MSW energy density for a number of countries.

Only in unusual circumstances can the HHV of a fuel be captured. Usually the latent heat contained in water vapor is lost to the atmosphere rather than being captured by the power plant. There are two possible sources of that water vapor loss: moisture in the wastes and hydrogen in the (dry) waste, which can react with oxygen to form water. Using 2,440 kJ/kg as the latent heat of vaporization for water

TABLE 23

Estimated Fraction of MSW by Weight and by HHV for Various Countries

Country	Paper (%)	Metals (%)	Glass (%)	Food (%)	PLR[a] (%)	HHV (MJ/kg)
Australia	38	11	18	13	0.1	8.1
Colombia	22	1	2	56	5	9.1
Czechoslovakia	13.4	6.2	6.6	41.8	4.2	6.4
England	37	8	8	28	2	9.4
France	30	4	4	30	1	7.8
Germany	20	5	10	21	2	5.7
India	3	1	8	36	1	2.9
Iran	17.2	1.8	2.1	69.8	3.8	8.5
Japan	21	5.7	3.9	50.0	6.2	9.0
Kenya	12.2	2.7	1.3	42.6	1.0	5.0
Sweden	50	7.0	8.0	15	8.0	13.4
U.S.A.	25.2	12.8	6.0	17.1	15.1	13.9

[a] PLR means plastic, leather, and rubber.

Source: Based on Khan and Abu-Ghararah, 1991, except for United States, which uses MSW discard data from U.S. EPA, 2006a.

(at 25°C), each kg of moisture that is vaporized and lost contains 2,440 kJ of energy. In addition, each kg of hydrogen in the waste itself can produce another 9 kg of water vapor. The total energy lost in vaporized water is therefore

$$Q_L = 2{,}440\,(W + 9H) \tag{6}$$

where

Q_L = latent heat of water vapor released (kJ)
W = kg of moisture in the waste
H = kg of dry waste that is hydrogen

If we assume all of that latent heat is lost out of the stack during combustion, the net energy derived from the material burned, called the *Lower Heating Value* (LHV), is

$$\text{LHV} = \text{HHV} - 2{,}440\,(W + 9H) \tag{7}$$

EXAMPLE 11 Lower Heating Value of MSW

Typical MSW has a moisture content of around 20 percent, and roughly 6 percent of the dry mass of MSW is hydrogen. Using the High Heating Value of MSW found in Example 10, estimate the latent heat of vaporized water and the net Lower Heating Value.

Solution In 1 kg of waste, there will be 0.2 kg of moisture, leaving a dry weight of 0.8 kg. In that dry mass, there will be $0.8 \times 0.06 = 0.048$ kg of hydrogen.

Using (6) the energy content of the water vapor released when 1 kg of waste is burned is

$$Q_L = 2{,}440 \times (0.2 + 9 \times 0.048) = 1{,}542 \text{ kJ}$$

If that heat is not recovered, the 13,902 kJ/kg found in Example 10 is reduced to

$$\text{LHV} = 13{,}902 - 1{,}542 = 12{,}360 \text{ kJ/kg}$$

Applying the useful conversion factor of 2.326 kJ/kg = 1 Btu/lb gives

$$\text{LHV} = 5{,}314 \text{ Btu/lb}$$

Table 24 shows typical moisture and hydrogen contents of various components of solid waste. The values given for the Lower Heating Value are based on (7).

Mass Burn and Refuse-Derived Fuel

Combustion for energy recovery is typically done in one of two ways: Either MSW is sent directly to a mass-burn incinerator, or it is preprocessed to produce a more homogeneous product called *refuse-derived fuel* (RDF) that has much better combustion characteristics. About one-third of the capacity additions of late have been RDF facilities and two-thirds have been mass-burn incinerators.

Figure 27 shows a sketch of a mass-burn, rotary-kiln furnace receiving refuse directly from a packer truck. Heat recovery from the waterwalls, convection tubes, and superheater provides steam that can be used for process heat or space heating in nearby industrial facilities, or the steam can be used to generate electricity, which can more easily be transported over long distances.

TABLE 24

High and Low Heating Values of Combustible Components of MSW

Material	Moisture[a] (%)	H[b] (%)	HHV (kJ/kg)	LHV[c] (kJ/kg)
Mixed paper	10.2	5.8	15,800	14,400
Mixed food waste	72.0	6.4	5,500	3,400
Mixed green yard waste	62.0	5.6	6,300	4,300
Mixed plastics	2.0	7.2	32,800	31,200
Rubber	1.2	10.4	26,100	23,800
Leather	10.0	8.0	18,500	16,700
Textiles	22.0	6.4	18,700	17,100
Demolition softwood	7.7	6.2	17,000	15,600
Waste hardwood	12.0	6.1	15,000	13,500

[a] Percent moisture is as received, by weight.
[b] H is percent hydrogen in dry mass.
[c] LHV is computed from (7).
Source: Based on data given in Niessen, 1977.

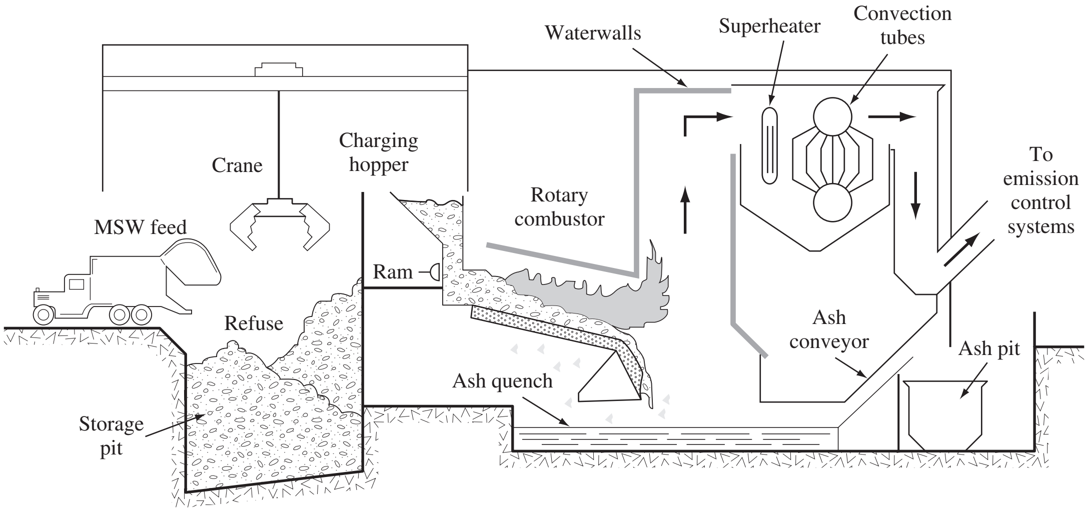


FIGURE 27 Mass burn incinerator featuring a rotary combustor.

EXAMPLE 12 U.S. Electricity Potential from MSW

The United States discards roughly 167 million tons of MSW (Table 21). The High Heating Value of those discards is about 6,000 Btu/lb (Example 10). A mass-burn waste-to-energy facility can convert those wastes to electricity with a heat rate of 17,000 Btu of thermal input per kWh of electrical output (roughly 20 percent efficiency). Estimate the electrical energy that could be produced per year if all of our discards were used in this type of WTE system. Compare it with the total that is now generated, which is about $4{,}000 \times 10^9$ kWh/yr.

Solution

$$\text{Electrical output} = \frac{167 \times 10^6 \text{ ton/yr} \times 2{,}000 \text{ lb/ton} \times 6{,}000 \text{ Btu/lb}}{17{,}000 \text{ Btu/kWh}}$$

$$= 118 \times 10^9 \text{ kWh/yr}$$

$$\text{Fraction from MSW} = \frac{118 \times 10^9 \text{ kWh/yr}}{4{,}000 \times 10^9 \text{ kWh/yr}} = 0.0295 = 2.95\%$$

Refuse-derived fuels have had most of the noncombustible portion of MSW removed, and after shredding and sorting, a relatively uniform product is produced. An example flow diagram for a resource recovery facility that yields RDF is shown in Figure 28. With the noncombustibles removed, the energy content of RDF is obviously higher than that of unprocessed MSW. The typical range of heat values for MSW is roughly 3,500 to 6,500 Btu/lb (8,000 kJ/kg to 15,000 kJ/kg), while the RDF range is typically 6,000 to 7,800 Btu/lb (14,000 kJ/kg to 18,000 kJ/kg). Even

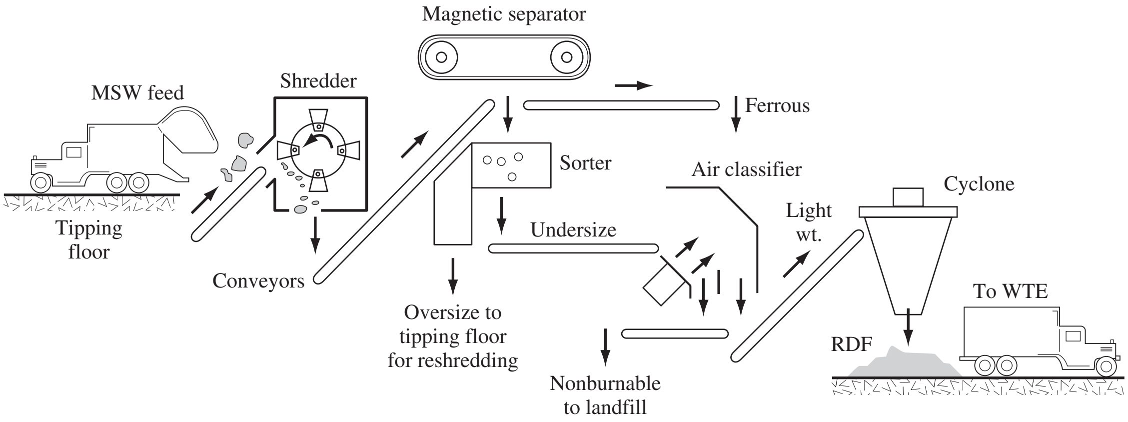


FIGURE 28 An RDF processing system.

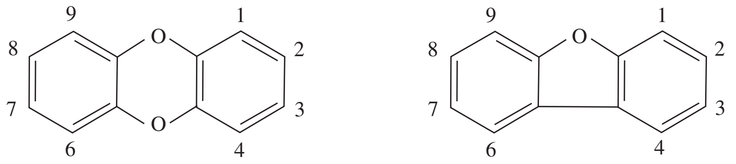
with these higher heating values, RDF is often used as a supplemental fuel in coal-fired facilities that have been suitably modified. Such *co-firing facilities* tend to use up to 20 percent RDF, with the remaining portion being higher energy-content fossil fuel. Fluidized-bed combustors are well suited to refuse-derived fuels since they are able to burn a wide variety of fuels with combustion efficiencies that exceed 99 percent. Japan has more than 100 such incinerators, but only recently have they been introduced in the United States.

Environmental Impacts of WTE

There can be significant environmental impacts associated with the solids and gases that are produced during combustion. Solids are in the form of ash, which is made up of bits of glass and metal, unburned carbon particles, and other inert substances such as silica and alumina. Bottom ash is the residue that falls through grates in the combustion chamber, or which drops off the walls of the boiler. Fly ash is made up of solid particles that are carried out of the boiler by combustion gases. Fly ash should be collected by particulate control systems such as scrubbers, electrostatic precipitators, and/or baghouses. Since combustion is involved, there are all of the usual criteria air pollutants, including nitrogen oxides, sulfur oxides, and carbon monoxide to worry about, and in addition there can be emissions of various acid gases such as hydrochloric and hydrofluoric acid. Of special concern are emissions of especially toxic substances such as dioxins, furans, polychlorinated biphenyls (PCBs), metals, and polycyclic aromatic hydrocarbons (PAHs).

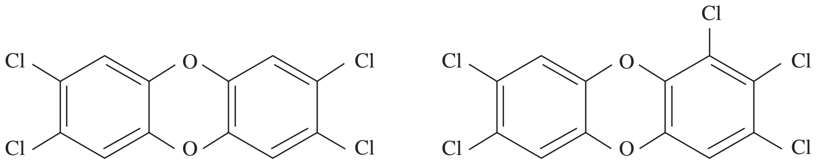
Dioxins and Furans. Dioxins and furans are chlorinated compounds that have somewhat similar chemical structures. They both consist of two benzene rings that are linked together with oxygen bridges. Dioxins have two oxygen atoms joining the benzene rings, while furans have only one oxygen link, the other link being a direct carbon bond. Figure 29 introduces the generic structures of dioxins and furans, showing the conventional numbering system for the chlorine bonding sites.

(*a*) Benzene (*b*) Simplified representations



(*c*) Dioxins (*d*) Furans

FIGURE 29 Dioxins and furans consist of two benzene rings linked by oxygen bridges. (*a*) The benzene ring. (*b*) Simplified representations of the ring. (*c*) The two oxygen bridges that characterize dioxins, along with the conventional site numbering system. (*d*) The single oxygen bridge of furans.



2,3,7,8-tetrachlorodibenzo-*p*-dioxin 1,2,3,7,8-pentachlorodibenzo-*p*-dioxin

FIGURE 30 Two of the most toxic dioxins, 2,3,7,8-TCDD and 1,2,3,7,8-PCDD.

With chlorine added to the basic dioxin and furan structures shown in Figure 29, two families of chemicals are created. There are 76 possible isomers of dioxin, 2 of which are shown in Figure 30, and 136 possible isomers of furan. The most common, and most toxic, of the dioxins is 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, or 2,3,7,8-TCDD, shown in Figure 30.

Dioxins are highly toxic substances that are produced as byproducts of incineration, chemical processing, chlorine bleaching of paper and pulp, the burning of diesel fuel, and the use of some herbicides. They are not manufactured for any commercial purpose and they appear to be entirely the result of human activities. Dioxins have become rather well known to the public as a result of several incidents in the 1970s. They were identified as the chemical that caused the death of numerous birds, cats, and horses when dioxin-laced waste oil was used as a spray to control dust in Times Beach and other locations in Missouri in 1971. TCDD was a contaminant of the herbicides 2,4-D and 2,4,5-T, which were commonly used for weed control in the United States, and which were mixed together and used as the defoliant Agent Orange in Vietnam. In 1976, an explosion at a chemical plant in Seveso, Italy, released a cloud of dioxin that exposed workers and residents. These episodes, tragic as they were, did help to provide data on carcinogenicity and other health effects of exposure.

Dioxins do cause cancer in laboratory animals and, in 2003, EPA classified 2,3,7,8-TCDD as a human carcinogen with no known safe dose, or threshold, below which it will not cause cancer. Dioxins also cause reproductive, developmental, and immune system problems in fish, birds, and mammals at concentrations that are close to those humans experience. The ability of dioxins to affect cell growth and confuse biological signals early in fetal development indicates the human embryo may be susceptible to long-term impairment of immune function from in-utero exposure. Most of human exposure to dioxin is the result of airborne emissions that settle on plants and grasses, and which make their way up the food chain into the beef, dairy, pork, and poultry products that we eat. When dioxins land on water, they are rapidly bioaccumulated in the food chain, which makes some freshwater fish the most heavily contaminated human food source.

Most dioxins and furans are created during combustion, so incineration of municipal solid waste (as well as hazardous and medical wastes) has to be done carefully if emissions are to be minimized. When combustion occurs at temperatures of over 900°C, with residence times of over one to two seconds, dioxin and furan emissions from the furnace itself can be reduced to nondetectable levels. However, if dioxin-furan precursers, such as hydrogen chloride, phenols, chlorophenols, and aromatic hydrocargons, have not been completely destroyed during combustion they can react in the presence of fly ash to form new dioxins and furans as the exiting flue gases cool. Complete combustion, combined with downstream emission control systems, are therefore essential to achieve acceptable emission limits.

It is encouraging to note that dioxin levels in the fatty tissue of the average American dropped from 18 parts per trillion in 1976 to less than 5 parts per trillion in 1994. These gains are thought to have been achieved as a result of the ban on PCBs in 1978; the ban on the herbicide 2,4,5-T in 1979; as well as improved control of the production, use, and disposal practices for a number of other important chlorinated hydrocarbons.

Heavy Metals. Metals such as lead, zinc, cadmium, arsenic, and mercury are part of the municipal wastestream and, when burned, become part of the gases and particles that leave the combustion chamber, or end up as part of the ash residue. In either case, their potential to contaminate the environment requires that they be most carefully controlled. Removal prior to combustion, as well as after combustion, are both necessary parts of the control strategy.

The temperature at which these metals vaporize and condense is an important characteristic that affects the ease with which they can be removed from the incinerator. During combustion, most metals vaporize into gases, which are difficult to control. When those flue gases are cooled, however, metals with high volatilization temperatures condense and adsorb onto fly ash particles, which simplifies their removal in particulate control systems. Metals condense and adsorb more readily onto small particles, typically below 2 μm in size. Most metals are thus best controlled by a combination of flue-gas cooling to below the condensation temperature, followed by an emission control system that effectively removes small particles. A heat exchanger, called an *economizer,* recovers heat from flue gases after they leave the boiler, which not only increases the overall energy recovery rate but also cools and causes condensation of most metals. The emission control system of choice for control of metals is the baghouse since it outperforms electrostatic precipitators at removing very small particles.

The most problematic metal in flue gases to control is mercury, which has the lowest boiling point of all the metals. It is hard to cool flue gases enough to cause mercury vapor to condense, so it may not be effectively removed by particulate control systems. This makes it even more important to focus attention on removing mercury from the wastestream before incineration. By far, the biggest source of mercury in municipal solid waste has been dry cell batteries, but that problem began to ease after Congress passed the *Mercury-Containing and Rechargeable Battery Management Act* of 1996. Other mercury sources include fluorescent lamps and mercury thermometers.

Portable dry-cell batteries that power our flashlights, cell phones, laptops, and every other portable electronic device contain a number of heavy metals in addition to mercury, including cadmium, nickel, silver, and zinc. Americans purchase and discard over three billion of these batteries each year, which works out to about ten per capita per year. Rechargeable batteries last longer and can significantly reduce the rate at which batteries enter our wastestream. Lithium-ion batteries in our laptops and potentially in our future battery-powered vehicles are considered to be nonhazardous waste. Attempts at community battery collection programs have met with limited success but are certainly to be encouraged.

MSW Incinerator Ash. The ash from incineration of municipal solid waste is collected as *fly ash* and *bottom ash,* both of which are contaminated with heavy metals and dioxins/furans. There are two major environmental concerns with respect to disposing of this ash. For one, it is the nature of ash that it can easily become airborne, which increases the risk of exposure due to ingestion or inhalation. Secondly, when ash is disposed of in landfills it becomes exposed to acidic conditions that enhance the ease with which metals dissolve into the leachate. Of special concern are the heavy metals, lead and cadmium, which are highly toxic and are especially common in fly ash.

Until recently, fly ash and bottom ash were typicaly combined and disposed of in conventional clay or plastic lined landfills, which is a much cheaper option than disposal in landfills designed for hazardous materials. This practice was based on Subtitle C, Section 3001(i) of the Resource Conservation and Recovery Act (RCRA), which exempts municipal incinerators from Federal hazardous waste regulation as long as they do not burn hazardous wastes. In 1994, however, that section was interpreted by the U.S. Supreme Court to apply only to the waste itself, not the resulting ash. Under their ruling, ash must now be tested and if it fails the toxicity test it must be managed as a hazardous waste. Roughly 90 percent of an incinerator's ash is bottom ash that may not require disposal as a hazardous waste, but the remaining 10 percent, which is fly ash, is highly toxic and will almost always require special disposal. Mixing them together means the entire ash production may require expensive handling and disposal as hazardous wastes, so that practice is being discontinued.

Of course, if metals are not in the wastestream, they won't end up in incinerator ash. The need to keep toxic metals out of municipal solid waste, especially in areas that utilize incineration, has become especially important.

14 Landfills

The image of the local dump, complete with flies, rats, gulls, odors, airborne bits of paper and garbage, and a black cloud of smoke is well engrained in the consciousness of most of the public. The NIMBY reaction (Not In My Backyard) to proposals for new sites, or expansion of existing sites, comes as no surprise then, even though modern, engineered facilities bear little resemblance to that image just described. In fact, one of the major impacts of the Resource Conservation and Recovery Act (RCRA) on solid waste management has been the prohibition of such open dumps as of 1984.

There are three classifications for landfills. Class I landfills, or *secure* landfills, are those designed to handle hazardous wastes; Class II landfills, or *monofills,* handle so-called designated wastes, which are particular types of waste, such as incinerator ash or sewage sludge, that are relatively uniform in characteristics and require special handling; and Class III landfills, or *sanitary* landfills, are engineered facilities designed to handle municipal solid waste. This chapter will deal with Class III, MSW landfills.

While landfilling continues to be the primary means of MSW disposal in the United States, the "3Rs" of *reduction, reuse, recycling* are beginning to have some impact, and the amounts landfilled have been on the decline since the late 1980s (refer to Figure 6). The number of MSW landfills in the United States has been declining rapidly of late as well. In 1988, there were 8,000 landfill sites around the country; in 1995, there were only 3,200; and in 2005, that total had declined to just 1,650. Most of the sites closed were small, publicly owned facilities that were unable to meet new RCRA Subtitle D landfill regulations that went into effect in 1994. What those statistics mask is the fact that newer facilities tend to be much larger than the smaller ones they replace, so the overall capacity of landfills has actually increased.

With closure of so many local, substandard landfills, coupled with the difficulty in siting new ones, it is not surprising that there are parts of the country that face disposal capacity problems. New RCRA rules, additional maintenance requirements, and tight capacity constraints in the early 1990s caused a surge in tipping fees charged when a truck drops its load at a landfill site. As shown in Figure 31, these fees, which vary considerably from one part of the country to another, seem to have stabilized somewhat in recent years. Notably, the average tipping fee for incinerators is about 80 percent higher than that charged at landfills. Figure 32 shows typical landfill operations.

Basic Landfill Construction and Operation

Moisture in a landfill is critically important if the wastes are to decompose properly, so it is an aspect of landfill design that receives considerable attention. The initial

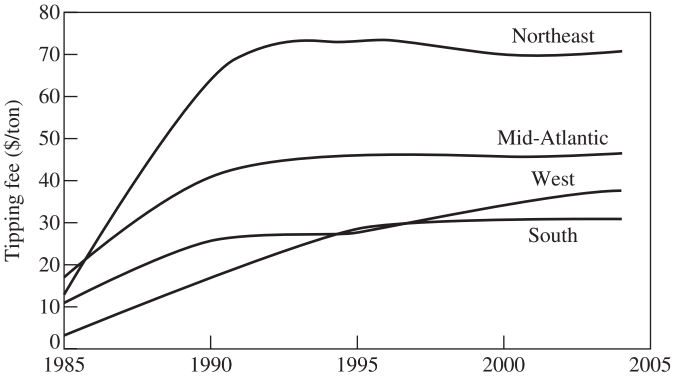


FIGURE 31 Regional MSW landfill tipping fees in the United States.
(*Source:* Data from Repa, 2005.)

FIGURE 32 Disposal of MSW at a local landfill.
(*Source:* British Plastics Federation, www.bpfwindowsgroup.com.)

moisture contained in the wastes themselves is rather quickly dissipated, so it is water that percolates through the surface, sides, and bottom that eventually dominates the water balance in the landfill. Water percolating through the wastes is called leachate, and its collection and treatment is essential to the protection of the local groundwater.

RCRA Subtitle D regulations require MSW landfills to have composite liners and leachate collection systems to prevent groundwater contamination. The composite liner consists of a flexible membrane liner (FML) above a layer of compacted clay soil. Leachate is collected with perforated pipes that are situated above the

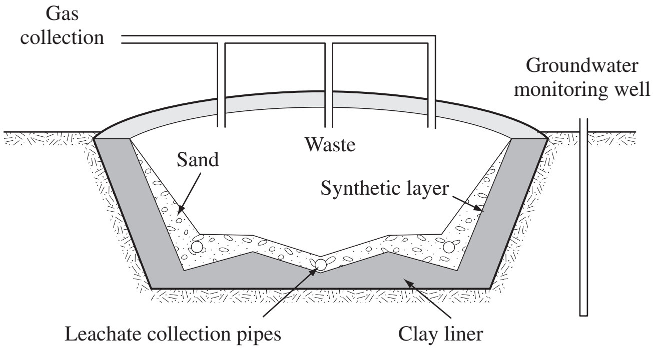


FIGURE 33 Basic features of a Subtitle D municipal solid waste landfill.

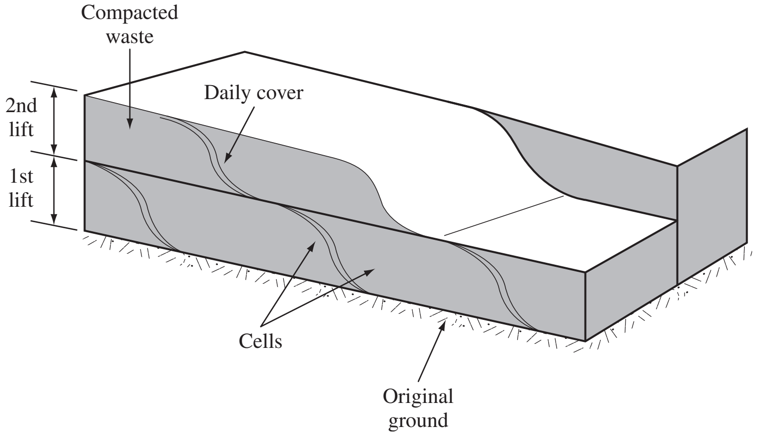


FIGURE 34 Cells, daily cover, and lifts in a sanitary landfill.

FML. A final cover over the completed landfill must be designed to minimize infiltration of water. During waste decomposition, methane gas is formed, so completed landfills need collection and venting systems. The essential features of such a landfill are shown in Figure 33.

The operational phase of a sanitary landfill is organized around the concepts of *cells, daily covers,* and *lifts*. Each day's wastes are received and compacted into cells, which are then covered at the end of the day with a thin layer of soil or other material. The size of a cell depends on the daily volume of refuse to be buried, but typically they are on the order of 3 m (10 ft) thick (including daily cover) and their individual area is determined by the amount and density of the compacted refuse. Cells are covered each day, or more often if necessary, to prevent windblown spread of refuse, reduce odor, control the amount of water entering the cell, and to control rodent, bird, and fly access to the garbage. When a given active area of the landfill is filled with cells, other layers, called lifts, can be added on top. Figure 34 illustrates an active landfill with a second lift being added to the first.

Sizing a landfill requires estimates of the rate at which wastes are discarded and the density of those wastes when they are compacted in the fill. The density of waste depends on how well it is spread and compacted by heavy machinery making multiple passes over the refuse. For municipal solid waste, the density in a landfill will usually be somewhere between 800 and 1,200 lb/yd^3, with a reasonable average estimate of about 1,000 lb/yd^3.

EXAMPLE 13 Estimating Landfill Requirements

Estimate the landfill area needed to handle one year's MSW for a town of 100,000 people. Assume per capita national average discards of 3 lbs per day, no combustion, a landfill density of 1,000 lb/yd^3, and one 10-foot lift per year. Assume 20 percent of the cell volume is soil used for cover.

Solution From the data given, the landfill volume of refuse for 100,000 people would be

$$V_{\text{MSW}} = \frac{3 \text{ lbs/day-person} \times 365 \text{ day/yr} \times 100{,}000 \text{ people}}{1{,}000 \text{ lb/yd}^3} = 109{,}500 \text{ yd}^3\text{/yr}$$

Since only 80 percent of a cell is landfill, the volume of cell needed is

$$V_{\text{cells}} = \frac{109{,}500 \text{ yd}^3}{0.8} = 136{,}875 \text{ yd}^3\text{/yr}$$

The area of lift per year at 10-ft/yr of depth is

$$A_{\text{lift}} = \frac{136{,}875 \text{ yd}^3\text{/yr} \times 27 \text{ ft}^3\text{/yd}^3}{10 \text{ ft/yr}} = 369{,}560 \text{ ft}^2\text{/yr} \quad (8.5 \text{ acre/yr})$$

The actual sizing of a landfill would include a number of additional factors, such as additional area requirements for access roads and auxiliary facilities, reduction in landfill volume as biological decomposition takes place, and increases in compaction as additional lifts are included.

Decomposition in Landfills

When MSW is deposited in landfills, microbial decomposition breaks down the wastes creating gaseous end products, such as carbon dioxide, methane, and various volatile organic compounds (VOCs), as well liquid leachate. The main focus of environmental engineers has traditionally been on keeping leachate from contaminating groundwater, but more recently attention has shifted to the global warming implications of keeping potent methane emissions from reaching the atmosphere, capturing those emissions for clean power generation, and the potential for long-term carbon storage in landfills (Figure 35). To understand these issues, we need to learn more about waste decomposition.

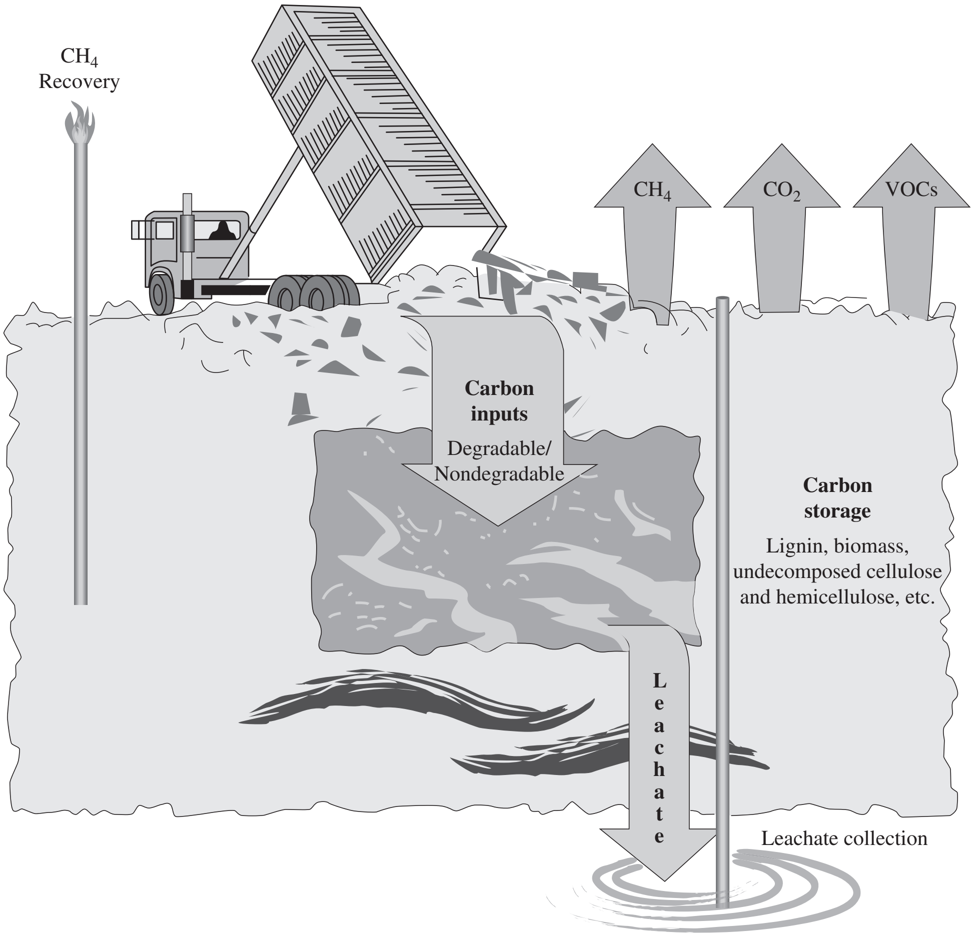


FIGURE 35 Attention to environmental impacts of landfills has shifted from concern for groundwater contamination by leachates to global warming impacts of greenhouse gas emissions and long-term carbon storage.
(*Source:* U.S. EPA, 2006b.)

The decomposition of landfill materials can be thought of as a four-stage process. As shown in Figure 36, these stages are:

I. *Aerobic Phase:* When wastes are placed in a landfill, there is enough entrained oxygen to allow aerobic decomposition to take place for the first few days. Oxygen levels drop, and at the end of this phase, anaerobic conditions begin.

II. *Acid Phase:* During this phase, anaerobic conditions prevail, and a two-step process begins. First, hydrolyzing-fermentative organisms produce enzymes that break down complex organics such as cellulose and starch into simpler products that can be fermented into hydrogen, carbon dioxide, fatty acids, and alcohols. In the second step, those products are converted by bacteria, called *acetogens,* into simpler organic acids, typified by acetic acid (CH_3COOH). As these acids form, the pH of the leachate drops, which can

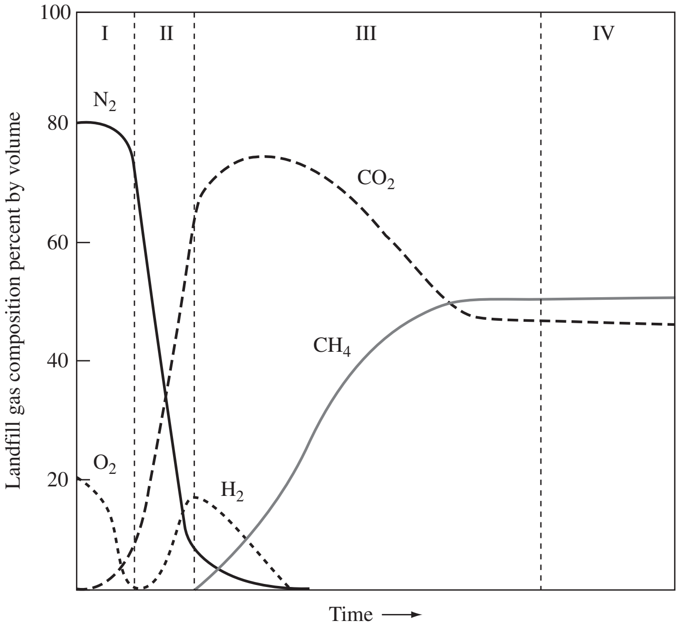


FIGURE 36 Changes in the composition of landfill gases help define a four-phase decomposition process. Phase I Aerobic; Phase II Acidogenesis; Phase III Unsteady methanogenesis; Phase IV Steady methanogenesis.
(*Source: Solid Waste Landfill Engineering and Design,* by McBean & Rovers, © 1995. Reprinted by permission of Prentice-Hall, Upper Saddle River, NJ.)

allow heavy metals to be solubilized. The CO_2 concentration in the waste rises and small amounts of hydrogen gas H_2 are produced.

III. *Methanogenesis, unsteady:* Another group of microorganisms, called *methane formers* or *methanogens,* convert the organic acids into CH_4 and CO_2. The pH begins to return toward more neutral conditions and the release of heavy metals into the leachate declines. This phase can last for months.

IV. *Methanogenesis, steady:* The duration of each phase depends on the availability of moisture and nutrients, but typically on the order of a year or so after a landfill cell is completed the generation rate of CH_4 and CO_2 settles down into nearly equal percentages, which is the characteristic of Phase IV. After many years, perhaps several decades, the decomposition process and the rate of production of methane decline significantly.

The rate of landfill decomposition is affected by a number of factors, including moisture, available nutrients, pH, and temperature. The amount of moisture available for decomposition is a critical factor, which has recently led to increasing interest in recirculating leachate and other liquids into the landfill in order to optimize moisture content to accelerate decomposition and facilitate methane collection.

Under the assumption that CO_2, CH_4, and NH_3 are the principal gases liberated during decomposition, an equation that describes the complete decomposition of organic materials under anaerobic conditions is given by the following:

$$C_aH_bO_cN_d + n\,H_2O \rightarrow m\,CH_4 + s\,CO_2 + d\,NH_3 \qquad (8)$$

where

$$n = (4a - b - 2c + 3d)/4$$

$$m = (4a + b - 2c - 3d)/8$$

$$s = (4a - b + 2c + 3d)/8$$

To use (8) to determine methane production from waste requires that we first determine a chemical formula for that waste. Niessen (1977) has developed estimates of the mass percentage of the chemical elements carbon, hydrogen, oxygen, and nitrogen that make up fully dried components of solid waste. Such a breakdown by elements is called an *ultimate analysis,* or *elemental analysis*. The following example illustrates the process.

EXAMPLE 14 Estimating the Methane Potential of Discards

Suppose "as delivered" MSW discards include 67.3 percent decomposable materials, such as paper, yard trimmings, food waste, textiles, and wood. Of that, 32.3 percent is moisture. An elemental analysis of the dried decomposable components yields the following mass percentages:

Element:	C	H	O	N	Other	Total
Dry Mass Percentage:	44.17	5.91	42.50	0.73	6.69	100%

Find the chemical formula for the C, H, O, N portion of the decomposables. Then find the energy content of the methane that would be generated per kg of discards. The HHV of methane is 890 kJ/mol.

Solution Starting with 1,000 g of as received discards, 673 g are decomposable. Of that, 32.3 percent is moisture, leaving a dry mass of

$$\text{Dry mass of decomposables} = (1 - 0.323) \times 673 = 456.1 \text{ g/kg discards}$$

which gives for each element:

Element	% of Dry Mass	Mass (g)
C	44.17	$0.4417 \times 456.1 = 201.5$
H	5.91	$0.0591 \times 456.1 = 27.0$
O	42.50	$0.4250 \times 456.1 = 193.9$
N	0.73	$0.0073 \times 456.1 = 3.34$
		$C_aH_bO_cN_d = 425.7$ g/mol

Thus, for example, 1 g-mol of $C_aH_bO_cN_d$ has 201.5 g of carbon, so using an atomic weight of 12 for carbon, we can compute the moles of *a* as follows:

$$12 \text{ g/mol} \times a \text{ mol} = 201.5 \text{ g}, \quad \text{so } a = 201.5/12.01 = 16.8 \text{ mol}$$

Similarly, using atomic weights of H, O, and N of 1, 16, and 14, respectively, gives us

H:	$1 \text{ g/mol} \times b \text{ mol} = 27.0 \text{ g}$	$b = 27.0 \text{ mol}$
O:	$16 \text{ g/mol} \times c \text{ mol} = 193.9 \text{ g}$	$c = 12.1 \text{ mol}$
N:	$14 \text{ g/mol} \times d \text{ mol} = 3.34 \text{ g}$	$d = 0.24 \text{ mol}$

So the chemical formula is $C_{16.8}H_{27.0}O_{12.1}N_{0.24}$.

We can now write the equation that describes the complete decomposition of this material using Eq. (8)

$$C_aH_bO_cN_d + n\,H_2O \rightarrow m\,CH_4 + s\,CO_2 + d\,NH_3 \tag{8}$$

where

$$n = (4a - b - 2c + 3d)/4 = (4 \times 16.8 - 27.0 - 2 \times 12.1 + 3 \times 0.24)/4 = 4.17$$

$$m = (4a + b - 2c - 3d)/8 = (4 \times 16.8 + 27.0 - 2 \times 12.1 - 3 \times 0.24)/8 = 8.65$$

$$s = (4a - b + 2c + 3d)/8 = (4 \times 16.8 - 27.0 + 2 \times 12.1 + 3 \times 0.24)/8 = 8.14$$

So, complete decomposition can be described by

$$C_{16.8}H_{27.0}O_{12.1}N_{0.24} + 4.17\,H_2O \rightarrow 8.65\,CH_4 + 8.14\,CO_2 + 0.24\,NH_3$$

Notice that 8.65 mol out of 17.03 mol of gas produced (8.65 + 8.14 + 0.24 = 17.03) is methane. That is, 8.65/17.03 = 0.51 = 51 percent by volume of the gas produced is CH_4, 48 percent is CO_2, and about 1 percent is NH_3. That is, landfill gas is about half methane and half carbon dioxide.

One mole of dried discards yields 8.65 mol of CH_4. At 890 kJ/mol, that equals

$$\text{HHV of } CH_4 \text{ released} = 8.65 \text{ mol} \times 890 \text{ kJ/mol} = 7{,}698 \text{ kJ}$$

To summarize, 1 kg of as received moist discards has 425.7 g of C, H, O, and N, which makes up 1 mol of $C_{16.8}H_{27.0}O_{12.1}N_{0.24}$, and which produces 8.65 mol of methane having an HHV of about 7,700 kJ.

The calculation performed in Example 14 assumes that everything that can decompose, does decompose. This is an overly optimistic assumption. Excavations of old dumps routinely find newspapers that can still be read a decade or two after disposal. One reason for the slow decomposition is that landfills dry out, and without the moisture assumed in (8) the process proceeds slowly, if at all. From a climate perspective, that is good since it locks up CO_2 and CH_4 that would have normally been released into the atmosphere. Also, the calculation doesn't provide any insight into the rate at which methane is generated. Some models suggest one-half of the total methane production occurs over the first 10 years, while 30 years are required to generate 90 percent of the ultimate volume (McBean et al., 1995).

Methane is explosive when it is in the presence of air in concentrations between 5 and 15 percent, so some attention must be paid to its management.

Within the landfill itself, there isn't enough oxygen to cause such an explosion, but when it outgases from the landfill there is some potential for it to cause problems if it is allowed to accumulate under buildings and other enclosed spaces that are located nearby. Methane is lighter than air, so it tends to rise. If there is an impermeable cover over the landfill, the rising gas creates pressure, which can cause it to move horizontally hundreds of meters before emerging from the soil. Pressure relief can be obtained passively, with simple perforated vent pipes that penetrate the cover and extend into a layer of gravel above the wastes, or actively, with collection systems that suck the gases out of the landfill. Passive venting is no longer allowed in large landfills because of strict air emission regulations.

Collected landfill gases are usually flared, which wastes the energy that could be recovered, but it does help control odors and it does convert the potent greenhouse gas, methane, into carbon dioxide, which contributes less to global warming.

Only a few percent of MSW landfills currently take advantage of the energy in landfill gases, but that fraction is growing. One drawback to use of landfill gases is the relatively low heating value (about 500 Btu/ft^3[18,000 kJ/m^3]), which is about half the heating value of natural gas. This results, of course, from the fact that only about half of landfill gas is methane, while natural gas is almost entirely methane. Raw landfill gas can be used as is, but it is highly corrosive, so some treatment is usually required. It can be upgraded to a medium-energy-content gas by removing moisture and some of the contaminants, while leaving the CO_2 intact. In that form, it can be used in steam or gas turbines to produce on-site electricity. Further upgrading is possible by removing the carbon dioxide, resulting in pipeline-quality gas that can be sold for off-site use.

PROBLEMS

1 The Fort Lauderdale International Airport (FLL) reported the following recycling rates for 2005:

Aluminum cans	45 ton/yr
Corrugated cardboard	163 ton/yr
Glass	58 ton/yr
Magazines	207 ton/yr
PET plastic	39 ton/yr
Newspapers	424 ton/yr

(a) If all of these materials would normally have been landfilled, how many metric tons of carbon equivalents (MTCE) per year would the recycling program save?

(b) If the tipping rate at the local landfill is \$50/ton, how much money would be saved on tipping fees?

(c) If these recycled products were to be sold at prices indicated in Table 18, how much revenue would be generated?

(d) If a carbon tax of \$50 per metric ton of carbon-equivalents is charged, how much tax would be avoided by the carbon savings due to recycling?

2 Consider the energy side of the airport recycling program described in Problem 1:
   (a) Find the annual energy savings for society created by this program.
   (b) If energy is valued at \$5 per million Btu, what is the annual dollar savings society would realize as a result of this recycling program.
   (c) Suppose the government were to offer a rebate based on the dollar value of the energy saved by society for recycling. What amount per ton of these mixed wastes would be justified based on the savings found in b?

3 A large corporation generates the following amounts of recyclable materials:

Aluminum cans	200 ton/yr
Corrugated cardboard	2,200 ton/yr
Glass	400 ton/yr
PET plastic	2,000 ton/yr
Newspapers	500 ton/yr

   (a) Suppose they currently pay \$120 per ton for pick up and disposal of their trash, including these recyclable materials. How much would they save if they could sell these recyclables to a scrap dealer who would pick them up and pay them half the Table 18 market price for recyclables?
   (b) If a carbon tax is announced of \$10 per ton of CO_2, how much money would be saved in carbon taxes because of the recycling program?
   (c) If the corporation spends \$400,000 per year to run this recycling program, what is the annual net benefit, including avoided carbon taxes?

4 Suppose the average SUV is driven 10,000 miles per year with a fuel efficiency of 20 mpg. Gasoline contains 5.22 lbs of carbon per gallon, all of which ends up as CO_2 when burned, and it has an energy density of 125,000 Btu/gal. Suppose all of the annual cardboard recovered in Table 10 would have gone to landfills each year.
   (a) How many tons of CO_2 will be emitted per SUV per year?
   (b) How many Btus of energy will have been consumed?
   (c) The carbon savings from cardboard recycling is equivalent to carbon emissions from how many SUVs?
   (d) How many "average SUVs" of energy are saved by cardboard recycling?

5 A 0.355-L (12 oz) aluminum can has a mass of about 16 g. Assume a 70 percent recycling rate and, using data for the energy intensity of various can production processes given in Table 13, find the energy required to produce cans per liter of drink.

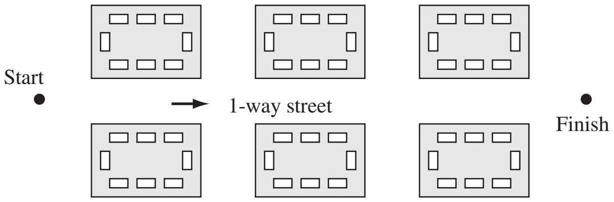
6 Aluminum cans in the 1970s were considerably heavier than they are now. For a 0.355-L (12-oz) aluminum can, with mass 0.0205 kg, and using a 25 percent recycling rate, find the primary energy required to produce the aluminum in one such can. What percent reduction in primary energy has resulted from the lightweighting of cans and the higher recycling rate illustrated in Example 4?

7 In the United States in the mid-1990s, 1.8 million tons/yr of aluminum was used in beverage cans, 63 percent of which was recovered and recycled from old cans. Using estimates from Table 12, determine the following:
   (a) The total primary energy used to make the aluminum for those cans.

(b) The total primary energy that would have been needed if no recycling had occurred.

(c) The reduction in CO_2 emissions that result from that recycling.

8 The following neighborhood has a one-way street running east to west. In general, trucks can only pick up from one side of the street at a time.



(a) Assuming pickups can be made from both sides of the street at the same time on the one-way street, devise a heuristic route in which no house on the route is passed twice (no deadheading).

(b) If trash can only be picked up from one side of the one-way street at a time (it's a busy street), modify your heuristic pickup route with a minimum of deadheading and turns.

9 How long would it take to fill a 30-yd^3 packer truck that compresses waste to 750 lb/yd^3 if it travels 100 ft between stops at an average of 5 mph and it takes 1 minute to load 200 lb of waste at each stop? If each stop services four homes and two collection runs are made per day, how many customers could be provided with once-per-week service by this single truck (assuming a five-day week)?

10 Consider the following data for a municipal waste collection system:

Travel time, garage to route:	20 min
Travel time, route to disposal site:	20 min
Time to unload at disposal site:	15 min
Time from disposal site to garage:	15 min
Time spent on worker breaks:	40 min/day
Packer truck volume:	25 yd^3
Compaction ratio:	4
Curb volume per service:	0.2 yd^3/customer
Travel time between stops:	30 sec
Customers served per stop:	4
Time loading per stop:	1 min

(a) How many hours per day would the crew have to work if it fills the truck twice per day?

(b) Making two runs per day, how many customers would be served per truck if each home has once-per-week service and the truck is used five days per week?

(c) Suppose the cost of a crew for one truck is \$40 per hour for the first eight hours per day, plus \$60 per hour for any hours over that amount. Assume that the crew works 52 weeks per year: Furthermore, suppose a packer truck has an annualized cost of \$10,000 + \$3,500/yd^3. What would be the annual cost of service (crew plus truck) per customer?

11 To avoid overtime pay, the crew in Problem 10 are to work only eight hours per day. They still make two runs to the disposal site each day, but their truck is not always full. How many customers would the truck operated 5 days/week provide once-per-week service to now?

12 To avoid overtime pay, the crew in Problem 10 are to work only 8 hours per day, which means the truck can be smaller. What minimum-size truck would be needed to serve this route? If the cost of a crew for one truck is $40 per hour for 52 weeks per year, and if packer trucks have an annualized cost of $10,000 + $3,500/yd^3, what would be the annual cost of service (crew plus truck) per customer? You might want to compare this with the results found in Problem 10.

13 Suppose each customer puts out 0.25 m^3 of waste each week. Packer trucks with a compaction ratio of 4 take 0.4 minutes per customer to collect the waste. Two trucks are being considered: one that makes two trips per day to the disposal site, and another that makes three.

Trips per day to disposal site	2	3
Truck volume (m^3)	27	15
Annualized truck cost ($/yr)	120,000	70,000
Time driving, unloading, breaks (min/day)	160	215

(a) Operating five days per week, with once-per-week pick-up, how many customers would each truck service?

(b) How many hours per day would each truck and crew have to operate to fill the trucks each day?

(c) With the cost of crew being $40 per hour, what is the total annual cost of trucks and crew for each system? From that, find the annual cost per customer. Which system is less expensive?

14 Suppose a packer truck costs $150,000 to purchase; it uses 2 gallons of fuel per mile at $2.50 per gallon and it is driven 10,000 miles per year; and annual maintenance costs are estimated at $20,000.

(a) If it is amortized over an eight-year period at 12 percent, what is the annualized cost to own and operate this truck?

(b) If labor costs $25 per hour each and the truck has a crew of two who work 40-hour weeks all year, what is the annualized cost of labor?

(c) If this truck hauls 10 ton per day, 260 days per year, what is the cost per ton?

15 Rework Examples 5–7 to confirm the costs per ton of waste given in Table 17 for

(a) One trip to the disposal site per day

(b) Three trips to the disposal site per day

16 Suppose a transfer station that handles 200 tons per day, five days per week, costs $3 million to build and $100,000 per year to operate. Transfer trucks cost $120,000, carry 20 tons per trip, and have an annual cost for maintenance and driver of $80,000 per year. If individual trucks make four trips per day, five days per week from transfer station to disposal site and if trucks and station are amortized at 10 percent interest over 10 years, what is the cost of this operation ($/ton)?

17 Suppose the cost of a direct haul system is given by

$$\text{Direct haul cost (\$/ton)} = 40 + 30t_1$$

where t_1 is the time (hrs) required to make a one-way run to the disposal site (or transfer station). Furthermore, suppose the cost of owning and operating a transfer station with its associated long-haul trucks is given by

$$\text{Transfer station cost (\$/ton)} = 10 + 10t_2$$

where t_2 is the time (hrs) to make a one-way run from the transfer station to the disposal site. Suppose the disposal site is 1.5 hr from the collection route so that $t_1 + t_2 = 1.5$ hr (see Figure P17).

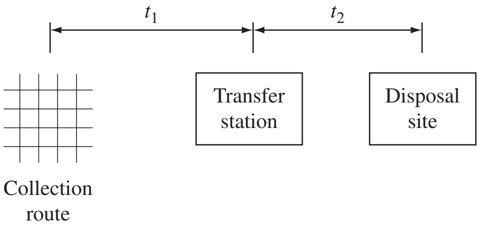


FIGURE P17

(a) What is the cost ($/ton) of a direct haul to the disposal site?

(b) What would be the total cost of the collection system if a transfer station is located 0.3 hr from the collection route?

(c) If the transfer station location is under consideration, what minimum separation (one-way hours) must there be between the transfer station and disposal site to justify having a transfer station?

18 "As received" newsprint is 5.97 percent moisture and has an HHV of 18,540 kJ/kg. When dried, 6.1 percent is hydrogen. Find the lower heating value of this waste.

19 "As received" corrugated boxes have 5.2 percent moisture and an HHV of 16,380 kJ/kg. When dried, 5.7 percent is hydrogen. Find the LHV.

20 A 2-L PET bottle has a mass of 54 g. Polyethylene is roughly 14 percent hydrogen by mass and has an HHV of 43,500 kJ/kg. Find the net energy that might be derived during combustion of one such bottle.

21 One way to estimate the energy content of wastes is based on an empirical equation described in Rhyner et al. (1995), which utilizes an elemental analysis (also called an ultimate analysis) of the material in question:

$$\text{HHV (kJ/kg)} = 339(\text{C}) + 1{,}440(\text{H}) - 139(\text{O}) + 105(\text{S})$$

where (C), (H), (O), and (S) are the mass percentages of carbon, hydrogen, oxygen, and sulfur in dry material (for example, cardboard is 43.73 percent carbon, so C = 43.73). For the following materials, find the higher heating values (HHV) of the dry waste and the HHV of "as received" waste accounting for the moisture content.

Material	Moisture (% as Received)	Dry Weight Percentages			
		C	H	O	S
(a) Corrugated boxes	5.20	43.73	5.70	44.93	0.21
(b) Junk mail	4.56	37.87	5.41	42.74	0.09
(c) Mixed garbage	72.0	44.99	6.43	28.76	0.52
(d) Lawn grass	75.24	46.18	5.96	36.43	0.42
(e) Demolition softwood	7.7	51.0	6.2	41.8	0.1
(f) Tires	1.02	79.1	6.8	5.9	1.5
(g) Polystyrene	0.20	87.10	8.45	3.96	0.02

Source: Niessen, 1977.

22 Draw the chemical structures of the following dioxins and furans:

(a) 1,2,3,4,7,8-hexachlorodibenzo-*p*-dioxin

(b) 1,2,3,4,6,7,8-heptochlorodibenzo-*p*-dioxin

(c) Octachlorodibenzo-*p*-dioxin (8 chlorines)

(d) 2,3,4,7,8-pentachlorodibenzofuran

(e) 1,2,3,6,7,8-hexachlorodibenzofuran

23 The United States, with 300 million people, sends about 129 million tons of municipal solid waste to landfills. What landfill area would be required for one year's worth of MSW if the landfill density is 800 lb/yd^3, cell depth is 10 ft with one lift per year, and 80 percent of the cell is MSW? What is the area in acres/yr per 1,000 people (1 acre = 43,560 ft^2)?

24 Suppose a city of 50,000 people generates 40,000 tons of MSW per year. At current recovery and recycling rates, 22 percent of that is recovered or recycled and the rest goes to a landfill. Suppose also that the landfill density is 1,000 lb/yd^3, cell depth is 10 ft, and 80 percent of the cell is MSW.

(a) What lift area would be required per year?

(b) If the current landfill site covers 50 acres, including 10 acres needed for access roads and auxiliary facilities, and two more lifts are envisioned, how long would it take to complete this landfill?

25 If the city described in Problem 24 increases its recovery and recycling rate to 40 percent, how many years would it take to complete the landfill?

26 One kilogram of as received yard trimmings is made up of approximately 620 g of moisture, 330 g of decomposible organics represented by $C_{12.76}H_{21.28}O_{9.26}N_{0.54}$, and 50 g of other constituents.

(a) At 0.0224 m^3/mol of CH_4, what volume of methane gas would be produced per kilogram of as received yard trimmings?

(b) At 890 kJ/mol, what is the energy content of that methane?

27 Food wastes are estimated to be 72 percent moisture, with the remaining portion containing 45 percent C, 6.4 percent H, 28.8 percent O, 3.3 percent N, and 16.5 percent other constituents.

(a) Write a chemical formula for the C, H, O, N portion of the waste.

(b) Write a balanced chemical reaction showing the production of methane.

(c) What fraction of the volume of gas produced is methane?

(d) At STP (1 atm and 0°C), 1 mol of gas occupies 22.4×10^{-3} m^3. What volume of methane is produced per kilogram of food waste?

(e) Find the HHV value of the methane in kilojoules per kilogram of food waste.

REFERENCES

Alter, H., 1991, The future course of solid waste management in the U.S., *Waste Management and Research,* 9:3–10.

American Forest and Paper Association, 1993, *Recovered Paper Statistical Highlights 1992,* Washington, DC.

Agarwal, J. C., 1991, Minerals, energy, and the environment, *Energy and Environment in the 21st Century,* J. W. Testor (ed.), MIT Press, Cambridge, MA.

Atkin, P., 2006, *Trash Landings: How Airlines and Airports Can Clean Up Their Recycling Programs,* Natural Resources Defense Council, Washington, DC.

Atkins, P. R., H. J. Hittner, and D. Willoughby, 1991, Some energy and environmental impacts of aluminum usage, *Energy and Environment in the 21st Century,* J. W. Testor (ed.), MIT Press, Cambridge, MA.

Ausubel, J. H., 1989, Regularities is technological development: An environmental view, *Technology and the Environment,* J. H. Ausubel and H. E. Sladovich (eds.), National Academy Press, Washington, DC.

Bukholz, D. M., 1993, Aluminum cans, *The McGraw Hill Recycling Book,* H. F. Lund (ed.), McGraw-Hill, New York.

Chapman, P. F., and F. Roberts, 1983, *Metal Resources and Energy,* Butterworths, Boston.

Corbitt, R. A., 1990, *Handbook of Environmental Engineering,* McGraw-Hill, New York.

Denison, R. A., 1996, Environmental lifecycle comparisons of recycling, landfilling and incineration: A review of recent studies, *Annual Review of Energy and Environment,* Vol. 21, Annual Reviews, Inc., Palo Alto.

Denison, R. A., and J. Ruston (eds.), 1990, *Recycling and Incineration,* Environmental Defense Fund, Island Press, Washington, DC.

Franklin Associates, Ltd., 1990, *Estimates of the Volume of MSW and Selected Components in Trash Cans and Landfills,* for the Council for Solid Waste Solutions, Washington, DC.

Franklin Associates, Ltd., 1994, *The Role of Recycling in Integrated Solid Waste Management to the Year 2000,* Keep America Beautiful, Inc., Stamford, CT.

Graedel, T. E., and B. R. Allenby, 1995, *Industrial Ecology,* Prentice Hall, Englewood Cliffs, NJ.

Graham, B., 1993, Collection equipment and vehicles, *The McGraw-Hill Recycling Handbook,* H. F. Lund (ed.), McGraw-Hill, New York.

Hannon, B. M., 1972, Bottles cans energy, *Environment,* March, 11–21.

Hocking, M. B., 1991, Relative merits of polystyrene foam and paper in hot drink cups: Implications for packaging, *Environmental Management,* 15(6):731–747.

International Petroleum Industry Environmental Conservation Association, 1992, *Climate Change and Energy Efficiency in Industry,* London.

Keoleian, G. A., D. Menerey, B. Vigon, D. Tolle, B. Cornaby, H. Latham, C. Harrison, T. Boguski, R. Hunt, and J. Sellers, 1994, *Product Life Cycle Assessment to Reduce Health Risks and Environmental Impacts,* Noyes Publications, Park Ridge, NJ.

Khan, M. Z. A., and Z. H. Abu-Gharah, 1991, New approach for estimating energy content of municipal solid waste, *Journal of Environmental Engineering,* 117(3):376–380.

Kreith, F., 1994, *Handbook of Solid Waste Management,* McGraw-Hill, New York.

Lawson, B., 1996, *Building Materials Energy and the Environment, Towards Ecologically Sustainable Development,* The Royal Australian Institute of Architects, Sydney, Australia.

McBean, E. A., F. A. Rovers, and G. J. Farquhar, 1995, *Solid Waste Landfill Engineering and Design,* Prentice Hall, Englewood Cliffs, NJ.

Miller, C., 1993, The cost of recycling at the curb, *Waste Age,* October, 46–54.

Moore, B., 1994, Breakthrough in plastics recovery, *EPA Journal,* Fall 1994, U.S. EPA, Washington, DC.

Newman, A., 1991, The greening of environmental labeling, *Environmental Science and Technology,* 25(12).

Niessen, W. R., 1977, *Handbook of Solid Waste Management,* D. G. Wilson (ed.), Van Nostrand Reinhold Co., New York.

Office of Technology Assessment, 1992, *Green Products by Design: Choices for a Cleaner Environment,* OTA-E-541, U.S. Government Printing Office, Washington, DC.

Pirages, S. W., and J. E. Johnston, 1993, Municipal waste combustion and new source performance standards: Use of scientific and technical information, *Keeping Pace with Science and Engineering, Case Studies in Environmental Regulation,* M. F. Uman (ed.), National Academy of Engineering, Washington, DC. 1993.

Rabasca, L., 1995, Recycling came of age in 1994, *Waste Age,* April, 213–222.

Repa, E. W., 2005, NSWMA's 2005 tip fee survey, *NSWMA Research Bulletin 05-03.*

Rhyner, C. R., L. J. Schwartz, R. B. Wenger, and M. G. Kohrell, 1995, *Waste Management and Resource Recovery,* Lewis Publishers, New York.

Sellers, V. R., and J. D. Sellers, 1989, *Comparative Energy and Environmental Impacts for Soft Drink Delivery Systems,* Franklin Associates, Ltd., Prairie Village, KS.

Stearns, R. A., 1982, Measuring productivity in residential solid waste collection systems, *Residential Solid Waste Collection,* GRCSA: 3-1/3-19.

Steuteville, R. A., 1994, Duales system on firmer ground in Germany, *BioCycle,* June, 61–63.

Steuteville, R., 1995, The state of garbage in America, *BioCycle,* April, 54–63.

U.S. Department of Energy, 2006, *2006 Buildings Energy Data Book,* Energy Efficiency and Renewable Energy, Washington, DC.

U.S. EPA, 1989, *Recycling Works! State and Local Solutions to Solid Waste Management Problems,* Office of Solid Waste, Washington, DC.

U.S. EPA, 1991, *Assessing the Environmental Consumer Market,* Office of Policy Planning and Evaluation, 21P-1003, Washington, DC.

U.S. EPA, 1992, *Used Dry Cell Batteries, Is a Collection Program Right for Your Community?,* Solid Waste and Emergency Response, OS-305, Washington, DC.

U.S. EPA, 1994a, *Characterization of Municipal Solid Waste in the United States: 1994 Update,* Office of Solid Waste, Washington, DC.

U.S. EPA, 1994b, *Composting Yard Trimmings and Municipal Solid Waste,* Solid Waste and Emergency Response, EPA530-R-94-003, Washington, DC.

U.S. EPA, 1998, *Characterization of Building-Related Construction and Demolition Debris in the United States,* Office of Solid Waste, EPA530-R-98-010.

U.S. EPA, 2002, *Solid Waste Management and Greenhouse Gases*, Office of Solid Waste, May.

U.S. EPA, 2006a, *Municipal Solid Waste in the United States: 2005 Facts and Figures*, Office of Solid Waste, EPA530-R-011, October.

U.S. EPA, 2006b, *Solid Waste Management and Greenhouse Gases: A Life-Cycle Assessment of Emissions and Sinks,* 3rd ed., Washington, DC.

White, P. R., M. Franke, and P. Hindle, 1995, *Integrated Solid Waste Management: A Lifecycle Inventory,* Blackie Academic & Professional, New York.

World Resources Institute, 1992, *World Resources 1992–93,* Oxford University Press, New York.

Young, J. E., 1992, *Mining the Earth,* Worldwatch Technical Paper 109, Worldwatch Institute, Washington, DC.

This page intentionally left blank

Index

Page references followed by "f" indicate illustrated figures or photographs; followed by "t" indicates a table.