Drinking Water Quality and Treatment

Dan Askenaizer
Montgomery Watson Engineers

GLOSSARY

Disinfection Water systems add disinfectants to destroy microorganisms that can cause disease in humans. Primary methods of disinfection include chlorination, chloramines, chlorine dioxide, ozone, and ultraviolet light.

Disinfection byproducts Side reactions can occur in water when chemical oxidants such as chlorine and ozone are used to control potentially pathogenic microorganisms. These reactions can form low levels of disinfection byproducts, several of which have been regulated for potential adverse human health effects.

Filtration Filtration is the process of removing suspended solids from water by passing the water through a permeable fabric or porous bed of material. The most common filtration process employs a granular media (e.g., sand, anthracite coal). Filtration is usually a combination of physical and chemical processes.

Membrane filtration Membrane separation processes use semipermeable membranes to separate impurities from water. The membranes are selectively permeable to water and certain solutes. A driving force is used to force the water to pass through the membrane, leaving the impurities behind as a concentrate. The amount and type of material removed depends upon the type of membrane, the type and amount of the driving force and the characteristics of the water.

Ozone Ozone is a colorless gas that is extremely unstable and is a strong oxidizing agent that is capable or reacting with a wide variety of organic and inorganic solutes in water. Effectiveness of ozone disinfection is a function of the pH, temperature of water and method for ozone application.

Ultraviolet light Ultraviolet light is electromagnetic energy that is located in the electromagnetic spectrum at wavelengths between those of X-rays and visible light. UV light that is effective is destroying microbial entities in located in the 200- to 310-nm range of the energy spectrum. Most typical applications of UV at water treatment plants apply UV light in the wavelength range of 250 to 270 nm.

WATER is the most abundant compound on the surface of the earth. The physical and chemical properties of water are important issues with regard to water supply, water quality, and water treatment processes. With the increasing growth of urban areas and activities that can possibly introduce contaminants into drinking water sources, source water protection has become an increasingly important facet of providing safe drinking water. Water treatment professionals deal with a wide range of water qualities and they have a growing array of treatment methods at their disposal. During the 1970s and 1980s there was a growing interest and concern with groundwater contamination due to organic chemicals such as solvents and pesticides. In the early 1990s and in to the new millennium, there is a growing awareness of the need to balance the risks of the need to disinfect the water to reduce the threat of disease from microorganisms against the potential health risks from disinfection byproducts that are formed as a results of adding a disinfectant. Microorganisms such as *Giardia* and *Cryptosporidium* present challenges to regulators and water treatment engineers. The purpose of this chapter is to provide an overview of drinking water quality and treatment methods.

I. OVERVIEW OF DRINKING WATER SOURCES

The vast majority of fresh water in the world is provided by precipitation resulting from the evaporation of seawater. This transfer of moisture from the sea to land and back to the sea is referred to as the hydrologic cycle. Figure 1 presents a depiction of the hydrologic cycle.

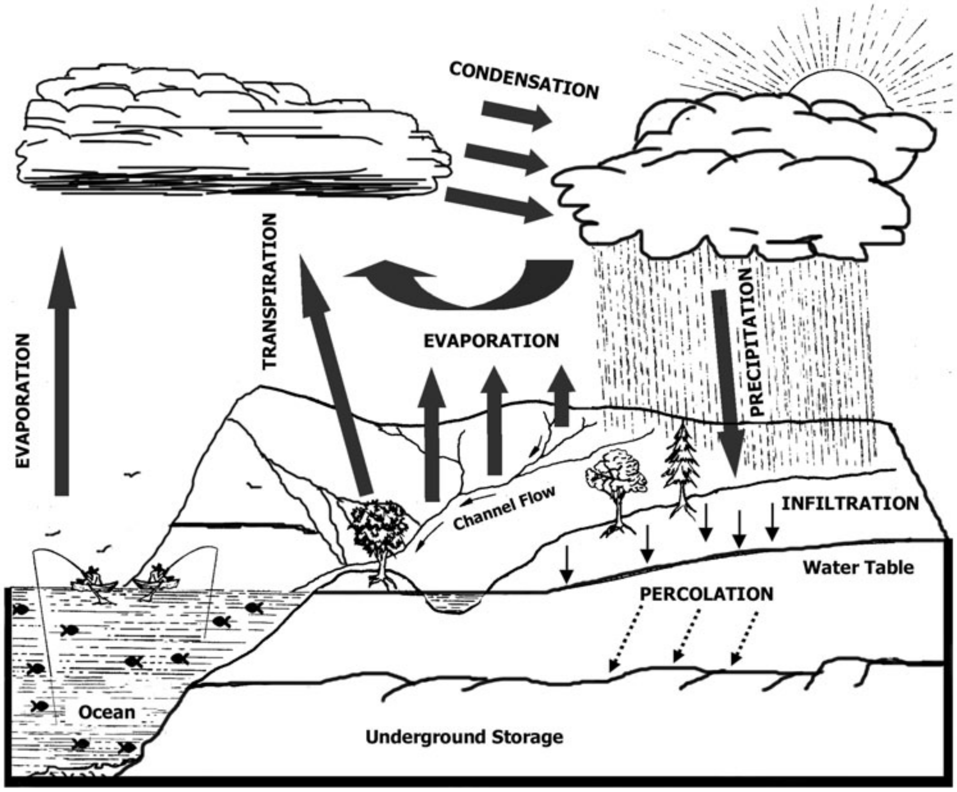


FIGURE 1 Hydrologic cycle.

About two thirds of the precipitation which reaches land surface is returned to the atmosphere by evaporation from water surfaces, soil, and vegetation, and through transpiration by plants. The remainder of the precipitation returns ultimately to the ocean through surface or underground channels.

The following section presents a brief overview of three sources of water: surface water sources, groundwater sources, and the use of reclaimed water. In the United States around 35% of the population served by community water systems drink groundwater while nearly 65% of the population served by community water systems receive water taken primarily from surface water sources. Increasingly, communities are looking to reclaimed water to meet a portion of their needs.

A. Surface Water

Surface waters sources for drinking water include lakes, rivers, canals, runoff, and impounding reservoirs. Water quality conditions in streams and rivers can change dramatically over a short period of time. Surface water sources such as streams and rivers are susceptible to chemical spills and accidental releases.

There are three stages (trophic levels) in the life cycle of a body of water. These are oligotrophic (low nutrients, minimal microbiological activity), mesotrophic (moderate nutrients, moderate microbiological activity), and eutrophic (high nutrients, high microbiological activity).

Water quality in lakes and reservoirs can change throughout the year as the water can stratify during warmer months. Thermal stratification can be a significant process in many lakes. During the warmer months of the year the warmer water (and therefore less dense water) will stay near the surface of the water body while the cooler and more dense water is trapped below. In the absence of strong winds there will be little mixing of the colder, denser water below with the warmer, less dense water near the surface. Under certain conditions where adequate nutrients are present this can lead to a depletion of oxygen in the lower parts of the water body and can thus cause water quality issues such as taste and odor problems and problems with iron and manganese (which will have increased solubility under the reducing conditions). As temperatures cool and the temperature of the surface of the lake cools, this together with wind action can cause mixing throughout the reservoir.

Depending upon nutrient, temperature, and carbonate conditions the upper regions of a lake or reservoir can be susceptible to algal blooms (which can cause changes in source water turbidity, alkalinity, taste, odor, and pH) and can make it difficult to treat the water near the surface of the lake.

B. Groundwater

Groundwater sources are beneath the land surface and include springs and wells. As can be seen from the hydrologic cycle, when rain falls to the ground, some water flows along the land to streams or lakes, some water evaporates into the atmosphere, some is taken up by plants, and some seeps into the ground. As water seeps into the ground, it enters a zone referred to as the unsaturated zone or vadose zone. Water moves through the unsaturated zone into the saturated zone, where the interconnected openings between rock particles are filled with water.

Groundwater quality is typically constant over time; however, changes in hydrogeological conditions can lead to differences in water quality over a relatively short distance. The chemistry of the groundwater is influenced by the composition of the aquifer and by the chemical and biological processes that occur as water infiltrates through the aquifer.

C. Reclaimed Water

An additional source of water that must be included in any discussion of sources is the use of reclaimed water for nonpotable and potable uses. Reclaimed water sources include desalination of brackish water or seawater, and reuse or recycling of wastewater through the application of appropriate treatment technology. In some cities, reclaimed wastewater has been used to irrigate golf courses and parks. In addition, several communities have seriously evaluated and studied the use of reclaimed wastewater to augment the drinking water supply (e.g., adding highly treated wastewater meeting drinking water standards under the Safe Drinking Water Act to a reservoir that is the source for raw water for the community's drinking water treatment plant). This drive toward utilizing a greater amount of reclaimed water comes from growing urban populations and constraints on the development of new water sources. Some public health authorities are reluctant to support or endorse the planned use of reclaimed water to augment a drinking water supply. However, it is already the case that there are many surface water sources (river and streams) that are subject to sewage contamination prior to their use as a potable drinking water supply. In these instances, in effect, the cities are practicing unplanned indirect potable reuse.

II. PROTECTION OF SOURCE WATER

Natural waters acquire their chemical characteristics by dissolution and by chemical reactions with solids, liquids, and gases with which they have come into contact during the various parts of the hydrological cycle. An example

TABLE I Water Quality Results for Three Different Sources[a] (from Snoeyink, Jenkins, 1980)

Constituent (mg/L)	Source 1[b] (reservoir)	Source 2[c] (river water)	Source 3[d] (groundwater)
SiO_2	9.5	1.2	10
Fe(III)	0.07	0.02	0.09
Ca^{2+}	4.0	36	92
Mg^{2+}	1.1	8.1	34
Na^+	2.6	6.5	8.2
K^+	0.6	1.2	1.4
HCO_3^-	18.3	119	339
SO_4^{2-}	1.6	22	84
Cl^-	2.0	13	9.6
NO_3^-	0.41	0.1	13
Total dissolved solids	34	165	434
Total hardness as $CaCO_3$	14.6	123	369

[a] (From Snoeyink, V. L., Jenkins, D. (1980). "Water Chemistry," John Wiley and Sons, New York.)

[b] (Source 1. Pardee Reservoir. East Bay Municipal Utility District, Oakland, CA, 1976.)

[c] (Source 2. Niagara River, Niagara Falls, New York.)

[d] (Source 3. Groundwater, Dayton, OH.)

of this would be weathering reactions, which are caused by the interaction of water and atmosphere with the crust of the earth. Table I presents examples of water quality for three different sources

Beginning in the early 1970s there was a growing interest in the presence of and increased understanding of health effects associated with low levels of organic compounds in water. Organic compounds in water can occur due to (1) degradation of naturally occurring organic material (i.e., leaves), (2) human activities such as handling and disposal of chemicals, and (3) chemicals reactions during the treatment of water (i.e., the production of disinfection byproducts).

Protecting sources of drinking water has become an increasingly important aspect of providing safe drinking water. Source water quality management is the first step toward ensuring an adequate supply of safe drinking water.

Potential sources of contamination or water quality problems in source water include the following:

- Climate (primarily precipitation causing increased levels of sediment, turbidity, and other contaminants)
- Temperature (can affect biological activity, oxygen saturation, and mass transfer coefficients)
- Watershed characteristics (steep slopes, vegetative cover, wildlife)
- Geology (e.g., mineral content)
- Presence of nutrients (can stimulate microbiological growth)
- Saltwater intrusion (increased salinity)
- Wastewater discharges (bacterial contamination, depletion of dissolved oxygen increased levels of inorganic and organic contaminants)
- Industrial discharges (accidental or planned discharges of chemical contaminants)
- Hazardous waste facilities (release of toxic, reactive, corrosive, or flammable contaminants)
- Mine drainage (acid discharges, increases in sediments, turbidity, color)
- Agricultural runoff (pesticides, herbicides, fertilizers)
- Livestock (microbial contamination, increased erosion increased nitrates)
- Urban Runoff (petroleum products, metals, salts, silts, and sediments)
- Land development (increased erosion and sediment loading, increased human activities that can release contaminants to the environment)
- Atmospheric deposition (acid rain)
- Recreational activities (swimming, boating, camping)

Activities to protect and enhance the quality of surface water sources include conducting sanitary surveys, programs to monitor source water quality and activities to provide watershed control. Other activities that can be undertaken to protect source water include storm-water management, development and implementation of emergency response procedures (to contain and clean up spills to prevent contamination of source water).

A sanitary survey is an on-site review of the water source, facilities, equipment, operation, and maintenance of a public water system to evaluate the adequacy of the source, facilities, equipment, operation, and maintenance for producing and distributing drinking water.

Monitoring programs can be conducted of both chemical and microbiological parameters at locations throughout a source of supply. A thorough monitoring program can provide valuable information toward an understanding and identification of possible changes in source water quality.

For groundwater sources, the potential effectiveness of a groundwater management program depends upon the degree to which potential contamination is accurately identified and the practicality of the response, remediation, and protection measures that are developed. All residential and commercial development and industrial and agricultural activities within the well field zone of influence and upstream of the general direction of groundwater flow should be investigated, and monitoring programs can be implemented to detect and control contaminants that could be introduced into the groundwater. For surface water sources, source protection can involve such activities as storm water management and controls on activities in a watershed that could impact the quality of the source water, including activities such as fishing, boating, swimming, hunting, and camping.

III. OVERVIEW OF BASIC DRINKING WATER TREATMENT PROCESSES

The amount and type of treatment applied by a given public water system will vary depending upon the source type and quality. Many, if not most, groundwater systems can provide adequate treatment that involves little or no treatment of the source. Surface waters, however, are exposed to the atmosphere and surface runoff and are more likely to contain contaminants. Surface water systems, therefore, must implement a greater level of treatment to provide safe and potable drinking water.

Water utilities can use a variety of treatment processes together at a single treatment plant to remove contaminants, remove turbidity, and provide disinfection. The most common physical processes used at public water systems with surface water supplies include coagulation, flocculation, sedimentation, and filtration. The following sections provide descriptions of physical and chemical processes that can be used to treat drinking water. In addition to a description of some basic water treatment methods (coagulation, flocculation, sedimentation and filtration, slow sand filtration, lime-soda softening, granular activated carbon), the following section also provides description of some typical water quality issues that can be addressed through treatment (iron and manganese removal, taste and odor problem, corrosion control) and additional treatment methods (membranes and ion exchange).

A. Coagulation

Colloidal suspended particles in water have like electrical charges that tend to keep them in suspension. Coagulation is defined as the destabilization of the charge on colloids and suspended solids, including bacteria and viruses, by use of a coagulant. The most commonly used coagulants are metal salt coagulants such as aluminum sulfate, ferric chloride, and ferric sulfate. In water, metal salts undergo hydrolysis. The products of this hydrolysis readily adsorb to colloid particles and cause the destabilization of their electrical charge. An important parameter in determining the effectiveness of a given coagulant is the pH of the water.

Synthetic polymers such as polydiallyl dimethyl ammonium (PDADMA) are also used as coagulants. Organic polymers can be used as the primary coagulant or as a coagulant aid. Polymers are classified as anionic, cationic, or nonionic. Anionic polymers ionize in water to form negative sites along the polymer molecule. Cationic polymers ionize to form positively charged sites, while nonionic polymers exhibit only slight ionization.

Flash or rapid mixing is an important part of coagulation. The purpose of flash mixing is to quickly and uniformly disperse water treatment chemicals throughout the water. Effective flash mixing is especially important when using metal salt coagulants, since their hydrolysis occurs within a second and subsequent adsorption to colloidal particles is almost immediate. Rapid mix processes can typically be accomplished in just a few minutes.

B. Flocculation

Flocculation is a gentle mixing phase that follows the initial rapid mix step. During the flocculation step the chemically treated water is sent into a basin where the suspended particles can collide and form heavier particles called floc. Gentle agitation and appropriate detention times are used to allow this process to occur. Typical time for the flocculation step could be on the order of 15–30 min. After the flocculation step the water can then move into the sedimentation step.

C. Sedimentation

Also known as clarification, the purpose of sedimentation is to remove a majority of the settleable solids by gravitational settling. By removing the majority of the settleable solids in the sedimentation step this will maximize downstream unit processes such as filtration. During the sedimentation step, the velocity of water is decreased so that suspended material can settle out of the water stream by gravity. The key to effective sedimentation is proper coagulation and flocculation of suspended material in the raw water. Removal and disposal of the sludge from the sedimentation basin are important parts of the treatment process.

D. Filtration

Filtration is the process of removing suspended solids from water by passing the water through a permeable fabric or porous bed of material. The most common filtration process employs a granular media (e.g., sand, anthracite coal). Filtration is usually a combination of physical and chemical processes. Mechanical straining removes some particles by trapping them between the grains of the filter medium (such as sand). Adhesion is an equally important process by which suspended particles stick to the surface of filter grains or previously deposited material. The average filtration rate in the United States is 5–6 gal/min/ft^2 of filter area. At a conventional treatment plant the filters are

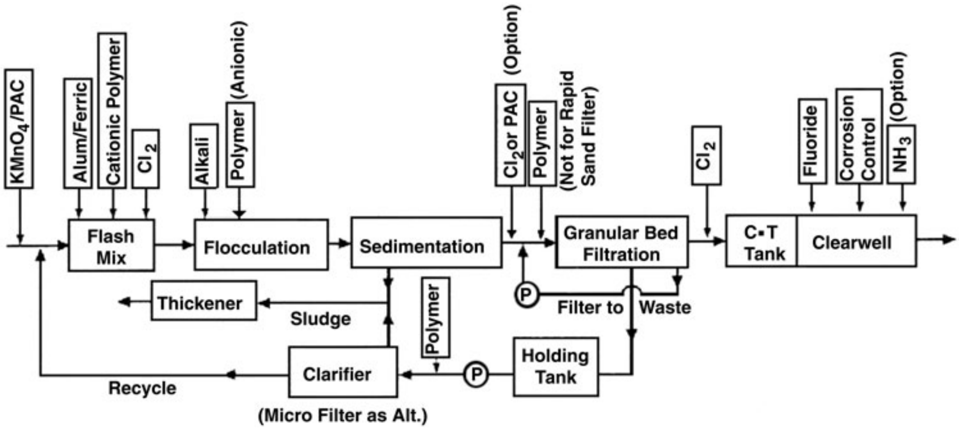


FIGURE 2 Conventional treatment process.

preceded by coagulation, flocculation, and sedimentation. At a direct filtration plant, the filters are preceded by coagulation and flocculation only; with the floc being removed directly by the filters. Figure 2 presents a schematic for a conventional treatment plant together with examples of potential application points for chemical addition. Figure 3 presents the schematic for a direct filtration plant.

1. Slow Sand Filters

Slow sand filters are operated at very low filtration rates without the use of coagulation. Slow sand filters are a simple, reliable and easy to operate system. The filtration rate for slow sand filters is typically 50–100 times slower than that of granular media filters. Therefore, a much larger area is needed for the filter bed to produce an equivalent amount of water. Contaminants are removed from water through a combination and physical straining and microbiological processes in a slow sand filter. When in operation, the surface of the filter bed is covered by a thin layer of medium, known as the "schmutzdecke." This layer contains a large variety of microorganisms and enables these filters to remove large numbers of bacteria. Slow sand filters do not require highly trained operators, have minimal power requirements, and can tolerate reasonable hydraulic and solids shock loadings. Some of the disadvantages of slow sand filters include the large amount of land they require; the filters can be easily clogged by excessive amounts of algae; they are not very effective at removing color; and intermittent operation of the filters may degrade the quality of the filter effluent by promoting anaerobic conditions within the filter bed. The filters must be periodically cleaned by scraping off a thin layer of sand from the surface of the filter bed.

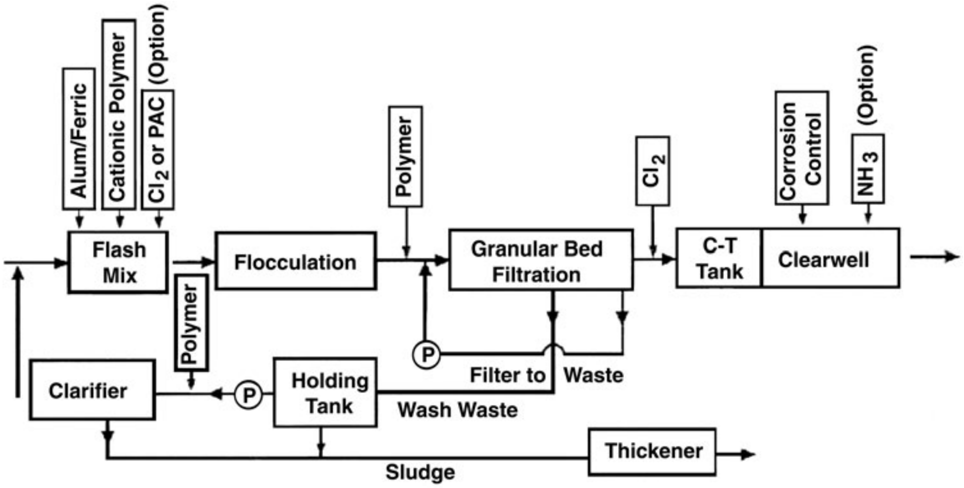


FIGURE 3 Direct filtration process.

TABLE II General Classification of Water Hardness[a]

Level of hardness (mg/L)	Classification
0–75	Soft water
75–150	Moderately hard water
150–300	Hard water
>300	Very hard water

[a] Hardness is sometimes expressed as grains per gallon, where 1 grain of $CaCO_3$/gallon is equivalent to 17.1 mg/L as $CaCO_3$.

E. Lime-Soda Softening

The hardness of water is defined as the concentrations of divalent metallic cations in water and is expressed as mg/L of $CaC0_3$. Table II presents a general classification of water hardness. The principal divalent metallic cations that contribute to the hardness of a water are calcium and magnesium, with contributions from iron, manganese, and strontium. The main purpose of water softening is to reduce the levels of calcium and magnesium in the water to reduce the hardness of the water. In the United States there are over 1000 treatment plants were softening is practiced. Historically, softening was important due to the high consumption of soap by hard water. With today's synthetic detergents this is no longer a major issue. However, there are other benefits to softening including removal of heavy metals, metallic elements and organic compounds; effective destruction of bacteria, viruses and algae; and improvement in boiler feed water and cooling waters.

While there are many variations, the primary method of softening a water is the addition of lime (calcium hydroxide) and soda ash (sodium carbonate) to the water. The purpose of adding these compounds to a water is to change the hardness compounds such that they become insoluble and precipitate (e.g., calcium and magnesium are converted, respectively, to calcium carbonate ($CaCO_3$, partially soluble) and magnesium hydroxide ($Mg(OH)_2$). In addition to chemical treatment, ion exchange resins and membranes can be used to soften a water.

F. Granular Activated Carbon

Granular activated carbon (GAC) has been used as a substitute for granular filter media and as an additional process in conventional treatment plants for the removal of organic compounds including compounds producing taste and odors, pesticides and other synthetic organic compounds. GAC can be manufactured from a large variety of materials including wood, nuts, shells, coal, peat, or petroleum residues. GAC used in water treatment plants is typically manufactured from bituminous or lignite coal by heating the coal under anaerobic conditions. The product of this process is then activated by exposure to a mixture of steam and air at a temperature of 1500°F, which oxidizes the surface of the carbon pores and allows the surface to attract and hold organic compounds. After being used at a water treatment plant, the spent GAC can be regenerated through steam, thermal regeneration and chemical means.

G. Iron and Manganese Removal

The presence of iron and manganese is drinking water have been associated with unpleasant taste and odors, staining of laundry and fixtures as well as causing the formation of mineral deposits. Iron and manganese are commonly found in soil in insoluble forms. When water contains carbon dioxide (or is an acidic water) then the ferric iron can be reduced to the ferrous form (which is soluble in water) and manganese is reduced to a form that is also soluble in water. Treatment to remove iron and manganese includes oxidation (aeration, chlorination, chlorine dioxide, potassium permanganate, ozone) followed by clarification and filtration; ion exchange; stabilization through use of a sequestering agent, and lime softening.

H. Taste and Odor Control

The most frequent causes of taste and odor in a drinking water are metabolites of algae (most commonly blue-green algae), actinomycetes (filamentous bacteria), and decaying vegetation. Other potential causes of taste and odor issues are hydrogen sulfide, agricultural runoff, industrial chemical spills, and sewage pollution.

The most common odor-producing compounds are geosmin and 2-methlyisoborneal (MIB) which can impart objectionable odor at very low concentrations. These compounds are responsible for the earthy-musty odors in water and have been isolated from actinomycetes (*Actinomyces*, *Nocardia*, *Streptomyces*) and from blue-green algae (e.g., *Anabaena* and *Oscillatoria*). Control methods for taste and odor include: prevention at the source (reservoir mixing, aquatic plant control, reservoir management), removal of a particular constituent at the treatment plant (aeration, oxidation, adsorption), and control within the distribution system (minimization of dead-ends, use of blow-off and cleanout assemblies, distribution system flushing).

I. Corrosion Control

All waters are corrosive to some degree. Corrosion can reduce the life of a pipe by reducing wall thickness until there are leaks, it can result in encrustations that reduce the effective carrying capacity and can result in corrosion by-products at the consumer's tap that have public health implications (i.e., lead and copper). The tendency of a water to be corrosive will depend on its physical and chemical characteristics as well as on the nature of the material it comes into contact with. The most common types of materials used in distribution systems include cast iron, ductile iron (cast iron containing a small amount of alloying elements such as magnesium), asbestos-cement, steel, copper, galvanized iron, and plastics.

The mechanisms of corrosion in a water distribution system are typically a complex and interrelated combination of physical, chemical, and even biological processes. The basic principles that affect corrosion of materials include solubility, described by chemical equilibria among materials and constituents in the water; and the rate of dissolution, which is described by chemical and electrochemical kinetics. Electrochemical corrosion occurs where two different metals have an electropotential between them are immersed in a common body of water. All waters can act as an electrolyte, but the degree to which they do so depends on the dissolved chemicals present. High-velocity water flow can cause pitting and erosion of surfaces due to cavitation. Certain types of bacteria, sulfate-reducing bacteria, and iron bacteria can also can cause internal corrosion in piping material. Other factors that can influence the corrosivity of a given water include the concentration of dissolved salts in the water, level of dissolved gases in the water, water temperature, and stress and fatigue.

In terms of lead solubility, the most important water quality parameters are pH, alkalinity, dissolved inorganic carbonate and orthophosphate levels. In general, low pH levels have been associated with higher lead levels at the tap. Soft waters that are low in pH and alkalinity are often corrosive toward lead and other metals. Water treatment for controlling and/or reducing lead leaching include pH adjustment, carbonate adjustment, use of corrosion inhibitors, calcium carbonate deposition, as well as painting, coating, and usage of cathodic protection systems.

J. Membrane Filtration Technology

Membrane separation processes use semipermeable membranes to separate impurities from water. The membranes are selectively permeable to water and certain solutes. A driving force is used to force the water to pass through the membrane, leaving the impurities behind as a concentrate. The amount and type of material removed depends upon the type of membrane, the type and amount of the driving force, and the characteristics of the water. Important issues involved with the operation of membrane systems include membrane fouling and disposal of the concentrate.

There are two classes of membrane treatment systems. These include low-pressure membrane systems (such as microfiltration (MF) and ultrafiltration (UF)), and high-pressure membrane systems (such as nanofiltration (NF) and reverse osmosis (RO)). Low-pressure membranes are operated at pressures ranging from 10 to 30 lb/in.^{2} (psi), whereas high-pressure membranes, including nanofiltration are operated at pressures ranging from 75 to 250 psi. Figure 4 presents a general description of various membranes types and their ability to remove impurities from water.

MF can remove particles that are greater than 0.5 μm in diameter. UF is capable of removing colloids, bacteria, viruses, and high-molecular-weight organic compounds.

Some advantages of using low-pressure membranes include small waste stream, limited chemical usage, a

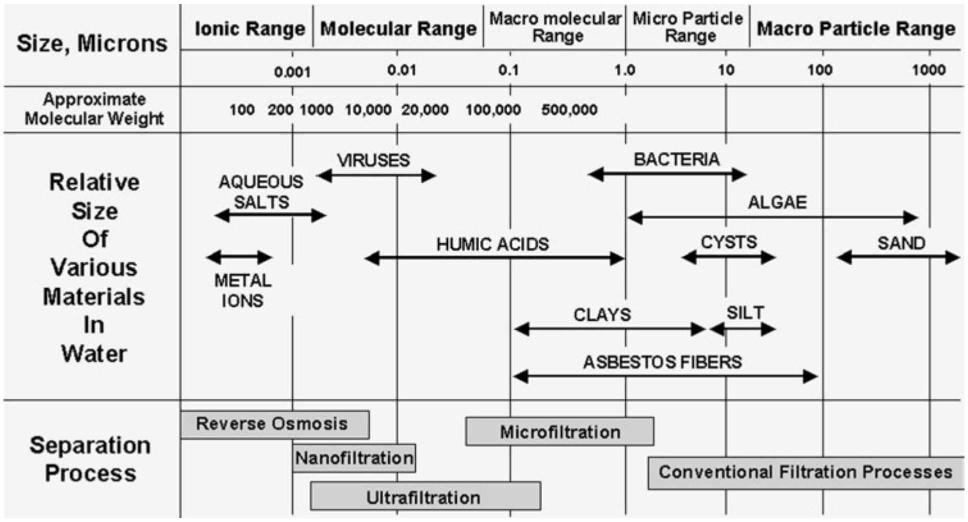


FIGURE 4 Membrane size ranges.

relatively small area needed for equipment, good pathogen reduction, and no disinfection by-product formation. Low-pressure membranes, however, are ineffective for the removal of dissolved organic matter. Therefore, color-causing organic matter, taste-and-odor causing compounds such as Geosmin and methylisoborneol (MIB), and man-made chemicals can pass through the membranes into the treated water.

NF membranes have been used successfully for groundwater softening since they achieve greater than 90% rejection of divalent ions such as calcium and magnesium. NF membranes are also capable of removing greater than 90% of natural organic matter present in the water. Therefore, they are also excellent candidates for the removal of color, and also DBP precursor material.

RO membranes have long been used for the desalination of seawater around the world. These membranes can consistently remove about 99% of the total dissolved solids (TDS) present in the water, including monovalent ions such as chloride, bromide, and sodium.

K. Ion-Exchange Technology

Ion-Exchange (IX) technology removes undesirable ions from raw water and exchanges them for desirable ions. The two most common applications of ion exchange are for water softening (Ca^{2+} and Mg^{2+} removal), either at the water treatment plant or as a point-of-entry (POE) treatment process, and for industrial applications, such as the production of fully demineralized water.

Examples of ions that can be removed using IX include nitrate, arsenic, selenium, barium, radium, lead, fluoride, and chromate.

In the IX process water passes through a resin bed where contaminant ions present in the water are exchanged with ions on the resin surface, thus removing the contaminant ions from the water and concentrating them on the resin. The resin is frequently regenerated to remove the contaminant from the resin surface and replenish the resin with the original exchange ion. There are four primary types of IX resins:

1. Strong Acid Cationic (SAC) Resin
2. Weak Acid Cationic (WAC) Resin
3. Strong Base Anionic (SBA) Resin
4. Weak Base Anionic (WBA) Resin

SAC and WAC resins are used to remove cations from water (e.g., Ca^{2+}, Mg^{2+}, Ra^{2+}, Ba^{2+}, Pb^{2+}), while SBA and WBA resins are used to remove anions from water (e.g., NO_3^-, SO_4^{2-}, ClO_4^-, $HAsO_4^{2-}$, SeO_3^{2-}). During water softening, SAC resins can remove both carbonate and noncarbonate hardness, whereas WAC resins can only remove carbonate hardness. On the other hand, WAC resins are easier to regenerate than SAC resins and do not result in sodium concentration increase as SAC resins do.

An important issue for the application of IX technology is the waste stream produced by the process. The volume of the waste stream is not large on the order of 2 to 5% of the water volume treated. However, the waste stream contains a high concentration of acid (HCl), base (NaOH), or salt (NaCl). The waste stream also can contain a high concentration of the contaminant removed from the water (e.g., NO_3^-, Pb^{2+}, etc. . .).

IV. DISINFECTION

Disinfection is an important step in ensuring that water is safe to drink. Water systems add disinfectants to destroy microorganisms that can cause disease in humans. Primary methods of disinfection include chlorination, chloramines, chlorine dioxide, ozone, and ultraviolet light.

From a historical perspective, the Chick–Watson model has been the predominant model used to describe the kinetics of using disinfectants to inactivate microorganisms. Chick's law (1908) expresses the rate of destruction of microorganisms using the relationship of a first-order chemical reaction. Watson (1908) refined the equation to produce an empirical relationship that reflected changes in the disinfectant concentration. The Chick–Watson model can be expressed as follows:

$$\log N/N_o = -k\mathrm{C}^n T,$$

where $N_o =$ the initial concentration of bacteria, $N =$ the concentration of surviving bacteria at time T, $k =$ rate constant (coefficient of specific lethality), C = concentration of the disinfectant, and $n =$ concentration of the dilution (empirically derived). The Chick–Watson model leads to a correlation between the level of inactivation and the product of the concentration of the disinfectant (C) and the contact time (T). The United States Environmental Protection Agency's (USEPA) Surface Water Treatment Rule includes tables that associate specific CT values (mg*min/L) with a given level of inactivation of *Giardia* and viruses.

A. Chlorine

Chlorination of potable water has been practiced in the United States since 1903. Chlorine can be applied by drinking water treatment plants as chlorine gas, sodium hypochlorite solutions, or as solid calcium hypochlorite. Free chlorine reacts rapidly with many substances in the water, including microorganisms. The effectiveness of chlorine to provide disinfection is affected by many variables including amount of oxidizable substances in the water (that exert a demand on the chlorine), concentration of particulate matter, pH, temperature, contact time, and the level of residual chlorine. The formation of disinfection byproducts including trihalomethanes (THMs) is one of the major concerns with regard to the use of chlorine disinfection.

1. Chlorine Gas

Elemental chlorine is a toxic, yellow-green gas at normal pressures. At high pressures, it is a liquid. Chlorine gas is typically supplied as a liquid in high pressure cylinders. Chlorine gas is released from the liquid chlorine cylinder by a pressure reducing and flow control valve operating at pressures less than atmospheric. The gas is led to an injector in the water supply pipe where highly pressurized water is passed through a venturi orifice creasing a vacuum that draws the chlorine into the water stream. Adequate mixing and contact time must be provided after injection to ensure complete disinfection of pathogens. Gaseous chlorine, when added to water, rapidly hydrolyzes to hypochlorous acid (HOCl) and hydrochloric acid (HCl) as follows:

$$\mathrm{Cl_2 + H_2O \rightleftarrows HOCl + H^+ + Cl^-}$$

Hypochlorous acid is subject to further reaction including disinfection, reactions with various organic and inorganic compounds or dissociation to hydrogen and hypochlorite ion ($\mathrm{OCl^-}$) as follows:

$$\mathrm{HOCl \rightleftarrows H^+ + OCl^-}$$

The relative concentrations of hypochlorous acid and hypochlorite ion are dependent on the pH and the temperature. Generally, hypochlorous acid is a better disinfecting agent than is hypochlorite ion.

2. Sodium Hypochlorite Solution

Sodium hypochlorite is available as a solution in concentrations of 5 to 15% chlorine. Sodium hypochlorite is easier to handle than chlorine gas or calcium hypochlorite. It is, however, extremely corrosive and should be kept away from equipment that could be damaged by corrosion. Hypochlorite solutions decompose and should not be stored for more than 1 month and must be stored in a cool, dark, dry area.

The sodium hypochlorite solution is diluted with water in a mixing/holding tank. The diluted solution is then injected by a chemical pump in to the water supply pipe at a controlled rate. Adequate mixing and contact time must be provided. Sodium hypochlorite can be generated on site by using electrolysis of sodium chloride solution. Hydrogen gas is given off as a by-product and must be safely dispersed.

3. Calcium Hypochlorite

Calcium hypochlorite is a white solid that contains 65% available chlorine and dissolves easily in water. Calcium hypochlorite is very stable and can be stored for an extended period of time. Calcium hypochlorite is a corrosive material with a strong odor. Reactions between calcium hypochlorite and organic material can generate enough heat to cause a fire or explosion. It must be kept away from organic materials such as wood, cloth, and petroleum products. Calcium hypochlorite readily absorbs moisture, forming chlorine gas.

Calcium hypochlorite can be dissolved in a mixing/holding tank and injected in the same manner as sodium hypochlorite. Alternatively, where the pressure can be lowered to atmospheric, such as at a storage tank, tablets of calcium hypochlorite can be directly dissolved in the free flowing water.

V. DISTRIBUTION SYSTEM

While beyond the scope of this chapter, a vital component toward ensuring the delivery of safe drinking water is the series of transmission mains, fire hydrants, valves, pump stations, booster chlorination stations, storage reservoirs, standpipes, and service lines that constitute the distribution system. The proper design, construction material, and maintaining the integrity of the distribution system and the individual components are important to maintain the microbiological safety of the drinking water.

VI. DISINFECTION BY-PRODUCTS

In 1974 trihalomethanes (THMs), were first identified in finished drinking water. In 1975 the USEPA conducted the National Organics Reconnaissance Survey of 80 cities in the United States and observed that the occurrence of THMs was widespread in chlorinated drinking water and was associated with the practice of using chlorine to disinfect the water. Later studies demonstrated that THMs continued to form in the distribution system.

THMs are a class of organic compounds where three hydrogen atoms in the methane molecule have been replaced with three halogen atoms (chlorine or bromine). The four THMs identified were chloroform, bromodichloromethane, dibromochloromethane, and bromoform. THMs were important to regulators initially as suspected human carcinogens. Recent information suggests that some disinfection by-products (DBPs) may have adverse developmental and reproductive impacts.

The production of THMs can be shown simply as follows:

$$\text{Free Chlorine} + \text{Natural Organics (precursors)} + \text{bromide} \rightarrow \text{THMs} + \text{Other byproducts}$$

The natural organic precursors in the raw water are generally humic and fulvic acids. The humic and fulvic acids enter water from the degradation of algae, leaves, bark, wood, and soil. Humic and fulvic acids comprise the major fraction of dissolved organic carbon in most natural waters. Factors that can influence the concentration of THMs in a given water include the concentration and type of precursor material, the concentration of free chlorine residual, contact time with chlorine, water temperature, water pH, and the concentration of bromide ion. The brominated species of THMs has been attributed to the presence of bromide in the raw water and the fact that hypochlorous acid can oxidize the bromide ion to hypobromous acid. Hypobromous acid can undergo addition and substitution reactions with various types of organic compounds in water to produce halogenated organics such as the THMs.

Since the initial observation on the occurrence and formation of THMs, additional research has shown that all chemical disinfectants (chlorine, monochloramines, ozone, and chlorine dioxide) can form various types of disinfection byproducts (DBPs).

Table III presents a list of disinfection byproducts that can be formed through the use of chemical disinfectants.

Approaches for controlling or limiting the formation of THMs include the following: use of an alternative source (such as groundwater versus a surface water with higher concentration of organic precursor material), improved flocculation and sedimentation (enhanced coagulation) to remove the DBP precursor material, moving the point of chlorine application within the treatment plant to minimize contact time, and the use of alternative disinfectants.

TABLE III Disinfection Byproducts

Trihalomethanes	Haloacetonitriles
Chloroform	Dichloroacetonitriles
Bromodichloromethane	Trichloroacetonitrile
Dibromochloromethane	Dibromoacetonitrile
Bromoform	Tribromoacetonitrile
	Bromochloroacetonitrile
Haloacetic acids	
Monochloroacetic acid	Halopicrins
Dichloroacetic acid	Chloropicrin
Trichloroacetic acid	Bromopicrin
Monobromoacetic acid	
Dibromoacetic acid	Chloral hydrate
Tribromoacetic acid	
Bromochloroacetic acid	Oxyhalides
Bromodichloroacetic acid	Chlorite
	Chlorate
Aldehydes	Bromate
Formaldehyde	
Acetaldehyde	Haloketones
Glyoxal	1,1-Dichloroacetone
Methyl glyoxal	1,1,1-Trichloroacetone

VII. ALTERNATIVES TO CHLORINATION

Over the years, public water systems have used various alternatives to chlorine disinfection. The more common alternatives include (1) chloramines, (2) Ozone, (3) Chlorine Dioxide, and (4) Ultraviolet irradiation.

A. Chloramines

Chloramines (referred to as combined chlorine) are formed when water containing ammonia is chlorinated. There are three inorganic chloramine species: monochloramine (NH_2Cl), dichloramine ($NHCl_2$), and trichloramine (NCl_3). The species of chloramines that are formed depends on factors such as the ratio of chlorine to ammonia–nitrogen, chlorine dose, temperature, pH, and alkalinity.

The principal reactions for chloramine formation are presented below:

$$NH_3(aq) + HOCl \rightleftharpoons NH_2Cl + H_2O \text{ (monochloramine)}$$

$$NH_2Cl + HOCl \rightleftharpoons NHCl_2 + H_2O \text{ (dichloramine)}$$

$$NHCl_2 + HOCl \rightleftharpoons NCl_3 + H_2O \text{ (trichloramine)}$$

At a typical water treatment plant, the dominant chloramine species will be monochloramines. Chloramine generating reactions are 99% complete within a few minutes. Chloramines are a weak disinfectant that are less effective against viruses or protozoa than free chlorine but produce fewer disinfection by-products. The use of chloramines as a DBP control strategy is well established in the United States. Chloramines are generated onsite at the treatment plant. Anhydrous ammonia and ammonia sulfate are examples of ammonia containing chemicals used by water systems to form chloramines. In most situations in the United States, chloramines are used as a secondary disinfectant to maintain a residual in the distribution system.

B. Ozone

Ozone is a colorless gas that is extremely unstable and is a strong oxidizing agent that is capable or reacting with a wide variety of organic and inorganic solutes in water. Ozone can also undergo a rapid, autocatalytic decomposition to form a variety of oxidant species, with the hydroxyl radical ($^{-}$OH) being the most notable. Both molecular ozone and the free hydroxyl radical are powerful oxidants. Ozone does not provide a lasting residual, therefore a secondary disinfectant, chlorine or chloramines, are typically used to provide a stable residual within the public water system distribution system.

The actual ozone dose utilized at a water treatment plant is determined by the quality of the raw water (i.e., how much ozone demand will be exerted) and what is the ultimate objective for using ozone. Plants using ozone for iron and manganese oxidation will need to apply a different dose than plants using ozone as a disinfectant for inactivation of protozoans such as *Cryptosporidium*.

Ozone must be manufactured and used onsite. Ozone is manufactured by passing air or oxygen through two electrodes with a high, alternating potential difference. The use of ozone to provide disinfection at a water treatment plant involves taking into account the concentration of ozone applied, competing ozone demands, and providing a minimum contact time. As indicated, ozone is a strong oxidant and requires a shorter reaction time as compared to free chlorine to accomplish the same results. The basic components of an ozone system at a water treatment plant include ozone generation equipment, feed gas preparation, ozone contacting in the water and a means to destroy ozone off-gas.

While ozone does not form many of the byproducts formed through the use of chlorine (e.g., trihalomethanes, haloacetic acids, etc.) it is important to note that ozone can produce a byproduct, bromate (BrO_3^-), which is a potential human carcinogen, when the water being treated contains bromide. In general, bromide concentrations in raw water greater than 50 μg/L may result in bromate formation at levels greater than the current USEPA maximum contaminant level (MCL) of 10 μg/L in the United States. Additional byproducts of ozonation include aldehydes, ketones, and carboxylic acids. However, these later compounds have as yet not been considered a concern for public health protection at the levels produced in drinking water treatment using ozone.

Ozone off-gas destruction is the final major component in the ozonation process. This system is required to remove ozone from spent off-gas streams, which are collected and treated prior to discharge into the atmosphere. Both catalytic and thermal destruction devices are used for this purpose.

C. Chlorine Dioxide

Chlorine dioxide (ClO_2) has been used as a primary disinfectant throughout Europe but only in a limited number of situations in the United States. Chlorine dioxide is a strong oxidant and has been used primarily to control taste and odors but it is also an effective bactericide, equal to or slightly better than hypochlorous acid. Chlorine dioxide is also used for iron and manganese oxidation. The use of chlorine dioxide can lead to the formation of chlorite (ClO_2^-) and chlorate (ClO_3^-) ions that have been a concern to regulators due to possible adverse health effects. When applied to surface water or groundwater, ClO_2 reacts rapidly with naturally occurring, oxidizable organic material. Up to 70% of the applied dose is reduced to ClO_2^- with the remainder being ClO_3^- or Cl^-. Once produced, chlorate ion, is not removed by conventional treatment. Clorite ion (ClO_2^-) can be minimized by optimizing the operation of the chlorine dioxide generator or by using a reducing agent such as ferrous iron.

Chlorine dioxide is unstable at high concentrations and must be generated on site at the water treatment plant. Chlorine dioxide is highly soluble and hydrolyzes very slowly in water.

D. Ultraviolet (UV) Irradiation Technology

UV light is electromagnetic energy that is located in the electromagnetic spectrum at wavelengths between those of X-rays and visible light. UV light that is effective is destroying microbial entities in located in the 200- to 310-nm range of the energy spectrum. Most typical applications of UV at water treatment plants apply UV light in the wavelength range of 250 to 270 nm. Most lamps emit UV irradiation by passing an electrical arc between filaments in a pressurized gas or vapor (typically mercury vapor). Ultraviolet dosage is commonly measured as milliWatt-second per square centimeter (mW-s/cm^2) or milliJoule per square centimeter (mJ/cm^2).

Typically a UV process is designed such that water flows in a narrow region around a series of UV lamps. Microorganisms in water are inactivated through exposure to the UV light. In general, a molecule in the ground state absorbs electromagnetic energy from the UV source and the bonds in the molecule are transformed to an excited state, and chemical and physical processes become thermodynamically possible. The process works on the principle that UV energy disrupts the DNA of the microorganisms and prevents it from reproducing.

There are four types of UV technologies of interest to the water industry. They include (1) low-pressure, low-intensity UV technology; (2) low-pressure, medium-intensity UV technology; (3) medium-pressure, high-intensity UV technology; and (4) pulsed-UV technology. Unlike using disinfectants such as ozone, chlorine, or chlorine dioxide, UV irradiation does not provide oxidation for color, taste, and odor control because UV light is not a strong oxidant.

A UV treatment process is comprised of a series of UV lamps enclosed inside a quartz sleeve. The UV light passes through the quartz sleeve and into the water. Due to the high energy emitted by the UV lamps, the temperature of the quartz sleeve can rise substantially causing the precipitation of various scales on the surface of the sleeve, thus blocking the passage of the UV light into the water and dramatically reducing the efficiency of the process. The scales are commonly caused by the precipitation of calcium, iron, or magnesium salts. Preventing the buildup of this scale is a major operational challenge for the use of UV. One of the current problems facing the use of UV irradiation is determining the actual UV dose the water receives, because measuring a residual is not possible.

VIII. ADVANCED TREATMENT METHODS

The term Advanced Oxidation Processes (AOPs) was first used to describe a process that produces hydroxyl radicals ($^{-}$OH) for the oxidation of organic and inorganic water impurities. AOPs can have multiple uses in water treatment. Examples include oxidation of synthetic organic chemicals (SOCs), color, taste-and-odor causing compounds, sulfide, iron, and manganese, and destruction of DBP precursor material prior to the addition of chlorine.

A. Peroxone

As an example of an AOP process, the following discussion on ozone with hydrogen peroxide is presented (referred to as “Peroxone”). The reactions of ozone can be divided into two types, direct ozonation reactions and free radical decomposition reactions. When hydrogen peroxide (H_2O_2) is added to ozonated water, it reacts with molecular ozone to accelerate the formation of hydroxyl radicals. Therefore, in an ozone–H_2O_2 process, the goal is to increase the concentration of hydroxyl radicals, which are a stronger oxidizer than molecular ozone, and reduce the concentration of molecular ozone. Therefore, hydrogen peroxide is added to an ozone process if it is used as an oxidation process, but not if ozone is intended to be used in a disinfection process that relies on the prevalence of a high concentration of molecular ozone. The Peroxone process is used for the destruction of taste and odor causing compounds, color removal, and the destruction of micropollutants, such as VOCs, pesticides, and herbicides.

Currently, the conventional design of an ozone–H_2O_2 treatment process is one where hydrogen peroxide is fed as a liquid to the influent water and an ozone-rich gas is fed through fine-bubble diffusers at the bottom of a contactor.

IX. NEW EMERGING ISSUES FOR DISINFECTION

For many years public water systems and regulatory agencies, such as the USEPA, have approached the idea of changing water treatment practices to limit the formation of DBPs with caution due to the possible implications for control of microbial risks in drinking water. Balancing acute risks from microbial pathogens such as viruses and protozoans against adverse health risks from DBPs has been a challenging situation to address. Selecting the appropriate disinfectant involves not only consideration of its ability to form potentially harmful DBPs but also its ability to effectively inactivate waterborne pathogens among other water quality considerations. For example, the use of chloramines can be an effective strategy for limiting the production of THMs, however, they are a weaker disinfectant than chlorine.

Giardia lamblia and *Cryptosporidium parvum* are two protozoan pathogens that have emerged to present significant challenges to public water systems and regulators for the delivery of *safe*, potable drinking water.

A. Giardia

Giardia lamblia is a flagellated protozoan that has a trophozoite form and a cyst form. The length of the trophozoite form ranges from 9 to 21 μm, the width from 5 to 15 μm and a thickness from 2 to 4 μm. In an unfavorable environment the parasite encysts, and the cysts are characteristically oval or ellipsoid in shape (slightly asymmetric) with a length of 8–14 μm and a width of 7–10 μm.

Giardiasis is an acute, self-limiting diarrheal disease caused by ingestion of Giardia cysts. Giardiasis is one of the most commonly identified waterborne intestinal diseases in the U.S. *Giardia lamblia* has been identified as the causative agent of numerous outbreaks involving public and private drinking water systems.

Outbreaks of Giardiasis have occurred due to the use of untreated surface water, contaminated water distribution systems and treatment deficiencies. Wild and domestic mammals can be significant sources of *Giardia* cysts. Human wastewater is another source of *Giardia* contamination. A person infected with *Giardia lamblia* may shed on the order of 10^8 cysts/day. Giardia cysts can be effectively removed from drinking water through the use of physical treatment such as filtration. The cysts can also be inactivated using disinfectants such as chlorine.

B. Cryptosporidium

Cryptosporidium is a protozoan parasite that reproduces within the gut of an animal host. The life cycle of

Cryptosporidium is extremely complex and includes the formation of resistant oocysts. *Cryptosporidium* oocysts can originate from a number of animal hosts, including cattle, swine, horses, deer, chicken, ducks, fish, turtles, guinea pigs, cats, and dogs. *Cryptosporidium* is transmitted by ingestion of oocysts that have been excreted in the feces of infected humans or animals. *Cryptosporidium parvum* has been recognized as a human pathogen since 1976 when the first case of cryptosporidiosis (the disease caused by *Cryptosporidium*) was diagnosed.

Like Giardiasis, Cryptosporidiosis is an acute self-limiting diarrheal disease. The oocysts can remain infective in water and moist environments for several months and are resistant to high concentrations of chlorine. This, together with the small size of the oocysts (4–6 μm in diameter) can lead to oocysts occasionally passing through conventional water treatment plants. Not all infected individuals develop symptoms but the illness can be serious or life-threatening for individuals with a compromised immune system (e.g., AIDS patients). Cryptosporidiosis outbreaks have been documented in the United States and around the world.

Table IV presents an overview of cryptosporidiosis outbreaks and the suspected source of *Cryptosporidium*

TABLE IV Waterborne Outbreaks of Cryptosporidiosis[a]

Location	Date	Water source	Number of cases (estimated)	Treatment	Suspected source of contamination
Braun Station, TX	July 1984	Well	47–117 (2006)	Chlorine	Raw Sewage
Sheffield, England	May–June 1986	Surface water	49 (537)	Unfiltered	Agricultural, nonpoint source pollution during heavy rainfall
Bernalillo County, NM	July–Oct. 1986	Surface water	78	Untreated	Runoff from livestock grazing areas
Carrolton, GA	Jan.–Feb. 1987	River	(13,000)	Conv., chlorine	Raw sewage and runoff from cattle grazing areas
Ayrshire, Scotland	Apr. 1988	Reservoir	(13,000)	N/A	Cross-connection to sewage contaminated source in a distribution system tank
Swindon and Oxfordshire, England	Mar. 1989	River	516	N/A	Agricultural nonpoint source pollution during heavy rainfall event
Berks County, PA	Aug. 1991	Well	(551)	Chlorine	Septic tank influence
Jackson County, OR (Talent and Medford)	Jan.–June 1992	Spring/river	(15,000)	Conv., chlorine	Surface water, treated wastewater or runoff from agricultural areas
Milwaukee, WI	March 1993	Lake	(403,000)	Conv., chlorine	Cattle wastes, slaughterhouse wastes and sewage carried by tributary rivers
Yakima County, WA	April 1993	Well	3 (7)	Untreated	Infiltration of runoff from cattle, sheep or elk grazing areas
Cook County, MN	Aug. 1993	Lake	5 (27)	Pressure filter, chlorine	Backflow of sewage or septic tank effluent into distribution, raw water inlet lines or both
Clark County, NV	Dec. 1993	Lake	78	Direct, chlorine	Treated wastewater, sewage from boats
Walla Walla, WA	Aug. 1994	Well	86	Chlorine	Treated wastewater
Alachua County, FL	July 1995	N/A	72	N/A	Backflow from a wastewater line in a camp's drinking water system
Collingwood, CANADA	1996	39(150)	N/A	N/A	Source of contamination unclear. 100 year storm prior to outbreak

[a] (Craun, G. F., Hubbs, S. A., Frost, F., Calderon, R. L., Via, S. H. (1998). "Waterborne Outbreaks of Cryptosporidiosis," *J. Am. Water Works Assoc.* **90,** 81–91.)

Solo-Gabriele, H., Neumeister, S. (1996). "US Outbreaks of Cryptosporidiosis," *J. Am Water Works Assoc.* 76–86.

Frey, M. M., Hancock, C., Logsdon, G. S. (1997). "*Cryptosporidium*: Answers to Questions Commonly Asked by Drinking Water Professionals," Amer. Water Works Assoc. Research Foundation.

that occurred between 1984 and 1996. The following presents brief description of two of those outbreaks of cryptosporidiosis that were significant in the United States (Milwaukee, WI; Las Vegas, NV).

1. Milwaukee Outbreak

The Milwaukee outbreak occurred during March and April 1993. Estimates are that 403,000 people had diarrhea (estimates are that around 100 people died) due to ingestion of *Cryptosporidium parvum*.

At the time of the outbreak, the city of Milwaukee operated two treatment plants. Both of the plants treated Lake Michigan water using conventional treatment. The intake for the treatment plant associated with the outbreak was located 42 ft below the surface of the lake. Chlorine and/or potassium permanganate were occasionally added at the intake to control taste and odors and to control zebra mussels.

Historically, alum had been used as the coagulant. In September 1992, the facility switched to polyaluminum chloride (PACl) as the coagulant. The switch was made after consulting with other water agencies treating Lake Michigan water, regulatory authorities and the chemical manufacturer. By converting to PACl, the intent was to achieve a higher finished water pH (for corrosion control), reduce sludge volume and improve coagulation effectiveness in cold water conditions.

After the rapid mix stage, the water passed into one of four 1 million-gallon parallel baffled coagulation basins. The water then passed through eight sand and gravel media filters. Flow rates and head loss (i.e., the pressure drop between two point along the path of a flowing liquid) were monitored across each filter and were the basis for initiating each filter cleaning procedure. Coagulant dosages were adjusted as needed on the basis of laboratory jar tests.

Between March 18 and April 8, the turbidity of the raw water into the treatment plant associated with the outbreak ranged from 1.5 nephelometric turbidity units (NTU) to 44 NTU. Levels of total coliforms ranged from less than 1 colony forming unit (cfu) per 100 mL to about 3200 cfu per 100 mL. Plant staff responded to the fluctuations in turbidity throughout the treatment processes by adjusting coagulant dosages. Coagulation dosage adjustments were also made to compensate for coagulation demands resulting from treatment to control taste and odor. The coagulant dosages were based on jar-test data and consultation with the chemical supplier.

Several times during the period of the incident, filter effluent turbidity levels exceeded turbidity values achieved in previous months. On April 2, 1993, the plant switched back to alum as the primary coagulant. From March 18 through April 8 (the day the plant was shut down) the effluent turbidity was highly variable and ranged between 0.1 and 2.7 NTU. The ultimate source of the *Cryptosporidium* oocysts was not determined. As indicated previously, nearly 400,000 people had diarrhea during the outbreak and estimates are that nearly 100 people died.

2. Las Vegas Outbreak

Between January and May 1994, Las Vegas experienced an outbreak of cryptosporidiosis in the severely immunocompromised population of the city. The treatment plant in question had a capacity of 400 million gallons per day and employed direct filtration. At this treatment plant, filter backwash water (e.g., the water used to clean the filter beds) was recycled to the beginning of the treatment plant after sedimentation in clarifiers. The source water was Lake Mead and the intake was located 130 ft below the surface. The average raw water turbidity from January 1993 through June 1995 was 0.14 NTU, with a high of 0.3 NTU and a low of 0.1 NTU in December 1993 and April 1994.

In March 1994, the local health department observed an increase in cryptosporidiosis in the HIV-infected community in the city. From June 28 through December 31, 1993, there were 9 cases of cryptosporidiosis reported. From January 1 through March 19, 1994, there were 49 cases of cryptosporidiosis reported. In June 1994, the outbreak ceased. There had been no changes in the treatment process of the distribution system.

In March 1995 the U.S. Centers for Disease Control and Prevention (CDC) released the results of their investigation. The data indicated that HIV-infected patients who drank tap water were at greater risk of contracting cryptosporidiosis than if they drank bottled or filtered water. CDC found no deficiencies in the treatment plant or in the distribution system. CDC was unable to locate a specific source of *Cryptosporidium* or evidence of *Cryptosporidium* in the water supply. During January 1 through April 30, 1994, seventy-eight people became infected, 65 adults and 13 children. Of the 65 adults, 61 had HIV infections. Of the 13 children, 2 were HIV infected. There was no geographic clustering of the cases of cryptosporidiosis observed.

X. DRINKING WATER REGULATIONS

A. United States Regulations

The Safe Drinking Water Act (SDWA) was enacted by the United States Congress in 1974. Through the SDWA, the federal government gave the USEPA the authority to set standards for contaminants in drinking water supplies. The EPA was required to establish primary regulations for the control of contaminants which affect public health and secondary regulations for compounds which affect the taste or aesthetics of drinking water.

The first step taken by the EPA to establish MCLs is to determine a maximum contaminant level goal (MCLG) for the target compound. The MCLG represents the estimated concentration in which no known or anticipated adverse health effects occur, including an adequate margin of safety. MCLGs are strictly a health-based number and do not take into account issues associated with analytical methods, treatment technology or economics. The MCLs, however, are then set as close to the MCLG as is technically and economically feasible. Under the SDWA, EPA can regulate compounds by requiring utilities to institute specified treatment techniques to limit its concentration in drinking water (the Surface Water Treatment Rule, the Lead and Copper Rule are examples of treatment techniques). The SDWA was reauthorized and amended by Congress in 1986 and again in 1996. Under the SDWA, individual states are allowed to adopt and implement drinking water regulations. To do so, states must adopt drinking water regulations that are no less stringent than the regulations published by USEPA (state regulations can be more stringent than federal regulations).

1. THM Regulation

The THM regulation was promulgated in 1979 and applied to all public water systems serving populations greater than 10,000. The THM regulation established an MCL of 100 μg/L for total trihalomethanes (TTHMs) in the distribution system. [Total trihalomethanes include the sum of chloroform, bromodichloromethane, dibromochloromethane, and bromoform.] Systems must collect a minimum of four distribution system samples per treatment plant on a quarterly basis. Compliance with the MCL is based on a running annual average concentration of four quarterly monitoring periods.

In 1998 USEPA published a DBP regulation that revised the 100-μg/L standard for THMs to 80 μg/L as well as establishing MCLs for HAAs (60 μg/L), bromate (10 μg/L), and chlorite (1000 μg/L).

EPA has also established maximum residual disinfectant levels (MRDLs) to limit the applied dose of chlorine, chloramines and chlorine dioxide during drinking water treatment:

Chlorine	4.0 mg/L
Chloramines	4.0 mg/L as total chlorine
Chlorine Dioxide	0.8 mg/L

a. Enhanced coagulation. Since only approximately 50% of the total organic halides produced during disinfection of drinking water have been identified, the EPA established a requirement for conventional treatment plants to remove total organic carbon (TOC) as a surrogate for DBP precursor material. The reasoning was that by removing TOC this would reduce the formation of unidentified DBPs with potential adverse health effects.

TABLE V Enhanced Coagulation TOC Removal Requirements

Source Water TOC	Source Water Alkalinity, mg/l 0 to 60	>60 to 120	>120
>2.0 to 4.0 mg/l	35%	25%	15%
>4.0 to 8.0 mg/l	45%	35%	25%
>8.0 mg/l	50%	40%	30%

Performance criteria for TOC removal based on raw water TOC levels and alkalinity have been established to serve as a definition of enhanced coagulation. Table V presents a matrix identifying the percent TOC removal required based on raw water TOC and alkalinity levels. If a conventional treatment plant can not achieve the percent reductions listed in Table V, the system can use an alternative criteria for evaluating enhanced coagulation. This alternative criteria requires a system to conduct bench- or pilot-scale studies to identify optimum coagulation conditions.

2. Surface Water Treatment Rule and Interim Enhanced Surface Water Treatment Rule

The Surface Water Treatment Rule (SWTR) was published in June 1989. The SWTR was promulgated to control the levels of turbidity, *Giardia lamblia*, viruses, *Legionella*, and heterotrophic plate count bacteria in U.S. drinking waters. In 1998, the USEPA published an Interim Enhanced Surface Water Treatment Rule (ESWTR) that added additional restrictions to the SWTR and addressed the issue of *Cryptosporidium* removal.

The SWTR requires all utilities utilizing a surface water supply or a ground water supply under the influence of a surface water supply, to provide adequate disinfection and under most conditions, to provide filtration. (Surface water supplies can avoid the requirement to filter their supply if they meet specific requirements for water quality and the utility can demonstrate control of the watershed.)

a. General requirements. The SWTR and the Interim ESWTR includes the following general requirements in order to minimize human exposure to microbial contaminants in drinking water.

- Utilities are required to achieve at least 99.9% removal and/or inactivation of *Giardia lamblia* cysts (3-log removal), a minimum 99.99% removal and/or inactivation of viruses (4-log removal), and a 2-log removal of *Cryptosporidium*. The required level of removal/inactivation must occur between the point where the raw water is not longer subject to surface water runoff and the point at which the first customer is served.
- The disinfectant residual entering the distribution system must not fall below 0.2 mg/L for more than 4 hr during any 24-hr period.
- A disinfectant residual must be detectable in 95% of distribution system samples. A heterotrophic plate count (HPC) concentration of less than 500 colonies per milliliter can serve as a detectable residual if no residual is measured.
- Each utility must perform a watershed sanitary survey at least every 5 years.
- The combined filtered water turbidity must be less than or equal to 0.3 NTU in 95% of samples taken during each month (measured every four hours), never to exceed 1 NTU.
- Systems are required to achieve 2-log reduction in *Cryptosporidium*.
- If a system meets the turbidity performance standard of 0.3 NTU in 95% of the samples taken each month (never to exceed 1 NTU) than the system gets credit for achieving the 2-log reduction in *Cryptosporidium*.
- Systems must monitor the turbidity for individual filters, and any filter that is not performing per specified criteria must be evaluated.

b. Removal credit. The level of removal credit given a utility for both *Giardia lamblia* and viruses is determined by the type of treatment process used. For a conventional water treatment plant the SWTR provides a 2.5-log removal credit for *Giardia lamblia* and a 2.0-log removal credit for viruses. As described above, water treatment plants meeting the turbidity performance standard of 0.3 NTU in 95% of the monthly measurements also get credit for a 2-log reduction in *Cryptosporidium*.

c. Disinfection credit. Disinfection during conventional treatment (assuming all operational criteria and performance standards are met), must achieve 0.5-log inactivation of *Giardia lamblia* and 2.0-log inactivation of viruses. As a substitute to actually measuring the inactivation of *Giardia lamblia* and viruses achieved at a treatment plant, the SWTR established the concept of CT to evaluate inactivation. CT is the product of the concentration of disinfectant remaining at the end of a treatment process ("C" in milligrams per liter) and the contact time in which 10% of the water passes through the treatment process ("T" or "T_{10}" in minutes). The SWTR provides tables which identify the log removal of both *Giardia lamblia* and viruses achieved for a calculated CT value based on the type of disinfectant, the water temperature, and pH.

3. Total Coliform Rule

The Total Coliform Rule (TCR) was promulgated by the EPA in June 1989. A specific MCL value was not established for total or fecal coliforms under the TCR. Instead, there are three potential scenarios in which an MCL is violated that depend on the presence and/or absence of total coliforms (and fecal coliforms or E. *coli*). Public water systems monitor for total coliforms throughout the distribution system to determine compliance with the TCR.

TABLE VI WHO Guidelines for Bacteriological Quality of Drinking Water[a]

Organisms	Guideline value
All water intended for drinking	
E. coli or thermotolerant coliform bacteria[b,c]	Must not be detectable in any 100-mL sample
Treated water entering the distribution system	
E. coli or thermotolerant coliform bacteria[b]	Must not be detectable in any 100-mL sample
Total coliform bacteria	Must not be detectable in any 100-mL sample
Treated water in the distribution system	
E. coli or thermotolerant coliform bacteria[b]	Must not be detectable in any 100-mL sample
Total coliform bacteria	Must not be detectable in any 100-mL sample. In the case of large supplies, where sufficient samples are examined, must not be present in 95% of samples taken throughout any 12-month period.

[a] Immediate investigative action must be taken if either *E. coli* or total coliform bacteria are detected. The minimum action in the case of total coliform bacteria is repeat sampling; if these bacteria are detected in the repeat sample, the cause must be determined by immediate further investigation.

[b] Although *E. coli* is the more precise indicator of fecal pollution, the count of thermotolerant coliform bacteria is an acceptable alternative. If necessary, proper confirmatory tests must be carried out. Total coliform bacteria are not acceptable indicators of the sanitary quality of rural water supplies, particularly in tropical areas where many bacteria of no sanitary significance occur in almost all untreated supplies.

[c] It is recognized that, in the great majority of rural water supplies in developing countries, fecal contamination is widespread. Under these conditions, the national surveillance agency should set medium-term targets for the progressive improvement of water supplies.

TABLE VII WHO and USEPA Limits for Inorganic Chemical Constituents[a]

	WHO guideline value (mg/L)	USEPA limit (mg/L)
Constituent		
Antimony	0.005 (P)[b]	0.006
Arsenic	0.01 (P)	0.05
Asbestos	—	7 MFL
Barium	0.7	2
Beryllium	—	0.004
Boron	0.3	—
Cadmium	0.003	0.005
Chromium	0.05 (P)	0.1
Copper	2 (P)	TT
Cyanide	0.07	0.2
Fluoride	1.5[c]	4.0
Lead	0.01	TT
Manganese	0.5 (P)	—
Mercury	0.001	0.002
Molybdenum	0.07	—
Nickel	0.02	—
Nitrate (as NO_3^-)	50	45
Nitrite (as NO_2^-)	3 (P)	—
Selenium	0.01	0.05
Radium 226 and 228	—	5 pCi/L[d]
Gross alpha particle activity	0.1 Bq/L[c]	15 pCi/L
Beta particle and photon activity	1 Bq/L	4 mrem (annual dose equivalent)[e]

[a] MFL = million fibers per liter greater than 10 μ in length.
[b] (P) = Provisional.
[c] Bq/L = Becquerel per liter.
[d] pCi/L = picoCuries per liter.
[e] mrem = millirem.

4. Lead and Copper Rule

The Lead and Copper Rule (LCR) was published by the EPA on June 7, 1991. The objective of the LCR is to minimize the corrosion of lead and copper-containing plumbing materials in public water systems (PWS) by requiring utilities to optimize treatment for corrosion control. The LCR establishes "action levels" in lieu of MCLs for regulating the levels of both lead and copper in drinking water. The action level for lead was established at 0.015 mg/L while the action level for copper was set at 1.3 mg/L. An action level is exceeded when greater than 10% of samples collected from the sampling pool contain lead levels above 0.015 mg/L or copper levels above 1.3 mg/L.

5. Arsenic

Arsenic is the 20th most abundant element in nature and the 12th most abundant in the human body. Arsenic occurs in both the organic and inorganic forms. The inorganic forms are arsenite (As^{+3}) and arsenate (As^{+5}). These are the most toxic forms of arsenic, and arsenite appears to be the most toxic. Human exposure to arsenic comes primarily from food sources (shellfish and grain raised in arsenic-laden soils) and is mostly in the organic forms. EPA is in the process of lowering the MCL for arsenic, perhaps significantly lower than the current standard. Historically, arsenic has been regulated in drinking water based on its potential to cause skin cancer (usually a nonfatal disease). Information indicates that arsenic may be associated with internal cancers such as lung, kidney and bladder.

Potential treatment processes for reducing arsenic levels include activated alumina, activated carbon, ion exchange, adsorption/coprecipitation with Fe/Al oxides, and membrane processes.

B. WHO Regulations

The World Health Organization (WHO) has published guidelines for drinking water quality. According to WHO,

TABLE VIII USEPA and WHO Limits for Volatile Organic Chemicals in Drinking Water

Contaminant	Federal MCL (mg/L)	WHO guideline (mg/L)
1,1-Dichloroethylene	0.007	0.030
1,2-Dichloroethylene	—	0.050
1,1,1-Trichloroethane	0.2	2 (P)
1,1,2-Trichloroethane	0.005	—
1,2-Dichloroethane	0.005	0.030
1,2-Dichloropropane	0.005	0.020 (P)
1,3-Dichloropropylene	—	0.020
1,2,4-Trichlorobenzene	0.07	—
Benzene	0.005	0.010
Carbon tetrachloride	0.005	0.002
cis-1,2-Dichloroethylene	0.07	—
Dichloromethane	0.005	0.020
Ethylbenzene	0.7	0.30
Monochlorobenzene	0.1	0.30
1,2-Dichlorobenzene	—	1
1,3-Dichlorobenzene	0.6	—
1,4-Dichlorobenzene	0.075	0.30
Trichlorobenzene (total)	—	0.020
Styrene	0.1	0.020
Tetrachloroethylene	0.005	0.040
Toluene	1	0.70
trans-1,2-Dichloroethylene	0.1	—
Trichloroethylene	0.005	0.070 (P)
Vinyl chloride	0.002	0.005
Xylenes (total)	10	0.50

TABLE IX USEPA and WHO Limits for Synthetic Organic Chemical in Drinking Water[a]

Contaminant	Federal MCL (mg/L)	WHO guideline value (mg/L)
2,3,7,8-TCDD (Dioxin)[b]	3×10^{-8}	—
2,4-D	0.07	0.030
2,4,5-T	—	0.009
2,4,5-TP (Silvex)	0.05	—
Acrylamide	TT	0.0005
Alachlor	0.002	0.020
Aldicarb	(a)	0.010
Aldrin/dieldrin	—	0.00003
Atrazine	0.003	0.002
Bentazone	—	0.030
Benzo(a)pyrene	0.0002	0.0007
Carbofuran	0.04	0.005
Chlordane	0.002	0.0002
Chlorotoluron	—	0.030
Dalapon	0.2	
Di(2-ethylhexyl)adipate	0.4	0.080
Di(2-ethylhexyl)phthalate	0.006	0.008
Dibromochloropropane (DBCP)	0.0002	0.001
Dichlorprop	—	0.10
Dinoseb	0.007	—
Diquat	0.02	—
DDT	—	0.002
Edetic acid (EDTA)	—	0.20
Endothal	0.1	—
Endrin	0.002	—
Epichlorohydrin	TT	0.0004
Ethylene dibromide	0.00005	—
Fenoprop	—	0.009
Glyphosate	0.7	—
Heptachlor	0.0004	0.00003
Heptachlor epoxide	0.0002	0.00003
Hexachlorobenzene	0.001	0.001
Hexachlorobutadiene	—	0.0006
Hexachlorocyclopentadiene	0.05	—
Isoproturon	—	0.009
Lindane	0.0002	0.002
MCPA	—	0.002
Mecoprop	—	0.010
Methoxychlor	0.04	0.020
Metolachlor	—	0.010
Molinate	—	0.006
Nitrilotriacetic acid	—	0.20
Oxamyl (vydate)	0.2	—
Pendimethalin	—	0.020
Pentachlorophenol	0.001	0.009 (P)
Permethrin	—	0.020
Picloram	0.5	—
Polychlorinated biphenyl (PCB)	0.0005	—
Propanil	—	0.020
Pyridate	—	0.10
Simazine	0.004	0.002
Tributyltin oxide	—	0.002
Trifluralin	—	0.020
Toxaphene	0.003	—

[a] P = Provisional guideline value. This term is used for constituents for which there is some evidence of a potential hazard but where the available information on health effects in limited or where an uncertainty factor greater than 1000 has been used in the derivation of the tolerable daily intake. Provisional guideline values are also recommended (1) for substances for which the calculated guideline value would be below the practical quantification level, or below the level that can be achieved through practical treatment methods; or (2) where disinfection is likely to result in the guideline value being exceeded.

TT = Treatment Technique.

[b] The USEPA MCLs for aldicarb (MCL = 0.01 mg/L), aldicarb sulfoxide (MCL = 0.01 mg/L), aldicarb sulfone (MCL = 0.04 mg/L) are not enforced.

these guidelines are intended to be used as the basis for the development of national standards that will ensure the safety of drinking water supplies.

According to WHO, the judgement of safety—or what is an acceptable level of risk in particular circumstances—is a matter in which society as a whole has a role to play. The final judgement as to whether the benefit resulting from the adoption of any of the guideline values given here as standards justifies the cost is for each country to decide.

The guideline values presented by WHO represent the concentration of a given constituent that does not result in any significant risk to the health of the consumer over a lifetime of consumption. Short-term deviations above the guideline values do not necessarily mean that the water is unsuitable for consumption. The amount by which, and the period for which, any guideline values can be exceeded

TABLE X USEPA and WHO Limits for Disinfectants

Disinfectant	USEPA MRDL[a] (mg/L)	WHO guideline value (mg/L)
Chorine[b]	4.0	
Chloramine[b]	4.0	
Monochloramine	—	3
Di- and trichloramine	—	5
Chlorine dioxide	0.8	

[a] MRDL = maximum residual disinfectant level.

[a] Measured as free chlorine.

without affecting public health depends upon the specific substance involved. In developing national drinking water standards based on these guideline values, it will be necessary to take account of a variety of geographical, socioeconomic, dietary, and other conditions affecting potential exposure. This may lead to national standards that differ appreciably from the guideline values.

C. Current USEPA Regulations and WHO Guideline Values

Table VI presents the WHO guidelines for bacteriological quality of drinking water. Tables VII, VIII, and IX contain the USEPA standards and WHO Guideline Values for inorganic, volatile organic and synthetic organic chemicals, respectively. Table X and XI presents the USEPA and WHO Guidelines Values for disinfectants and disinfection byproducts. Table XII presents the USEPA limits (secondary standards that are not enforceable at the federal level) and the WHO Guideline Values for constituents affecting the taste and odor of the water.

TABLE XI USEPA and WHO Limits for Disinfection By-Products[a]

	USEPA MCL (mg/L)	WHO guideline value (mg/L)
Constituent		
Bromate	0.010	0.025 (P)
Chlorite	1.0	0.20
2,4,6-chlorophenol	—	0.20
Formaldehyde	—	0.90
Total THMs	0.080	—
Bromoform	—	0.10
Dibromochloromethane	—	0.10
Bromodichloromethane	—	0.06
Chloroform	—	0.20
HAA5	0.060	—
Dichloroacetic acid	—	0.050 (P)
Trichloroacetic acid	—	0.10 (P)
Chloral hydrate	—	0.010 (P)
Dichloroacetonitrile	—	0.090 (P)
Dibromoacetonitrile	—	0.10 (P)
Trichloroacetonitrile	—	0.001 (P)
Cyanogen chloride	—	0.070

[a] P = Provisional guideline value. This term is used for constituents for which there is some evidence of a potential hazard but where the available information on health effects in limited or where an uncertainty factor greater than 1000 has been used in the derivation of the tolerable daily intake. Provisional guideline values are also recommended (1) for substances for which the calculated guideline value would be below the practical quantification level, or below the level that can be achieved through practical treatment methods; or (2) where disinfection is likely to result in the guideline value being exceeded.

TABLE XII USEPA and WHO Limits for Constituents Affecting Taste and Odor

	USEPA limit (mg/L)	WHO guideline value (mg/L)[a]
Color	15 CU	15 TCU
Turbidity	—	5 NTU
Aluminum	0.05–0.2	0.2
Ammonia	—	1.5
Chloride	—	250
Copper	1.0	1
Fluoride	2.0	—
Hydrogen sulfide	—	0.05
Iron	0.3	0.3
Manganese	0.05	0.1
Sodium	—	200
Sulfate	250	250
Total dissolved solids	500	1000
Zinc	5	3
Touene	—	0.024–0.170
Xylene	—	0.020–1.8
Ethylbenzene	—	0.002–0.2
Styrene	—	0.004–2.6
Monochlorobenzene	—	0.010-0.12
1,2-dichlorobenzene	—	0.001–0.010
1,4-Dichlorobenzene	—	0.0003–0.030
Trichlorobenzene (total)	—	0.005–0.050
Chlorine	—	0.6–1.0
2-Chlorophenol	—	0.0001–0.010
2,4-Dichlorophenol	—	0.0003–0.040
2.4.6-Trichlorophenol	—	0.002–0.3

[a] The levels indicated are not precise numbers. Problems may occur at lower or higher values according to local circumstances. A range of taste and odor threshold concentrations is given for organic constituents. TCU = time color unit. CU = color unit. NTU = nephelometric turbidity unit.

SEE ALSO THE FOLLOWING ARTICLES

DRINKING WATER QUALITY AND TREATMENT • ENVIRONMENTAL MEASUREMENTS • ENVIRONMENTAL TOXICOLOGY • POLLUTION, AIR • POLLUTION, ENVIRONMENTAL • SOIL AND GROUNDWATER POLLUTION • TRANSPORT AND FATE OF CHEMICALS IN THE ENVIRONMENT • WASTEWATER TREATMENT AND WATER RECLAMATION • WATER POLLUTION • WATER RESOURCES

BIBLIOGRAPHY

American Water Works Association (1990). Pontius, F. (technical ed.). "Water Quality and Treatment," McGraw-Hill, New York.

Faust, S. D., and Aly, O. M. (1998). "Chemistry of Water Treatment," Ann Arbor Press, MI.

Fox, K. R., and Lytle, D. A. (1996). "Milwaukee's crypto outbreak: Investigation and recommendations," *J. Am. Water Works Assoc.* **88,** 87–94.

Glaze, Wh. H., Kang, H. W., and Chapin, H. D. (1987). "The chemistry of water treatment processes involving ozone, hydrogen peroxide and ultraviolet radiation," *Ozone Sci. Engrg.* **9,** 335.

Goldstein, S. T., *et al.* (1996). Cryptosporidiosis: An Outbreak Associated with Drinking Water Despite State-of-the-Art Treatment. *Ann. Int. Med.* **124,** 459–468.

Karamura, S. (1991). "Integrated Design of Water Treatment Facilities," Wiley, New York.

MacKenzie, W. R., *et al*. (1994). "A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply," *New Engl. J. Med.* **331** (3), 161–167.

Montgomery, J. M. (Consulting Engineer) (1985). "Water Treatment Principles and Design," Wiley, New York.

National Research Council (1998). "Issues in Potable Reuse. The Viability of Augmenting Drinking Water Supplies with Reclaimed Water," National Academy Press. Washington, D.C.

Singer, P. (1999). "Formation and Control of Disinfection By-Products in Drinking Water," American Water Works Association. Denver, CO.

Twort, A., Ratnayaka, D., and Brandt, M. (2000). "Water Supply," (5th edition), Arnold and IWA Publishing, London.

USEPA (1999). National Primary Drinking Water Regulations. "Code of Federal Regulations," **40,** Sections 141.1–141.175.

USEPA (1998). National Secondary Drinking Water Regulations. "Code of Federal Regulations," **40,** Sections 143.1–143.5.

WHO (1993). "Guidelines for Drinking-Water Quality," (2nd edition), Volume 1, Recommendations, World Health Organization, Geneva.

Environmental Geochemistry

Konrad B. Krauskopf
Keith Loague
Stanford University

GLOSSARY

Acid rain Rain that is more acidic than normal because raindrops have dissolved acid gases and/or dust particles from the atmosphere; the principal gases responsible for increased acidity are oxides of sulfur and nitrogen. Generally, rain with a pH below about 4.5 is considered environmentally harmful.

Adsorption Adherence of gas molecules or of ions or molecules in solutions to the surfaces of solids with which they are in contact. If the surfaces are thought of as being inside an object or objects, the phenomenon is often called absorption. The general term *sorption* refers to either.

Aerosol Suspension of small solid particles and/or liquid droplets in a gas, usually air.

Crust of the earth Outer part of the solid earth, commonly defined as the material above a major seismic discontinuity having a depth of about 35 km under the continents and 6 km under the floor of the ocean.

Greenhouse effect Trapping of heat from the sun by the atmosphere, in the same manner as the sun's heat is trapped by the glass walls and roof of a greenhouse. The atmosphere, like the glass, is largely transparent to the sun's radiation, but it absorbs the longer-wavelength radiation from the earth's surface into which the sun's radiation is converted. The principal gases responsible for the absorption are carbon dioxide, water vapor, and ozone.

Hydrolysis Reaction of water with metal ions to form hydroxy complexes and free H^+, or with anions to form weak acids and free OH^-.

Incongruent dissolution Dissolving of a compound in which dissolution is only partial, which leaves behind a different solid and gives dissolved material in proportions different from those in the original substance.

Ozone layer Layer in the stratosphere between $\sim$15 and 30 km above the earth's surface in which ozone is relatively concentrated (amounts greater than 10^{12} molecules/cm^3). The ozone is produced by the action of solar ultraviolet radiation on oxygen and is important in screening the earth's surface from potentially harmful short-wavelength radiation.

Photosynthesis Reaction catalyzed by light leading to a combination of substances. When used without qualification, it refers to the combining of CO_2 and H_2O in the green leaves of plants to form organic compounds and oxygen.

Radioactive waste Waste material containing radioactive elements in amounts greater than those normally present in the environment. Such waste is generated in large amounts by nuclear reactors used for production of electric power or of plutonium for weapons manufacture. Much low-level waste also results from uranium and phosphate mining and milling, industrial processes, laboratory research, and discarded materials that were used in medical diagnosis and therapy.

Redox potential Measure of the ability of a substance (or more generally of an environment) to oxidize or reduce another substance. Numerically it is the potential (in volts) that would result if the substance or the environment were made part of a cell in which the opposite electrode was the standard hydrogen electrode.

Solute transport Transport of chemical species as dissolved aqueous phase liquids or organic liquids that are only sparingly solube in water and move as a separate phase.

Steady state State of balance in a system of reactions in which concentrations of reacting substances remain constant or nearly so because they are forming at about the same rate as they are reacting. Equilibrium is a special kind of steady state in which the balance is perfect and is maintained indefinitely; other steady states may lead ultimately to change in one or more constituents, but concentrations remain nearly constant for long periods.

Stratosphere Layer of the atmosphere above the troposphere, up to $\sim$50 km from the surface of the earth.

Troposphere Lower part of the atmosphere, up to an altitude of $\sim$16 km near the equator and $\sim$10 km near the poles.

Vadose zone Region between the land surface and the water table, where the voids in the porous or fractured media are occupied by both the water and air phases, and the fluid pressure is less than atmospheric.

Weathering Disintegration and chemical decomposition of rocks by slow reaction with air and water at or near the earth's surface.

ENVIRONMENTAL GEOCHEMISTRY refers to the chemical makeup of rocks, soil, water, air, and organisms near the surface of the earth and the chemical reactions that take place among them. The reactions form a system of interrelated processes that maintain an approximate steady state in which concentrations of major substances remain constant or nearly so for long periods. Environmental geochemistry, as a study, is concerned with (1) the mechanisms by which this steady state is preserved, (2) the reasons for the concentrations of principal substances in the environment, and (3) possible influences that might change the concentrations and disturb the steady state. It deals not only with natural processes, but also with interactions between humans and their environment—both the effect of the chemical environment on human activity and the possible long-term effects of human enterprise on the reactions in the steady-state system.

I. INTRODUCTION

The present steady state of the earth's chemical environment is the product of a long history of change from very different conditions on the primitive earth. Very slow overall change may be continuing, but reaction rates are now generally well adjusted so that most constituents of the environment are being regenerated about as fast as they are consumed. For example, oxygen continually disappears in the processes of decay and slow oxidation of iron minerals, but it is constantly replenished by photosynthesis in the green leaves of plants. Environmental reactions proceed at widely different rates. Some are fast enough that an approach toward chemical equilibrium can be discerned: for example, calcium carbonate in contact with water at the earth's surface is commonly in approximate equilibrium with its ions and the carbon dioxide of the adjacent air. Other reactions are so slow that even in geologic times equilibrium is not approached: the nitrogen and oxygen of air, for example, would be in equilibrium with a fairly concentrated solution of nitrate or nitric acid, yet they remain in contact with seawater containing very little nitrate and show no sign of reaction. Most environmental reactions have rates fast enough to make the equilibrium "model" useful in predicting their general course, but slow enough that kinetic factors are commonly more important determiners of consequences at a particular time and place.

Humankind is well adjusted to present steady-state conditions and is necessarily concerned with possible changes in these conditions that can be brought about by natural causes or by human activity. It is apparent that the steady state is reasonably stable, because it has maintained an environment much like the present all through human history. Although our ancestors disturbed the environment in many ways (e.g., by setting fires, cutting forests, and polluting streams), other reactions in the steady-state system have compensated for these disturbances. In modern times the population is larger and thus our ability to modify the environment has prodigiously increased. Important questions to consider are whether the chemical reactions of the steady-state complex are still adequate to compensate for changes that humankind can produce, and whether the system is stable enough to resist changes that might be brought about by predictable natural events.

Such problems loom large in a survey of environmental geochemistry. To understand them, we must first look at some of the individual reactions of the steady-state system, then see how changes in constituents of the reactions are compensated for at present, and finally try to predict consequences of some of the possible environmental modifications that can be foreseen in the future. Because our concern is mainly with the environment in which we spend most of our lives, this discussion is focused on the land surface of the continents. Other parts of the earth's environment, for example, the oceans or the high atmosphere, are mentioned only to the extent that they enter the global geochemical system that influences our immediate surroundings.

II. SOLID MATERIALS

Continental surfaces are everywhere underlain by solid rock, and other materials at the surface of the earth are mainly derived from rock. From the behavior of seismic waves we know that ordinary rock perisists downward into the earth for a few tens of kilometers, but that it changes rather abruptly into a different kind of rock at a depth averaging about 35 km. The material above this depth is called the earth's crust. The crust also extends under the ocean basins to a depth of about 6 km, but this part of the crust is not of interest here. Two kinds of materials make up the bulk of continental crust: compounds of silicon and compounds of carbon. Our concern is with the nature of these compounds and their reactions with air and water at the earth's surface.

A. Silicon, Silica, and Silicates

Various kinds of rocks that form the crust of the earth include igneous rocks, such as granite and basalt; metamorphic varieties, such as schist and quartzite; and layered sedimentary rocks, such as sandstone and shale. Most of them consist chiefly of minerals that are compounds of silicon. Much of the inorganic geochemistry of our surroundings is linked to the properties of this one element. Of all the common rock types, only limestone and dolomite are exceptions to the rule that rocks are predominantly aggregates of silicon compounds.

Atoms of silicon are small and have four valence electrons, so that four strong bonds can be formed with adjacent atoms. In the crystal structures of rock-forming minerals, the adjacent atoms are always oxygen, so that each Si atom is surrounded by four O atoms at the corners of a tetrahedron. Each oxygen in the SiO_4^{4-} group has an unsatisfied valence that can link the group to other atoms, either other atoms of silicon or atoms of metals. If all the bonds are formed with metals, a simple crystal structure can result, with alternating SiO_4^{4-} groups and metal cations. But if some of the oxygen bonds are formed with other silicon atoms, the SiO_4^{4-} tetrahedra can link themselves into many kinds of structures: single chains, double chains, rings, and three-dimensional frameworks, generally with metal ions forming linkages between the structures. It is the unique capacity of SiO_4^{4-} groups to join in many patterns that gives rise to the extraordinary variety of silicon–oxygen–metal compounds that compose ordinary rocks.

In the common mineral quartz, the tetrahedra are linked in a three-dimensional structure with no metals, thus the formula of the mineral is simply SiO_2, or silicondioxide, commonly called silica. The metal-containing minerals are collectively called silicates; formulas of many silicates are very complex because metals can substitute extensively for one another in the crystal structures. If somewhat simplified formulas are used, a few examples include

feldspar, $(K, Na)AlSi_3O_8$ and
$(Na, Ca)((Si, Al)AlSi_2O_8$;
amphibole, $Ca_2(Mg, Fe)_5Si_8O_{22}(OH)_2$;
pyroxene, $Ca(Mg, Fe)Si_2O_6$;
mica, $KAl_2(AlSi_3O_{10})(OH)_2$ and
$K(Mg, Fe)_3AlSi_3O_{10}(OH)_2$;
olivine, $(Mg, Fe)_2SiO_4$;
kaolinite, $Al_4Si_4O_{10}(OH)_8$; and
serpentine, $(Mg, Fe)_6Si_4O_{10}(OH)_8$.

The first five on this list, along with quartz, are the major constituents of igneous rocks; except for olivine, they are also common is metamorphic rocks. Kaolinite is common in sedimentary rocks, serpentine in some metamorphic rocks. The minerals range in hardness from feldspar, nearly as hard as quartz, to kaolinite, which is soft and chalky. The five silicates that contain iron are dark colored (green to black), and the others are light colored. Such a sampling gives only a small glimpse of the variety of properties found in silicate minerals. One property they have in common is that all are resistant to chemical attack and react only slowly to powerful reagents.

B. Composition of the Earth's Crust

The minerals in the above list, plus quartz, make up a considerable part of the bedrock under all parts of the earth's surface. The mineral formulas include only a small number of elements, and these elements are thus among the most abundant constituents of the earth's crust. In any

TABLE I Average Abundances of Major Elements (wt %) in the Earth's Crust, Five Common Rocks, and Soil[a]

Element	Crust	Granite	Basalt	Shale	Sandstone	Limestone	Soil
O	46.3	48.5	44.1	49.5	49.2	49.4	49.0
Si	27.7	32.3	23.0	23.8	32.7	3.2	33.0
Al	8.1	7.7	8.4	8.8	4.3	0.9	7.1
Fe	5.6	2.7	8.6	4.7	2.9	1.7	4.0
Ca	4.4	1.6	7.2	2.5	3.1	34.0	1.5
Na	2.4	2.8	1.9	0.9	1.0	0.1	0.5
Mg	2.4	0.4	4.5	1.4	1.2	0.6	0.5
K	2.0	3.2	0.8	2.5	1.5	0.3	1.4

[a] Source for sandstone, limestone, and soil: Bowen, H. J. M. (1979). "Environmental Geochemistry of the Elements," Academic Press, London. Main soruces for crust, granite, basalt, and shale: Wedepohl, K. H., ed. (1969–1974). "Handbook of Geochemistry," Springer-Verlag, Berlin; Taylor, S. R. (1964). "Abundance of elements in the continental crust." *Geochim. Cosmochim. Acta* **28,** 1273–1286.

estimate of the average composition of the continental crust, the eight elements O, Si, Al, Fe, Mg, Ca, Na, and K stand out as more abundant than all others (Tables I and II).

To arrive at an average composition (such as that given in the second column of Tables I and II) is not simple. It requires combining analyses of samples of many different rock types and making an estimate of the relative abundance of them in the part of the earth accessible to direct observation. Because granite and basalt are the two most common igneous rocks, and because most sedimentary and metamorphic rocks are ultimately derived from these two, it is customary to calculate an average analysis by giving roughly equal weight to granite and basalt analyses (the weights used depend on the prejudices of the compiler). Because of uncertainties in the weighting and because of differences of opinion about the rocks to be included under the headings "granite" and "basalt," average analyses prepared by different authors are usually not identical. Generally values assigned to the major elements are much alike, but values for some of the minor elements can differ by a factor of 2 or 3.

Tables I and II clearly show some of the major chemical differences between basalt and granite. A rock consisting chiefly of quartz and feldspar, granite shows high Si, Na and K values; basalt, which is made up chiefly of Fe–Mg silicates and Ca–Na feldspar, is rich in Fe, Mg, and Ca. (Since the oceanic crust is largely basalt, its inclusion in the average analysis would result in larger values for these latter three elements.) Minor elements in the two rocks tend to follow major elements of similar chemical properties: thus Ba, Rb, and La are more abundant in granite, while V, Cr, and Ni predominate in basalt. Some elements, for example, Al, Ga, and Nb, show little difference from one rock to the other and are also the same in the crustal average.

Notable in the crustal average is the great abundance of only a few elements. Oxygen and silicon alone make up nearly 75% of the weight of the crust, and including the six metals mentioned above, $>$98% is accounted for. Such common and important elements as H, C, S, Cu, and Pb make up small parts of the small remainder. Only because some of the minor elements have been concentrated locally by geochemical processes do we find them in the quantities needed for civilized living.

C. Carbon Chemistry

An all-important minor element, from a human standpoint, is carbon, the chief constituent of materials produced by life processes. Because plants and animals are so conspicuous and ever-present in the world around us, it is difficult to realize that their principal element makes up a bare 0.05% of the weight of the earth's crust. Organic materials, living and dead, form only a thin and discontinuous veneer on the outer surface of the silica-rich materials that make up most ordinary rocks. The rocks themselves are part of a layer that is thin in comparison with the earth's diameter. Thus carbon composes a very minor portion of the planet's overall chemistry, but in the human environment at the surface it plays a major role.

The properties of organic materials are a reflection of the chemistry of this one element, much as the properties of rocks depend largely on the chemistry of silicon. Like silicon, carbon consists of small atoms with four valence electrons, but the atoms are so small that four oxygen atoms cannot fit comfortably around them. Thus, in contrast to silicon, carbon atoms are not linked with oxygen in tetrahedra, but rather form strong bonds with each other, a property that permits them to join into chains and rings of great variety. Attached to the chains and rings are the other elements that are common in living organisms, chiefly oxygen and hydrogen as well as small amounts of nitrogen, phosphorus, sulfur, and many others. Thus the complexity of silicate minerals can be traced to the ability of silicon–oxygen tetrahedra to join in many patterns with metals as linkages, and the complexity of organic compounds can be explained by the tendency of carbon

TABLE II Average Abundances of Minor Elements (ppm) in the Earth's Crust, Five Common Rocks, and Soil[a,b]

Element	Crust	Granite	Basalt	Shale	Sandstone	Limestone	Soil
Ag	0.07	0.04	0.1	0.1	0.2	0.1	0.05
As	1.5	1.3	1.7	10	1	1	6
Au	0.003	0.002	0.004	0.003	0.003	0.002	0.003
B	10	15	5	100	30	20	20
Ba	500	700	300	600	300	90	500
Be	3	5	0.5	3	<1	<1	0.6
Br	2	0.5	0.5	5	1	6	10
C	480	360	610	14,000	16,000	116,000	20,000
Cd	0.15	0.1	0.2	0.3	0.05	0.03	0.4
Ce	70	90	50	70	80	20	50
Cl	130	200	60	170	500	76	100
Co	22	3	40	20	0.3	0.1	8
Cr	100	20	200	130	35	11	70
Cs	3	5	1	7	0.5	0.5	4
Cu	50	12	90	50	30	5.5	30
Dy	6	7	4	5	7	2	5
F	950	1300	500	800	180	220	200
Ga	18	19	17	25	6	2.5	20
Hf	5	7	4	3	4	0.3	6
Hg	0.05	0.06	0.05	0.1	0.05	0.2	0.06
I	0.2	0.2	0.1	0.1	0.1	0.1	5
La	35	55	20	40	42	10	40
Li	20	30	12	60	38	7.5	25
Mn	1000	500	1500	850	460	620	1,000
Mo	1.5	1.5	1	2	0.2	0.2	1.2
N	20	20	20	600	120	80	2,000
Nb	20	20	20	15	0.05	0.3	10
Nd	30	40	20	30	38	9	35
Ni	75	5	150	80	9	7	50
P	1100	700	1400	750	440	700	800
Pb	14	24	4	20	10	6	35
pr	9	11	7	9	11	3	7
Rb	90	150	30	160	46	50	150
S	300	300	300	2,500	2,300	1,100	700
Sc	20	8	35	15	1	1	7
Se	0.05	0.05	0.05	0.6	0.05	0.03	0.4
Sm	7	9	5	7	8	2	5
Sn	2.5	3	2	6	0.5	0.5	4
Sr	375	300	450	400	20	610	250
Ta	2	3	1	2	0.05	0.05	2
Th	12	20	2	12	4	2	9
Ti	5600	2100	9000	4,500	3,500	300	5,000
U	3	5	0.5	3.5	0.5	2	2
V	150	50	250	130	20	45	90
Y	35	40	30	35	54	23	40
Zn	75	50	100	90	30	20	90
Zr	190	240	140	180	220	20	300

[a] Elements not listed in the table have abundances in the crust of <1.5 ppm.

[b] Source for sandstone, limestone, and soil: Bowen, H. J. M. (1979). "Environmental Geochemistry of the Elements," Academic Press, London. Main sources for crust, granite, basalt, and shale: Wedepohl, K. H., ed. (1969–1974). "Handbook of Geochemistry," Springer-Verlag, Berlin; Taylor, S. R. (1964). "Abundance of elements in the continental crust." *Geochim. Cosmochim. Acta* **28,** 1273–1286.

atoms to unite with each other in structures whose other components are largely nonmetals.

Organic compounds are usually less stable than silicates in environments at the earth's surface. Their reactions are generally slow, but most of them yield, in times that are measurable in human terms, to reactions involving oxidation or internal rearrangement of complex molecules to form simpler ones. Common products of oxidation reactions are CO_2 and H_2O; a common ultimate product of nonoxidative decay is methane, CH_4. Some organic compounds are exceptions to the general rule, and under the right circumstances may persist for geologic times. Coal is a rock formed by the burial of partly decayed vegetation from ancient swamps, and petroleum consists of organic compounds resulting from partial decay of the remains of tiny marine organisms that accumulated in ancient seas. Most sedimentary rocks contain at least ~1% of resistant organic materials that have persisted with little change since shortly after the rocks were formed.

One other group of rocks contains carbon in abundance: these are carbonate rocks, of which limestone (chiefly $CaCO_3$) and dolomite [chiefly $CaMg(CO_3)_2$] are the principal varieties. The carbonate minerals can form in many ways, but in large part they are an indirect product of organic processes. Most limestone consists of material originally deposited as the shells of myriads of marine organisms, and much dolomite is formed by replacement of this material. Some limestone is nearly pure $CaCO_3$, but most contains at least minor amounts of quartz, clay, and other silicate minerals. Mixtures of carbonate and silicate minerals are very common in sedimentary rocks and locally in metamorphic rocks, but carbonate minerals are scarce or absent in common igneous rocks.

D. Weathering

Rocks and organic materials exposed at the earth's surface are subject to attack by water and air. The attack involves a complex of slow reactions, partly physical and partly chemical, which are collectively called weathering. The chemical part of weathering is a slow approach to a new equilibrium for minerals and compounds that were formed in quite different environments. The silicate minerals of igneous and metamorphic rocks, for example, were generated under conditions of high temperature and often high pressure out of contact with the atmosphere. The carbonate minerals of limestone and dolomite were formed in bodies of water with compositions different from that of rain and fresh running water. Organic materials originate in the special chemical environment of living substances. When these materials are exposed to air and water at the earth's surface, they must adjust to conditions of temperature, pressure, and composition of adjacent materials that are different from those where they were formed. The process of adjustment gives rise to the discoloring and softening of most rock surfaces and to the disappearance of most organic matter. Because rocks are slowly disintegrated by weathering, the ulimate result is the formation of loose debris that remains as a layer of soil over the rock surface or is carried away by wind or running water to be deposited as sediment.

Many kinds of weathering reactions can be distinguished. Simplest is the ordinary dissolving of soluble compounds: when a rock containing salt or gypsum comes in contact with water, the reactions are

$$\begin{gathered} \mathrm{NaCl} \rightarrow \mathrm{Na}^{+} + \mathrm{Cl}^{-} \\ \mathrm{CaSO_4 \cdot 2\,H_2O} \rightarrow \mathrm{Ca}^{2+} + \mathrm{SO_4^{2-}} + 2\,\mathrm{H_2O}. \end{gathered} \tag{1}$$

Dissolution may be aided by hydrolysis, which means the reaction of one of the liberated ions with water forms a weak acid or a weak base to set free H^+ or OH^-. For example, carbonates dissolve to form a bicarbonate ion by the reaction

$$\mathrm{CaCO_3} \rightarrow \mathrm{Ca}^{2} + \mathrm{CO_3^{2-}} \tag{2}$$

followed by

$$\mathrm{CO_3^{2-}} + \mathrm{H_2O} \rightarrow \mathrm{HCO_3^-} + \mathrm{OH}^-, \tag{3}$$

and silicates dissolve the produce the even less dissociated silicic acid or its ion:

$$\mathrm{Mg_2SiO_4} + 3\,\mathrm{H_2O} \rightarrow 2\,\mathrm{Mg}^{2+} + \mathrm{H_3SiO_4^-} + 3\,\mathrm{OH}^-. \tag{4}$$

These reactions go further if the water contains acid to combine with the liberated OH^-. The most common acid present in water, responsible for weathering under natural conditions, is carbonic acid, H_2CO_3, which is formed by the dissolving of CO_2 from air:

$$\mathrm{CO_2} + \mathrm{H_2O} \rightleftharpoons \mathrm{H_2CO_3}. \tag{5}$$

This acid is so readily formed that nearly all the water at the earth's surface contains at least small concentrations. The acid is very weak, as shown by its dissociation constants, $K_1 = 10^{-6.4}$ and $K_2 = 10^{-10.3}$. Despite the weakness and the small amount generally present (solubility of CO_2 in pure water at 25°C is $\sim 10^{-5}\,M$ in equlibrium with the $10^{-3.5}$ atm of the gas in air), carbonic acid can greatly increase the solubility of carbonates and silicates:

$$\mathrm{CaCO_3} + \mathrm{H_2CO_3} \rightarrow \mathrm{Ca}^{2+} + 2\,\mathrm{HCO_3^-} \tag{6}$$

and

$$\mathrm{Mg_2SiO_4} + 4\,\mathrm{H_2CO_3} \rightarrow 2\,\mathrm{Mg}^{2+} + \mathrm{H_4SiO_4} + 4\,\mathrm{HCO_3^-}. \tag{7}$$

The effectiveness of the acid is greatly increased if weathering takes place (as it commonly does) in contact with soil, because decaying vegetation makes the concentration of CO_2 in the air in pores of the soil much higher than that in the surrounding atmosphere.

These reactions show simple dissolving of the minerals prompted by the tendency of their anions to form weak acids. The carbonate reaction is rapid enough that a close approach to equilibrium is often indicated by the measured concentrations of Ca^{2+} being similar to those calculated from dissociation constants.

More complicated are hydrolysis reactions of silicates with two or more metals in their formulas. Such a mineral commonly dissolves incongruently, which means that at first it only partially dissolves. A new mineral is then left behind, and the solution has constituents with a mole ratio different from that in the original mineral. A good example is the hydrolysis of aluminosilicate minerals such as the feldspars, which are particularly important since these reactions lead to the formation of clay minerals:

$$\begin{aligned} 2\,\underset{\text{feldspar}}{KAlSi_3O_8} + 2\,H_2CO_3 + 9\,H_2O \rightarrow \\ \underset{\text{kaolinite}}{H_4Al_2Si_2O_9} + 2\,K^+ + 4\,H_4SiO_4 + 2\,HCO_3^-. \quad (8) \end{aligned}$$

Note that this process is driven [as in Eq. (7)] by the reaction of an acid (H_2CO_3) with silica to form the undissociated silicic acid (H_4SiO_4). Here, however, only part of the silica reacts, some being left behind in the clay, and the ratio K : Si in solution is 1 : 2 in contrast to 1 : 3 in the original feldspar. Aluminosilicate minerals are common constituents of ordinary rocks, and their weathering produces many types of clays. The kind of clay produced depends on the nature of the original mineral and on the conditions of weathering, with the reactions often being much more complex than Eq. (8). One or more clay minerals, such as generally kaolinite, montmorillonite, or illite, is the usual product of aluminosilicate weathering in temperate climates, but under extreme weathering conditions (particularly in the tropics) the clay minerals break down by releasing silica to solution and leaving a hydrate of alumininum oxide [gibbsite, $Al(OH)_3$, or boehmite, AlOOH] as a solid residue.

Another common reaction of weathering is oxidation, which primarily affects minerals containing iron and/or sulfur:

$$Fe_2SiO_4 + O_2 + H_2O \rightarrow Fe_2O_3 + H_4SiO_4, \quad (9)$$

$$ZnS + 2\,O_2 \rightarrow Zn^{2+} + SO_4^{2-}, \quad (10)$$

and

$$2\,FeS_2 + 15/2\,O_2 + 4\,H_2O \rightarrow Fe_2O_3 + 4\,SO_4^{2-} + 8\,H^+. \quad (11)$$

The iron oxide product of reactions (9) and (11) is written as Fe_2O_3 (hematite), but it could as well be shown as a hydrated form, such as FeOOH (goethite) or the amorphous material commonly given the formula $Fe(OH)_3$. These forms of ferric oxide have a variety of colors ranging from red through purple and brown to yellow and orange; the color depends on both the particular mineral and the grain size. These are the chief coloring agents in rocks of ordinary environments and are responsible for the common browns, reds, and yellows of weathered rock surfaces and sediments derived from them. Reaction (10), the oxidation of pyrite, is particularly important, because it shows that upon weathering, this most abundant sulfide mineral gives an acid solution and abundant ferric oxide, two products that are usually conspicuous where sulfide ore deposits are exposed to air and water. Oxidation is also a major process in the weathering of organic materials, which leads to carbon dioxide and water as the principal products; for example,

$$\underset{\text{carbohydrate}}{C_6H_{12}O_6} + 6\,O_2 \rightarrow 6\,CO_2 + 6\,H_2O. \quad (12)$$

If nitrogen is present, it is usually released as N_2; sulfur is converted to sulfate in solution and phosphorus to phosphate.

Long-continued weathering ultimately produces solutions containing most of the original metal ions present in the rock and some of the silica, plus solid products consisting of newly formed compounds (ferric oxide and clay minerals) and residual solids that are little affected by weathering reactions (chiefly quartz and minor amounts of such resistant minerals as zircon and garnet). Because weathering is very slow and its rates are markedly different for different minerals, its products generally also include material that is only partly weathered. For example, feldspar in cloudy grains partly converted to clay minerals [Eq. (8)] is a common constituent of weathered debris. The dissolved metals are in large part carried by running water to the sea, although some may be precipitated en route. The solid debris is roughly sorted by moving water into materials of different grain sizes; it is then deposited, either in the ocean or in basins on land, as sand, silt, or clay sediment. If the sediments accumulate in sufficient quantity they may ultimately be compacted and cemented to form layers of sedimentary rock.

Thus weathering results in the conversion of igneous and metamorphic rock into sediments and dissolved ions in the water of streams and oceans, and in the disappearance of most dead organic matter. It is a process of decomposition of complex original compounds into simpler ones and a rough sorting of elements according to their chemistry: Si mainly into detrital sands because of the resistance of quartz to chemical and physical attack, Al into fine-grained sediments because of the stability of clay minerals, and Na and K into the ocean because of the solubility of most of their compounds. The new substances are nearer to chemical equilibrium with air and water than the original rock minerals were, but total equilibrium is seldom attained. Other overall chemical effects are the depletion of O_2 and CO_2 from the atmosphere, the net change in oxidation state of iron and sulfur, the neutralization of naturally occurring acid solutions, and the creation of slightly alkaline solutions by hydrolysis.

E. Formation of Soils

The solid products of weathering are not removed as fast as they form in most places and in part remain close to their point of origin as soil. Soil formation involves two major processes: (1) slow chemical alteration by water seeping through the weathered rock material after rains and (2) mixing of the rock material with organic debris produced by the decay of plants. The chemical changes are a continuation of the reactions involved in weathering of the original rock: the soluble alkali metals (Na and K) and alkaline earth metals (Ca and Mg) are leached out of the upper layers of the soil and partly removed altogether; Al goes into clay minerals that may be moved mechanically from upper to lower layers; and finally, Fe is dissolved from the upper part (where it is kept in the more soluble reduced form by organic matter) and precipitated as ferric oxide as it moves downward. By such processes soil commonly develops a roughly layered structure, which comprises (1) an upper layer (called the A horizon) from which much of the soluble material has been removed and in which the organic matter is concentrated, (2) a middle layer (B horizon) in which clay and ferric oxide are prominent constituents, and (3) a lower layer (C horizon) made up of little-altered rock debris.

Organic matter that is incorporated in soil as plant debris is subject not only to decay but to the biological activity of living bacteria and higher organisms. The soil is stirred by this activity, complex organic molecules are broken down into simpler ones, and CO_2 is liberated to form acid solutions for further decay of the rock material. The decomposing organic matter also provides surfaces that can trap some of the alkali and alkaline earth ions. Although these ions are in large part leached out of the soil, some are held by adsorption on both clay particles and organic matter—a fortunate circumstance, because it is from these adsorbed ions that plants obtain most of the mineral substances necessary for their growth.

Soils vary greatly in their structure and composition, with the differences depending on the rocks from which they are derived and on the prevailing climate. Surprisingly, climate is usually the more important of the two factors. In humid climates the three major horizons are commonly well developed, with an upper layer impoverished in the more soluble elements and a B horizon containing much clay and iron oxide, which creates a “hardpan.” In the tropics, where oxidation is rapid and organic matter decays quickly, ferric oxide may be prominent in all layers, which gives rise to the red soils called laterites. Where rainfall is less, there is not as much leaching from the A horizon and there may be some accumulation of Ca and Mg in the B horizon. Because the more soluble elements are essential for plant growth, these soils tend to be more productive than those of very humid regions. Leaching becomes less and less effective with increasing aridity, and the accumulation of alkali and alkaline earth elements at all levels produces the typical alkaline soils of desert regions. These generalizations apply to soils developed on most kinds of rock; in areas where soil formation has been going on for a long time, the soils covering different varieties of bedrock are often remarkably similar. Where rock compositions are extreme, soils developed on their surfaces do show more dependence on the character of the rock; soils on limestone, for example, can be unusually rich in calcium and those on serpentine rich in magnesium.

Another factor influencing the composition of a soil is its age. Any soil represents a steady state between processes of formation and processes of destruction: as new soil is being generated by weathering of bedrock below, some of the surface material is being removed by erosion. A relatively young soil that develops in areas where erosion is rapid and soil-forming processes have little time to operate (as on a steep hillside) would be expected to have a composition similar to that of the rock beneath it. An older soil, for example, over a flat area where the formation of soil horizons has been going on for a long time, would have a composition largely determined by climate. Because soil formation is exceedingly slow and because mature soils are better suited for agriculture than raw young soils are, the prevention of rapid soil erosion is an important part of environmental management.

Attempting to determine an average composition for soil has little value because any one soil differes from top to bottom and because soils differ greatly from place to place. Nevertheless, attempts have been made to give rough estimates of the overall chemistry of soils, and one such effort is included as the last column in Tables I and II. Although the particular numbers may be questionable, they provide a useful summary of the general chemical effects of soil formation. In comparison with such common rocks as granite and basalt, for example, soils are markedly depleted in the alkali and alkaline earth elements, while C and N are higher because of the added organic matter. Note also that the composition is similar to that of shale, a sedimentary rock formed largely from the fine-grained part of the debris produced by rock weathering.

F. Ore Deposits

The chemical reactions of weathering and soil formation, along with the more active processes of erosion, can lead to enrichment of elements or groups of elements in some deposits and depletion in others. Under special geological conditions, the enrichment of particular elements becomes so marked, and the amount of accumulation so large, that the elements can be extracted in commercially important quantities. Accumulations from which elements can be extracted at a profit are called ore deposits. Not all ore deposits orginate in the low-temperature processes of weathering and erosion, but some important ones do form in this way and have contributed greatly to the materials essential for industrial civilization.

The simplest example is placer deposits, which are formed when grains of a resistant mineral like gold, originally dispersed in small amounts through bedrock, are released through long periods of weathering of the more easily dissolved part of the rock and then accumulated by the action of running water or waves. In addition to being a source of gold, placers are important sources of tin, thorium, and the platinum metals. Rich hematite deposits, such as the exhausted high-grade ores of the Mesabi Range in Minnesota and similar deposits being exploited in Venezuela and Brazil, were formed from rocks containing abundant iron minerals through processes similar to those that can be observed segregating ferric oxide in sediments and soils. Important deposits of pure clays have been formed by weathering of rocks rich in feldspar, and the pure quartz sand needed for making glass comes from accumulations of quartz grains left after the more soluble parts of the original silica-rich rocks were dissolved or winnowed out by erosion. Some deposits of copper and uranium, and possibly also lead and zinc, have formed by the dissolution of these metals from lower-grade material in solutions similar to those involved in weathering and by their subsequent precipitation when conditions of acidity or oxidation changed.

Many ore deposits are formed by higher-temperature processes associated directly or indirectly with the intrusion of igneous magma into rocks of the crust. Many of these deposits consist of sulfide minerals, which are stable under reducing conditions at depths out of contact with the atmosphere. When the rock material that once covered such deposits is removed by weathering and erosion to directly expose the sulfide ore to air, the metal compounds undergo a variety of reactions. Examples have been given in Eqs. (10) and (11); two others are

$$\mathrm{H_2O} + \underset{\text{chalcolite}}{\mathrm{Cu_2S}} + 5/2\,\mathrm{O_2} + 2\,\mathrm{HCO_3^-} \rightarrow$$

$$\underset{\text{malachite}}{\mathrm{Cu_2(CO_3)(OH)_2}} + \mathrm{SO_4^{2-}} + \mathrm{H_2CO_3} \qquad (13)$$

and

$$17/2\,\mathrm{O_2} + 2\,\underset{\text{chalcopyrite}}{\mathrm{CuFeS_2}} + 2\,\mathrm{H_2O} \rightarrow$$

$$2\,\mathrm{Cu^{2+}} + \underset{\text{hematite}}{\mathrm{Fe_2O_3}} + 4\,\mathrm{SO_4^{2-}} + 4\,\mathrm{H^+}. \qquad (14)$$

In all such reactions the sulfur is oxidized to sulfate, the solution becomes more acid, and the metal ions are either set free in solution or precipitated as carbonates, oxides, sulfates, or other compounds formed with the common anions of waters at the earth’s surface. Because pyrite and other sulfide minerals containing iron are very common in sulfide deposits, hematite or a hydrated form of ferric oxide is nearly always present in large amounts at the surface around a weathering deposit of sulfide ore. The resulting patches of brown and yellow staining on surface rocks have long been used by prospectors as an indication of the possible presence of valuable ore beneath. As an ore deposit is opened up for the recovery of its metal(s), sulfide grains may be scattered widely over the surface and the acid water draining the deposit may carry abundant iron oxide far downstream. The resulting large area of discoloration and stunted vegetation is offensive in the eyes of environmentalists, and mining companies take expensive measures to limit the extent and unsightliness of such areas.

Ore deposits are an important part of the environment for the simple reason that their products are indispensable to civilized life as we know it. They are rare because their formation requires special combinations of geological factors, for example, the right temperature and pressure and fluids of the right compostion. In the search for new deposits, and in devising methods for controlling the environmental effects when known deposits are exploited, geochemistry has a major role to play.

G. Reversal of Weathering Reactions

Weathering reactions are part of the huge steady-state system that makes up our environment, which means that somewhere these reactions must be reversed. Proof that reversal takes place lies in the simple fact that high mountains still exist despite continuous weathering of the continental surface for more than 4 billion yr. If there were no compensation for the effects of weathering, the continents would long ago have become level plains mantled with thick, weathered debris. We can be sure that compensation occurs, but just how it occurs and whether it is complete in all details are questions subject to lively debate.

Weathering is essentially a destructive process in which complex minerals in igneous and metamorphic rocks are broken down by reaction with water and gases from the air into simpler substances that are deposited as sediments and ultimately converted to sedimentary rocks. The overall process can be summarized in a symbolic reaction:

$$CaSiO_3 + CO_2 \rightarrow CaCO_3 + SiO_2 \qquad (15)$$

or

$$CaSiO_3 + H_2O + 2\,CO_2 \rightarrow Ca^{2+} + 2\,HCO_3^- + SiO_2\,, \qquad (16)$$

where $CaSiO_3$ stands for all the complex silicate minerals in the original rocks, SiO_2 for the new solid materials produced by weathering that ultimately become clastic sediments, and $CaCO_3$ (or $Ca^{2+} + 2\,HCO_3^-$) for material dissolved during weathering and partly precipitated as chemical sediment. Somewhere among the many processes that shape the earth's surface are the reversals of such reactions.

The most significant such process is a part of plate tectonics. In subduction zones great masses of rock and sediment are dragged down into the earth's upper mantle by moving crustal plates and are squeezed and heated to form metamorphic equivalents or melted to produce liquid rock (magma) that rises up through the crust in volcanoes or "freezes" under the surface of the earth as granite. In addition, the gases that take part in weathering are recycled, because the water and carbon dioxide that are added to sedimentary material [according to Eqs. (8) and (11)] reappear as the major gases of volcanic eruptions. In effect, high-temperature processes in the earth's interior serve to reverse the low-temperature weathering reactions symbolized by Eqs. (15) and (16): silica combines with simple metal compounds to re-form the complex silicates of metamorphic and igneous rocks, and other elements in the metal compounds are released to maintain the supply of CO_2, H_2O, and other gases in the atmosphere. On this large scale the destructive work of weathering is adequately compensated by the synthesis of silicates and liberation of gases brought about by plate tectonics.

On a smaller scale, it is less certain that parts of the overall scheme remain in balance, or that some of the compensating reactions are fast enough to prevent progressive and continuing changes in the geochemical environment. The formation of limestone, for example, requires that CO_2 be taken from the atmosphere. It is difficult to assess whether all limestone is recycled through subduction zones, in view of the great quantities of this rock that have been preserved on the continents for at least 500 million yr. It is possible that CO_2 is being permanently removed from air and locked up in $CaCO_3$ far from any present or likely future subduction zone. A similar possibility exists for oxygen, which is continually reacting with minerals of iron and sulfur. It is uncertain whether these reactions are adequately reversed by subduction and photosynthesis and whether oxygen is being permanently lost to sediments. We look at such problems in Sections III.E and IV.E following a discussion of atmospheric and aqueous chemistry.

The problem of reversibility is particularly important regarding ore deposits, because both natural and human processes are destroying these concentrations of unusual materials that were formed slowly over long periods. It would be useful to know to what extent new accumulations of ore are now being formed that can make up for dispersal of those currently known. Although ore deposits are surely forming today as they have in the past, the rate of formation is so slow that in human terms it has no significance. Here the balance that probably exists in nature between formation and destruction is being upset by human influence, and ore deposits that are being exhausted now will not be replenished at any times that would interest our descendants.

Thus on the basis of imperfect knowledge we can guess that the earth's present chemical environment is being maintained in a generally stable steady state, with the obvious reactions of weathering being reversed by less visible processes deep within the earth. There are some details of the steady state that are uncertain. The dispersal of metals and nonmetals from ores is one example of a part of the steady state in which human interference has grossly disturbed the natural balance of compensating reactions.

III. THE ATMOSPHERE

The earth's atmosphere is a mixture of many gases, some essential to life and some continuously active in the weathering of rocks. Like the solid materials of the crust, the gases in air are maintained in a steady state, some of them being rapidly consumed and as rapidly generated, in a balance that has persisted without major change for long periods of geologic time. As for rocks and minerals, important geochemical data about the atmosphere relate to the mechanisms by which the balance is maintained, the stability of the balance, and the possible long-term effects of human activites.

A. Composition and Structure

The part of the atmosphere of direct concern to the environment is the relatively dense lower layer called the troposphere, ranging in thickness from about 10 km at the

TABLE III Composition of Dry Air (vol %)[a,b]

Gas	vol %
N_2	78.08
O_2	20.95
Ar	0.93
CO_2	0.033
Ne	0.0018
He	0.00052
CH_4	0.00015
Kr	0.00011
H_2	0.00005

[a] Other important minor gases, all <0.0001 vol %: Xe, N_2O, NO, NO_2, NH_3, H_2S, SO_2, CO, O_3, $CClF_3$, and CCl_2F_2.

[b] Main source: Wedepohl, K. H., ed. (1969–1974). "Handbook of Geochemistry," Springer-Verlag, Berlin.

poles to 16 km over the equator. Clouds and storms are limited to this part of the atmosphere. Beyond the troposphere is the much less dense stratosphere, and beyond the stratosphere rarefied gases persist for many tens of kilometers into space. The stratosphere constituent of most immediate interest to the environment is a layer of ozone which shields the earth's surface from high-energy ultraviolet radiation from the sun.

The troposphere is kept so sufficiently stirred that its composition is constant from place to place, except for variable amounts of water vapor and local additions of gas from volcanic activity and industrial operations. As shown in Table III, ordinary dry air consists of about 80% nitrogen and 20% oxygen by volume, plus ~1% argon and minor amounts of many other gases. Of the minor gases, carbon dioxide (0.03%) is the most important. Water vapor can amount to 4% in warm humid air or can be nearly absent in air over deserts. In addition to gases, air contains variable amounts of tiny solid particles and liquid droplets (aerosols): the water droplets of clouds, the mineral and organic particles of windblown dust, soot particles from combustion, and salt crystals from the evaporation of ocean spray. Such fine-grained material eventually settles to the surface but may remain suspended for months or years. Especially persistent is the material ejected in the eruption clouds of volcanoes in the form of solids and acid liquids, which sometimes remains in the stratosphere and troposphere in amounts large enough to affect weather patterns for several years.

B. Oxygen and Carbon Dioxide

From a cosmic perspective, the strangest constituent of the earth's atmosphere is free oxygen, O_2. This active gas is never present in more than traces in the atmospheres of other planets and all other known objects in the universe. The earth, however, has oxygen as a major constituent of its gaseous envelope. The reason for this is the reaction called photosynthesis, the process by which green plants are able to use the energy of sunlight to produce organic compounds from carbon dioxide and water. The overall reaction may be symbolized

$$6\,CO_2 + 6\,H_2O \rightarrow C_6H_{12}O_6 + 6\,O_2, \qquad (17)$$

but the mechanism is much more complicated than this simple equation suggests. The reaction takes place in green leaves, with CO_2 coming from the surrounding air and H_2O from the soil through the plant's roots. It is catalyzed by the green leaf pigment chlorophyll, which absorbs sunlight and transfers the energy to the reacting molecules. Plants have the further ability to manufacture other organic molecules from those first formed and thus to prepare the foods on which all animal life depends.

Oxygen, which is a by-product of the photosynthesis reaction, is important to animals as a source of energy through respiration. The reaction involved in respiration can be symbolized as the reverse of Eq. (17), with the understanding that it applies to many other organic molecules in addition to the simple carbohydrate shown. The reverse reaction also shows what happens when organic materials are burned, as symbolized by Eq. (12), or when they are oxidized by the slower processes of decay. Thus living things maintain a balance between CO_2 and O_2: plants use the former to produce the latter, and animals accomplish the reverse (Fig. 1). It is no accident that the earth is unique not only in the abundance of O_2 in its atmosphere but also in the flourishing of life on its surface.

Photosynthesis is not the only source of molecular oxygen in a planetary atmosphere. Small amounts can be

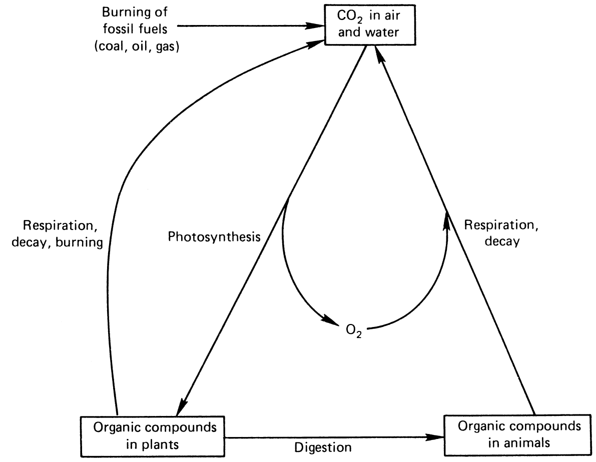


FIGURE 1 Major short-time controls of the CO_2 concentration in air (periods of years or decades).

formed in the stratosphere by the action of high-energy radiation on water molecules, and circulation of the atmosphere brings it to lower levels. Oxygen thus produced would have only a transitory existence because it would soon react with other gases or solids, but it may have played a role in the earth's early history before life became abundant. Steps in the development of life from its primitive beginnings in an atmosphere nearly devoid of oxygen to the establishment of photosynthesis as a major reaction can be traced only speculatively. But at the present time, and for a long while in the geological past, it is certain that the earth's supply of oxygen is maintained by this one reaction.

Carbon dioxide also has sources other than the cycle represented by Eq. (17) and its reverse (Fig. 2). It is one of the major gases emitted during volcanic eruptions, and it is produced in large amounts by the precipitation of $CaCO_3$ from bicarbonate solutions:

$$Ca^{2+} + 2\,HCO_3^- \rightarrow CaCO_3 + H_2O + CO_2. \qquad (18)$$

Note that this reaction is the reverse of Eq. (6). Unlike oxygen, CO_2 is a relatively unreactive gas that is found in the atmospheres of planets other than the earth. It may have been part of the earth's primitive atmosphere and has certainly been added by volcanoes and the precipitation of calcium carbonate sediments throughout the earth's history. It is currently maintained at a low concentration, probably chiefly by photosynthesis, but it has almost surely been more abundant at times in the earth's past. Because of its low concentration, CO_2 is one gas whose amount can be perceptibly influenced by human activity: the ever-increasing production of CO_2 by the burning of fossil fuels (coal, pertroleum, and natural gas) during the past century has raised the concentration in air by ~10%.

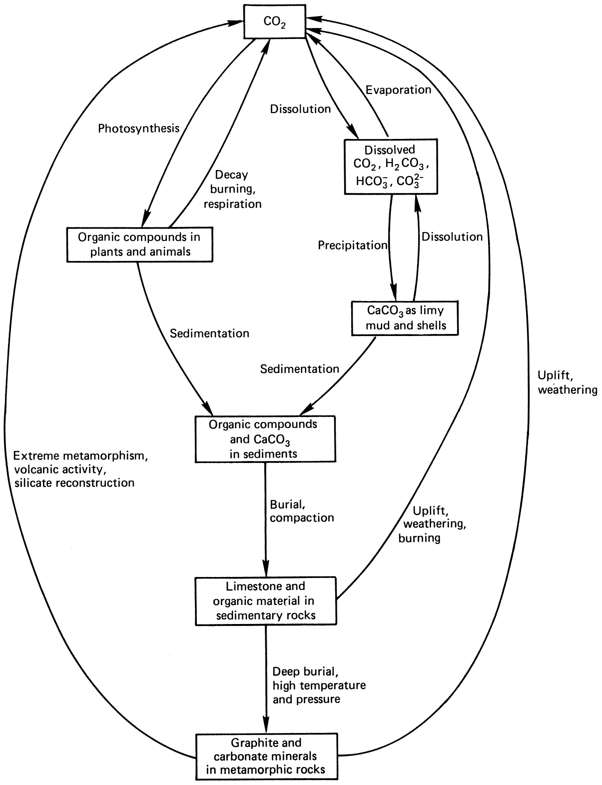


FIGURE 2 Major long-time controls of the CO_2 concentration in air (thousands to millions of years).

C. Nitrogen

The other principal constituent of air, N_2, is a less active element than O_2. It exists in the atmospheres of some other planets and most likely was part of the earth's primordial atmosphere before life began. In the present world nitrogen takes part in major organic processes, and these processes play a role in stabilizing its concentration in air.

As an essential constituent of protein, nitrogen is important to all living things. The nitrogen that plants use to form protein (which is the ultimate source of all animal protein as well) comes not from the atmosphere but from nitrogen compounds dissolved in water that enters plants through their roots. The principal form of nitrogen dissolved in soil water is nitrate, NO_3^-. This is derived from several sources, but mainly from oxidation of ammonia and ammonium compounds generated by the decay of organic matter. Not all the nitrogen from the protein of decaying organisms is converted to nitrate. In the steps leading from protein to ammonia and then to nitrate, some of the nitrogen is set free as the stable diatomic molecule N_2, which returns to the atmosphere. Thus the cycle of nitrogen (Fig. 3) from living organisms to dead organisms and back to living is not complete, since some of the element is lost in the form of N_2, which plants cannot use. Without other methods of generating compounds containing nitrogen (so-called fixed nitrogen), the usable nitrogen in soil water would gradually disappear and plant life would be impossible.

Fortunately, nitrogen compounds enter the soil in other ways. One important source is the activity of a particular kind of bacteria that lives in nodules on the roots of legumes. These bacteria have the unique ability to convert N_2 from air to compounds (amino acids) that the plant can use directly. A few other kinds of bacteria that are free-living in soil and water, plus blue-green algae, are capable of changing N_2 to ammonia (NH_3) and ammonium compounds. These compounds are oxidized by still other varieties of bacteria to NO_3^-, the form of fixed nitrogen used by most plants. Another natural source of combined nitrogen is atmospheric electricity: lightning discharges lead to reaction between N_2 and O_2 to form nitrogen oxides (chiefly NO and NO_2), which are carried to the soil

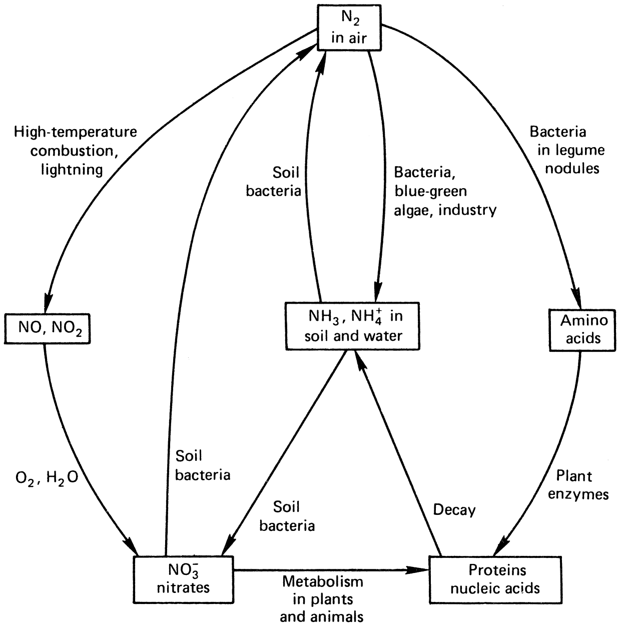


FIGURE 3 Principal reactions of the nitrogen cycle.

dissolved in rainwater and further oxidized to nitrate ion. To supplement these natural sources, manufacturers produce nitrogen fertilizers in large amounts, chiefly by making NH_3 through the direct union of N_2 and H_2 and from it forming ammonium salts. Nitrogen compounds are also major constituents of explosives, so that the "fixing" of nitrogen from the atmosphere has become a large industrial enterprise.

The reservoir of free N_2 in the atmosphere is so vast that human activity cannot appreciably alter its amount, and adequate means are at hand for ensuring plentiful supplies of fixed nitrogen for agricultural and industrial purposes. The major environmental question concerning nitrogen at present relates to the role of nitrogen oxides generated by industry and automobile exhaust as contributors to the increased acidity of rainfall.

D. Minor Constituents

Of the minor constituents of air (see Table III), the most abundant is argon, which makes up nearly 1% of the total. Argon and related elements (He, Ne, Kr, and Xe) have little environmental significance because of their chemical inertness. Their amounts do not change, except that one isotope of argon (^{40}Ar) is produced by the radioactive decay of potassium (^{40}K) and hence has steadily increased through geologic time. Helium also is a product of radioactive decay, but its concentration does not increase because its atoms are light enough to escape from the earth's gravitational field. The small amount in ordinary air represents a balance between its production from uranium and thorium in rocks and its escape into space from the upper atmosphere.

Other minor constituents of air are produced by a variety of processes, for example, decay of organic matter with or without the aid of bacteria, volcanic eruptions, the chemical industry, and automobile exhaust. In addition to the molecules listed in Table III, the trace constituents include free radicals formed by molecular dissociation, generally in photochemical reactions promoted by low-wavelength radiation from the sun. Examples of such reactions include

$$O_2 + \text{radiation} \rightarrow O + O, \qquad (19)$$

$$H_2O + \text{radiation} \rightarrow H + OH, \qquad (20)$$

and

$$NO_2 + \text{radiation} \rightarrow NO + O. \qquad (21)$$

These radicals, which are chemically very reactive, have only a brief existence in the laboratory. But in air and at high altitudes where molecules are far apart and collissions are infrequent, some of the radicals persist for appreciable times and serve as intermediaries in reactions that would otherwise be extremely slow. For example, the formation of ozone from oxygen can be written

$$3\,O_2 \rightarrow 2\,O_3, \qquad (22)$$

but the collision and rearrangement of three O_2 molecules suggested by the equation would be immeasurably slow. In air the reaction takes place through the agency of oxygen atoms produced by photochemical dissociation of O_2 or NO_2 [Eqs. (19) and (21)]:

$$O + O_2 \rightarrow O_3. \qquad (23)$$

The existence of numerous free radicals makes the chemistry of minor atmospheric constituents exceedingly complex.

Gases containing carbon, other than CO_2, include principally methane (CH_4) and carbon monoxide (CO). Both are produced in nature by bacterial decay under low-oxygen conditions, and both are by-products of some industrial processes: CO is also a product of incomplete combustion in gasoline motors. Methane is the major constituent of natural gas and leaks to the surface in small amounts in areas of oil and gas accumulations. Methane is known to escape from rocks in other areas as well and is a minor constituent of some volcanic gases. It is a much-debated question whether the amount of methane in natural gas can be accounted for by buried organic matter or whether some may have a source in the earth's deep interior. Both CH_4 and CO have only an ephemeral existence in air because they react with oxygen (through free-radical intermediaries) to form CO_2. Ordinarily, neither gas attains sufficient concentration to be of much environmental significance, although amounts of CO from automobile exhaust can be hazardous on crowded streets.

The two sulfur gases H_2S (hydrogen sulfide) and SO_2 (sulfur dioxide) are also products of both natural and industrial processes and are short-lived in air because of reaction with O_2. The product of oxidation is SO_3, which unites with any available water to form sulfuric acid, H_2SO_4. This may appear as a constituent of raindrops, or it may be deposited on the ground's surface where it reacts with soil or rock to form sulfates. Where SO_2 is produced in abundance, as it is in many industrial operations, the resulting H_2SO_4 in the atmosphere can have noticeable effects on nearby vegetations, metal surfaces, and human lung tissue. Smelters for sulfide ores are especially efficient generators of SO_2 and are commonly surrounded by wide areas barren of plants if emission of the gas is uncontrolled. On a larger scale, the windblown sulfur gases from industrialized areas are a major source of acid rain that has become a well-publicized hazard to fish life in the lakes and streams of eastern North America and northern Europe.

Ozone is an unstable and reactive form of oxygen produced in air by the action of ultraviolet rays in sunlight on O_2 molecules [Eqs. (19) and (23)]. The reaction is aided by catalysts, and materials that can act catalytically are abundant in the automobile exhaust of large urban areas. The smog that sometimes covers such areas is partly a product of the ozone that is generated. Another source of O_3 is electric discharges; ozone's peculiar odor is often noticeable when lightning strikes nearby. Ozone molecules do not last long, reverting in days or weeks to ordinary O_2, but during their brief existence they can be destructive because they oxidize metals and organic matter more rapidly than diatomic molecules do. The embrittlement and cracking of rubber is a familiar effect of oxidation by ozone.

While ozone has unpleasant effects in the surface environment, it plays an important role high in the stratosphere. The lifetime of an O_3 molecule is longer (a few months on the average) in the attenuated gas of the stratosphere because collisions with other molecules are infrequent; thus the stratosphere has a permanent layer with relatively high concentrations (up to 10^{-4}%) of this active gas. The O_3 absorbs short-wavelength ultraviolet light from the sun, screening the earth from this potent radiation. Without the protection provided by the ozone layer, skin cancer would be more prevalent and some forms of plant and animal life would be damaged. Thus ozone in the troposphere is a nuisance, but the ozone in the stratosphere provides an important benefit to the environment.

Free hydrogen is a minor constituent of air. Small amounts are produced by bacterial decay in swamps, but it cannot accumulate because it reacts slowly with oxygen to form water. Any H_2 molecules that survive this reaction would rise to the top of the atmosphere and then, like helium, would escape from the earth's gravitational field. All other atmospheric gases are held firmly to the earth, but the molecules of these two are so small and fast-moving that the earth can hold them only temporarily.

The gaseous nitrogen compounds include ammonia (NH_3), nitrous oxide (N_2O), and the higher nitrogen oxides, chiefly NO and NO_2, but often symbolized NO_x because proportions of the oxides are variable. All these compounds are produced in nature by various kinds of bacteria; the higher oxides are also produced by electric discharges. NO_x has an additional source in some industrial processes and the exhaust from gasoline motors. Reaction of the oxides with O_2 (a reaction that also takes place by means of free radicals produced by photochemical dissociation) generates nitric acid:

$$\begin{aligned} 2\,NO + 3/2\,O_2 + H_2O &\rightarrow 2\,HNO_3 \\ 2\,NO_2 + 1/2\,O_2 + H_2O &\rightarrow 2\,HNO_3. \end{aligned} \tag{24}$$

The reaction is slow enough that the oxides persist for long distances downwind from their source, and the resulting acid (like the acid from sulfur gas) can thus become part of the rainfall over a wide area. In nature this could be beneficial, since fixed nitrogen is thereby added to soil as a plant nutrient. But where the nitrogen oxides are produced in large quantity, as they are in some urban and industrial areas, the acid can become a serious contaminant of surface waters. It is interesting that if NO and HNO_3 were at equilibrium with N_2 and O_2, their concentrations should be very large and HNO_3 or its ion NO_3^- would be a major constituent of seawater. Because this is not the case, which means that equilibrium does not exist, the reaction of Eq. (24) must be very slow and the take-up of nitrate by plants very efficient.

E. Change in Atmospheric Composition over Time

Over geologic time the nature of the earth's gaseous envelope has unquestionably changed. The atmosphere of the primitive earth must have been similar to the atmospheres known to be present on some other objects in the universe: at first there was most likely an abundance of hydrogen and helium, then as these gases escaped, a mixture of compounds that are stable under reducing conditions, chiefly CH_4, NH_3, H_2O, and H_2S. Later, or initially if conditions were not so reducing, the important gases would have been CO_2, SO_2, H_2O, and N_2. Along with the development of green plants came profound changes toward the oxygen-rich atmosphere we have today. The timing and details of the changes are subjects for lively argument, but the long paleontological record of living forms similar to those of the present suggests that the major changes were completed by early Paleozoic time. For many millions of years the atmosphere has had close to its present composition, maintained by the balance between processes by which the principal gases are added to air and processes by which they are removed.

This balance is not always perfect. One example is argon: this inert gas has increased in amount very slowly through geologic time and is still increasing because its addition to air by the decay of radioactive potassium is not compensated by any process of removal. More environmentally interesting are the possible fluctuations in carbon dioxide. If all the carbon now locked up in carbonate rocks and such organic materials as coal and oil came originally from CO_2 in the atmosphere (as seems probable), the amount that would have been required is much greater than the amount now available in air. Either the earth at some time in its history had a very dense atmosphere of CO_2 that has gradually shrunk to its present small concentration, or more likely, carbon has been gradually added to the atmosphere from the earth's interior in the form of either CO_2 or CH_4. If this is taken as a reasonable hypothesis, the addition of carbon need not have been at a uniform rate, but may have varied from one geological period to another. Such variation in the concentration of atmospheric CO_2 is also suggested by other kinds of geological evidence. If CO_2 has changed, concentrations of other gases may have changed as well, especially that of oxygen, which is closely linked to CO_2 through the reaction of photosynthesis [Eq. (17)]. Such changes in the latter part of the earth's history would probably not be large, and for other minor constituents the current processes of addition and removal would probably have kept their concentrations close to present-day values. For the atmosphere, as for the solid materials of the crust, the overall composition seems held within narrow limits by a steady state of complexly interrelated reactions. This generalization may not hold, however, for gases that are produced or removed in large amounts by human activity. The possible long-term effects of human interference with atmospheric processes are the subject of Section V.

IV. WATER IN THE ENVIRONMENT

A. Analyses

Waters at and near the earth's surface have a complex and variable composition that is difficult to express in a few analyses. Some representative values are given in Tables IV (major elements) and V (minor elements), but particular samples can deviate widely from the suggested values. Table IV shows the expected low concentration of solutes in rainwater, the high concentrations of some elements in seawater, and the generally intermediate values for river water and groundwater.

The dissolved material in rainwater comes largely from dust particles and grains of sea salt with which raindrops have come in contact, plus a small amount of CO_2 from the surrounding air that normally makes rainwater slightly

TABLE IV **Average Composition (ppm) of Rain, River Water, Groundwater, and Seawater**[a]

		River water				
Solute	Rain	Average	Range	Granite groundwater	Basalt groundwater	Seawater
Na^+	2.0	6.3	2–200	3–20	10–150	10,770
K^+	0.3	3.3	1–12	0.5–4	3–15	398
Mg^{2+}	0.3	4.1	1–40	0.3–7	1–20	1,290
Ca^{2+}	0.6	15.0	3–120	1–40	2–50	412
Cl^-	3.8	7.8	3–200	1–15	4–55	19,500
SO_4^{2-}	2.0	11.2	2–300	1–30	4–40	900
HCO_3^-	0.1	58.4	10–300	7–130	70–290	28
SiO_2	—	13.1	2–30	6–45	30–80	2
F^-	—	—	0.05–2.7	0–1	0.2–8	1.3
HPO_4^{2-}	—	—	0.001–0.3	—	—	0.06
pH	5–7	6–8		7–9	8–10	8.0–8.3

[a] Sources: For rain, Garrels, R. M., and Mackenzie, F. T. (1971). "Evolution of Sedimentary Rocks," Norton, New York; for river water average, Livingston, D. A. (1963). "Chemical composition of rivers and lakes." *U.S.G.S. Prof. Paper* 440-G; for ranges of river water and groundwater, Bowen, H. J. M. (1979). "Environmental Geochemistry of the Elements," Academic Press, London; for seawater, Broecker, W. S., and Peng, T.-H. (1982). "Tracers in the Sea," Eldigio Press, Lamont-Doherty Geol. Observ., Columbia Univ., New York.

TABLE V Average Concentrations (ppb or μg/L) of Some Minor Elements in Freshwater and Seawater[a]

	Freshwater		
Element	Average	Range	Seawater
Al	50	8–3500	0.8[b]
As	1.7	0.2–230	1.7
B	18	7–500	4,500
Ba	60	3–150	14
Br	20	0.05–55	67,000
Cd	0.1	0.01–3	0.08
Co	0.2	0.04–8	0.002[b]
Cr	1	0.1–6	0.2
Cs	0.03	0.005–1	0.3
Cu	10	0.2–30	0.3
Fe	40	10–1400	0.06[b]
I	6	0.5–7	56
Li	12	0.07–40	170
Mn	8	0.02–130	0.3
Mo	0.5	0.03–10	10
N (as NO_3)	—	—	420
Ni	2	0.02–27	0.5
Pb	1	0.06–120	0.002
Rb	1.5	0.6–9	120
Se	0.2	0.02–1	0.1
Sr	60	3–1000	7,600
U	0.2	0.002–5	3
Y	1	0.01–20	1
Zn	30	0.2–100	4

[a] Source for average values in freshwater and seawater: Broecker, W. S., and Peng, T.-H. (1982). "Tracers in the Sea," Eldigio Press, Lamont-Doherty Geol. Observ., Columbia Univ., New York. Source for range in freshwater: Bowen, H. J. M. (1979). "Environmental Geochemistry of the Elements," Academic Press, London.

[b] Values highly uncertain.

acid. The amount of sea salt carried inland from ocean spray decreases with increasing distance from the coast, thus rain samples can have concentrations of some elements, particularly Na and Cl, that are different from those in the average analysis in Table IV. On reaching the ground, rainwater picks up additional dissolved material from soil and rocks, and its pH generally increases because of hydrolsis of silicate and carbonate rocks [Eqs. (2)–(4)]. Water that seeps into the ground has more intimate contact with rock, and its solute content is further increased. The sea is the ultimate repository for all water from the land, and it becomes a concentrated solution of some constituents because of long evaporation. The concentrating effects of evaporation, however, are offset by reactions of the dissolved materials with organisms, sediments on the seafloor, and rock beneath the sediments. Only in restricted basins does evaporation lead to extreme concentrations and the precipitation of all or nearly all of the salts.

The principal dissolved ions in most natural waters are Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^-, SO_4^{2-}, and HCO_3^- (Tables IV and V). In addition, most waters contain at least a few parts per million of silica, chiefly in the form of the neutral molecule H_4SiO_4 (commonly reported in analyses as SiO_2). Only rarely is water encountered in which some other element (e.g., F, P, or Sr) has concentrations comparable to those of the eight major constituents. The reason for the predominance of these eight substances is simply that they are fairly abundant in ordinary rock material (Table I) and hence accessible to water and that they do not form insoluble compounds with each other or with common substances in rocks and soil.

In addition to containing the dissolved substances, most natural waters contain at least a little suspended material. This can be in the form of clay or silt, which slowly settles out if the water is undisturbed, or finer-grained colloidal particles that are invisible to the unaided eye and can remain suspended indefinitely. Common colloidal materials are ferric oxide and aluminum oxide; the Fe and Al listed in Table V are usually in this form. Colloidal silica is often present in addition to dissolved H_4SiO_4. If water contains organic matter, it is commonly present in colloidal form, although some may exist in true solution. The distinction between colloidal suspension and true solution is not always easy to make because particle sizes grade imperceptibly from one to the other.

B. Streams and Lakes

Running water dissolves substances from both the atmosphere and the rocks and soil over which it flows. Some of the CO_2 that dissolves combines with water to form carbonic acid [Eq. (5)]. This is a weak acid and present in only minor amounts ($\sim 10^{-5}\ M$ in equilibrium with air at ordinary temperatures), but it is nevertheless an important agent of weathering [see Eqs. (6)–(8)]. It dissociates in two steps:

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^-, \qquad K_1 = 10^{-6.4}, \qquad (25)$$

and

$$HCO_3^- \rightleftharpoons H^+ + CO_3^{2-}, \qquad K_2 = 10^{-10.3}, \qquad (26)$$

where K is the equilibrium constant (the product of concentrations of products divided by concentrations of reactants). The resulting H^+ [chiefly from Eq. (25)] gives a pH of ~5.7 in water with no other dissolved material. Tending to counteract this slight acidity is the hydrolysis of carbonate or silicate minerals in the streambed, so that measured pH values in stream water are commonly between 6 and 8. River water in warm humid regions where weathering is deep may have pH values of <6, and water in arid climates may develop pH values of >8.

The dissociation reactions [Eqs. (25) and (26)] indicate that river water contains three carbonate species (H_2CO_3, HCO_3^-, and CO_3^{2-}) plus CO_2 that has not reacted with water. The carbonate species predominating in a given water depends on pH: at pH <6.4 the major species is H_2CO_3, between 6.4 and 10.3 it is HCO_3^-, and >10.3 it is CO_3^{2-}. Because pH values in stream water are commonly in the range from 6 to 9, dissolved carbonate is symbolized by HCO_3^-, but the other two species are always present in small amounts. Similarly, dissolved silica exists in natural waters as both molecules of H_4SiO_4 and ions formed by dissociation of this acid. Since H_4SiO_4 is a weaker acid than H_2CO_3 $K_1 = 10^{-9.9}$ the concentration of the ions is very small: $H_3SiO_4^-$ becomes appreciable only at pH values >9, which are uncommon in nature, and $H_2SiO_4^{2-}$ is completely negligible. Thus the appropriate symbol for dissolved silica is H_4SiO_4.

Besides CO_2, the other active gas that streams dissolve from air is O_2, which gives stream water the capability of oxidizing minerals containing reduced iron or sulfur. The oxidizing ability can be expressed quantitatively in terms of redox potential relative to the H_2–H^+ couple, which is arbitrarily assigned a potential of zero:

$$\tfrac{1}{2}\,H_2 \rightleftharpoons H^+ + e^-, \qquad E^\circ = 0.00\,V. \qquad (27)$$

In reference to this couple, oxygen has a potential expressed as

$$H_2O \rightleftharpoons \tfrac{1}{2}O_2 + 2\,H^+ + 2e^-, \qquad E^\circ = +1.23\ V. \quad (28)$$

The voltages refer to equilibrium conditions (E°) when all gases are at 1 atm pressure and dissolved substances have an activity of 1, but they are also useful in a qualitative sense to compare the oxidizing characteristics of different environments. Free oxygen dissolved in stream water, for example, is a powerful oxidizing agent capable of oxidizing ions or compounds of Fe(II) to the ferric state,

$$Fe^{2+} \rightleftharpoons Fe^{3+} + e^-, \qquad E^\circ = +0.77\ V, \quad (29)$$

and

$$2\,Fe_3O_4 + H_2O \rightleftharpoons 3\,Fe_2O^3 + 2\,H^3 + 2e^-, \qquad E^\circ = +0.22\ V, \quad (30)$$

or of oxidizing the sulfur of sulfide minerals to sulfates

$$ZnS + 4\,H_2O \rightleftharpoons Zn^{2+} + SO_4^{2-} + 8\,H^+ + 8e^-, \qquad E^\circ = +0.34\ V. \quad (31)$$

These potentials are lower than that of the H_2O–O_2 couple in Eq. (28), which means that dissolved O_2 is capable of oxidizing Fe^{2+}, Fe_3O_4, and ZnS. Results of such reactions are evident in streams that drain areas where sulfide ores are being mined: the water is brown from suspended ferric oxide, is markedly acidic, and may contain high concentrations of metals from the ore.

Dissolved O_2 is also effective in oxidizing the dissolved or suspended organic matter that a stream may carry. This reaction can be effective, within limits, in clarifying stream water and in removing toxic organic chemicals. If organic matter is abundant (e.g., where a stream flows sluggishly through a swamp), however, its dissolved oxygen may be entirely consumed and the water may become reducing rather than oxidizing.

Water from near the surface of freshwater lakes is similar to that of streams, as would be expected since much of it is supplied by streams and since it is also continuously exposed to CO_2 and O_2 in the atmosphere. Below the surface where the water is out of contact with air, the dissolved O_2 is depleted by reaction with organic matter that drops from the surface. How completely the O_2 is removed depends on how fast the lake water circulates. In a shallow lake, wave action and currents produced by inflowing streams may be sufficient to keep the entire lake aerated. Deeper lakes may have bottom waters with little or no circulation where dissolved O_2 is completely exhausted and where partly decayed organic matter accumulates. In many lakes of temperate regions the development of stagnant bottoms is prevented by annual or semiannual overturning of the water due to temperature changes, which brings bottom water to the surface where its organic content may be oxidized. Where nutrient material favorable to plant growth is supplied to a lake in large amounts (e.g., by intensive use of fertilizer on adjacent fields), aquatic plants may grow so luxuriantly that the entire lake may become choked with organic debris, even if in its natural state the circulation would be sufficient to keep it oxidized. This process of eutrophication, common in areas of intensive agriculture and some kinds of chemical industry, can be a menace to the use of lakes for water supplies and recreation.

In lakes of arid regions, dissolved materials become concentrated because of long-continued evaporation, often to an extent that plant life is inhibited and the water remains beautifully clear. Extreme evaporation leads to precipitation of some of the dissolved ions as salts, and the nature of the salts depends on the extent of evaporation and the composition of the lake waters. Common precipitates are calcite ($CaCO_3$), gypsum ($CaSO_4 \cdot 2\,H_2O$), and halite (NaCl). Because most arid-zone lakes are in isolated basins, their compositions can vary widely according to the nature of the rock in different basins, and salts different from these familiar ones are often found: Na_2CO_3, Na_2SO_4, $MgSO_4$, and occasionally salts containing less common elements such as B, Sr, Li, and Br. Waters that contain abundant HCO_3^- may become very alkaline during evaporation because of escape of CO_2:

$$2\,HCO_3^- \rightleftharpoons CO_3^{2-} + H_2O + CO_2. \qquad (32)$$

The liberated carbonate ion is partly precipitated as $CaCO_3$ or Na_2CO_3, and the part remaining in solution hydrolyzes to generate OH^-:

$$CO_3^{2-} + H_2O \rightleftharpoons HCO_3^- + OH^-. \qquad (33)$$

Thus the chemical behavior of streams and lakes can be understood largely in terms of reactions among the cations Na^+, K^+, Ca^{2+}, and Mg^{2+} and the anions Cl^-, SO_4^{2-}, and HCO_3^-, along with organic matter and the gases CO_2 and O_2 dissolved from the surrounding air.

C. Groundwater and the Vadose Zone

Much rainwater does not run off the ground surface immediately to join streams, but seeps into the soil and underlying rock to become part of the great reservior of underground water that supplies water to wells and feeds the springs and seepages that maintain the flow of rivers through periods of dry weather. Water in the subsurface is divided into two zones: the vadose zone and the saturated zone. The water table divides the vadose zone from the saturated zone. In the vadose zone, the pores are only partially filled with water, whereas in the saturated (or groundwater) zone, the pores are completely filled with water. Groundwater normally does not form definite streams underground (although it may do so locally in caves) but moves very slowly through the interstices of soil and rock. Measured subsurface flow rates range from a few centimeters to tens of meters per year, so that the water remains in intimate contact with soil and rock surfaces for long periods. Subsurface water is subject to the same sort of chemical controls as water at the surface, except that it soon loses much of its initial dissolved O_2 and CO_2 and can more nearly approach equilibrium with surrounding soil and rock because of its prolonged contact.

Rainwater that moves underground gives up its dissolved CO_2 through reactions with silicate and carbonate minerals [see Eqs. (6)–(8)], and its pH is further raised by hydrolysis. Its dissolved O_2 reacts with organic matter in soil and with minerals containing Fe(II) or reduced sulfur in rock. How fast and how completely these reactions occur depends on many variables, particularly the rate of movement and the kind of rock. Thus water moving through pores of clean quartz sandstone may remain dilute, slightly acidic, and oxygen-rich for a long time, while water encountering an organic-rich soil overlying an iron-rich rock like basalt would quickly become alkaline and reducing and would acquire a substantial content of dissolved material. As a general rule, most groundwater is alkaline rather than acid and is depleted in O_2; its alkalinity and reducing capacity generally increase with depth, but there is much variability. The dissolved cations and anions reflect to some extent the composition of the surrounding rock, although differences in the waters from different rocks are not as great as might be expected (cf. columns 4 and 5 in Table IV).

As subsurface water moves from one rock to another, one depth to another, or one temperature regime to another, its composition slowly adjusts itself to the new environment. The adjustment is accomplished by dissolution of material from the adjacent rock or by precipitation of dissolved constituents in the cracks and interstices through which the water moves. An extreme example of dissolution is the formation of limestone caves, which form by water containing H_2CO_3 (and rarely other acids) moving through cracks and dissolving $CaCO_3$ [Eq. (5)]. Examples of precipitation include the cementing of grains of sand and silt to form solid rock and the deposition of material in cracks to form veins. The two substances that are dissolved and precipitated in greatest abundance are SiO_2 and $CaCO_3$, which are the constituents of quartz veins and calcite veins, respectively, and the commonest cementing materials of sedimentary rocks.

One kind of chemical reaction that is common in subsurface water but less so in surface waters is the process of adsorption (mentioned in Section II.E), which is the ability of some mineral surfaces to temporarily hold ions taken out of solution in water moving past them. Thus subsurface water containing Mg^{2+} that is moving through clay can lose most of its Mg^{2+} because the ions are adsorbed onto surfaces of the clay minerals. The ions are not strongly held, because if a later flow of subsurface water deficient in Mg^{2+} moves through the same clay, some of the adsorbed ions will return to solution. Adsorbed ions are held to mineral surfaces in a variety of ways. The attraction may be electrical, because charges on the ions are partly neutralized by residual charges on the surface, or ions may be concentrated on corners or edges of mineral crystals where charges are greater than on flat crystal faces. Ions from solution may replace particular ions in the mineral structure, or the attachment may be the result of a chemical reaction. Some kinds of adsorption are stronger than others, but generally the attachment is less firm than in a normal chemical bond; adsorbed ions ordinarily can be removed by a simple change in conditions, such as a change in subsurface water composition. Adsorption is nevertheless an important part of subsurface water chemistry, often greatly delaying the transport of particular ions as subsurface water moves through sorbent material. It is less important in the chemistry of waters in streams and lakes because these waters have less exposure to mineral surfaces.

Both dissolved and colloidal organic matter can be picked up as subsurface water moves through soil and can persist in water moving at shallow levels through underlying soil and rock. Organic compounds are subject to slow oxidation (often aided by bacteria) and to adsorption on mineral surfaces, so that at deeper levels organic matter is generally absent or nearly so. The natural organic matter of shallow subsurface water can be augmented by compounds leaking from underground tanks or from areas where industrial organic wastes have been disposed near the ground surface. This kind of subsurface water contamination has become a serious problem, especially since some of the hazardous compounds produced by the modern chemical industry are not easily oxidized or attacked by bacteria and can persist in the environment almost indefinitely. The problem is particularly difficult in comparison with contamination of stream water, because the movement of contamination underground is hard to trace and very expensive to control even where tracing is possible.

Another sort of subsurface water contamination, sometimes accompanying organic contamination, is that produced by toxic metals. Cadmium, mercury, lead, and arsenic have been particularly troublesome. In nature such metals can find their way into subsurface water in the vicinity of ore deposits, especially deposits that have been exposed or nearly exposed at the earth's surface by erosion. Oxidation of the ore and release of its metals to solution is often aided and spread over a wide area by mining operations. Other sources of metal contamination include smelting operations and manufacturing procedures in which a metal or its compounds are used in large amounts. The best method of controlling contamination by both metals and organic compounds is to make sure that subsurface water is protected from contamination initially; however, arranging for such protection is often expensive and time-consuming.

Still another kind of subsurface water contamination is that produced by salt, either by the incursion of salt water from the sea when freshwater is pumped too intensively from wells in coastal areas or by the dissolution of salt in water that is in contact underground with deposits of rock salt. Salt can also become troublesome in the subsurface water of arid regions, where repeated use of irrigation water leads to slow concentration by evaporation during each use.

Increasingly since the mid-1970s, with the advent of high-speed digital computers, process-based mathematical models of coupled subsurface fluid flow, solute transport, and geochemistry have been used to assess subsurface water contamination impacts.

D. Seawater

The oceans serve as gigantic basins for the dissolved and suspended material carried by rivers and for solid debris erupted from volcanoes and scoured from the land surface by winds and glaciers. As might be expected, the major ions in seawater are the same as those in rivers: Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, and HCO_3^-. Evaporation over geologic time has concentrated these ions but not to the extent that might be expected. This suggests that the ions are steadily being removed from seawater as well as being added. Some of the removal processes are obvious, for example, the use of some elements by the abundant organisms that live in the sea and the carrying of some elements to the bottom as precipitates or adsorbed on solid particles. Less obvious removal processes include adsorption on bottom sediments and reaction with underlying basalt as heated seawater circulates through cracks. Organic processes are particularly effective in controlling Ca^{2+} and SiO_2, the former by precipitation of $CaCO_3$ in the shells of marine organisms whose remains cover large areas of the seafloor, and the latter by precipitation of SiO_2 in the form of opal by other species of small plants and animals. Not all the controlling processes are definitely known, but analyses of samples from many areas and many depths show that chemical controls and distribution by waves and currents are effective in keeping the concentrations of major ions within narrow limits (see Table IV). The pH of seawater is also kept within a small range (8.0–8.3 near the surface, slightly lower at depth), and circulation is rapid enough to bring dissolved CO_2 and O_2 from the surface to great depths in nearly all parts of the ocean.

The concentrations of minor elements are less uniform, many of them depending on organic activity. Silica, for example, is held to very low values (1–3 ppm) in large parts of the ocean by the activity of radiolaria and diatoms, but in places where these organisms do not flourish its concentration may rise to values like those in freshwater (10–20 ppm). Phosphorus is commonly rich in bottom water, where it is set free as phosphate during decay of organic remains, but its concentration is low near the surface, where it is used by organisms whose growth depends on sunlight. The concentrations of some minor metals vary considerably with depth, and the values given in Table V are only rough averages.

In two ways the sea plays an important role in the overall system of chemical reactions that maintain a steady state in land environments. One is the addition of salt particles to air as windblown spray evaporates; such salt can be carried far inland and is a major source of the Na^+, K^+, Mg^{2+}, Cl^-, and SO_4^{2-} in stream water. The second, and probably more important, influence of the sea on land chemistry is its contribution to regulating the CO_2 concentration in air. The controls on this concentration are complex. As noted in Section III.B, part of the control is a balance between the consumption of CO_2 in photosynthesis [Eq. (17)] and the release of the gas by respiration, combustion, and decay. The life processes of marine plants and animals are included in this balance.

But beyond the strictly organic production and consumption of CO_2, the sea exerts control in other ways. The dissolving of CO_2 at the sea surface establishes equilibria among CO_2, H_2CO_3, HCO_3^-, and CO_3^{2-}, as represented by Eqs. (5), (25), and (26). Any increase of CO_2 in the atmosphere would drive all these reactions to the right, which would thus increase the concentrations of the ions and make the water more acid, while a decrease would cause the reactions to reverse and generate CO_2 to make up for the deficiency. In addition, CO_2 takes part in the equilibrium represented by Eqs. (6) and (18):

$$CO_2 + H_2O + CaCO_3 \rightleftharpoons Ca^{2+} + 2\,HCO_3^- .$$

As shown by this equation, an excess of CO_2 would lead to the dissolving of $CaCO_3$ in sediments and the formation of HCO_3^-, and a deficiency would be compensated by the reverse reaction of this ion with Ca^{2+} to produce more $CaCO_3$. Several other equations can be written representing more complicated equilibria involving CO_2 and the constituents of seawater and its sediments; all of them would be capable of generating CO_2 if its concentration in air would drop or be capable of using it if the concentration would rise. Thus the great reservior of the ocean is a major source and sink for atmospheric CO_2 and a major control on its global concentration.

E. Long-Term Changes in Natural Waters

Much of the water at the earth's surface probably came initially from the interior, and additions out of the interior probably have continued throughout geological history. How large the later additions have been is a matter of argument, but present opinion is that the great bulk of our present water supply was created early in the earth's history, shortly after the earth's surface became cool enough for oceans to accumulate. Water is coming to the surface today in volcanic eruptions, and presumably has done so in all past ages, but most of this is probably recycled water from sedimentary rocks that have been carried down into the mantle by subducting plates and heated to release their contained water. If volcanic water contains any additions from the deep interior, they are probably minor. The best current guess is that the total amount of surface and near-surface water has been virtually constant for millions of years and will remain constant in the future.

The composition of streams and lakes is necessarily responsive to changes in climate, and such changes have been recorded in human history as the climate of particular regions has shifted from humid to arid or vice versa. The changes are slow, hardly perceptible in a single lifetime. If the CO_2 content of the atmosphere continues to increase as it has during the last century, a slight, long-term increase in the overall acidity of surface waters would be an expectable result.

That seawater has not changed its composition markedly during the past 0.5 billion yr is demonstrated by the fact that marine life similar to that in the present ocean has existed for at least that length of time. Closer bounds on the variation in composition have been established by study of evaporite minerals deposited at various times in the geological past. A future increase in atmospheric CO_2 would lead to progressive change in the composition of seawater as well as freshwater, but the increase would be resisted by the reactions discussed in Section IV.D and for a long time the compositional change would be slight.

V. EFFECTS OF HUMAN ACTIVITY

Chemical reactions in soil, air, and water at the earth's surface are part of a huge system that long ago reached a steady state. The system is remarkably stable toward natural events that might disturb it. An excess of any component sets in motion reactions that tend to cut down the excess, and a deficiency automatically leads to reactions that seek to restore the balance. In the operation of natural processes there are few reactions that might result in permanent long-term changes. The decay of radioactive isotopes and the loss of hydrogen and helium from the upper atmosphere are two examples of processes that do produce uncompensated change, but effects on the environment are insignificant. The oxygen content of air is possibly increasing because photosynthesis produces more than is removed by oxidation reactions. Also, carbon dioxide is being removed through the formation of carbonate rocks faster than it is being generated by the heating of rocks carried down in subduction zones. Effects of such changes, however, would be important only over long periods of geologic time. Human beings have modified the chemical environment in obvious ways by clearing forests, cultivating the soil, and extracting metals from ores, but for most of human history the effects on the whole steady-state system have been slight and often temporary.

With the huge growth in population over the last century, however, and with the development of ever larger and more efficient machines, the ability of humans to influence their environment has increased enormously. It may be that some kinds of human activity are producing changes that in the short term seem trivial but whose cumulative effect over a few generations may modify the environment to an extent not readily compensated for by the reactions that normally maintain the steady-state system. Some of these possible effects have been noted in preceding paragraphs, and here they are examined in more detail.

A. Increase in Carbon Dioxide: The Greenhouse Effect

The extent of the earth's heating by the sun depends on certain characteristics of the atmosphere. Radiation comes to us from the sun in many different wavelengths, which are partly reflected back into space by materials of the surface and partly converted to the long wavelengths of heat radiation (infrared light) that is responsible for warming the lower atmosphere. If the atmosphere did not exist, the infrared radiation would escape into space; the earth's surface would be intensely hot during the day and bitterly cold at night, as on the airless surface of the moon. But the molecules of some gases in the air absorb the radiation and are themselves heated; by collisions with other molecules, the heat is distributed through the lower layers of air. The moving molecules retain some of the heat overnight and thus keep temperatures from plummeting far below zero as soon as the sun sets. Just as the glass of a greenhouse admits many wavelengths of light but prevents the longer wavelengths of heat radiation from escaping, so does the atmosphere serve to trap the longer wavelengths and warm the surface environments.

Of the molecules that absorb infrared rays, CO_2 is the most efficient. Others include some of the variable constituents, particularly H_2O and CH_4; the effectiveness of H_2O is evident in the strong contrast between day and night temperatures in the dry air of deserts as compared with the more even temperatures of humid regions. If the CO_2 content of air should increase, suface temperatures would be raised, and a long-continued increase would necessarily affect world climates. Just such an increase is the probable result of many human activities: the burning of wood, coal, and oil; the smelting of metals; and the clearing of forests. Such activities have greatly accelerated over the past century, and during this period the CO_2 content of air has increased $>10\%$. Further increase seems ensured, and the question arises as to how much of an effect on climate can be expected.

To calculate the effect seems simple: the rise in temperature due to a given concentration of CO_2 is known, the excess amount of CO_2 produced annually can be estimated, and multiplication of these two figures should give the amount by which the earth's average temperature increases each year. But the calculation turns out to be much less straightforward. The difficulty stems from the fact that in the earth's steady-state system a rise in CO_2 concentration immediately sets in motion the natural reactions that tend to counteract the increase (see Section IV.D). Some of the excess CO_2 would be used for additional photosynthesis of organic compounds in the green leaves of plants; more of the excess would dissolve in seawater, and some of this would be consumed in the dissolution of $CaCO_3$ from sediments. An added complication is that rising temperature would lead to greater evaporation of water and thus to an increase in cloud cover that would shield the surface from the sun's rays and would moderate the temperature increase. Most of these counterreactions are slow, and it is uncertain whether they will keep pace with the global increase in CO_2 concentration. The increase in temperature will be less than it would be without the compensating reactions, but how much less remains a disputed point. A majority of researchers who have attempted the calculation suggest a worldwide temperature rise of 1 or 2°C by the year 2050. This is a small change, to be sure, but it is sufficient to alter climates appreciably. One consequence will be the melting of much of the ice in Greenland and Antarctica and a consequent rise in worldwide sea levels. The change in climate will be so gradual as to be hardly perceptible in a single lifetime, but if it continues for several generations, the positions of coastlines and the distribution of deserts and forested lands will become markedly different.

B. Acid Rain

In recent years the number of fish in lakes and streams of the northeastern United States and the eastern provinces of Canada has reportedly shown a marked decrease. The decrease has been correlated with an increase in the acidity of the water, which in turn is thought to be due to a change in the character of rain in this area caused by acid emissions from industries that are concentrated south of the Great Lakes and in the eastern United States. Prevailing winds from the west and southwest carry the acidified water droplets into the affected area. This process lowers the pH of surface water during rains and ultimately lowers the pH of the groundwater that maintains stream flow between storms. Not only the decline in fish life but increasing damage to vegetation has been attributed to the acid rain. Similar effects have been reported from Scandinavia, the acid rain there coming from the heavily industrialized parts of Germany, England, and the Low Countries. If the correlation of damage to aquatic life and forests with industrial emissions is correct, this is another example of large-scale disturbance of environmental geochemistry by human activities.

Although the correlation seems to be well established, argument remains as to the culpability of industry relative to acid emissions from automobile exhausts and possible changes in natural processes. The uncertainty stems from the difficulty of obtaining a meaningful average pH value over a large area and from the inadequacy of data on pH values of surface water and groundwater before the expansion of industry in modern times. Because efforts to control the acidity of rain will require the expenditure of billions of dollars, the determination of the actual magnitude of damage and the assignment of responsibility for it are subjects of current intensive research and political controversy.

The acid emissions, whatever their source, consist chiefly of oxides of sulfur and nitrogen. The principal sulfur gas is SO_2, which in the presence of air and water oxidizes slowly to form sulfuric acid:

$$SO_2 + \tfrac{1}{2}\,O_2 + H_2O \rightarrow 2\,H^+ + SO_4^{2-}. \qquad (34)$$

Nitrogen forms a number of oxides (see Section III.C) and so that the difficulty of identifying individual ones can be avoided, they are often symbolized NO_x. Two major ones, NO and NO_2, react with water and oxygen to form nitric acid, as shown in Eq. (24). These oxidation reactions take place slowly, in a series of steps involving free radicals, and the equations give merely a simplified representation of the final result. The two strong acids produced, if present in any quantity, would certainly be capable of causing the kind of damage ascribed to acid rain.

Even in their natural state, surface waters in the humid parts of the North Temperate Zone are somewhat acidic. In part this results simply from dissolution of CO_2 out of the surrounding air [Eq. (4)]; pure water in contact with air acquires a pH of $\sim$5.7. Soils of these regions, however, commonly contain enough decaying organic matter to produce a soil atmosphere much richer in CO_2 than ordinary air is (often as great as 0.1 atm partial pressure), and the resulting H_2CO_3 plus small amounts of organic acids cause the pH of groundwater to drop to 4 or even lower. In places where the bedrock is limestone the pH would not be this low because $CaCO_3$ is effective in neutralizing acid [Eq. (5)]. Thus the pH values in nature are controlled by a complex of reactions, and they vary a good deal from place to place and from season to season. Addition of acid from rainfall sets in motion the natural reactions that would try to neutralize it, especially the dissolution of carbonate and the replacement of adsorbed metal ions by H^+. Because of these buffering reactions, a substantial amount of acid can be added to a soil or a lake without greatly affecting its pH. Only when the quantity of excess acid becomes large enough to exceed the capacity of the natural buffers will the pH fall much below its usual value. That rainfall over a long period has indeed supplied quantities of acid sufficient to change the pH markedly is indicated by measurements reported from Canadian lakes of pH values as low as 3.

How serious the problem is and how effectively it could be remedied by controlling industrial emissions remain matters of dispute. Certainly some reduction in the activity of rain could be accomplished by installing devices to trap or lower the rate of emission of SO_2 and NO_x, but the amount of reduction is hard to estimate. It might prove necessary to control emissions from automobile exhausts and industrial smokestacks; some kinds of industry, or industrial plants in particular locations, might require more stringent controls than others. The problem is enormously complicated both technically and politically, and a final solution awaits additional research.

C. Ozone in the Stratosphere

As mentioned in Section III.D, ozone is formed in the atmosphere by photochemical reactions and is particularly abundant in one layer of the stratosphere. The ozone layer absorbs much of the low-wavelength, high-energy ultraviolet radiation from the sun, shielding the earth's surface from this damaging radiation. Just how harmful the radiation would be if the ozone layer did not exist is a matter of dispute, but it is known to be damaging to some organisms and to cause skin cancer in humans. Thus preservation of the ozone layer is an important concern, and there was much consternation a few years ago when it appeared that some volatile products of human industry might be effectively destroying it. In a state of nature the unstable O_3 molecules that make up the layer are being continuously created and destroyed, and the fear was that man-made substances might distrub this natural balance by accelerating the reactions leading to destruction.

The compounds in question are halogenated hydrocarbons, for example, $CFCl_3$ and CF_2Cl_2. These have had widespread use in ordinary life as refrigerants and propellants for spray cans. The compounds are gases or volatile liquids at usual temperatures, so when released they quickly mix with air. In contrast to most organic compounds they are very stable, resisting oxidation by O_2 for periods of years or decades. Once in the atmosphere, some of the molecules eventually find their way to the stratosphere, where Cl atoms are liberated by photochemical reactions and react with O_3 to destroy it:

$$Cl + O_3 \rightarrow ClO + O_2 \qquad (35)$$

and

$$ClO + O \rightarrow Cl + O_2. \qquad (36)$$

Although any individual use of a device containing one of the halogenated hydrocarbons would release only trivial amounts, the total of all projected future releases by the world's population could be sufficient to seriously diminish the amount of ozone in the stratosphere.

Recognition of this danger has caused widespread substitution of other propellants for halogenated hydrocarbons in spray cans, so the possibility of damage to the ozone layer is greatly reduced. Also a revision of the original calculations suggests that the danger was at first somewhat exaggerated. Nevertheless, the episode serves as a warning that possible effects on the chemical balance of the ozone layer must be one consideration in the extensive use of the new organic chemicals that industry is providing in ever-increasing numbers.

D. Mine and Tailings Drainage

Ores of many metals (e.g., Cu, Pb, Zn, Ag, Hg, Mo, and Sb) consist largely of sulfide minerals, which commonly are associated with abundant pyrite (a sulfide of iron, FeS_2). On exposure to weathering at the earth's surface, the sulfur of such ores is oxidized to SO_4^{2-} and hydrolysis of the metal ions makes infiltrating water acid [Eqs. (12) and (13)]. Pyrite is particularly susceptible to weathering, and the resulting Fe(III) is strongly hydrolyzed because of the extreme insolubility of ferric oxide [as Fe_2O_3, FeOOH, or $Fe(OH)_3$; see Eq. (11)]. In a state of nature much of the ferric oxide remains in place where it forms, accumulating as a red, brown, or yellow capping over the deposit (the "iron hat" or "gossan" in the jargon of miners), while most of the SO_4^{2-} and the ions of other metals are carried off in solution. (In arid climates some remain behind as oxide or carbonate minerals; often conspicuous are the green and blue colors of the copper carbonate minerals malachite and azurite.) Ordinarily the weathering process is slow and the contamination of surface water and groundwater by metal ions and acid is not serious.

When a deposit is opened by a mining operation, however, sulfides are exposed to air and water in large quantity and weathering is greatly accelerated. Waters draining the deposit, both on the surface and underground, can become highly acid and heavily contaminated with toxic metals. Ferric oxide precipitates not only at the deposit itself, but in drainage water as its acidity is gradually neutralized. Exposed pieces of metal are attacked by the acid and their surfaces become coated with iron oxide. Abandoned mines are often conspicuous eyesores in the landscape because of the large area of discolored rock, the heavily rusted machinery, and the brown color of streams that drain them. They can also be a nuisance to nearby communities by contaminating water supplies and making streams and lakes unsuitable for fish life, even to distances of many miles downstream.

Some measure of control is possible, for example, by using limestone to neutralize the acid and precipitate some of the metals, but the best insurance against damage to scenery and water is care in the original mining operation to minimize the dispersal of weathering sulfides and to control the escape of drainage water.

Similar problems arise in the mining of coal, because most coals contain at least a little iron sulfide. Abandoned coal mines, like abandonted metal mines, are often made conspicuous by iron oxide staining and the acidity of water that drains them.

Hazards to adjacent areas can also be associated with debris from mining and milling operations that do not involve sulfide minerals. A well-publicized example is the huge accumulation of tailings from the milling of uranium ore in many parts of the western United States. The tailings consist of material from which most of the uranium has been extracted, but in which are left all the highly radioactive elements resulting from the long-term decay of uranium in the original ore. These elements, of which radium is the best-known example, posed no danger to the environment as long as the ore remained in place deep underground because little groundwater could reach them. In piles of loose debris at the surface, however, they are exposed to infiltrating rainwater and hence can make their way into water that is used for human consumption. The element radium is hazardous for another reason as well: its decay produces atoms of the gaseous radioactive element radon (Rn), which decays quickly to a radioactive isotope of lead. From the loose materials of a tailing pile, the Rn escapes into air and the radioactive lead it produces becomes part of the dust that is blown into nearby communities and can be inhaled. Quantities of both the dissolved elements and the escaping Rn are exceeding small, so small that the seriousness of the hazard is open to question, but it is generally agreed that some effort at control is needed.

Thus the mining of metals and coal, a human activity that is essential to industrial civilization, presents chemical hazards to the environment that are sure to increase in the future as the demand for resources grows larger.

E. Disposal of Chemical Waste

Still another impact of human enterprise on the chemistry of the environment results from efforts to dispose of the huge amounts of waste that modern civilization generates. In preindustrial times, when the world's population was smaller, waste disposal was a relatively minor problem: garbage accumulated locally and temporarily in unpleasant amounts, but it seldom contained hazardous substances and was quickly rendered harmless by the natural processes of oxidation and bacterial decay. The natural environment has a large capacity for disposing of waste automatically. One can put substantial quantities of waste chemicals and toxic metals on the surface or in streams or burial pits and be assured that within a reasonable time most of the material will have disappeared by oxidation or be so widely dispersed in surface and underground water that it is no longer a hazard. It is only when the quantity of waste becomes too large for the earth's natural system of compensating reactions to handle it expeditiously, or when substances hazardous to health make up a sizable fraction of the waste, that the problem of disposal demands attention.

The magnitude of the current problem in the United States is suggested by figures recently published by the Environmental Protection Agency: municipal and industrial solid waste is being generated at a rate of $\sim$400 million tons/yr, of which 60 million tons may have hazardous ingredients. Liquid industrial wastes amount to 10 trillion gal/yr. Much of the waste is harmless and can be managed by the conventional techniques of incineration, bacterial digestion, reverse osmosis, shallow land burial, or addition to surface waters. But the waste containing toxic materials, now being handled at $>$1000 sites throughout the United States, cannot be so easily dismissed.

The problem becomes especially serious when materials in the waste dissolve in groundwater, as often happens when rain falls on accumulations at the surface or in shallow burial pits and sinks into the ground. In surface waters contamination is generally apparent at once, and remedial measures can be taken. But contamination underground can continue for a long time undetected and is difficult or nearly impossible to correct at a reasonable cost.

The most troublesome contaminants of groundwater and surface water are organic compounds, produced either directly or as by-products in many industrial operations. More than 10,000 such substances are known to be toxic, and their number increases yearly. Many, of course, are valuable materials that we consider indispensable in modern life, but as uncontrolled constituents of waste they can do enormous harm. Some of them, probably the majority, are fairly unstable compounds that are decomposed by bacteria ("biodegradable") or that oxidize on contact with air into innocuous substances and hence can be disposed of with only minimal care. Those that cause real concern are the stable compounds that persist unchanged for decades or centuries, that appear to be immune to attack by atmospheric gases or bacteria, and that can accumulate to dangerous levels in water that drains disposal sites. Widely publicized examples are dichlorodiphenyltrichloroethane (DDT) and a group of compounds called polychlorobiphenols (PCBs). A major difficulty is that many such compounds have been discovered so recently that their properties are not well known: both their degree of toxicity and their long-term behavior in contact with air and water are still subjects of active research. Studies have also shown that some familiar compounds, long considered harmless and widely used in households and industry, can be dangerously toxic on long-continued exposure; examples are benzene, carbon tetrachloride, and formaldehyde.

Several methods are available for reducing the amount of organic chemicals in waste and drainage water. Most of these chemicals can be oxidized or broken down by controlled incineration, some can be transformed by enzymatic reactions promoted by bacteria, and some can be adsorbed onto mineral sufraces or particulate organic matter in soils and sediments. But the tailoring of methods for particular kinds of waste can be difficult and expensive, especially for wastes containing a variety of chemicals. One of the worst problems is the treatment of old dump sites, places where large amounts of waste were disposed of before the toxicity of its constituents was recognized or before the movement of groundwater in the vicinity had been studied.

Other important toxic contaminants that may come from waste disposal sites are some of the metals and their compounds. Arsenic, cadmium, mercury, and lead are well-known examples, but several others can be damaging if present in unusual amounts; most of the heavy metals, in fact, are known to be toxic, and even the familiar metal aluminum is harmful to plant life if too much is carried in solution. The metals are toxic only if their concentrations exceed a threshold value, which is different for each metal, thus effective control means keeping the concentration below this limit. For some of the toxic metals it is not necessary that concentrations be reduced to zero, because in very small amounts they are essential for life processes. Selenium is a good example: it is a well-known poison for cattle if present in large amounts in plants, but small quantities are necessary in the diet of animals, including humans. Fortunately the compounds of most potentially toxic metals are nearly insoluble in waters whose pH is close to neutral, as it is in much surface water and groundwater. Exceptions are chromium and selenium, which under oxidizing conditions are soluble in the form of stable anions, but their concentrations would be small in groundwater at low redox potentials. In nature the concentrations of toxic metals are seldom high enough to be harmful, but metallic contaminants of waste can become troublesome where infiltrating water is slightly acid. Control is often possible by neutralizing acidity with $CaCO_3$ or by adding substances to precipitate or adsorb the metal ions.

By concentrating large quantities of long-lived organic substances and compounds of toxic metals in waste dumps, and by allowing some to be dissolved in groundwater, humans are altering parts of the earth's chemical environment in ways that natural processes will not be able to correct for a long time in the future.

F. Disposal of Radioactive Waste

Chemical waste that contains radioactive elements poses particularly difficult problems of disposal. The difficulties are partly technical and partly political: the radioactive constituents that find their way into groundwater and surface water must be kept to exceedingly small amounts for very long times, and fear of the mysterious effects of radiation makes politicians reluctant to consider disposal of radioactive waste in the area of their constituents.

Radioactive elements are widespread in materials of the earth's crust and, in small amounts, are part of the living tissues of all plants and animals. Radioactive atoms are produced in air by cosmic rays, to which the environment is continuously exposed. Additional radiation exposure to much of the human population comes from the common use of X-rays in medical diagnosis and therapy. Thus, radiation is simply a part of our environment, and it can hardly be counted harmful in and of itself. Even in regions of the earth where radioactive minerals are more abundant than normal and radiation is somewhat higher than the world average, no harm to the inhabitants has been demonstrated. The worrisome feature of radioactive waste is not that it generates radiation but that the *amount* of the radiation is so large. Such concentrated radiation is known to be destructive to living organisms. Furthermore, the damage from radiation is at least in part cumulative, so that repeated exposure to small doses may lead to illness many years later. Exactly how much radiation can be tolerated without long-term effects is not known, but in the absence of this knowledge it is obviously desirable that radioactive waste be isolated as effectively as possible from the environment in which we live.

The radioactive material in waste generally includes many elements, some of which decay quickly to nonradioactive daughters and thus becomes harmless, and others that remain dangerously radioactive for long times, up to many hundreds of thousands of years. The problem of isolation is difficult chiefly because of these long-lived elements: certainly barriers can be erected to keep the radioactive elements out of the environment for decades or centuries, but no man-made structure can be relied on to endure for 10,000 or 100,000 yr. Ultimately, whatever is done to dispose of the waste, the possibility exists of breakdown of the artificial barriers and exposure of the waste to natural environmental agents—particularly groundwater, since plans for disposal generally envision placement well below the earth's surface. The geochemical problem of radioactive waste disposal, therefore, is the location of places where groundwater is so slowly moving, where its composition is so noncorrosive, and where the adjacent rock contains so much sorbent material, that no more than minute quantities of radioactive material will dissolve in the groundwater and be carried by it to surface environments, even over very long times. Much research has been devoted to seeking such places, and several likely candidates have been found, but some slight uncertainty always remains as to how accurately the behavior of rocks and groundwater can be predicted for 100,000 yr into the future.

Many kinds of radioactive waste have been generated, ranging from such low-level materials as slightly contaminated laboratory equipment and clothing to the intensely radioactive fuel rods of nuclear reactors that have been removed after many months of neutron bombardment of their contained uranium. The different types require different methods of disposal. Low-level waste, particularly if most of its radioactive materials are elements having short half-lives, can be safely disposed of in shallow covered trenches near the surface. High-level waste, however, requires burial several hundreds of meters underground. The different disposal methods are designed to give assurance that radiation reaching the environment from any of the waste repositories will be no greater than that from ore deposits of the radioactive elements, in other words, no more than humans have been exposed to for many millennia.

By generating nuclear power from the fission of uranium atoms and then by burying the radio active waste that is produced, humans will be altering their environment in complex ways. From a geological perspective, the alteration is slight but permanent: mining uranium and exposing it to neutron bombardment in reactors removes from the earth some of the uranium with which our planet was originally endowed. This means that in the remote futur, after all the newly formed radioactive elements have decayed, the total amount of radiation to which living organisms are exposed will be reduced. In our own immediate future, parts of the environment could have temporary greatly increased exposure if waste containing the new elements is not properly disposed of. But if the earth's surface is kept shielded for tens of thousands of years by putting the waste deep underground, the overall alteration of surface environments will be imperceptible.

G. Agrochemicals and Non-point-source Pollution

Non-point-source (NPS) pollutants are now recognized as the single greatest threat to the quality of surface and subsurface waters. A significant cause of NPS contamination is the regional-scale use of synthetic organic agrochemicals (i.e., fertilizers, herbicides, and pesticides). Characteristically, NPS pollutants (a) are widespread, which makes remediation efforts extermely complex and difficult; (b) have the potential for maintaining a relatively long active presence in the ecosystem; and (c) may result in long-term, chronic effects on human health and aquatic degradation. Historically, NPS pollutants have received less attention than that received by point-source pollutants (i.e., pollutants originating from a single "point" such as a landfill) because point-source pollutants are usually highly toxic, possessing an immediate threat to human health. However, public concern has recently shifted to NPS pollutants because point-source pollutants are more easily identifiable (i.e., their location and identity are usually known), which makes them less of an unknown threat, whereas NPS pollutants originate from multiple sources and can have a cumulative effect that persists for decades. Lower detection limits and increased reliance on groundwater in recent years have helped to focus attention on the NPS pollution problem. It has recently been estimated that 30–50% of all surface and subsurface water resources are impacted by NPS pollutants. The complexity of the earth's surface and near-subsurface makes the problem of assessing NPS pollutants a data-intensive task. The volume of physical, chemical, and biological information needed for even the simplest of assessments, at regional NPS scales, can be staggering. Nevertheless, assessing the environmental impact of NPS pollutants is fundamental to protecting water resources into the future, as well as for maintaining large-scale, chemical-dependent, modern agriculture.

H. Depletion of Metal and Fossil-Fuel Resources

There is one other human activity that significantly affects the environment: the ever-increasing exploitation of ore deposits and accumulations of gas, oil, and coal for the materials and energy on which modern civilization depends. Such concentrations of useful materials, accumulated over geologic ages, are not being renewed by any reactions fast enough to be of human interest. How long the existing accumulations will last is a much-argued question, but it seems certain that within a few decades or centuries the high-grade ores and major oil pools will be gone. Lower-grade deposits will serve to maintain supplies for a longer period, but the readily available cheap metals and cheap energy that we have enjoyed will not be available to future generations. It is often said that if global war reduces humankind to a state of barbarism, civilization will never reappear because the easily obtainable metals on which its early growth depends will have disappeared. In this respect, humankind has effected a profound and irreversible change in the earth's chemical environment.

VI. CONCLUSION

In the earth's development through geologic time, a complicated system of interlocking reactions has developed among the chemical constituents of the crust, atmosphere, and water bodies that keeps surface environments in a steady state. If one reaction temporarily goes faster than normal, other reactions respond to slow it. If one substance is added to or subtracted from the atmosphere or oceans, reaction rates change so as to cut down the excess or make up the deficiency. Thus the concentrations of materials in the atmosphere and water bodies are maintained at remarkably constant values. Although some of the concentrations apparently have changed over long periods of geologic time, changes occurring today are so slow as to be nearly imperceptible.

Among known astronomical bodies, the earth is unique in having an abundance of liquid water, abundant free oxygen in its atmosphere, and abundant life on its surface. These three characteristics are interrelated. Oxygen is maintained by a reaction between water and carbon dioxide in the leaves of living plants. Carbon dioxide is kept at a constant level by this reaction and several others, including dissolution in seawater and the dissolution and precipitation of carbonate minerals. Rocks are attacked by water, oxygen, and carbon dioxide, their complex silicate minerals breaking down into simpler minerals and other substances in solution, which ultimately become part of sedimentary rocks. The three weathering agents are restored by reactions in the earth's mantle, where gases are released and complex silicate minerals are newly formed. Along with these major reactions are many subsidiary ones that keep all parts of the chemical environment in delicate balance.

Throughout most of history, humankind has had little effect on this system of compensating reactions, but in the last two centuries the growing population and the increasing use of energy have produced effects that may mean long-term or even permanent changes in the chemistry of the environment. Notable among these possible changes are an increase in the carbon dioxide content of air, an increase in the acidity of rainfall, a decrease in the ozone of the stratosphere, the contamination of much surface and underground water, and the depletion of concentrations of metallic ores and fossil fuels. The steady-state reactions of the global system will partly compensate for these changes, but they cannot undo them entirely.

SEE ALSO THE FOLLOWING ARTICLES

ADSORPTION (CHEMICAL ENGINEERING) • ENVIRONMENTAL RADIOACTIVITY • GEOCHEMISTRY, ORGANIC • GEOLOGIC TIME • GEOMORPHOLOGY • METEOROLOGY, DYNAMIC • PLATE TECTONICS • POLLUTION, AIR • RADIOACTIVE WASTE DISPOSAL • SOIL PHYSICS

BIBLIOGRAPHY

Alley, W. M. (1993). “Regional Ground-Water Quality,” Van Nostrand-Reinhold, New York.

Barbash, J. E., and Resek, E. A. (1996). “Pesticides in Ground Water,” Ann Arbor Press, Chelsea, Mich.

Berner, E. K., and Berner, R. A. (1996). “Global Environment,” Prentice-Hall, Upper Saddle River, N.J.

Birkeland, P. W. (1999). “Soils and Geomorphology,” 3rd ed., Oxford Univ. Press, New York.

Corwin, D. L., Loague, K., and Ellsworth, T. R. (1999). “Assessment of Non–Point Source Pollution in the Vadose Zone,” Am. Geophys. Union, Washington, D.C.

Drever, J. I. (1982). “Geochemistry of Natural Waters,” Prentice-Hall, Upper Saddle River, N.J. Environmental Studies Board, National Research Council (1983). “Acid Deposition: Atmospheric Processes in Eastern North America,” Natl. Acad. Press, Washington, D.C.

Environmental Studies Board and Commission on Natural Resources, National Research Council (1982). “Causes and Effects of Atmospheric Ozone Reduction: An Update,” Natl. Acad. Press, Washington, D.C.

Fetter, C. W. (1999). “Contaminant Hydrogeology,” 2nd ed., Prentice-Hall, Upper Saddle River, N.J.

Holland, H. D. (1984). “The Chemical Evolution of the Atmosphere and Oceans,” Princeton Univ. Press, Princeton, N.J.

Environmental Measurements

William L. Budde
U.S. Environmental Protection Agency

GLOSSARY

Accuracy The degree of agreement between the measured concentration of a substance in a sample and the true value of the concentration in the sample.

Analysis The process of investigation of a sample of the physical world to learn about its chemical components, composition, structure, or other physical or chemical characteristics. Generally only samples are analyzed, and individual elements, compounds, and ions are separated from one another, identified, measured, or determined. A pure compound or multi-element ion is analyzed only when it is investigated to determine its components, composition, structure, or other physical or chemical characteristics.

Analyte A general term for any element, compound, or ion that is present in a sample or is targeted for determination in a sample.

Analytical method The complete process used to determine an analyte or analytes in a sample. The analytical method documents all the individual steps in the process from sampling to reporting the results.

Analytical method attributes Measures of the quality, reliability, and uncertainty of the determinations obtained with an analytical method. Typical analytical method attributes are selectivity, sensitivity, detection limits, signal/noise, recovery, accuracy, bias, precision, and validation. **Analytical method attributes** are sometimes called **figures of merit.**

Bias The systematic error in a measurement of the amount or concentration of an analyte in a sample.

Congeners Compounds or ions that are members of a series of related substances that differ only by the number of hydrogens that have been substituted by the same atom.

Derivatives Compounds or ions that are produced by chemical reactions of analytes. An analytically useful derivative has physical or chemical properties that are not possessed by the analyte but that can be employed to determine the analyte.

Detection limit The minimum quantity or concentration of an analyte that can be detected with an analytical method or technique. There are no generally accepted standard criteria for detection and detection limits often depend on the sample matrix. Therefore, detection limits must include the criteria for detection and the nature of the sample matrix. A technique with a higher analyte sensitivity does not always provide a lower detection limit because interfering sample matrix components may also be observed with higher sensitivities.

Determination The identification and the measurement of the concentration of an analyte in a sample.

Isomers Two or more compounds or multielement ions that have the same elemental composition but different structures.

Precision The degree of random variation in repetitive measurements of the concentration of an analyte in a sample. Precision is usually measured by the standard deviation or the relative standard deviation of the measurements.

Procedure A specific part of an analytical method that is concerned with one aspect of the method, for example, the liquid–liquid extraction of groups of similar analytes from a water sample.

Qualitative analysis The process of just identifying the analytes in a sample.

Quantitative analysis The process of both identifying and measuring the concentrations of the analytes in a sample.

Recovery The amount of analyte measured a sample matrix as a fraction of the amount of the same analyte that was added to the sample. If the analyte is present in the sample before the addition, the native amount is subtracted from the measured quantity before calculating the recovery.

Sample matrix The general nature of the sample and its components that can have a significant impact on the performance of an analytical method. For example, seawater and freshwater sample matrices are significantly different and this difference can impact the performance of an analytical method.

Selectivity A qualitative estimate of how well the analyte identification procedure is able to distinguish an analyte in a sample from one or many similar analytes with similar, or even some of the same, physical or chemical properties.

Sensitivity The electronic or other measurable signal produced by the analytical method or measurement technique per unit amount of analyte.

Signal/noise The ratio of analyte electronic or other measurable signal to the mean background matrix signal.

Technique A specific way of manipulating a sample or substance or measuring a substance. One or more techniques may be used within each procedure, and several procedures may be used within an analytical method.

Validation The determination of the attributes, or figures of merit, of an analytical method for one or more analytes in one or more sample matrices by one or more analysts in one or more analytical laboratories, and the acceptance of the attributes as reasonable and useful by the users of the data. There are many levels of analytical method validation ranging from the validation of a method for a single analyte in a single matrix by a single analyst in a single laboratory to a multianalyte, multimatrix, multianalyst, and multilaboratory validation.

ENVIRONMENTAL CHEMICAL DETERMINATIONS are the identifications and measurements of the concentrations of elements, compounds, or ions in environmental samples. Air, water, soil, and sediment are broad categories of environmental samples, and each of these can be divided into several subcategories—for example, ambient air, indoor air, and vapor emissions from mobile or stationary combustion sources. Body fluids and tissue are often analyzed to determine human, animal, and plant exposure to environmental chemicals.

I. SIGNIFICANCE OF ENVIRONMENTAL CHEMICAL DETERMINATIONS

Accurate and precise identifications and measurements of specific chemical substances in the environment are fundamental to environmental research and protection programs. Determinations are required to understand natural background concentrations of chemicals in the environment, the nature and extent of environmental pollution by anthropogenic chemicals, trends in concentrations of these substances, the transport and fate of chemical substances, and the causes of variations of concentrations in time and space. Accurate and precise determinations are also required to assess human health and ecological risks caused by exposure to natural and anthropogenic substances, establish air and water quality standards, develop pollution control strategies, evaluate the effectiveness of pollution prevention and treatment technologies, and monitor compliance with and the effectiveness of discharge and other government regulations. These data are also needed to develop, calibrate, and verify mathematical models used to predict the impact of changes in concentrations of specific substances in the environment. Environmental chemical determinations are also required to set priorities and make cost estimates for the remediation of abandoned hazardous waste sites.

In addition to specific chemical substances, a variety of other determinations are very important and are often required for environmental research and protection programs. These include identifications and measurements of bacteria, viruses, protozoa, and minerals such as asbestos fibers. Measurements of meteorological conditions, particulate matter in air by size, water turbidity, biological oxygen demand, chemical oxygen demand, and radioactivity are also very important. These and other similar determinations are beyond the scope of this article and the reader should consult other reference books for information about these topics.

II. CHEMICAL ANALYSIS STRATEGIES

A wide variety of chemical analytical strategies are used in environmental analysis programs. In order to develop an appropriate strategy for a specific program, the objectives of the program must be thoroughly defined and understood by all participants. An analytical strategy for the program can then be developed by selecting and combining the most appropriate analytical methods and other key elements that will provide the results needed to fulfill the objectives of the program. Some of the major issues and strategies that must be considered in developing a specific strategy for an analysis program are described in this section.

A. Samples and Sampling Strategies

The goals of the environmental program will usually define the specific types of samples that must be acquired and analyzed—for example, ambient air, drinking water, ground water, soil, sediment, blood, sweat, urine, and tissue. The sampling strategy is a broad but detailed plan for the acquisition of the samples needed for the project. Sampling techniques and procedures are the actual physical processes used to acquire the samples. Sample preservation protects the integrity and validity of the samples before, during, and after the actual sampling process. Sampling strategies, techniques, and sample preservation are major factors contributing to the success or failure of environmental chemical determinations.

The sampling strategy defines in detail what, where, when, and how the environmental samples will be acquired. This plan should specify all details including the number of sampling stations for air, water, soil, sediment, or fish samples; the locations of the sampling stations; the time of the day of sampling; the frequency of sampling; the depth of sampling for water, soil, and sediment samples; the meteorological conditions and altitude for air sampling; and the need for replicate samples taken at the same time and place. For body fluids and tissue samples, a myriad of details must be considered including the number of persons contributing or the kinds of wildlife-related samples that must be collected. The plan should specify a series of discrete samples, a composite sample from contributions at various intervals, or a continuous flow of sample over a period of time. A sampling strategy that does not provide appropriate and representative samples seriously jeopardizes the value of the chemical determinations.

The sampling techniques and procedures define the type of sampling equipment, the containers used for the samples, the procedures used to clean the sample containers and sampling equipment, the calibration of sampling equipment, and sample compositing procedures if composite samples are required. Sample preservation must ensure that the chemical composition of the sample at the time of analysis is the same as it was at the time and place of sampling, and is not the result of physical or chemical changes caused by the conditions of shipment and storage prior to chemical analysis. Significant research has been conducted to find sample containers that resist adsorption or degradation of sample components. Samples are often shipped and stored at low temperatures, and treated with various chemical reagents, to retard microbiological degradation of the analytes or to prevent chemical reactions that can change the composition of the sample. Appropriate sampling techniques and preservation procedures are required to provide valid samples and valid chemical determinations.

B. Determination of Total Elements or Total Related Substances

A standard analysis strategy is the determination of the total amount of an element in a sample where the element is present in several or more elemental forms, compounds, ions, oxidation states, or physical phases. This strategy was developed before chromatographic and other techniques were available to separate the individual substances containing the element of interest. It is widely used in environmental research and government regulatory programs. However, the meaning of *total* is variable because some analytical methods do not include all the physical phases or chemical forms in the determination. The sample preparation procedures in individual analytical methods define exactly which elemental forms, compounds, ions, oxidation states, and physical phases are included in a total measurement. A similar strategy is the determination of the total amount or total concentration of a group of closely related compounds or ions without specifying the relative or absolute concentrations of the individual substances. Closely related substances usually have some common physical or chemical properties, elemental composition, or structure.

The broadest definition of, for example, total mercury (Hg) in a sample is the sum of the elemental Hg^0, the Hg in all inorganic compounds of Hg^{1+}, the Hg in all inorganic compounds of Hg^{2+}, and the Hg in all organic compounds in which Hg is bonded to C, O, N, S, or some other element. All phases are included because some species may be present in the vapor phase (Hg^0), some soluble in water, and some insoluble in water or present in the particulate phase of an air sample. More limited definitions of total are often used in analytical methods that separate physical phases or chemical forms that contain the element of interest.

If a water sample contains both soluble and insoluble manganese compounds and ions, and it is filtered to separate the dissolved and insoluble fractions, and the filtrate and insoluble residue are analyzed separately, the results can be expressed as total dissolved Mn and total suspended or insoluble Mn. Phosphorus (P) can be determined colorimetrically as the *ortho*-phosphate ion, PO_4^{-3}, in aqueous samples after a reaction that forms an intensely blue-colored derivative. However, polyphosphate ions and other ions and compounds containing P do not form this derivative. Total P in a sample can be determined with the same colorimetric procedure after acid hydrolysis and oxidation of all ions and compounds containing P to PO_4^{-3}. In some elemental analyses the sample is treated with reagents designed to make available for measurement some fraction of an element or elements but not the total amount. For example, a soil sample may be treated with water at pH 3 to simulate the leaching process of acid rain. A total elemental analysis of the filtrate provides information about just those elements solubilized by the mild acid treatment. This can be called the determination of total mild acid leachable elements.

The measurements of total organic carbon (TOC) and total organic halogen (TOX) in a sample are used to assess many types of environmental samples. The analytical methods for TOC employ procedures to physically separate the inorganic carbon—that is, carbonate, bicarbonate, cyanide, and other inorganic substances containing C—from the organic compounds and ions in the sample. The TOC is then measured by oxidation of the organic compounds and ions to carbon dioxide, which is determined by one of several techniques. The analytical methods for TOX use procedures to separate inorganic halogen-containing ions from halogen-containing organic compounds and then determine the total halogens in the organic compounds.

Some analytical methods have been developed for the determination of groups of closely related compounds or ions. The classic example is the measurement of combined phenolic materials colorimetrically after a reaction that forms a red derivative with many phenols. However, different phenols form derivatives with somewhat different visible absorption spectra and various phenols have different reactivities with the derivatizing reagent. Therefore this method is calibrated with pure phenol and the measurement gives just an estimate of the total phenols in the sample. For this reason this method is probably no longer widely used and phenols are usually determined as individual compounds with other analytical methods. Subgroups of complex mixtures of congeners are sometimes measured together for convenience of interpretation or for government regulatory programs. These determinations give concentrations of, for example, total tetrachlorobiphenyls and total pentachlorobiphenyls. Similar determinations of chlorinated dibenzo-*p*-dioxins and chlorinated dibenzofurans at each level of chlorination are specified in some analytical methods.

C. Determination of Specific Substances

Before the development of efficient chromatographic separation techniques and selective and sensitive detectors, analytical methods for the determination of specific analytes in environmental samples were very limited. These methods depend on highly selective chemical reactions that are relatively rare and difficult to discover, or on very selective physical measurements such as atomic absorption or emission techniques for elemental analytes. Therefore only a relatively few analytical methods for the most common and amenable organic and inorganic compounds or multielement ions were developed. It would have been essentially impossible, and enormously costly, to develop, test, document, and implement a large number of nonchromatographic analytical methods for a wide range of often similar organic and inorganic analytes.

Vapor phase or gas chromatography (GC) and high performance liquid chromatography (HPLC) provided capabilities for the separation of microgram (10^{-6}) and smaller quantities of often similar individual substances in complex mixtures. A variety of GC and HPLC detectors were developed, and depending on the detector, the separated substances could be determined with good to excellent sensitivity and with fair to excellent reliability. It was soon recognized that determinations of a broad range of specific chemical substances was not only practical but also essential to achieve the goals of many environmental research and protection programs.

Gas chromatographic techniques were first applied during the 1960s to synthetic organic compounds and natural products that are often complex mixtures of organic compounds. The earliest applications of these techniques in environmental research were GC-based determinations of petroleum hydrocarbons in polluted air and chlorinated hydrocarbon pesticides in several types of samples. Analytical chemists and environmental scientists concerned about organic pesticides, industrial organic chemicals, and other organic compounds emphasized the development of analytical methods for the determination of specific organic compounds. In contrast, the emphasis in elemental and inorganic analyses during the 1960–1980s was, with a few exceptions, on determinations of the total amounts or concentrations of specific elements in a sample without regard to the specific compounds or ions containing those elements. With the recognition that toxicity and other environmentally significant properties varied widely with the specific compound or ion, inorganic analytical chemists later focused on chromatographic separations of individual species. The term *speciation* came into use, mostly by inorganic chemists, to distinguish this type of analysis from the conventional determinations of the total amounts or concentrations of the elements in a sample.

There are two general strategies for the determination of specific organic or inorganic compounds and ions in environmental samples. The target analyte strategy dominated analytical chemistry before the development of chromatographic separation techniques, and is by far the most commonly used with contemporary separation and detection techniques. The broad spectrum strategy became feasible with the development of high resolution chromatographic separation techniques and spectroscopic detectors.

1. Target Analyte (TA) Strategy

Target analytes are known substances with known chemical, physical, and other properties. They are either known or thought to be in samples, and they must be determined to meet the objectives of the environmental analysis program. The TA strategy is the analysis of the sample with an analytical method that is designed and optimized to determine the target analyte or a group of similar and separated target analytes. If the target analytes are a diverse group with sufficiently different chemical and physical properties, they are divided into subgroups according to their similar properties. Several optimized analytical methods and separate environmental samples are used to determine the analytes in the subgroups.

The TA strategy has many advantages that favor its widespread application. Sample preparation procedures can be designed to separate the target analytes from the sample matrix with maximum efficiency and to concentrate them in a suitable solvent for further chromatographic separation. Some interferences can be separated from the analytes during sampling or sample processing by pH adjustments, chemical derivatization, evaporation of nontarget substances, or other techniques. Chromatographic separation and detection techniques can be selected to give the best practical resolution of analytes, selective detection, and the lowest detection limits. The detector can be calibrated for quantitative analysis with standard solutions of the target analyte or analytes and the analytical method tested, perfected, and validated using test sample matrices fortified with known concentrations of the target analyte or analytes.

The target analyte strategy is widely accepted, understood, and used in analytical chemistry, environmental research, environmental protection programs, and in many other fields of investigation. It is used in most environmental quality surveys and government regulatory compliance monitoring programs. Many analytical methods for a variety of target analytes have been developed, tested, documented, and implemented. Cost estimates for analyses are readily made and analytical costs are not difficult to control. The target analyte strategy is used in the vast majority of chemical analyses reported in the scientific literature.

2. The Broad Spectrum (BS) Strategy

The objective of the BS strategy is to discover the substances present in the sample and to measure their concentrations without a predetermined list of target analytes. As a starting point the BS strategy may follow the general scheme of a target analytical method, but with minimum sample processing to allow a broad variety of generally similar substances to reach the chromatographic separation and the detector. Procedures to remove target analyte interferences are minimized or not used to avoid discarding interesting and potentially important components. If the sample is very complex, it is usually divided into fractions that are likely to contain components with similar properties. Several or more different types of chromatographic separations may be required for the various fractions depending on the components of the sample and the breadth of information desired. The chromatographic detectors are generally spectroscopic detectors that can provide information about the composition and structure of the sample components. This information is used to identify known substances, unexpected substances, and even unknown substances. The most important and widely used detector for the broad spectrum strategy is the mass spectrometer, but infrared, nuclear magnetic resonance, and other spectroscopic techniques are sometimes used.

The identification of all or most substances in one or more chromatograms, even with extensive spectroscopic data, can be a challenging and difficult process. The available data may not be sufficient to even tentatively identify all the components, especially if pure authentic samples of suspected substances are not available in the laboratory. Calibration of a broad spectrum method for quantitative analysis is delayed until the desired components are identified. For these reasons, and the general preoccupation with target analytes, the BS strategy is much less common than the target analyte strategy.

The BS strategy is obviously important for the discovery of unknown naturally occurring substances and anthropogenic chemicals in the environment. However, BS strategies are more difficult than TA strategies to develop, document, and implement. Cost estimates for analyses are uncertain because the number of substances found, identified, and measured is not known until after the samples are analyzed. Therefore analytical costs are difficult to control. Because of the potential costs, the instrumentation requirements, the technical skills needed, and time required, the BS strategy is not often employed especially by programs with limited objectives and budgets. A strategy sometimes used in environmental studies is to develop a target analyte method that can meet the objectives of the study and give some attention to other chromatographic peaks to identify potentially new or unexpected substances.

D. Single Analyte and Multianalyte Methods

Traditional analytical methods, which were generally developed prior to the widespread application of GC and HPLC techniques, were nearly always designed for a single target analyte—for instance, the colorimetric determination of PO_4^{-3}. With the development of GC and HPLC separation techniques, the determination of several or more similar compounds or ions in a sample was feasible, and multianalyte analytical methods were developed and documented. However, analytical chemists did not generally attempt to include more than about 10–20 target analytes in a method. This strategy was necessary because early GC and HPLC columns were not very efficient and most GC and HPLC detectors were either not selective or had limited selectivity. Sample preparation procedures designed to reduce or eliminate interferences continued to be very important. Analytical methods for chlorinated hydrocarbon pesticides with a GC separation and an electron capture detector require appropriate sample preparation. These multianalyte pesticide methods are sometimes called multiresidue methods because the pesticides are residues in crops and other samples.

As more efficient high resolution chromatographic separation techniques were developed, and spectroscopic detectors came into widespread use, the number of target analytes in multianalyte methods was increased. The separation of most or all target analytes is often feasible and spectroscopic detectors usually provide sufficient information to make correct identifications of target analytes even when some are not fully separated. Well tested and documented analytical methods for 80–100 or more analytes in some types of samples are presently available. Multianalyte methods significantly reduce the cost of an analysis on a per analyte basis, and add support to the already strong justification for high resolution chromatographic separations and spectroscopic detectors.

E. Remote Laboratory Analyses and Field Analyses

Chemical analyses can be conducted in a laboratory remote from the locations where the samples are taken or in the field near the sampling sites (on-site). On-site analyses can be conducted in a field laboratory, which may be a temporary building or a truck trailer, van, or recreational vehicle equipped with utilities services and analytical equipment. Another type of field laboratory is within a materials or fluids processing facility. Alternatively, field analyses can be conducted with mobile or portable instrumentation carried in a small van, sport utility vehicle, moved with a hand cart, or carried by a person. Each of these strategies has some advantages and some disadvantages.

The remote laboratory has the advantages of providing carefully controlled temperature, humidity, ventilation, and background conditions with adequate space and utilities to support a large array of major analytical instrumentation and a staff of skilled analytical chemists and technicians working in a convenient, comfortable, and safe environment. The major disadvantage is that environmental samples must be carefully preserved, shipped, and stored prior to analysis. Furthermore, the analytical results may not be available for several days or weeks because of the time required to transport the samples to the laboratory, incorporate the analyses into work schedules, and service the multiple clients of a remote laboratory.

Field determinations have the major advantage of greatly reducing the time between taking the samples and the availability of the analytical results. This may allow utilization of the results quickly, which can provide significant cost savings in the field operations that utilize the analytical information. For example, the rapid availability of results from field determinations can be used to fine-tune a sampling strategy to obtain the most significant samples for detailed remote laboratory analysis. Similarly, the results from field determinations can be used to direct the work of construction crews or well drillers using heavy and costly equipment. Mobile or portable instrumentation can provide rapid and low-cost results from a large number of samples taken over a broad geographic area in a short period of time.

The analytical methods that are feasible in the field may be significantly limited compared to what is feasible in a remote permanent laboratory. If a field laboratory is located in a temporary building or a large truck trailer, and adequate utilities and personnel are available, many of the kinds of analytical methods that are routinely implemented in a remote laboratory may be feasible in the field. However, because of space and power limitations, a broad variety of instrumentation is usually not available, and the number of different analytical methods that can be implemented is smaller than in a remote permanent laboratory. Mobile or portable instrumentation is usually more limited, and generally sample analyses are less complete and detailed than in a field or remote laboratory.

F. Discrete Samples and Continuous Monitoring

Field analyses can be conducted with discrete or composited samples similar to the samples used in a remote laboratory or by continuous monitoring of substances in a flowing stream of gases or liquids. Continuous monitoring is required when the results are needed within a period of time that is shorter than the time required to take and analyze conventional discrete samples. Continuous monitoring is often required to determine substances in a processing facility and provide rapid feedback of results that are used in process control strategies. If continuous determinations are made with sufficient speed to permit changes in sampling or other strategies while the determinations are in progress, these are called real-time analyses.

Continuous environmental monitoring in the field is needed when the substances present and their concentrations are changing rapidly. Discrete samples taken at inappropriate times will give results that do not correctly assess the variable conditions in the atmosphere or a flowing stream. Sample compositing may provide a better assessment, but continuous monitoring and integration of the determinations over time gives the best assessment. The determination of the sources of fugitive emissions that are rapidly dispersed in the atmosphere or in a flowing water stream requires continuous and sometimes mobile continuous monitoring. The concentrations of air pollutants such as CO, NO_x, SO_2, and O_3, that vary widely with sunlight, automobile traffic, wind speed, wind direction, and other meteorological conditions are monitored continuously.

Sampling and analytical equipment used in processing facilities and in the field have some of the same requirements. However, equipment used in a facility may have fewer constraints in regard to size, weight, and power requirements than mobile and portable field equipment. Standard chromatographic separation techniques are generally too slow for continuous measurements although GC separations, especially fast GC, are used for process control.

G. Analytical Quality Assurance and Control

Analytical quality assurance (QA) is a broad program of actions designed to ensure that the chemical determinations are of known and acceptable quality. The QA program encompasses all aspects of the chemical analysis from the design of the sampling strategy to the documentation of the results. Analytical quality control (QC) is the implementation of specific actions designed to control the quality of the determinations at some defined level of acceptance. All physical measurements have some degree of variability and uncertainty, and the QA/QC program should define the limits of these. The quality of analytical determinations is judged by the attributes, or figures of merit, previously defined, which include selectivity, sensitivity, detection limits, signal/noise, recovery, accuracy, bias, precision, and validation. These attributes are determined using a variety of special measurements and fortified samples (QA/QC samples) that are described in the analytical method. The time relationship between the analyses of the environmental samples and the determination of the quality attributes is critical in estimating the quality of the environmental determinations.

Published environmental chemical determinations are often used in ways that are not anticipated by the original investigators. The determinations may be used by other investigators to estimate human or ecological exposure, assess the level of industrial discharges, or to develop environmental quality standards or discharge limitations. Therefore, the results of the QA/QC measurements, and their time relationships to the analysis of environmental samples, should be permanently attached to the environmental sample results. These QA/QC data should demonstrate that the environmental data were obtained under controlled conditions that provide credibility to the results. Other users of the data, sometimes years in the future, should be able to assess the quality and applicability of the information, and take into account the variability and uncertainty in the determinations.

III. DEVELOPMENT AND DOCUMENTATION OF ANALYTICAL METHODS

The complete development and documentation of an analytical method usually occurs over a period of several years or more. The germ of most new methods is often a research project in which a determination is needed. Organizations that specialize in developing and manufacturing analytical instrumentation frequently play a major role in supporting the development of new analytical methods. But instrumentation alone does not constitute an analytical method. The general acceptance and widespread use of methods depends on many factors, including the needs for research or environmental monitoring of various substances, the cost and complexity of instrumentation, the required laboratory or field skills, and especially government regulations. There is a high degree of variability in the completeness, documentation, testing, and validation of published analytical methods.

A. Research Methods

The basic concepts and fundamentals of most analytical methods are usually first published in scientific journals—for example, The American Chemical Society publications *Analytical Chemistry*, *Environmental Science and Technology*, and *The Journal of The American Chemical Society*. These descriptions are usually brief and often merely summarize the techniques and procedures of the method. Research reported in scientific journal articles is often focused on a detailed investigation of a narrow subject area, and new analytical techniques and procedures developed for the research are rarely tested in a broader context. Analytical quality assurance and control are often minimal or not described. Techniques and procedures described in scientific research journals are the beginnings of analytical methods, but they usually require considerable development, modifications, and testing before they become widely accepted analytical methods.

B. Methods in Development

Academic research groups, industrial laboratories, and government agencies interested in conducting environmental surveys or monitoring programs often adapt research techniques and procedures to the broader needs of the survey or program. During this stage considerable experimentation is underway to evaluate equipment, instrumentation, chemical reagents, and other materials used in the method. The incipient method may be tested with many potential analytes and sample matrices. Modifications are implemented to minimize or eliminate problems discovered during the development process. Techniques and procedures are developed for identifications of analytes, calibration of instruments for quantitative analysis, and analytical quality assurance and control. Methods in development may be described in a series of draft versions that are not usually formally published but may appear in bound or unbound technical reports. These methods are described in more detail than in scientific journal articles and descriptions often contain detailed information about required equipment, supplies, reagents, instrumentation, and personnel skills. The technical reports and draft method descriptions are often distributed informally through personal contacts, scientific conferences, or the Internet.

C. Methods Published by Standard-Setting Organizations

Standard-setting organizations that publish analytical methods for environmental analyses include The American Society for Testing and Materials (ASTM), The American Public Health Association (APHA), The American Water Works Association (AWWA), The Water Environment Federation (WEF), The International Standards Organization (ISO), and the Association of Official Analytical Chemists (AOAC). These organizations assemble working committees of experts that consider developed, widely used, and generally accepted analytical methods for publication as standardized analytical methods. Requirements of individual organizations vary, and these may include a description of the method in a specific editorial format and the availability of multilaboratory validation data for the proposed analytical method. Some of these organizations also sponsor or participate in multilaboratory validation studies of proposed analytical methods. When an analytical method is published by one of these organizations, it usually has been thoroughly tested and used by many laboratories and analysts, and is widely known and accepted by specialists in that type of determination. Published methods are reviewed periodically, updated as needed, and may eventually be replaced by entirely new methods.

D. Methods Published or Referenced in Government Agency Regulations

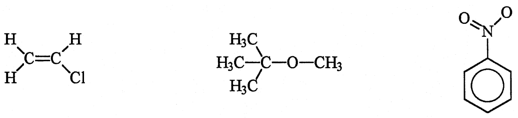
The U.S. Environmental Protection Agency (USEPA) and other federal, state, and local government agencies in the United States may require chemical analyses to determine compliance with air quality, water quality, liquid waste discharge, solid waste disposal, and other environmental regulations. The analytical methods specified for these purposes may be methods in development, methods published by standard-setting organizations, or new methods documented in the regulatory proposals. The USEPA has promulgated two types of analytical methods. Some regulations, like some drinking water and waste water regulations, require either analytical methods designated in the regulations or approved alternative test methods for compliance monitoring. Other regulations—some USEPA solid waste regulations, for example—include or reference analytical methods that are suggested or optional but allow any other appropriate analytical methods.

IV. CHARACTERISTICS OF ANALYTES, SAMPLES, AND SAMPLING TECHNIQUES

The physical and chemical properties of analytes and the nature of the sample have a major impact on, and often limit, the sampling and other procedures and techniques that can be employed in an analytical method. Major issues that must be considered when developing and analytical method are the volatilities, thermal stabilities, photochemical stabilities, polarities, water solubilities, and chemical reactivities of the sample components or target analytes; the physical state of the sample; and the nature of the sample matrix. Analytes, whether organic or inorganic, can be broadly divided into three categories based partly on vapor pressure, or volatility, at ambient temperature and on some other physical and chemical properties. There are major differences in the procedures and techniques used to acquire and process condensed phase and vapor phase samples. Sampling ground water requires expert selection of procedures and techniques to avoid significant analytical errors.

A. Volatile Analytes

Volatile analytes are usually defined as those having vapor pressures (VPs) greater than about 0.1 Torr at 25° and an external pressure of 760 Torr [1 Torr = 1 mm of Hg or 133 kilopascals (kPa)]. Figure 1 shows the structures, molecular weights (MWs), boiling points (BPs), and VPs of three representative volatile compounds of environmental interest. Most volatile analytes have MWs below 200, but a low MW does not guarantee that an analyte will be volatile. Many substances have MWs below 200, but they are not volatile because they are ionic or have polar groups of atoms or engage in hydrogen bonding with other molecules. Some compounds have high VPs—for example, MTBE in Fig. 1—but reduced volatility in water because of their high water solubilities. Boiling points of volatile substances range from below 0 to above 200°. Some compounds with BPs at the upper end of this range are surprisingly volatile. For example, nitrobenzene in Fig. 1 has a BP of 211° but still has a VP of 0.245 Torr at 25°. At normal room temperatures the distinctive odor of nitrobenzene vapor is readily detected by most people.



Vinyl chloride
MW 62.5
BP -13 C
VP 2980 Torr @ 25 C

Methyl *tert*.-butyl Ether (MTBE)
MW 88
BP 55 C
VP 249 Torr @ 25 C

Nitrobenzene
MW 123
BP 211 C
VP 0.245 Torr @ 25 C

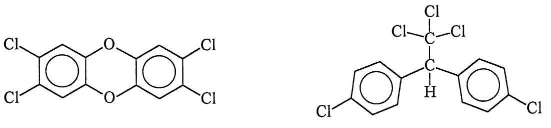
FIGURE 1 The structures, MWs, BPs, and VPs of three representative volatile compounds of environmental interest.

B. Semivolatile Analytes

Semivolatile analytes are usually defined as those having VPs in the range of 0.1 to about 10^{-9} Torr, but this range is approximate and some substances with VPs in this range are considered volatiles or even nonvolatiles (next section). For example, Hg^0 has a VP of 2×10^{-3} Torr at 25°, but is often considered a volatile analyte. Most semivolatile analytes have molecular weights in the range of 100–500, but a MW in this range does not guarantee that a compound is a semivolatile. However, since semivolatiles nearly always have higher MWs and contain more atoms than the volatiles, they have a significantly larger number of isomers, congeners, and chiral forms. Semivolatiles are usually devoid of structural groups that are susceptible to thermal decomposition below about 300°, or cause high polarity, or impart high water solubility, or are very chemically reactive. Figure 2 shows the structures, MWs, some melting points (MPs), and VPs of four representative semivolatile compounds of environmental interest. Both benzo[a]pyrene and 2,3,7, 8-tetrachlorodibenzo-*p*-dioxin have VPs at the low end of the semivolatile range, but they have little or no polarity and tend to behave like other semivolatiles with higher VPs. The VPs of semivolatile compounds are generally insufficient to give vapor concentrations at ambient temperatures that can be detected by a distinctive odor.

Benzo[a]pyrene
MW 252
MP 179 C
VP 5.49 x 10^{-9} Torr @ 25 C

3,3',5,5'-Tetrachlorobiphenyl
MW 292
VP 8.45 x 10^{-6} Torr @ 25 C



2,3,7,8-Tetrachlorodibenzo*p*-dioxin
MW 322
MP 305 C
VP 1.5 x 10^{-9} Torr @ 25 C

1,1-bis(4-Chlorophenyl)-2,2,2-trichloroethane (4,4'-DDT)
MW 354.5
MP 109 C
VP 1.60 x 10^{-7} Torr @ 20 C

FIGURE 2 The structures, MWs, some MPs, and VPs of four representative semivolatile compounds of environmental interest.

TABLE I The Number of Possible Chlorinated Biphenyl, Chlorinated Dibenzo-*p*-Dioxin, and Chlorinated Dibenzofuran Congeners and Isomers

Parent compound	Total congeners	Cl_1	Cl_2	Cl_3	Cl_4	Cl_5	Cl_6	Cl_7	Cl_8	Cl_9	Cl_{10}
Biphenyl	209	3	12	24	42	46	42	24	12	3	1
Dibenzo-*p*-dioxin	75	2	10	14	22	14	10	2	1		
Dibenzofuran	135	4	16	28	38	28	16	4	1		

The large number of potential congeners and isomers of some semivolatile compounds is illustrated in Table I, which shows the numbers of possible chlorinated biphenyl, chlorinated dibenzo-*p*-dioxin, and chlorinated dibenzofuran congeners and isomers. The total number of chlorinated congeners of each parent compound is the sum of the possible isomers at each level of chlorination. Dibenzofuran has more possible chlorinated congeners and isomers than chlorinated dibenzo-*p*-dioxin (Fig. 2) because it has just one ring oxygen and a less symmetrical structure. If F, Br, I, CH_3, or any other uniform substituent replaces Cl, the same numbers of possible congeners and isomers would exist. If different atoms or groups of atoms were mixed as substituents on these parent compounds, a significantly larger number of different substances, congeners, and isomers are possible.

C. Nonvolatile Analytes

All other compounds and essentially all ions are classified as nonvolatile. Substances in this category have VPs lower than about 10^{-9} Torr or have structural groups that are susceptible to thermal decomposition below about 300°, or that cause high polarity, or that impart high water solubility, or that are very chemically reactive. Figure 3 shows the structures, MWs, some MPs, and VPs of six representative compounds of environmental interest from this group. All these examples except the herbicide glyphosate have reported VPs in the semivolatile range, but they do not behave like typical semivolatile compounds. Nitroglycerine and N-nitrosodiphenylamine undergo thermal decomposition at temperatures below well below 300°. The pesticide carbofuran and the herbicide diuron are more thermally stable, but they also tend to decompose below 300° and on hot surfaces. The polarities of the functional groups of this class of compounds cause adsorption on polar surfaces and dipole–dipole interactions with other molecules. Nearly all carboxylic acids—for instance, the herbicide 2,4-D in Fig. 3—are susceptible to decarboxylation at elevated temperatures and they interact with basic substances or basic surfaces which reduces their volatility. Similarly, basic substances like amines and some other nitrogen compounds interact with acids and acidic surfaces, which reduces their volatility.

D. Condensed Phase Samples

If only semivolatile and non-volatile components are of interest, or the sample only contains these categories of analytes, sampling procedures for most condensed phase samples are not difficult and require only a few precautions. These include the materials used for sample containers, the cleaning of sample containers, and the preservation of the sample. Water samples for elemental analyses are treated with acid at the time of sampling, or well before analysis, to reduce the pH to $\lesssim 2$. This ensures the solubility of metal-containing ions, which can precipitate or adsorb on container walls at a pH $\gtrsim 2$. Glass containers

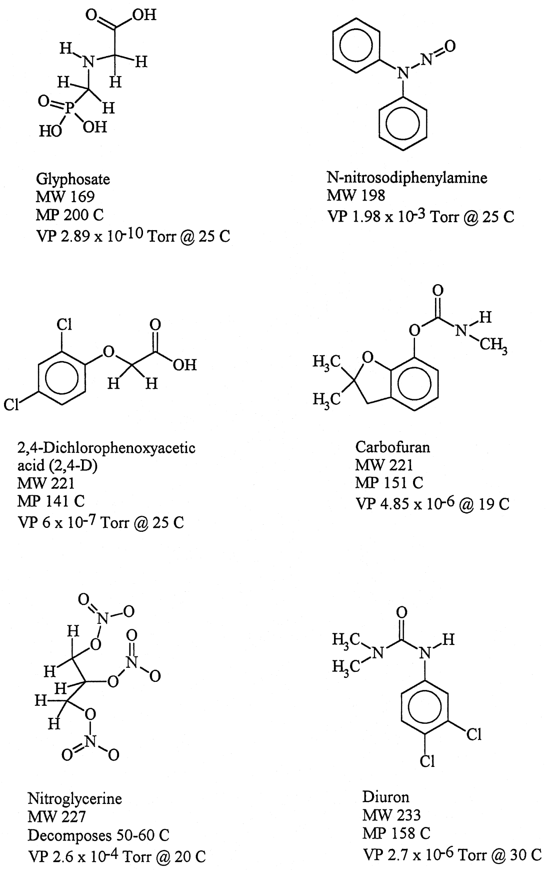


FIGURE 3 The structures, MWs, some MPs, and VPs of six representative nonvolatile compounds of environmental interest.

are not used for elemental analysis samples because aqueous acid solutions can leach trace elements from glass. Samples are taken in plastic containers that provide the added benefits of reduced weight compared to glass, little or no breakage during handling and shipping, and single-use containers that can be disposed of at a plastic recycling facility. However, samples for the determination of organic analytes are taken in glass containers to prevent background contamination of the samples by organic compounds that can leach from plastic materials. These samples are often preserved by addition of acid to make the sample pH $\lesssim 2$ to retard microbiological degradation of some analytes. Acid leaching of glass containers generally has no effect on the concentrations of organic analytes.

Condensed phase samples containing volatile analytes require special techniques since the volatile components are elusive and can be readily lost during sampling and sample processing. Water samples must be taken with little or no agitation of the sample and poured into the container so there is no air space, often called a headspace, between the top of the sample and the air tight seal of the container. This prevents vaporization of volatile components into the headspace during shipment and storage, which reduces the concentration in the aqueous phase. Aliquots of water samples are removed from sample containers by inserting a syringe needle through the inert airtight septum seal and drawing water into a gas tight syringe. Solid samples, for example soils and sediments, may also contain volatile components trapped in pores or dissolved in associated water and these components are easily lost during sample handling. Solid and semisolid samples are carefully and quickly placed in wide-mouth sample containers that are sealed in the field and not opened during any subsequent processing of the sample. Volatile components are usually partitioned into water, organic solvents, or the vapor phase, and the extracts are analyzed using techniques developed for liquid or vapor phase samples. Water or an organic solvent are added to the sample containers through inert and airtight septum seals with a syringe.

Groundwater sampling is a significant challenge because water must be lifted to the surface through a bore hole from depths ranging from a few meters to 75 meters or more. The lifting process can disturb the equilibrium between dissolved analytes and analytes associated with particulate matter and significantly change the temperature and pressure of the water sample. These factors can cause changes concentrations of analytes and raise questions about the representativeness of the sample. Volatile components of groundwater samples are particularly susceptible to losses caused by temperature and pressure changes and degassing of the sample. Several types of down-hole water samplers are available and a several types of pumps are used to purge a well and lift a stream of ground water to the surface. The materials of construction of the down-hole samplers, pumps, tubing used with pumps, and well casings can also impact the concentrations of analytes in the sample. Considerable research has been conducted to evaluate the materials and techniques used to sample groundwater, but uncertainties remain because there is no satisfactory groundwater standard of reference for comparison of various techniques.

Semivolatile analytes in vapor phase samples are often associated with particulate matter or aerosols that are collected on glass or quartz fiber filters in a flowing air stream. However, some semivolatile analytes that have higher VPs, for example, 3,3′,5,5′-tetrachlorobiphenyl in Fig. 2, can slowly vaporize from a particle trapped on a filter in the flowing air stream and will be lost unless captured by an in-stream sampling device. Small glass or metal tubes containing polyurethane foam are often used to capture vaporized semivolatile analytes. Other solid phase adsorbents, which are described in the next section, are also used to trap semivolatile analytes vaporized from particulate filters.

E. Vapor Phase Samples

There are two general approaches to sampling air, or vaporous emissions from stationary (stack) and mobile (automobile, truck, etc.) sources, for the laboratory determination of volatile analytes. Bulk vapor phase samples can be taken in the field in various containers and transported to a remote or field laboratory for analysis. Containers used for bulk vapor phase samples include flexible polyvinyl fluoride (Tedlar) bags, evacuated glass or metal reservoirs, and thermally insulated cryogenic collection vessels. Alternatively, the volatile analytes can be separated from the main components of air in the field and just the analytes and their collection devices transported to the laboratory. The principal techniques used to separate volatile analytes from air in the field are cryogenic traps, impingers, and solid phase adsorbents.

1. Bulk Vapor Phase Samples

Flexible plastic sample bags are generally limited to vapor phase samples that can be analyzed within a short time after sample collection. This limitation is due to potential losses of analytes by surface adsorption and surface chemical reactions. Transportation of inflated bags over long distances to a remote laboratory is cumbersome, and can result in total losses of samples due to punctures and other accidents. Plastic bags are used in some laboratory operations—for example, to collect automobile exhaust and vaporized fuel—and for samples that can be conveniently transported to a laboratory.

Evacuated stainless steel canisters are widely used collection devices for ambient air samples. Sample canisters have smooth and inert internal surfaces and few or no active sites that adsorb volatile analytes or catalyze chemical reactions. The 1–6 L canisters are easily transported to a remote or a field laboratory. Canisters are leak tested and cleaned in the laboratory prior to use, evacuated to about 5×10^{-2} Torr or less, and transported to the sampling site where samples are taken by opening the sampling valve. Composite samples can be taken over time and/or space, and an in-line pump can be used to pressurize the container with either additional sample air or pure air if sample dilution is required. Pressurized samples are useful when longer term composite samples are taken or when larger samples are needed to lower detection limits.

Condensation of an entire air sample with liquid nitrogen or liquid helium has been used for many studies. However, this technique is expensive to implement and requires specialized portable equipment for handling cryogenic fluids in the field.

2. Separated Analytes

Cold trapping is used to separate volatile analytes from the main components of air in the field. Air is drawn by a pump through an inert, often nickel, metal tube immersed in a fluid at a very low temperature as, for instance, $-150°$. The tube may be packed with some inert material such as Pyrex glass beads, and the temperature is sufficient to condense most analytes but insufficient to condense oxygen or nitrogen. Plugging of the condensation tube with ice or other solids can be a major problem when sampling large volumes of humid air. This problem is addressed by the employment of air dryers that trap moisture but allow the nonpolar volatile analytes to pass into the cold stage of the trapping system. However, the more polar and water-soluble analytes are also removed by efficient air drying systems. Another potential problem is that some analytes may react during subsequent processing with trapped ozone, nitrogen oxides, or other substances present in the air. A practical limitation of the cold trapping technique is the requirement for liquid nitrogen or liquid argon in the field during extended sampling periods.

Impingers are used to extract various substances from vapor phase samples. An impinger is a closed glass or metal vessel with an inlet tube that extends to near the bottom of a liquid and an outlet tube well above the surface of the liquid. The impinger may contain various aqueous or nonaqueous liquids including solutions of derivatizing agents. Vapor is drawn by a pump into the inlet tube and bubbled through the liquid, which dissolves soluble analytes or the analytes react with reagents in solution to form soluble compounds. The liquid may also condense various substances, including water vapor, and collect fine particulate material that passes through a coarse filter or another separation device. A important advantage of an impinger compared to some bulk air sampling techniques is that hundreds of liters of air can be drawn through the device over a period of several hours. Sampling trains with multiple collection devices in series are used to collect different fractions of a vapor phase sample. Components of sampling trains can include particulate filters of several types and sizes, particle-size separation devices, multiple impingers, and solid phase adsorbents.

Solid phase adsorbents are simple and inexpensive devices used to separate volatile analytes from the principal components of air in the field. A porus solid phase adsorbent is placed in a glass or metal tube that is taken to the field where air is drawn through the adsorbent to trap the analytes. The adsorbent tube is then sealed and returned to the laboratory for analysis. Many types of solid phase adsorbents are used including alumina, activated carbons of various types, charcoal, graphitized carbon black (Carbopack B and C), carbon molecular sieves, GC packing materials such as Chromosorb 101 and 102, ethylvinylbenzene–divinylbenzene copolymer (Porapak Q), styrene–ethylvinylbenzene–divinylbenzene terpolymer (Porapak P), silica gel, 2,6-diphenyl-*p*-phenyleneoxide (Tenax-GC), styrene–divinylbenzene copolymers (XAD-1, XAD-2, and XAD-4), and acrylic ester polymers (XAD-7 and XAD-8). Adsorbent trapping is applicable to a wide variety of nonpolar and some polar volatile analytes.

Some adsorbents do not bind analytes strongly, which is an advantage for subsequent processing, but may result in the gradual release of some volatile analytes during long vapor sampling periods. Long sampling periods are often required—for example with ambient air—because of generally lower analyte concentrations compared to vapor phase samples from stationary or mobile sources of air pollutants. Breakthrough of the more volatile polar and nonpolar analytes from the adsorbent is controlled by placing additional adsorbents in the tube or additional sampling tubes in a sampling train. Analytes adsorbed on solid phase materials may also react with substances in the air as, for instance, with nitrogen oxides and ozone, to produce products not in the original sample. Long air sampling times also risk oxidation of some adsorbents and the production of background substances and other artifacts. These potential problems are usually evaluated during analytical method development and controlled by using appropriate quality control procedures

V. PROCESSING OF SAMPLES BEFORE DETERMINATION OF THE ANALYTES

Some processing of samples before the determination of the analytes is often required to achieve optimum analytical method performance. Sample processing also defines exactly what elemental forms, compounds, ions, oxidation states, and physical phases are included in the determination of specific substances or in the total amount of an element or a group of related substances in a sample. The degree of sample processing that is either appropriate or needed depends on a variety of considerations including

- The chemical analysis strategy, that is, a total, target analyte, or broad spectrum determination
- The physical phases of the sample that are included in the determination
- Whether the samples will be analyzed in a remote or a field laboratory or with laboratory, mobile, or portable instrumentation
- The complexity of the sample matrix
- The number of target analytes
- The diversity of properties of the target analytes
- The nature and concentrations of analyte interferences
- The detection limits required
- The efficiency of the chromatographic separation used in the analytical method
- The selectivity of the chromatographic detector or the nonchromatographic measurement technique
- The willingness to risk severe contamination of a GC or HPLC column, or some other critical instrument component, by a concentrated or high background environmental sample

Some sample matrices, such as ambient air or drinking water, are analyzed without sample processing with some analytical methods. Volatile compounds in ambient air collected in canisters or volatile compounds trapped from ambient air on solid phase adsorbents are determined directly (Section VIA1). Similarly, the total concentrations of some elements in low turbidity drinking or surface water amples preserved at a $\text{pH} < 2$ can be determined without further sample processing (Section VIIA). Common inorganic anions, like Cl^-, Br^-, NO_3^-, etc., are determined in surface, ground, drinking, and some other water samples without sample processing. However, many sample matrices and analytical methods require at least some, and sometimes considerable, sample processing prior to the determination of the analytes.

Semivolatile organic analytes trapped on air filters or on solid phase adsorbents or in water, soil, and other solid samples are extracted with an organic solvent, or a solvent mixture, and the extracts are concentrated by evaporation of the solvent before the determination of the analytes. Solid samples are often extracted with the classical Soxhlet apparatus or with a variety of other techniques including several that use organic solvents at elevated temperatures and pressures. If a sample is highly concentrated, for example an industrial wastewater or a soil from a hazardous waste land fill, fractionation of the solvent extract before determination of the analytes is usually required. Fractionation is often accomplished with open column liquid chromatography or preparatory scale high performance liquid chromatography (Section VIA2). The determination of total elements in particulate matter collected on air filters, in most water samples, or in solids requires an acid digestion of the sample to ensure that the elements are in water-soluble chemical forms for the measurement. The vigor of the digestion also determines the recovery of the elements from sediment and other insoluble material.

Invariably, the analytical method developer is required to make compromises between the amount and complexity of the sample processing and the separating power, selectivity, and other attributes of the chromatographic or nonchromatographic determination. These compromises are often strongly influenced by the projected cost and time required for various method options and by the desired quality, detail, and reliability of the results. Major issues usually are the availability of laboratory or field equipment and instrumentation, the experience and skill of the staff in using the equipment, and other laboratory or field infrastructure required to complete the analyses of the samples. Most research and standard analytical methods contain many compromises that may not be clearly defined in the method description, but should be understood by the user.

VI. CHROMATOGRAPHIC ANALYTICAL METHODS

Nearly all contemporary analytical methods designed for specific compounds or ions in environmental samples employ some type of chromatographic separation. Gas chromatography is the dominant technique for the separation of volatile and semivolatile analytes (Figs. 1 and 2 and Table I). Reverse phase high performance liquid chromatography is used for the separation of nonionic analytes that are thermally unstable, nonvolatile, or reactive and not amenable to GC (Fig. 3). High performance ion exchange chromatography and capillary electrophoresis are used for the separation of ionic substances. Analytes, either fully or partially separated, are sensed by in-line chromatographic detectors that produce electronic signals that are usually converted into digital form and stored in computer data systems. The computer systems, which are used with nearly all contemporary chemical analytical instrumentation, typically have software to control the operation of the instrument, acquire and store raw data, and reduce the data to more usable analytical information.

A. Chromatographic Separation Techniques

1. Gas Chromatography

The basic components of a gas chromatograph are a carrier gas system, a column, a column oven, a sample injector, and a detector. Very pure helium is the near universal carrier gas for environmental and many other analyses. Open tubular GC columns are constructed of fused silica with low-bleed stationary phases of varying polarity chemically bonded to the silica surface. Columns are typically 30–75 m in length and have inside diameters (ID) in the range of about 0.25–0.75 mm. The column oven is capable of precise temperature control and temperature programming at variable rates for variable times.

Analytes are introduced into GC columns with several techniques. An aliquot of a relatively concentrated vapor or air sample—for example, from a plastic bag or a canister—can be introduced into a short section of tubing of known volume, called a sample loop, and subsequently purged with carrier gas into the GC column. Volatile analytes in ambient air samples in a canister or trapped on a solid phase adsorbent are usually concentrated and focused in a cryogenic trap or a secondary adsorbent trap, then thermally vaporized into the GC carrier gas stream. However, in some analytical methods, volatiles trapped on an absorbent are thermally desorbed directly into the GC column. Aliquots of organic solvent extracts from various aqueous and solid samples are usually injected with a syringe into the carrier gas stream in a heated injection port. Both manual and automated syringe injection systems (autoinjectors) are used and the latter are generally very reliable, precise, and have the capacity to process many samples unattended.

Mixtures of analytes are separated by repeated equilibrations between the vapor state, where the analytes are entrained in the flowing carrier gas toward the detector, and the absorbed or dissolved state, where they are attracted to the stationary phase on the wall of the column by generally weak molecular forces. Analytes that tend to favor the absorbed state move more slowly through the column to the detector than do analytes that favor the vapor state. Complete separations of a few nanograms (10^{-9}) or less of each of 40–50 or more analytes in 30 min or less is not uncommon.

2. High Performance Liquid Chromatography

The basic components of a high performance liquid chromatograph are a high pressure mobile phase delivery system, a metal column packed with fine particles containing the stationary phase, a sample injector, and a detector. A high pressure pump is used to force the mobile phase solvent or solvent mixture through the packed column. The term *high performance* is often used to distinguish this technique from open-column liquid chromatography conducted at atmospheric pressure with gravity flow of the mobile phase. Columns vary in diameter from <1–5 mm ID, or larger, and from a few cm in length to 30 cm or more. Column packings consist mainly of silica particles, usually 3–10 μm in diameter, coated with low-bleed stationary phases of varying polarity chemically bonded to the silica surface. Organic polymers are also used as stationary phases in analytical separations. Automated or manual syringe injections of aliquots of liquid samples and known-volume sample loops are used for sample injection.

The dominant HPLC technique is the reverse phase configuration in which a nonpolar or slightly polar stationary phase is used with a more polar mobile phase that is often water, methanol, acetonitrile, or mixtures of these solvents. The normal phase configuration, which was developed before reverse phase, uses a more polar stationary phase and a nonpolar mobile phase, but it is not often used in contemporary HPLC. A vast array of organic compounds containing a variety of functional groups are retained on nonpolar or slightly polar stationary phases, and are sufficiently soluble in more polar mobile phases to give excellent separations. Analytes are retained on the column by an equilibrium process in which the dissolved molecule is alternately associated with the stationary phase, through weak noncovalent bonding interactions, and the mobile phase where it is transported toward the detector.

Isocratic elution is the use of a mobile phase that has a constant composition throughout the elution of analytes from the column. Gradient elution is the gradual changing of the mobile phase composition as the analytes elute from the column. Gradient elution is frequently used to enhance analyte resolution and shorten the time required for a separation. The pH of typical mobile phases can be adjusted over a reasonable range to improve resolution and selectivity. Various buffer substances, salts, and ion pair reagents can be used to control pH and ionic strength, and to facilitate the separation of some analytes. Reverse phase HPLC is more complex than GC because of the number of operational parameters including column dimensions, column packings, mobile phases, gradient elution, and various mobile phase additives.

3. High Performance Ion Exchange Chromatography

Ion exchange is another form of HPLC that uses a stationary phase consisting of a cross-linked synthetic organic polymer, often called a resin, with —SO_3H or —NH_2 groups attached to phenyl or other aromatic rings on the polymer backbone. Mobile phases are usually water or water and a miscible organic solvent. In basic solution the SO_3H groups of a cation exchange resin are ionized and consist of $—SO_3^-$ groups and associated counterions, for instance, Na^+. Cationic analytes are retained on the cation exchange column by displacing the resin counterions in an equilibrium process. The cationic analytes are subsequently eluted with a mobile phase containing a high concentration of counterions or counterions of a higher charge, for example, Ca^{2+}. Isocratic and gradient elutions are used with gradients in both solvent composition and counterion concentration. Mobile phase pH is a very important operational parameter that can have a significant impact on the retention or elution of various ionizable analytes from a resin.

In an acid solution the amino groups of an anion exchange resin are protonated and exist as $—NH_3^+$ groups and associated counterions, for instance, Cl^-. Analyte anions are retained on the column by displacing the resin counterions in an equilibrium process. The anion analytes are subsequently eluted with a higher concentration of counterions or some type of gradient elution. Anion exchange chromatography is widely used for the separation of inorganic anions and chelated metal anions in aqueous samples. With some natural or industrial water samples that contain high concentrations of ions such as Na^+, Ca^{2+}, Mg^{2+}, Cl^-, and SO_4^-, cation and anion analytes may not be retained on the column, and this can result in poor analyte recoveries.

4. Other Chromatographic and Related Techniques

Supercritical fluid chromatography (SFC), capillary electrophoresis (CE), and several related separation techniques are occasionally used in environmental chemical determinations. The CE technique is very important in biochemistry and molecular biology because of the very high resolving power that can be achieved, the high speed of separations compared to HPLC, and its ability to separate charged species such as proteins, peptides, and deoxyribonucleic acid fragments. These techniques have considerable potential for environmental analyses and may emerge as very important techniques in the future.

B. Chromatography Detectors

On-line detectors for GC, HPLC, and other chromatographic separation techniques are conveniently divided into three general classes:

- Nonselective detectors
- Selective detectors
- Spectroscopic detectors

Nonselective detectors respond to most or all changes in the composition of the carrier gas or mobile phase, and are capable of detecting nearly all entrained or dissolved analytes. The GC flame ionization and thermal conductivity detectors and the HPLC refractive index detector are examples of nonselective detectors. Nonselective detectors provide no direct information about the identity of the analytes except the time of arrival, the time of maximum concentration, and the time of return to pure carrier gas or mobile phase. Thus the peak shape of the analyte is usually well-defined, the measurement of peak area or height can be precise, but the information needed to identify the analyte that caused the change is weak. Identifications are based on comparisons of measured retention times of separated analytes with retention times of expected analytes measured under the same chromatographic conditions. While measured retention times can be very precise, especially with fused silica capillary GC columns and multiple internal standards, there is a high probability of coelution of two or more analytes in most environmental samples. This is because the peak capacity of a chromatogram, which is the number of analytes that can be fully separated, is limited. For most environmental samples the peak capacity is much smaller than the number of possible analytes that respond to a nonselective detector. Therefore, there is a significant risk of misidentifications or a false positives.

Selective detectors respond to only certain classes of analytes and they are often used in environmental chemical determinations. The electron capture, photoionization, electrolytic conductivity, and flame photometric GC detectors are selective for limited groups of analytes. Selective HPLC detectors include electrochemical detectors and the single wavelength ultraviolet-visible (UV-VIS) absorption and fluorescence detectors. Preinjection or postcolumn on-line chemical reactions are used to convert nonresponsive analytes into derivatives that respond to a selective detector. The information produced by a selective detector about the nature of the analyte is improved, compared to the nonselective detector, but still is not strong because the most significant piece of information obtained is the retention time. Although the range of potential analytes that respond to a selective detector is much narrower than the range that responds to a nonselective detector, multiple potential analytes still can have the same retention time. The nature of the sample is a factor is considering whether multiple analytes could be present that may have the same retention time.

Spectroscopic detectors measure partial or complete energy absorption, energy emission, or mass spectra in real time as analytes are separated on a chromatography column. Spectroscopic data provides the strongest evidence to support the identifications of analytes. However, depending on the spectroscopic technique, other method attributes such as sensitivity and peak area measurement accuracy may be reduced compared to some nonselective and selective detectors. The mass spectrometer and Fourier transform infrared spectrometer are examples of spectroscopic detectors used on-line with GC and HPLC. The diode array detector, which can measure the UV-VIS spectra of eluting analytes is a selective spectroscopic detector because only some analytes absorb in the UV-VIS region of the spectrum. The mass spectrometer, which can use a variety of ionization techniques, is probably the most widely used GC and HPLC detector for environmental chemical determinations. All detectors respond, to some extent, to natural background substances in environmental samples and sample extracts. Background chemical noise is highest in the most contaminated environmental samples and at the highest instrument sensitivities, but spectroscopic detectors are best equipped to distinguish chemical noise from environmental analytes.

VII. NONCHROMATOGRAPHIC ANALYTICAL METHODS

Analytical methods that do not employ a chromatographic separation of analytes are widely used for some environmental determinations. These methods generally depend on highly selective physical measurements or selective chemical reactions. Atomic absorption spectrometry (AAS), atomic emission spectrometry (AES), and mass spectrometry (MS) are most often used for determinations of the elements, but electrochemical and colorimetric techniques are sometimes used. Derivatives of analytes, sometimes called complexes, that strongly absorb in the UV-VIS spectrum are employed in colorimetric methods for the elements and some compounds and multielement ions. Ozone, nitrogen oxides, and some other substances in air are determined by selective chemical reactions that produce measurable light emission (chemiluminescence). Fluorescence spectrometry is used in some methods, especially with fiber optics technology, for remote monitoring. Immunoassays have been developed for a few analytes and they are especially useful in field analyses.

A. Elemental Analysis

Atomic emission and atomic absorption techniques are highly selective because the number of elemental analytes is small, and each has a unique atomic spectrum. Frequencies are selected for measurements to maximize selectivity and sensitivity, minimize interferences, and correct for interferences when necessary. Mass spectrometry measures mass to charge ratios of elemental ions in the gas phase and also provides high selectivity, sensitivity, and the ability to correct for interferences when necessary. Colorimetric and electrochemical techniques were used more frequently for elemental determinations before the development of the three major spectroscopic techniques. Their use has decreased significantly because of inherent limitations in selectivity, sensitivity, susceptibility to interferences, and the need for chemical processing to prepare light absorbing derivatives.

1. Atomic Absorption

During the 1960s flame AAS became the dominant technique for determinations of the elements in low concentrations in water samples and aqueous extracts of other samples. Aqueous sample aerosols are injected directly into the flame and precise determinations of many elements can be made. However, instrument detection limits for some important elements, for example As, Cr, Pb, Sb, Se, and Tl, in water and other matrices are $>50\ \mu$g/L, and often much greater, and AAS is limited to measuring one element at a time. The high temperature graphite furnace sample introduction system provided instrument detection limits of $<5\ \mu$g/L for most elements, and allowed automation of sample processing, but was still limited to measuring one element at a time. Other AAS sample introduction techniques, especially chemical reduction and elemental Hg vaporization (cold vapor) and conversion of As and Se to volatile hydrides, provided similarly low detection limits. The graphite furnace and Hg cold vapor techniques are often referred to in the scientific literature as flameless AAS methods.

2. Atomic Emission

Elemental analysis with AES had been practiced since the mid-1930s using flames and arc or spark discharges to vaporize and atomize samples and excite the atoms for optical emission. The AES technique provided rapid simultaneous or sequential multielement determinations, but the flames and arc or spark discharges had significant limitations. Interferences from electrode and other sample components were not uncommon, elemental measurements were often imprecise, detection limits were not sufficiently low, and liquid samples were difficult to analyze. The increasing demand for rapid, selective, and sensitive multielement determinations led to the development of the inductively coupled plasma (ICP) sample vaporization, atomization, and excitation source that met the analytical requirements and gave new life to AES. The ICP allowed the direct injection of aqueous aerosols into a 5500–8000 K argon ion plasma and elemental measurements with minimal or no interferences from background components or other analytes. Analytical methods utilizing ICP/AES are widely used for the determination of multiple elements in environmental samples.

3. Mass Spectrometry

The ICP technique also produces gas phase elemental ions and the ICP was quickly adapted as an ion source for mass spectrometry. This combination became one of the most useful and important techniques for rapid multielement analyses of gases, liquids, and solids. The argon ICP is a very efficient ion source that produces mainly singly charged ions. It is estimated that 54 elements, all metals, are ionized with 90% or greater efficiency. Only C, H, N, a few electronegative elements, and the noble gases are ionized with efficiencies less than 10%. A small number of elements—for example, As, B, Be, Hg, I, P, S, Se, and Te—have estimated ionization efficiencies in the 10–90% range. Although Ar has a low ionization efficiency, it is present in the ion source in great quantity and gives a few significant ions at m/z 40 ($Ar^{+\cdot}$) and m/z 80 ($Ar_2^{+\cdot}$). Other background ions, such as m/z 41 (ArH^+) and m/z 56 ($ArO^{+\cdot}$), are formed from sample components, usually water, and mineral acids that are used to ensure dissolution of some analytes. These ions obscure ions from some elements, but techniques are available to circumvent most of these interferences.

B. Organic and Inorganic Compounds and Ions

Colorimetric determinations are selective for specific substances—for example, the PO_4^{-3} ion (Section IIB)—because of selective color-forming chemical reactions between reagents and target analytes. Several of these techniques continue to be important for the determination of a few inorganic anions. However, ion exchange chromatographic techniques are increasingly used for determinations of multiple ions in environmental samples. Some electrochemical techniques, like ion selective electrodes, are selective because the materials of construction, such as ion selective membranes, and operating parameters are carefully chosen so the devices respond only to specific analytes. Ion selective electrodes are widely used for the determination of pH and a few other inorganic ions, for example F^-. Immunoassay methods, which are very widely used and enormously important in clinical analyses, have been developed for a small number of organic compounds—for instance, several triazine pesticides. However, many chemicals of environmental interest have MWs and shapes that are too small and similar for effective application of immunochemical techniques.

The further development of nonchromatographic analytical methods based on these and other techniques has been impeded by fundamental limitations in selectivity, and sometimes sensitivity, for a wide variety of similar substances—for example, those in Table I and Figs. 1–3. The cost of developing, testing, and documenting a large number of specialized methods, often for just one or a few analytes, is generally prohibitive compared to the cost of chromatographic methods that allow the determination of multiple generally similar analytes in an environmental sample.

SEE ALSO THE FOLLOWING ARTICLES

ANALYTICAL CHEMISTRY • ENVIRONMENTAL GEOCHEMISTRY • ENVIRONMENTAL OBSERVATION AND FORECASTING SYSTEMS • ENVIRONMENTAL TOXICOLOGY • GAS CHROMATOGRAPHY • LIQUID CHROMATOGRAPHY • POLLUTION, AIR • POLLUTION CONTROL • POLLUTION, ENVIRONMENTAL • SOIL AND GROUNDWATER POLLUTION • WATER POLLUTION

BIBLIOGRAPHY

“Annual Book of ASTM Standards, Section 11, Water and Environmental Technology,” American Society for Testing and Materials, West Conshohocken, PA.

Bloemen, H. J. Th., and Burn, J., eds. (1993). “Chemistry and Analysis of Volatile Organic Compounds in the Environment,” Blackie Academic & Professional, London.

Budde, W. L. (2001). “Analytical Mass Spectrometry: Strategies for Environmental and Related Applications,” Oxford University Press, New York.

Clesceri, L. S., Greenberg, A. E., and Eaton, A. D., eds. (1998). “Standard Methods for the Examination of Water and Wastewater,” 20th ed., American Public Health Association, Washington, DC.

Keith, L. H., ed. (1996). “Principles of Environmental Sampling,” 2nd ed., American Chemical Society, Washington, DC.

Parker, L. V. (1994). “The effects of ground water sampling devices on water quality: A literature review,” *Ground Water Monitoring and Remediation* **14,** 130–141.

Schwarzenbach, R. P., Gschwend, P. M., and Imboden, D. M. (1993). “Environmental Organic Chemistry,” Wiley, New York.

Environmental Observation and Forecasting Systems

António M. Baptista
Oregon Graduate Institute of Science and Technology

GLOSSARY

Circulation forecasts Predictive description of circulation processes into the future, through numerical modeling.

Circulation hindcasts Retrospective description of circulation processes, through numerical modeling.

Environmental Information Technology (EIT) Ensemble of concepts and technologies designed to deliver quantifiably reliable environmental information at the right time and in the right form to the right users.

Environmental Observation and Forecasting Systems (EOFS) EIT tools that combine real-time sensor measurements with advanced numerical models to describe, with progressively increasing reliability and detail, critical aspects of the dynamics of complex environmental systems.

Numerical modeling The process of integration of numerical codes, system topology, forcings and quality controls toward the description of a physical or environmental process.

Visualization The process leading to the development of a mental or graphical description of static or dynamic processes.

PARADIGMS in modeling and visualization of environmental systems are shifting dramatically. Modeling is no longer perceived only as the rigorous but narrow process by which a set of differential equations and associated boundary conditions are transformed into systems of algebraic equations and solved by a computer. Visualization is no longer oriented primarily to specialists or restricted to a particular medium, and aesthetic sophistication is increasingly less important than contents and timeliness.

Modern modeling and visualization paradigms are being driven by the need to integrate quantifiably reliable scientific information in decision making, in operational procedures and in cross-disciplinary research, as well as by the opportunities resulting from fast-evolving computer, sensor, and information technologies. Environmental Observation and Forecasting Systems (EOFS) provide an excellent illustration of this influence. In EOFS, models produce critically important information contents, and visualization provides essential interfaces between information and multiple end users. But both modeling and visualization are integrated in an Environmental Information Technology (EIT) infrastructure, whose ultimate goal is delivering the right information to the right user, in the right format, at the right time and with bounded uncertainty. This article discusses resulting benefits, requirements and unresolved challenges.

I. INTRODUCTION TO ENVIRONMENTAL OBSERVATION AND FORECASTING SYSTEMS

The concept of systematic, human-computed weather forecasting dates back to the start of the twentieth century, when British meteorologist Lewis Richardson (1922) proved to be ahead of the technology of his time: it took him three months to predict weather for the next 24 hr. But the seed was planted, and with the post-World War II advent of electronic computers, systematic weather forecasting became a practical tool.

Arguably the original form of EOFS, weather forecasting is a pervasive commodity of modern life, influencing decisions as simple as an individual's choice of the day's clothing or as complex as the setting of regional energy strategies. In weather forecasting, time-sensitive data gathered from sensors or generated through computer models are integrated and interpreted toward a description of present and future atmospheric conditions. Established visual or audio interfaces routinely translate such description to both the general public and customized audiences such as farmers or mariners, leaving ample opportunity for the additional development of on-demand interfaces. Data and products are generated or flow through a worldwide network of information systems, managed by a disparate array of public and private entities.

Only in the later part of the twentieth century was the vision of systematic forecasting extended to other environmental systems. However, both society needs and advances in computer power and connectivity have in recent years fueled decisive advances in the development of real-time EOFS, in particular for ocean, coastal ocean, estuarine, and river systems (Parker, 1997; Baptista, 1999; Long, 1996; Glenn, 2000; Robinson, 1999). As they mature and self-organize, these EOFS are revolutionizing the way scientists share information about the environment and represent an unprecedented opportunity to break traditional information barriers between scientists and society at large.

EOFS have enormous potential for social payoff, and as with weather forecasting, success will lead to increased user expectations. As an example, the Marine Prediction Center of NOAA regularly generates forecast products for mariners, extending over large regions of the Pacific and Atlantic oceans, and describing combined atmospheric and sea conditions (e.g., Table I). These products fulfill U.S. responsibilities with the World Meteorological Organization and Safety of Life at Sea Convention, and have had direct daily impact on safety at sea, protection of life and property, and enhancement of economic opportunity. Yet, such products are often no longer considered sufficient. Coordinators of marine search and rescue operations or of oil spill response teams, for instance, seek to have detailed, on-demand information on marine currents and winds wherever and whenever an accident might occur, aided as needed by real-time computer simulations to narrow the search region or to track the spill. The payoff will be in additional lives saved, more effective ecosystem protection, and reduced operational costs. The necessary capabilities, including information quality control and uncertainty analysis, are within the technical reach of the next generation of EOFS.

Next-generation EOFS will also change the timing and subjectivity of regional environmental management decisions. For instance, decision makers contemplating changes in hydropower management strategies in a regulated system like the Columbia River (e.g., Fig. 1), will expect at least some of the impacts of alternative scenarios to be compared quickly and on demand, in preparation for and during their deliberations. These impacts are currently investigated through typically lengthy and often controversial studies. If EOFS are implemented with a long-term perspective and appropriate regional consensus, the payoff will be both in the timeliness of critical information, and in focus on what is indeed ambiguous or open to social choice. This focus is urgently needed: controversy is too often extended to aspects of the supporting science that are fundamentally unambiguous, preventing

TABLE I Message Issued by the Marine Prediction Center–Marine Forecast Branch of the U.S. National Weather Service[a]

MARINE INTERPRETATION MESSAGE FOR NORTH PACIFIC OCEAN
NATIONAL WEATHER SERVICE WASHINGTON DC
MARINE PREDICTION CENTER/MARINE FORECAST BRANCH
1203 PM PST MAR 4 2001

FORECAST DISCUSSION: MAJOR FEATURES/WINDS/SEAS/SIGNIFICANT WEATHER FOR NORTH PACIFIC N OF 30N AND E OF 150W.

HIGHLY AMPLIFIED PATTERN WILL CONTINUE ACROSS THE EPAC OVR THE NEXT COUPLE OF DAYS. UPR RIDGE WILL BUILD ACROSS ERN GLFAK DURING THE NEXT 48 HRS . . . WHILE A CLOSED UPR LOW WILL DIG SE OVR THE CENTRAL AND SOUTHERN CA WATERS. ANOTHER STRONG TROF IS ADVERTISED BY ALL THE MODELS TO MOVE INTO THE WRN PORTION OF THE MIM AREA AROUND 48 HRS. THE NOGAPS AND UKMET ARE QUICKER TO CLOSE THIS SECOND SYSTEM OFF . . . THEN THE OTHER MODELS . . . TOO QUICK IN MY OPINION. EVENTUALLY IT DOES CLOSE OFF DURING DAY 4 . . . JUST W OF THE N CA WATERS . . . WHILE AN UPR RIDGE MOVES INTO THE WRN PORTION OF THE MIM AREA.

A SFC LOW WILL CONTINUE TO MOVE ACROSS THE N WATERS TODAY . . . THEN BE ABSORBED INTO ANOTHER SYSTEM TO THE S. GALE CONDITIONS ARE EXPECTED TO CONTINUE TODAY . . . THEN WEAKEN AS THE LOW IS ABSORBED SOUTHWARD. THE SOUTHERN LOW COULD PRODUCE GALE CONDITIONS IN THE CENTRAL WATERS. THE WW3 CONTINUES TO BE TOO WEAK WITH THE SEAS ACROSS EPAC. THERE IS A SHIP W OF THE CENTRAL WATERS RPT 26 FT . . . ABOUT 6 FT OR SO ABOVE THE MODEL.

WARNINGS . . . GALE . . . CAPE LOOKOUT TO PT. ST. GEORGE TODAY INTO TONIGHT.
PT. ST. GEORGE TO PT. ARENA TODAY INTO TONIGHT.
PT. ARENA TO PT. CONCEPTION TODAY.

FORECASTER SHAW. MARINE FORECAST BRANCH

[a] Through the regular issue of messages like this, marine forecast centers across the world have changed navigation practice and saved extensive life and property.

emphasis on the objective analysis of the relative social merits of possible alternatives.

To meet the evolving vision and functionality introduced above, next-generation EOFS will be large-scale, extendable, multipurpose systems shared among many diverse applications, and will have unprecedented ability to respond quickly on demand, and to progressively accumulate knowledge from application to application. Many of these systems will be integrated in national or worldwide networks (e.g., Fig. 2). Continued advances in computational and visualization methods and in computer power and connectivity, complemented by new paradigms for research, training, and information access, will be necessary to fulfill this vision.

II. A PILOT ESTUARINE ENVIRONMENTAL OBSERVATION AND FORECASTING SYSTEM

Estuaries and adjacent coasts are poster cases for the need for new paradigms and technologies for interdisciplinary research, education, and information access. Indeed, these are interface ecosystems of great physical and ecological complexity, boasting strong variability at multiple scales in space–time. Estuaries are often the focus of intense and conflicting human activity: navigation, resource harvesting, effluent disposal, recreation, energy production, etc. They are also natural integrators of characteristics and impacts of adjoining watersheds, airsheds, and continental shelves. “Pristine” estuaries may be unexplored indicators of global climate change trends and their biological implications. Coastal plumes extend the influence of freshwater inputs beyond the mouth of the estuary. In the case of river-dominated systems, such as the Columbia River, this influence extends for hundreds of miles along the coast (Fig. 1).

Effective management of the nation’s estuaries will rely progressively more on real-time and prognostic information on physical characteristics, environmental stressors, and ecological indicators. EOFS are important enabling systems for this vision. Pilot implementations (Parker, 1997; Baptista, 1999; Glenn, 2000) have demonstrated

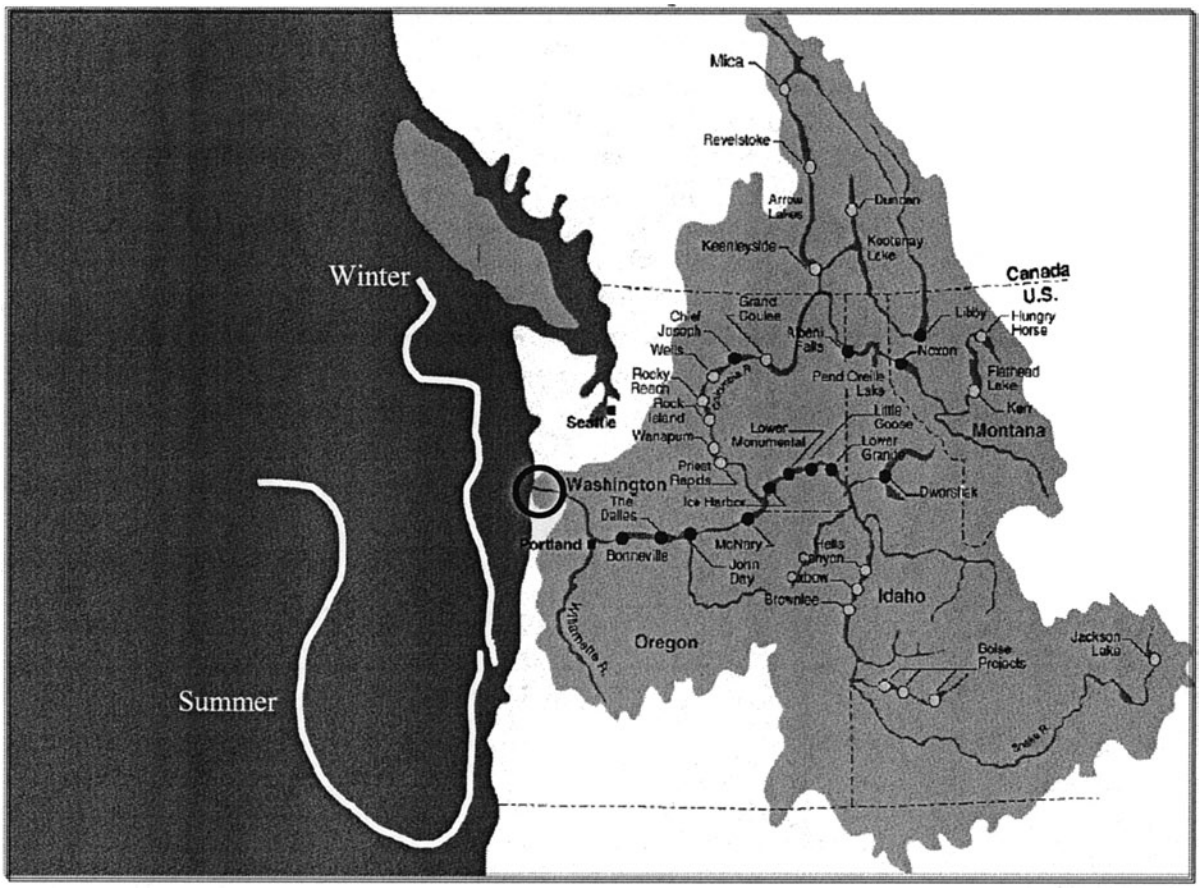


FIGURE 1 The Columbia River estuary (circle) is a small but important feature of the Columbia River system. This system features the second largest river discharges in the U.S., and experiences large tidal, seasonal, and annual variability. The contributing watershed spreads over multiple states/provinces in the U.S. and Canada. The river is highly regulated, with dams controlled by multiple agencies. The freshwater plume extends hundreds of miles into the ocean, often reaching northern California in the summer, and British Columbia in winter.

promising ability to combine real-time continuous data monitoring and computer modeling into detailed spatial and temporal descriptions of past, present, and future estuarine conditions. However, pilot EOFS are often limited by factors such as overspecialization, design rigidity, insufficient assessment of information quality and uncertainty, suboptimal monitoring, suboptimal trade-offs in accuracy/cost of forecasts, insufficient storage capacity, and high maintenance costs.

Intrigued by EOFS challenges and opportunities, the Center for Coastal and Land-Margin Research of the Oregon Graduate Institute initiated in 1996 the development of CORIE, a pilot system for the Columbia River estuary and plume. CORIE (Baptista, 1999; Baptista, 1998; REF?), was designed explicitly as multipurpose, multiuser regional infrastructure for science and management, thus differing from many other estuarine EOFS, which have historically been driven primarily by navigation concerns (Parker, 1997).

At the heart of CORIE are three integrated components (Fig. 3): a real-time observation system (Section III), a modeling system (Section IV), and an information management and visualization system (Section V). These three systems combine to automatically generate an array of primary products, including time-sensitive displays of sensor data and of computer forecasts, which are often openly distributed through the World Wide Web (henceforth, web). Data from sensors and models are also processed in customized fashion, either on- or off-line, for particular scientific, management, or operational applications (Section VI).

Modeling and visualization are integral aspects of CORIE, but special paradigms and requirements do apply. Modeling, observation, and information management and visualization are not disjoint components, conducted in isolation by separate teams. Rather, all components are closely integrated, requiring tight interaction among team members with different primary expertise: modelers are

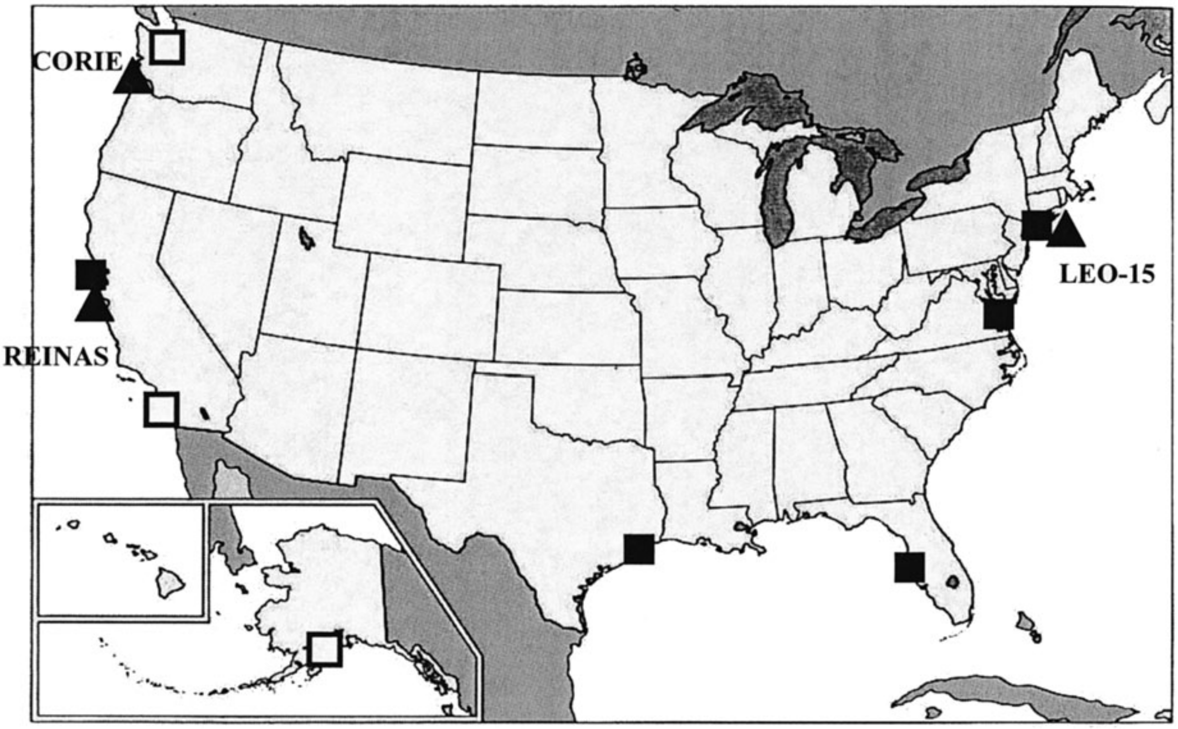


FIGURE 2 A network of estuarine and near-shore EOFS has "sprawled" over the U.S. coast. Several of these systems are part of the PORTS system, which is coordinated by NOAA (black squares if already installed, hollow squares if planned). CORIE, LEO-15, and REINAS (black triangles) are, however, examples of EOFS independently maintained by universities or university consortia.

often in the field, and field personnel participate in the analysis of modeling results.

Visualization is prevalent across the information domain, often as a mechanism for fusion of multiple data sources, and typically as the culmination of a time-sensitive, end-to-end flow of data from information producers (e.g., models, sensors) to a diverse range of human users. Visual sophistication is increasingly less important than contents, robustness, and timeliness. Quantification of uncertainty is highly desirable, but difficult to automate, thus being often unsatisfactory in its detail or lagging in time.

Modeling is more than the rigorous but narrow process by which a set of differential equations and associated boundary conditions are transformed into systems of algebraic equations and solved by a computer. Indeed, the modeling process often requires automated acquisition of external forcings, automated comparison against or fusion with field data, and code execution within predefined schedules. Models are typically required to run at least 5–10 times faster than real-time. While model results can always be examined in customized manners, an array of visual products is generated automatically as a part of default forecasting procedures.

III. OBSERVATION NETWORK

The CORIE observation network consists of multiple estuarine stations and one offshore station, each with a variable number of sensors. Most stations measure water levels, temperature, and salinity; several stations measure velocity profiles; and a few measure wind. Sampling intervals are in the 1- to 15-min range, with a sample consisting of the average over a short period (e.g., 30 sec) of measurements made at several Hertz. Most stations have telemetry, typically based on spread-spectrum radio. Selected stations have field computers, enabling local storage and facilitating two-way communication.

By design, the configuration and instrumentation of the observational network has and will continue to evolve over time. While a backbone of anchor stations are considered permanent, other stations or subnetworks appear and disappear as dictated by scientific, engineering, or management needs and by available funding. For instance, we are currently implementing two interrelated extensions of the network (Fig. 4), which became time sensitive due to an unusually dry hydrological year and the potential for record low river discharges during the 2001 spring freshet season. The first extension will enable the characterization

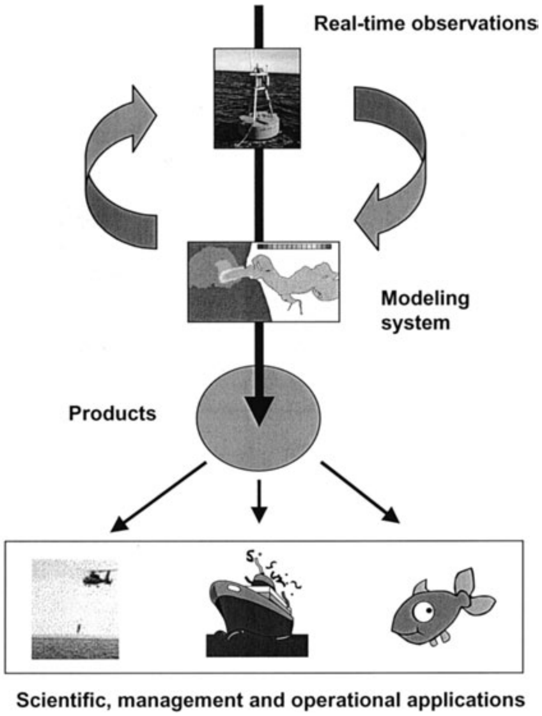


FIGURE 3 EOFS tightly couple real-time field observations with simulation models, and using the resulting model and sensor outputs to develop both generic, multipurpose products target at wide audiences, and customized products tailored to specific audiences.

of the vertical density structure in the two deep channels of the lower estuary, and is motivated primarily by the need for detailed validation of the CORIE circulation models. The second extension, concentrated on a lateral bay near the upstream end of salinity propagation, will provide anchor measurements of salinity, temperature, and pressure for third-party habitat surveys, in the context of investigations of the influence of the estuary on salmon survival.

Timely visualization of the data is an integral concept of CORIE, with deep implications for cost-effective system maintenance and for innovative data usage. For instance, the physical integrity of most stations, their power reserves, and the contents and quality of their data streams (e.g., Fig. 5) can be examined at least daily, enabling early detection of problems and their eventual resolution. Anyone, anywhere in the globe, can have free access to predetermined representations of most CORIE data within seconds of their collection, providing that they have web access. Reasons to want such access will vary: a researcher in an oceanographic vessel may use velocity, backscatter, and salinity data to track in real time the progression of estuary fronts or other estuarine features; a bar pilot may use water levels or velocities to plan a docking operation; a fisherman may look at salinity and temperature data to choose the day's fishing spot. Alternatively, users may choose to look at statistics and other forms of analysis of accumulated data. Timeliness is in this case measured not by time since collection, but by the time and/or effort required for obtaining the desired information in the desired form. An increasing number of CORIE products attempts to anticipate needs for access to processed data (Fig. 6).

Estuarine and offshore stations pose different logistical challenges, thus also eliciting different definitions of access timeliness. For instance, we are particularly concerned with the physical integrity of the offshore station, OGI01, which is often exposed to severe sea conditions during winter and is routinely exposed to ship traffic and seasonal fishing activities. Our telemetry strategy for this station consists of routine transmission, via satellite and at multiple times a day, of the location of the surface buoy where most instrumentation is mounted. Actual oceanographic data can be transmitted in the same fashion, but because of cost and bandwidth limitations, telemetry of these data is typically limited to periods where quasi real-time access is operationally valuable—e.g., to assist scientific cruises. For estuarine stations, inexpensive telemetry solutions based on spread-spectrum radio enable routine real-time transmission and web display of all the data being collected.

Associated with the notion of timely access is the challenge of timely quality control. A classic scientific paradigm is that of carefully extracting the most possible information from a limited data set, with modest consideration to the time that the analysis might take. CORIE and similar observation systems are adding a new paradigm: that of automated quality assessment of continuously flowing data streams from large arrays of sensors. We are finding that this challenge has to be addressed very differently for different sensors. Water levels, water temperature, and velocity profiles tend to be resilient data streams, for which failure modes are often abrupt and susceptible of early detection by automated quality assessment methods. By contrast, conductivity sensors are commonly subject to biofouling, often leading to slow, progressive signal degradation that mimics aspects of low-frequency natural variability (Fig. 7). Without novel methods for early detection of biofouling, automated quality assessment of conductivity (and by association, salinity) data will remain unsolved. Also, unsolved is the problem of eventual correction of biofouled records. Future methods to detect and correct biofouled data may involve identification of changes in temporal patterns of modeling error at sensor locations.

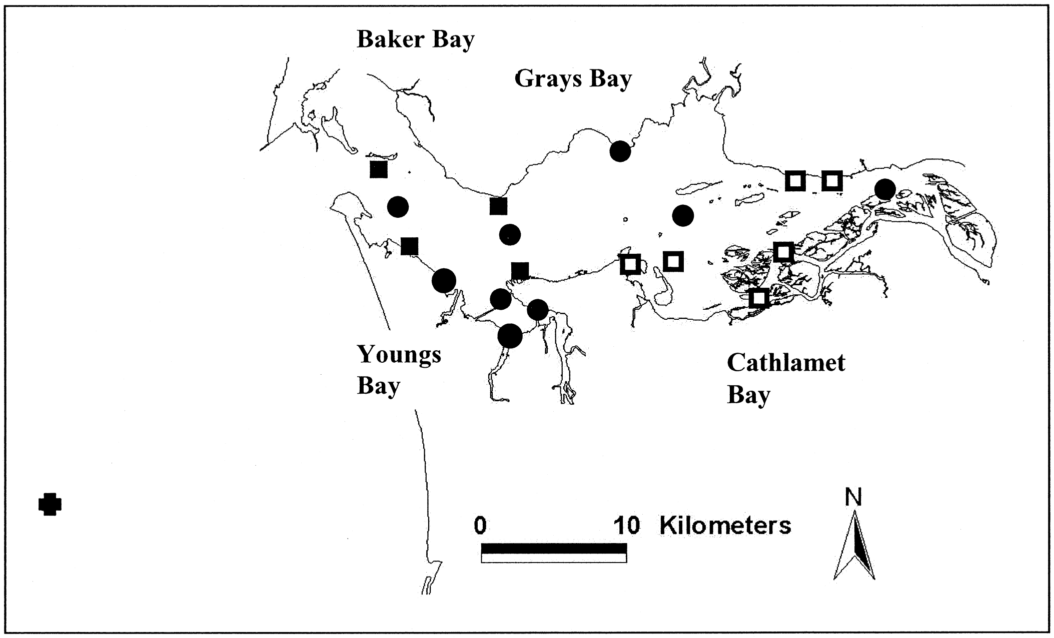


FIGURE 4 The stations of the CORIE observation network are concentrated on the estuary (circles and squares) up to the limit of the salinity propagation, with nominal presence offshore (cross). The network is an evolving system. For instance, four estuarine stations (black squares) are currently being modified to enable characterization of vertical stratification, and a subnetwork of stations (hollow squares) is being deployed at Cathlamet Bay to characterize physical habitat in support of research on salmon survival.

IV. MODELING INFRASTRUCTURE

Our approach to CORIE modeling reflects the philosophy that numerical modeling must, in the context of EOFS, be an end-to-end task. Our main objective is, at this stage of the development of CORIE, to produce (a) daily forecasts of circulation in the Columbia River estuary and plume, and (b) archival simulations of circulation, in a slightly larger domain, for modern, predevelopment, and future-scenario conditions.

Consistent with the above philosophy and objectives, we designed a computational infrastructure (Fig. 8) that includes five key components.

A. Numerical codes

Numerical codes solve for the governing equations describing an environmental process, given an appropriate set of topological inputs, initial conditions, and boundary conditions. All numerical codes used in the CORIE system are based on the solution of the two-dimensional (2D) or three-dimensional (3D) shallow water flow and/or transport equations, by either finite elements or finite volumes, on unstructured grids. Our design of the modeling system incorporates code redundancy as a mechanism of quality control, and similar purpose codes are conceptually interchangeable.

Table II identifies some of the main CORIE circulation codes, and highlights their differences. Because the modeling products include daily (and in the future, real-time) forecasts, and multiyear hindcasts, computational performance is very important: simulations that run about 10 times faster than clock time are a desirable target. Table III illustrates the computational efficiency of the codes that are currently used for generation of 2D and 3D circulation forecasts, respectively, and shows a clear improvement of computational performance: the most recent 3D code runs substantially faster than an older 2D code, using the same computer hardware.

The unstructured numerical grids used by CORIE circulation and transport models determine the spatial detail and much of the intrinsic numerical accuracy of the simulations. These grids are created with grid generation software developed specifically for estuarine and coastal applications. We have typically used semiautomatic grid generation (Turner, 1991), but recent developments (Zhang, 2000) suggest that fully automated, loosely adaptive generation may soon become the default procedure. The term "loosely adaptive" is used here to describe

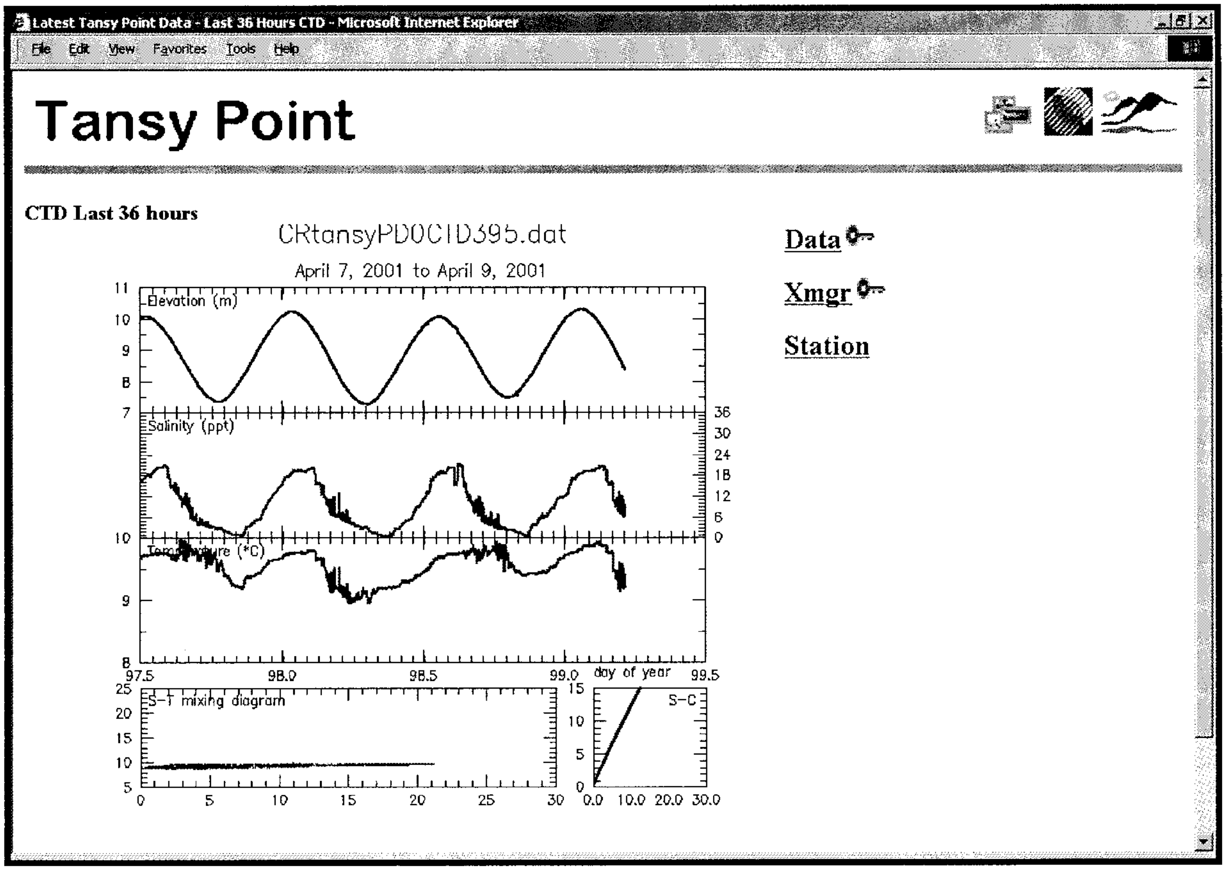


FIGURE 5 Data from CORIE sensors are displayed in real time on the web, using preselected representation formats. In the example, pressure, temperature, and salinity data for the 36 hr prior to a user-triggered request is shown for station Tansy Point. Direct access to the actual data, either in numerical or graphical form, is available to authorized users.

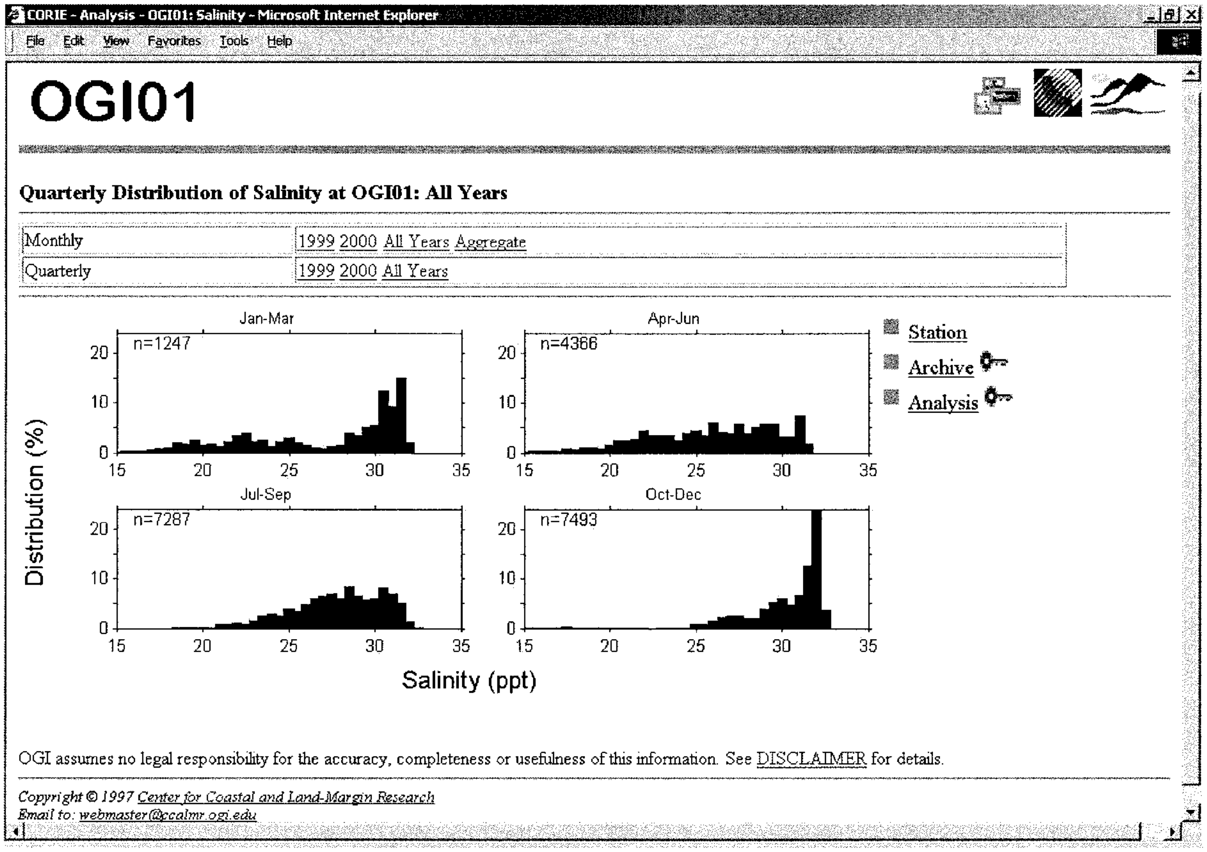
a process by which grids are optimized on a regular (e.g., seasonally) base, but not on every each time step.

B. Bathymetry

Modern and predevelopment bathymetries are substantially different in the Columbia River. Many of the changes have been imposed by human activity, in particular through (a) the installation of breakwaters to stabilize the entrance of the estuary, (b) the creation, maintenance and deepening of a 145-mile navigation channel, and (c) significant changes of upstream river discharge and sediment inputs via extensive flow regulation with a network of dams. In addition, bathymetry changes locally in response to events such as major storms, and also changes continuously, system-wide, in slow response to circulation processes.

Bathymetry is a key input to the numerical codes, and bathymetric uncertainties are responsible for a substantial part of the total uncertainty associated with model results. Dated survey data is very common in estuaries, in particular away from navigation channels. The last comprehensive, bank-to-bank bathymetric survey in the Columbia River dates back from the late 1950s. Channel bathymetry is surveyed multiple times per year by the U.S. Army Corps of Engineers, as a part of the agency's mandate.

Timely updates of bathymetry are becoming a critical limiting accuracy factor for CORIE modeling, and in general, for EOFS modeling. For channel bathymetry, the problem could technically be solved by automated downloads of the Corps of Engineers periodic bathymetric surveys, followed by semiautomatic procedures to load the new bathymetric data into the CORIE grids. In a foreseeable future, continuing advances in remote sensing and vessel-based survey techniques offer the best promise for timely updates of bank-to-bank bathymetry. The alternative of including a sediment transport module in the circulation codes, to enable the automated



OGI01

Quarterly Distribution of Salinity at OGI01: All Years

Monthly	1999 2000 All Years Aggregate
Quarterly	1999 2000 All Years

OGI assumes no legal responsibility for the accuracy, completeness or usefulness of this information. See DISCLAIMER for details.

Copyright © 1997 Center for Coastal and Land-Margin Research
Email to: webmaster@ccalmr.ogi.edu

FIGURE 6 Data from selected CORIE sensors is, after quality control, automatically aggregated in a variety of processed "data products." The example shown here is the seasonal climatology of salinities at the offshore station. Higher frequency of lower salinities differentiates the periods where the station is contained in the plume.

computation of bathymetry changes, is conceptually attractive but unlikely to become technically feasible in the short term, at useful accuracy levels.

C. Forcings

Circulation in the Columbia River estuary is primarily driven by tides, river discharges, and salinity- and temperature-induced density differences between freshwater and ocean water. In addition, wind and Eastern North Pacific Ocean currents are important factors in the dynamics of the Columbia River plume, and ocean-water heat exchange regulates temperatures in lateral bays in the estuary and at ocean scales. Arguably, current-wave interaction is also a significant factor at the entrance of the estuary.

No model of the Columbia River circulation will be realistic, in a multipurpose sense, without accounting for most of these various forcings. The CORIE infrastructure is designed to provide forcing conditions to both hindcast and forecast simulations. The latter simulations are most constraining, because of timing requirements. As illustrated in Fig. 8, we rely on a variety of sources, including various systems that would classify as EOFS, to derive the required forcing functions. In most cases, automatic scripts are responsible to capture the forcings from the source, and customize them for the CORIE codes.

D. Quality Controls

Field data provide natural controls for the quality of EOFS simulations. The sensors of the CORIE network of fixed, *in situ*, long-term stations (Section III) described are automatically compared against daily model forecasts, and are available for systematic comparison against long-term hindcasts. Long-term model-sensor comparisons, in either forecast or hindcast modes, shows exciting promise in advancing the state-of-the-art in identifying error sources and characterizing error trends. Application of pattern recognition techniques across the simulation database is, in particular, expected to reveal insightful correlations

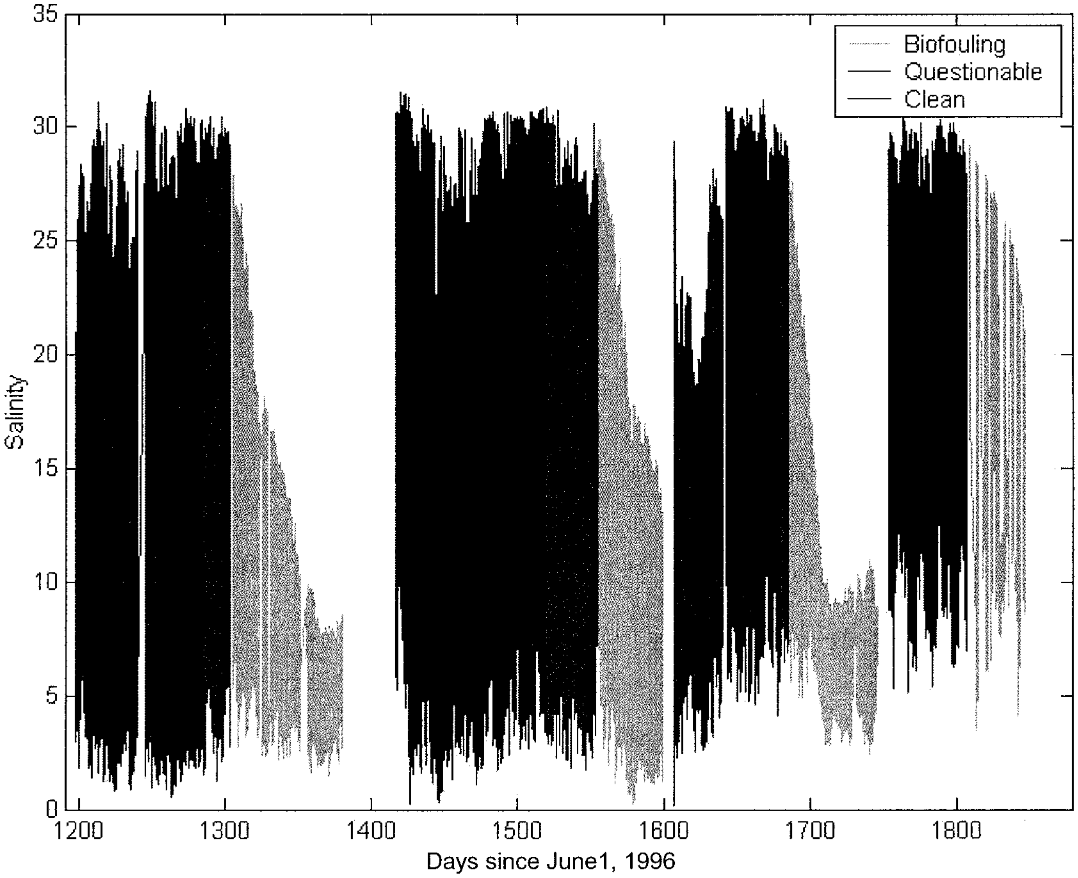


FIGURE 7 Conductivity sensors, used to characterize salinity, are severely affected by biofouling at times of active biological activity in the Columbia River estuary. The figure shows a long record of salinity data for the near-ocean station of Sand Island. A substantial percentage of the record is degraded, as seen by maximum salinities severely below ocean salinities. Research on the development of automatic classifiers of data quality shows substantial promise for early detection of the onset of biofouling. The prognosis is more reserved for recovery of biofouled data.

among model errors and indicators of internal dynamics or of variability of external forcings.

The broad range of scales of variability of estuarine processes requires that quality controls also include data types complementary to fixed *in situ* sensors. Of particular interest are data collected from on-board sensors during vessel surveys, and an increasing range of remote sensing techniques.

Vessel surveys are naturally amenable to collection of data along a vertical or transect, to capture spatial gradient scales that elude fixed sensors. We have recently installed a thermosalinigraph and a radio link as permanent, automated instrumentation in a 50 ft training vessel operated by the Clatsop Community College. The college uses the vessel for hands-on training of its students, typically three times a week during three quarters of the year. Every time this happens, transects of surface temperature and salinity data are collected and automatically added to the CORIE database. On selected occasions, class actually consists of performing vessel operations in support of scientific data collection. In those occasions, additional oceanographic instrumentation is added to the vessel (e.g., an acoustic Doppler current profiler, and a winch-operated assembly of conductivity, temperature and pressure sensors) to capture a broader range of variables and enable vertical profiling.

Remote sensing techniques are also particularly well suited to capture spatial gradients. Although with limitations in sampling frequency and subsurface penetration, remote sensing techniques have over vessel surveys the advantage of synopticity, higher dimensionality, and higher number of data points. Ongoing efforts are being directed to add a range of remote sensing products as quality controls for CORIE simulations. In particular, a X-band radar is being evaluated as possible permanent CORIE instrumentation, to detect surface manifestations of density fronts in the estuary entrance. High-resolution third-party Synthetic Aperture Radar (SAR) images will be explored as a tool for evaluation of near-field plume simulations. Low-resolution National Oceanic and Atmosphere

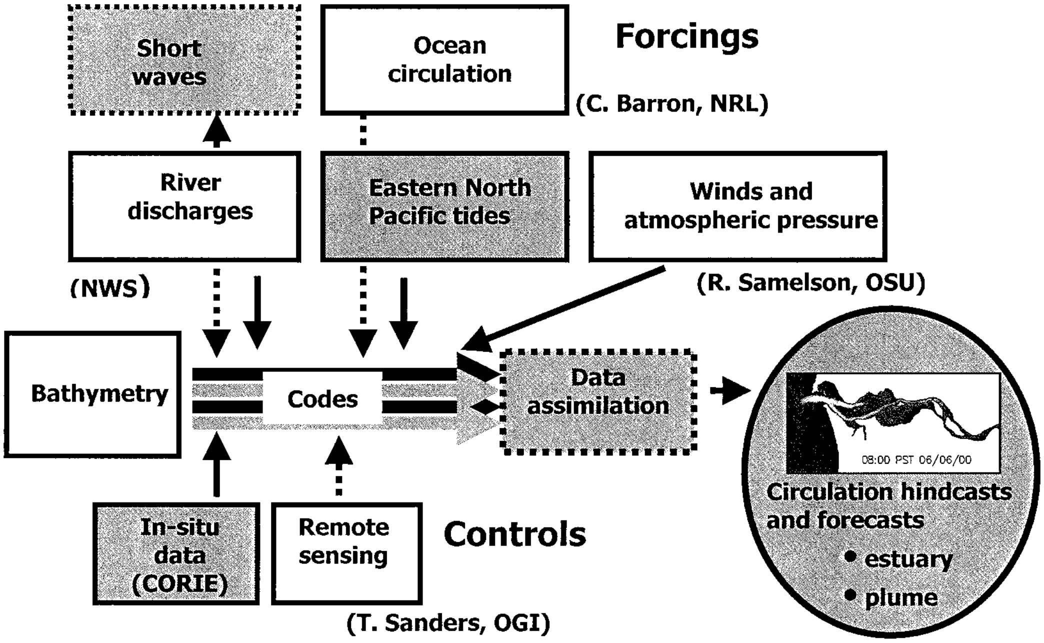


FIGURE 8 In addition to numerical codes, the modeling infrastructure of CORIE includes modules that specifically deal with forcings (bathymetry and river, atmospheric and ocean forcings), quality controls, and modeling products. While some modules and codes (shown in gray) are primarily developed in-house, many are developed externally and invoked as needed.

Administration/National Aeronautics and Space Administration (NOAA/NASA) Advanced Very High Resolution Radiometers (AVHRR) sea surface temperature data are being used to evaluate far-field plume simulations.

E. Modeling Products

CORIE modeling products can be divided in three major types: daily forecasts of circulation, long-term circulation hindcast database, and derivative products. A subset of the latter products is discussed in Section VI.

Daily forecasts of depth-averaged barotropic circulation have been produced regularly since 1997 (Fig. 9). These simulations ignore the effect of density gradients on circulation, and output variables are restricted to water levels and velocities. Experimental forecasts of 3D baroclinic circulation have been conducted since late 2000, and are expected to enter an operational stage in 2001. These

TABLE II Distinguishing Characteristics Among the Main Circulation Codes Used in CORIE

	ADCIRC[a]	QUODDY[b]	ELCIRC[c]
Shallow water equations	2D	3D	3D
Baroclinic terms	No	Yes	Yes
Wetting and drying	Yes	No	Yes
Unstructured horizontal grid	Yes	Yes	Yes
Vertical representation	N/A	σ coordinates	z-coordinates
Finite elements or finite volumes	FE	FE	FV
Continuity wave or primitive equations	CWE	CWE	PE
Advection treatment	Eulerian	Eulerian	Lagrangian

[a] Developed by Luettich (1991), ADCIRC is the base code for the CORIE forecasts of depth-averaged circulation. A 3D version of the code exists.

[b] Developed by Ip (1995), QUODDY was used in early CORIE hindcasts of 3D baroclinic circulation.

[c] Developed by Myers (2000), from the formulation proposed by Casulli (1998), ELCIRC is currently the base code for 3D CORIE forecasts and hindcasts.

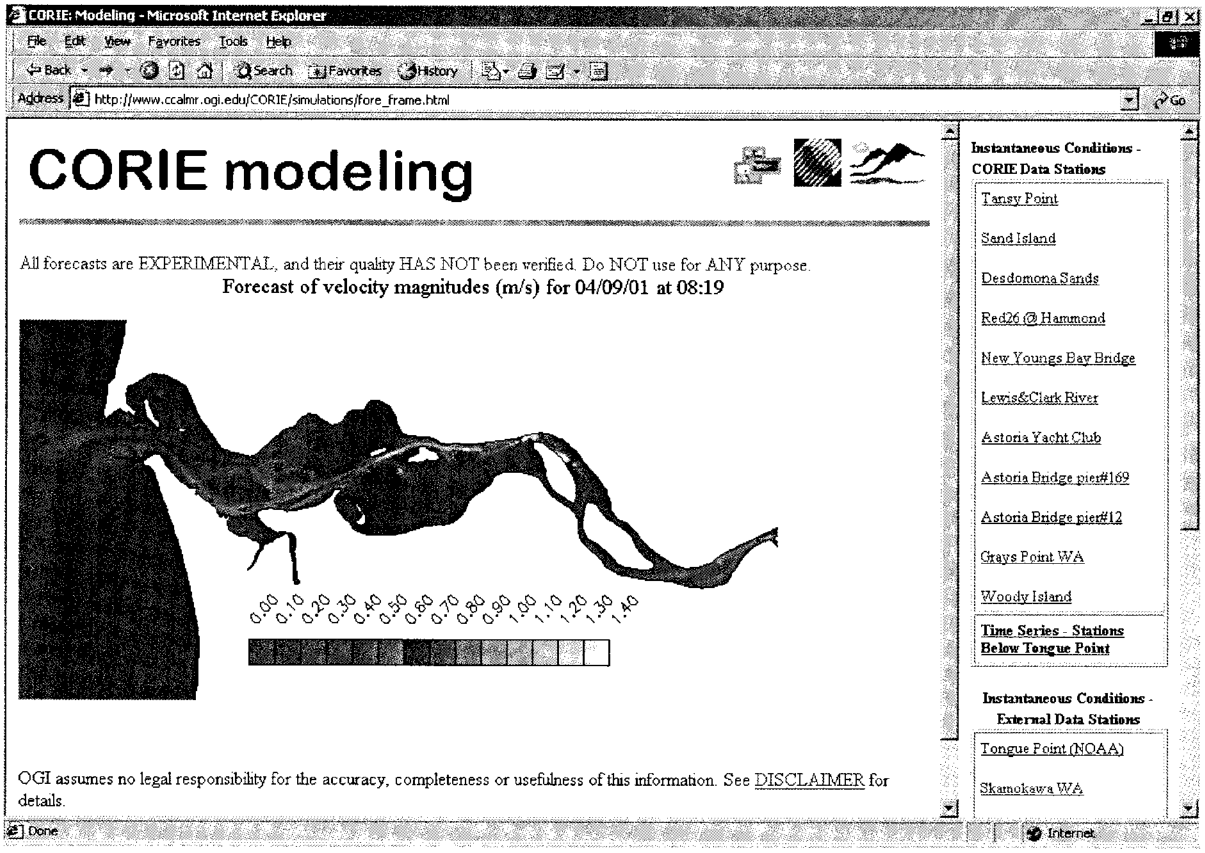
TABLE III Computational Performance of the Circulation Codes Used to Generate 2D and 3D CORIE Forecasts

	3D ELCIRC	2D ADCIRC
Grid size	19380 nodes	Same
Time step	15 min	1 sec[a]
Vertical levels	27	N/A
Computation time per time step	2 min	0.4 sec
Computation time for a 1 month simulation	4 days	12 days

[a] Required for stability.

simulations include salinity and temperature as additional variables, and account for the effect of density gradients on circulation. The 3D forecasts produce 300 million data points a day, creating the need to rethink in very fundamental ways the paradigms for storage, management and visualization of model outputs and model-sensor comparisons. This is an ongoing process.

Long-term hindcasts may include simulation of modern conditions, historical conditions, and possible future scenarios. Hindcasts of modern conditions differ from the forecasts in that all forcings that are measurable are imposed from data, rather than predictions. We have created an extensive database of depth-averaged barotropic simulations of modern conditions (1997–1999), and a more modest database of predevelopment historical conditions (Table IV). A new research project will start the development of an equivalent database of 3D baroclinic simulations. The availability of hindcast databases is proving invaluable for ecosystem applications, as described in Section VI. Those databases are also advancing our ability to understand modeling errors, both toward identification or error sources and in characterization of error dependencies on internal and external factors (e.g., mode-switching in the regime of density stratification in the estuary, or long-term variability of river discharges).



http://www.ccalmr.ogi.edu/CORIE/simulations/fore_frame.html

CORIE modeling

All forecasts are EXPERIMENTAL, and their quality HAS NOT been verified. Do NOT use for ANY purpose.

Forecast of velocity magnitudes (m/s) for 04/09/01 at 08:19

OGI assumes no legal responsibility for the accuracy, completeness or usefulness of this information. See DISCLAIMER for details.

Instantaneous Conditions - CORIE Data Stations

- Tansy Point
- Sand Island
- Desdomona Sands
- Red26 @ Hammond
- New Youngs Bay Bridge
- Lewis&Clark River
- Astoria Yacht Club
- Astoria Bridge pier#169
- Astoria Bridge pier#12
- Grays Point WA
- Woody Island
- **Time Series - Stations Below Tongue Point**

Instantaneous Conditions - External Data Stations

- Tongue Point (NOAA)
- Skamokawa WA

FIGURE 9 Daily forecasts of water levels and depth-averaged circulation have been produced and posted on the web since 1997. The main image in the web browser shown here represents the field of velocity magnitudes at the selected forecast time. The right-side panel gives the user access to model-sensor data comparisons that illustrate the accuracy of the forecasts. We have more recently started to generate equivalent 3D forecasts, which cover a wider range of parameters, including salinity and temperature.

TABLE IV Accumulated Simulation Database of 2D CORIE Simulations

1880	1980	1997	1998	1999
		January	January	January
		February	February	February
		March	March	March
	April	April	April	April
May	May	May	May	May
		June	June	
July		July	July	
		August	August	
		September	September	
		October	October	
		November	November	
December		December	December	

V. INFORMATION AND VISUALIZATION INFRASTRUCTURE

Information flow, management, and representation are integral and often tightly coupled concepts within EOFS. Information flows from producers (e.g., sensors, models) to consumers (e.g., models or end users) through a complex "transportation system" (e.g., Internet, Ethernet, or modem connections) complete with "parking lots" (e.g., storage devices)—e.g., Fig. 10. Emphasis is often on concepts such as timeliness, "end-to-end" data delivery, and relevance to consumer, frequently in detriment of sophisticated representation.

A researcher aboard a vessel, needing data from CORIE models or sensors to plan the next survey line but having access to limited communication bandwidth, will almost

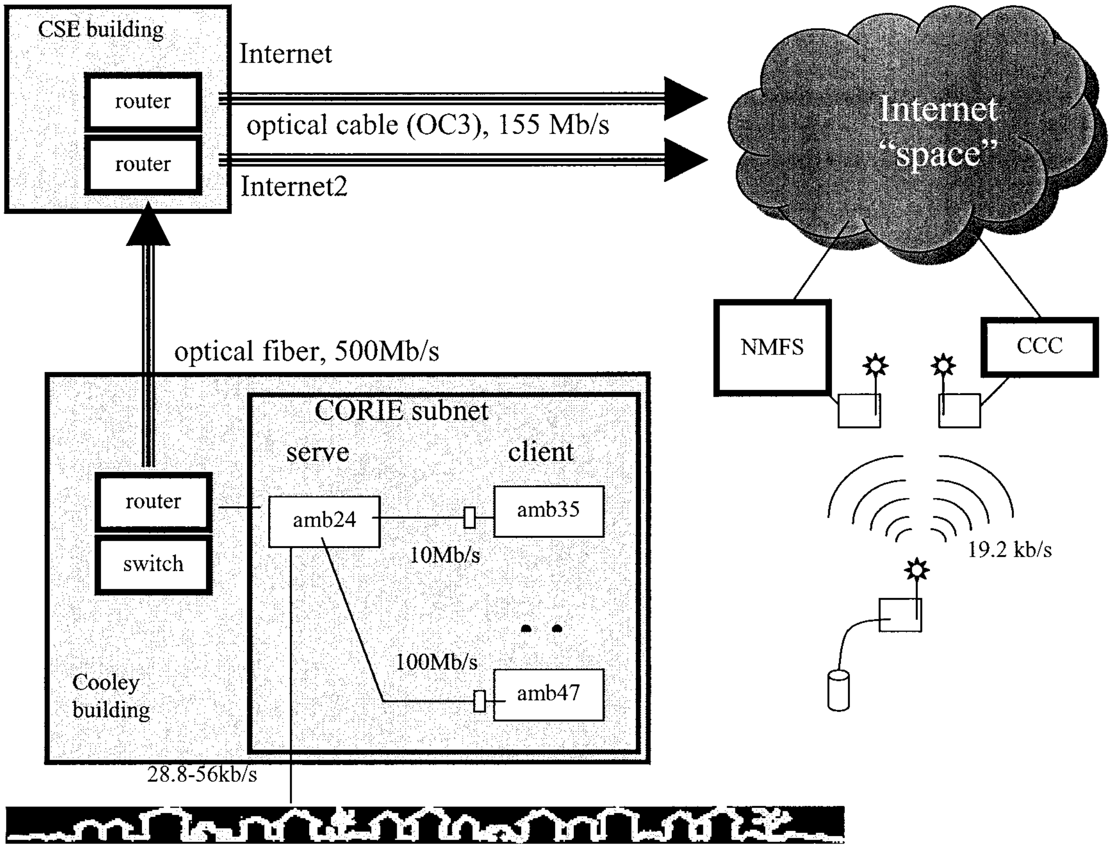


FIGURE 10 CORIE information flows from producers to consumers through an hybrid communication infrastructure. The figure illustrates the path followed by data since its collection in the field until its quasi-instantaneous access by a user in a residential area. Hidden from the user are the various steps, including radio transmission from field sensors to land-based receiving stations in Astoria, routing through the internet and then the Ethernet to the CORIE subnet, primary and tertiary storage, processing into a graphical image, and posting in the web server.

invariably give priority to contents over elaborated graphics. A fisheries researcher or a decision maker, even if possibly attracted by the aesthetics of a 3D animation of simulated estuarine circulation, can typically use circulation information only after it is mined toward a simple, objective summary of its implications for the decision on hand (e.g., Fig. 11). The pilot of a search-and-rescue helicopter will care to receive information that narrows the search area, given priority to timeliness and understandability: the simple reading over the radio of a table with geographic coordinates and times may be adequate. In the representation of CORIE daily forecasts, scheduling considerations often limit the time available for generation and web posting of plots, thus constraining the choices in software and visual complexity.

Our preference for visualization software supporting CORIE quickly moved away from sophisticated, high-dimension tools requiring extensive human intervention and resulting in very large image or animation files. Rather, we have tended to use relatively simple 2D visualization software, activated from schedulable scripts and enabling the timely production of reasonably sized images, which can be satisfactorily downloaded from a web browser even with moderate speed Internet access. We anticipate that the trend will change over time, as hardware and software evolve, but our experience suggests that functionality rather than sensorial sophistication will drive the design for computer–human interfaces that best match the philosophy of EOFS.

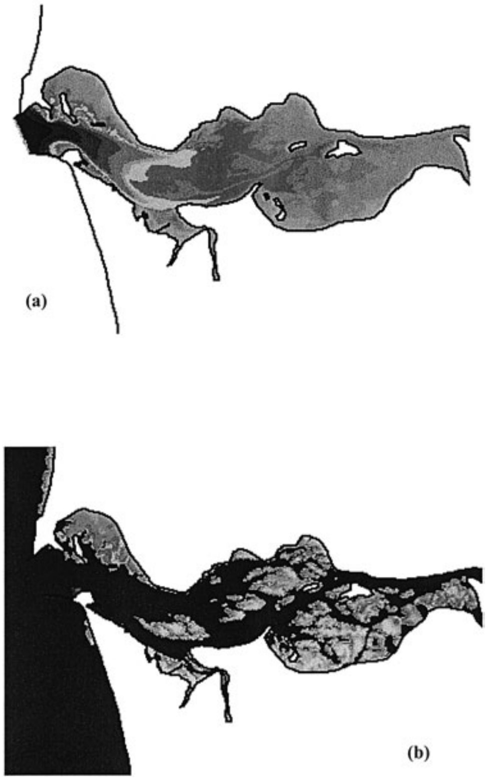


FIGURE 11 The CORIE database of depth-averaged barotropic simulations cover multiple years of modern conditions, and selected month of predevelopment conditions. The database has been interrogated in various forms, for different applications. Shown here are products developed in support of salmon survival research (Bottom, 2001). (a) Maps of residual velocities proved too far removed from the central fisheries question. (b) Maps of habitat opportunity, based on presumed preferences of salmon relative to physical variables, were a significant step to address to needs of the user. (c) However, it was the correlation of habitat opportunity versus river discharge, integrated over distinct regions of the estuary, that provided the type of synthesis that fisheries researchers needed.

VI. APPLICATIONS

Corie has supported a variety of scientific and regional applications, consistently with the notion of EOFS as multipurpose, multiuser infrastructure. Common to all applications is the need to understand Columbia River circulation. However, requirements for timeliness, period of coverage, and detail are widely different across different applications. We discuss below two applications that exemplify the contrast.

A common scientific use of CORIE is in support of vessel surveys. The surveying vessel functions as an Internet node, with ability for two-way exchange of information in real time. The chief scientist aboard the vessel has the ability to access as much or as little information from models and sensors as required, before, during, and after the survey. When the survey involves on-board decisions on how to sample specific aspects of a dynamic environmental front (e.g., the estuarine turbidity maximum), real-time access to reliable contextual data becomes critical to estimate the location of the front in enough detail to guide actual sampling. While CORIE has been able to guide field surveys in a coarser sense, this specific application scenario remains an elusive litmus test for our forecast capabilities, which we have over time become progressively closer to meet. Once the test is successfully met, the range of real-time applications that CORIE can confidently support will increase considerably. For instance, CORIE search-and-rescue products, for instance, could be upgraded from training to operational tools. Also, we would consider developing real-time oil-spill response tools and tools for on-deck support of large-vessel navigation.

Most ecosystem applications of CORIE have response times that are measured in months or years rather than minutes. Two major challenges emerge in this context. The first is to develop products that synthesize information in ways that are understandable to the target audience. The second is to quantify uncertainty, with similar contextual constraints. A recent application of CORIE to the characterization of physical habitat opportunity for juvenile

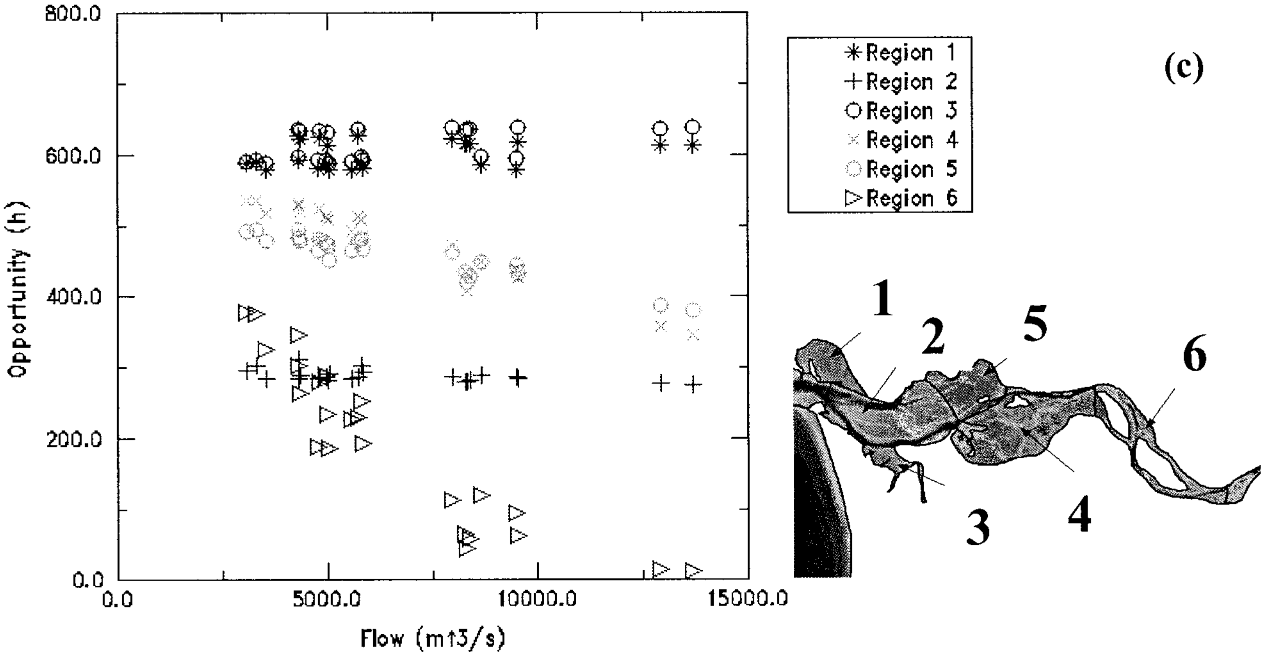


FIGURE 12

salmon exemplifies the dual challenge. While the database of depth-averaged barotropic simulations was interrogated in various forms (e.g. Fig. 11), the derivative products (e.g., Fig. 11a,b) were initially overwhelming in detail or targeted at parameters that were difficult to link to fisheries questions. Detail became useful only after a specific set of derivative products (Fig. 11c) enabled a synthetic characterization of broad system behavior, both for the modern and predevelopment systems. Once the synthetic derivative products were developed, the value of the approach of interrogating a complex circulation database to guide the interpretation of habitat opportunity variability across multiple time scales became intuitive to multiple audiences. The approach is now being used to assist the National Marine Fisheries Service in the development of biological surveys targeted at deepening the understanding of the sensitivity of salmon habitat to anthropogenic forcings, and will be extended to other fisheries species in support of the mandate of the United States Fish and Wildlife Service.

However, the characterization and clear communication of uncertainty levels in derivative products such as those of Fig. 11 remains an unresolved challenge. An Advisory Board was recently created for CORIE, involving multiple agencies with mandate on Columbia River issues. A priority for that group will be to create a consensus on the process toward certification of the CORIE models as regional infrastructure. An integral component of such process will be accepted ways to quantify error and uncertainty, across a variety of primary and derivative products based on models and sensors.

VII. PARADIGMS IN TRAINING AND EDUCATION

Current EOFS are already starting to introduce information-overload problems of unprecedented proportions: valuable information is commonly discarded, ignored, or lacking in elementary quality control. Next-generation EOFS will likely aggravate the problem, unless appropriate education programs are designed to prepare both developers and users of EOFS for massive increases in data and for continuously evolving information-flow needs and technology.

The education challenges posed by EOFS are complex, because (a) there are many constituencies involved, including very diverse user types, very distinct information producers, and multiple requirements of disciplinary expertise within teams of EOFS designers and caretakers; and (b) both key technologies and environmental information needs are in constant evolution. However, EOFS-related educational challenges are not unique; rather, they exemplify the challenges that we will face educating the next generation of environmental scientists, engineers, managers, and regulators. These professionals will need some level of end-to-end understanding of the process by which environmental information is generated, quality-controlled, moved, and ultimately used. Often, such understanding must complement, rather than replace, expertise in one or more traditional areas of environmental science, engineering, management, and law.

We believe that there is a strong need for a new type of environmental education programs, one that enables students to combine understanding of environmental processes with mastery of sensing, modeling, and information technology. New cross-departmental educational tracks in EIT are being designed at OGI to help meet this demand, at various levels of expertise: Ph.D., M.Sc., and certificate degrees. We predict a national trend toward similar educational programs.

We anticipate that most EIT Ph.D. graduates will become environmental scientists, engineers, and managers. Some of these graduates will help fill the demand for high-level technical expertise in the development and maintenance of EOFS, in academic or federal government positions with responsibilities such as project directors, leading modelers, or leading data analysts. A few might form or join companies focusing on EOFS services or supporting technologies (e.g., environmental sensors). The bulk of the graduates, however, might focus mostly on the science or engineering of the environment. Whether in academia, government, or consulting companies, these graduates will have, over colleagues with a more traditional environmental education, the significant advantage of higher sophistication and expanded but realistic expectations in the use of and interfacing with EOFS data and products. This advantage should lead to improved research and engineering productivity and to innovation in educational activities and student advising.

A significant percentage of EIT M.Sc. graduates will probably work directly in EOFS teams. Whether in academia (typically as research staff), government, industry, or consulting companies, these graduates will represent the highly sought-out, cross-trained, midlevel implementation expertise required to develop, maintain, and quality control EOFS. Other EIT M.Sc. graduates will become sophisticated users of EOFS, either for direct benefit or as “interpreters” for larger groups of less sophisticated users. Certificate programs are likely to be in high demand among professionals who have the need for expert use of a particular EOFS, but are unable or unwilling to invest the time and resources required by a M.S. or Ph.D.

VIII. OUTLOOK

It is difficult to predict how far EOFS will push environmental modeling and visualization paradigms. However, it is uncontroversial that EOFS are slowly but effectively contributing to place science at the core of environmental decision making and of operational activities in complex environments. In the process, it is also uncontroversial that many modeling and visualization paradigms are and will continue to change:

- Quality requirements are becoming higher, as the social cost of uncertainties grows and pressure mounts for quality standardization.
- Novel model-sensor coupling paradigms offer increasing opportunities to identify and minimize model uncertainty, both in real-time and in archival modes.
- Priorities in algorithmic changes are often directly driven by failure to meet timeliness or accuracy requirements in the representation of increasing complex environmental processes.
- Visualization, defined as the graphical or mental representation of results, is being increasingly focused on the needs of end-users, leading to opportunities in the development of a broad range of derivative modeling products that would have been difficult to imagine a decade ago.
- Massive datasets, in particular those generated from models, are pushing the limits of existing primary storage hardware.
- Training paradigms are changing, with modeling becoming slowly a base skill that should be acquired by broad groups of future professionals.

ACKNOWLEDGMENTS

While this paper reflects the author’s individual perspectives, these perspectives have been heavily influenced by his collaborative experience in the development and application of CORIE. Thanks are due to Michael Wilkin, Paul Turner, Edward Myers, Calton Pu, Jon Walpole, Todd Leen, Edmundo Casillas, Andre Fortunato, and many other colleagues, collaborators and students over the years.

Early development of CORIE (June 1996 to September 1998) was partially funded by the Office of Naval Research (Grant N00014-96-1-0893). Applications of CORIE have been partially funded by the National Science Foundation (LMER, EGB, SGER, and ITR programs), Bonneville Power Administration, National Marine Fisheries Service, Defense Advanced Research Projects Agency (Software Enabled Control Program), and Office of Naval Research (Modeling and Prediction Program).

The development and maintenance of a system like CORIE requires strong community support. Thanks are due the Clatsop Community College, U.S. Coast Guard, U.S. Geological Survey, Oregon Department of Transportation, Coastal Studies and Technology Center, U.S. Army Corps of Engineers, Port of Portland, City of Astoria, Columbia Pacific Community Information Center, and Capt. R. Johnson (Columbia River Bar Pilots). The Northwest River Forecast Center, Roger Samelson, and Charlie Barron have graciously provided forcings for CORIE, respectively from their own EOFS for river, atmospheric, and ocean systems.

SEE ALSO THE FOLLOWING ARTICLES

CLIMATOLOGY • COASTAL METEOROLOGY • ENVIRONMENTAL MEASUREMENTS • IMAGE PROCESSING • IMAGING THROUGH THE ATMOSPHERE • METEOROLOGY,

DYNAMIC (STRATOSPHERE) • METEOROLOGY, DYNAMIC (TROPOSPHERE) • OCEAN-ATMOSPHERIC EXCHANGE • WEATHER PREDICTION, NUMERICAL

BIBLIOGRAPHY

Baptista, A. M., *et al.* (1998). *In* "Ocean Community Conference '98," Baltimore, MD.

Baptista, A. M., *et al.* (1999). *In* "Earth System Monitor," pp. 1–2, 4–5, 16. Environmental Information Services, National Oceanic and Atmospheric Administration: Washington, DC.

Bottom, D. L., *et al.* (2001). "Salmon at River's End The Role of the Estuary in the Decline and Recovery of Columbia River Salmon," National Marine Fisheries Service.

Center for Coastal and Land-Margin Research. *http://www.ccalmr.ogi.edu/CORIE: Towards a Nowcast-Forecast System for the Columbia River Estuary*.

Glenn, S. M., *et al.* (2000). "Operational observation networks for ports, a large estuary and an open shelf," Oceanography **13,** 12–23.

Long, D., *et al.* (1996). *In* "Conference on Software Engineering and Knowledge Engineering."

Parker, B. B. (1997). *In* "Estuarine and Coastal Modeling" (M. L. Spaulding and A. F. Blumberg, eds.), pp. 355–373. ASCE.

Richardson, L. F. (1922). "Weather Prediction by Numerical Process," Cambridge University Press.

Robinson, A. R. (1999). *In* "Coastal Ocean Prediction" (C. N. K. Mooers, ed.), pp. 77–100. American Geophysical Union, Washington, DC.

Turner, P. J., and Baptista, A. M. (1991). "User's Manual: Software for Semi-Automatic Generation of Two-Dimensional Finite Element grids," OGI-CCALMR Software Documentation Series SDS2-91/2.

Zhang, Y., and Baptista, A. M. (2000). *In* "7th International Conference on Numerical Grid Generation in Computational Field Simulations," Whistler, BC.

Environmental Radioactivity

Brian D. Amiro
Canadian Forest Service

GLOSSARY

Bq (becquerel) The SI unit of radioactivity for measuring the rate of decay of a radioactive substance, equivalent to one disintegration per second.

Gy (gray) The SI unit of absorbed dose for ionizing radiation, equal to 1 J of radiation energy absorbed in 1 kg of matter.

Half-life The length of time required for one-half the original radioactive material to decay to new atoms.

Ionizing radiation Particulate radiations (alpha, electron, neutron, proton) and electromagnetic radiations (X-rays and gamma rays) that have the ability to disrupt molecular bonds (ionize) when they interact with matter.

Radionuclide A radioactive form (isotope) of an element.

Radionuclide abbreviations Superscript prefixes denote the molecular weights of specific radionuclides, for example, ^{137}Cs identifies the radionuclide of cesium with molecular weight 137 g/mol.

Specific activity The amount of radioactivity per unit mass (i.e., Bq/mol). This can be calculated as follows: (ln 2) * (Avogadro's number)/(half-life in seconds).

Sv (sievert) The SI unit of dose equivalent, equal to the product of 1 Gy and a radiation weighting factor that depends on the type of radiation.

THE SUBJECT of environmental radioactivity includes the physical, chemical, and biological aspects of ionizing radiation. It involves the environmental cycling of natural radionuclides as well as the dynamics of anthropogenic (human-created) radioactivity. Of particular interest is the effect of ionizing radiation on humans and other biota.

I. INTRODUCTION

Radioactivity is an integral part of the earth's environment. We are continuously bombarded by high-energy electromagnetic radiation from outer space, and long-lived radionuclides are part of the earth's crust. Ionizing radiation includes particle emissions from various energetic decays, such as alpha particles, electrons (including beta particles), protons, and neutrons. It also includes electromagnetic energies in the form of X-rays and gamma rays. These radiations have sufficient energy to ionize materials and break molecular bonds. The particle and photon emissions are caused by radioactive decay of atoms that disintegrate spontaneously and transform into different atoms. The rate of this decay is defined by the half-life of the radionuclide (i.e., radioactive isotope), but the types of energies emitted depend on the specific radionuclide. Most of the lighter radionuclides (e.g., ^{14}C, ^{40}K) emit only electrons or photons, whereas many of the heavier radionuclides also emit alpha particles (e.g., ^{226}Ra, ^{238}U). The decaying radionuclides create new progeny, which may be stable or may decay also. In some cases, decay chains can be long, with many radionuclides being formed during the decay process, each having a particular half-life and emission type of its own. Alpha particles do not penetrate far into most solid materials, but this feature also causes them to release high amounts of energy in a localized area. In contrast, electrons and photons are more penetrating, and some energetic photons can require many centimeters of lead shielding to reduce their energy. Some of the more common environmental radionuclides are listed in Table I.

Most environmental radioactivity originates from natural sources, but radionuclides and ionizing radiation are also created through human activities. Fission is the process whereby an atom is split, which releases energy and creates fission products (i.e., new atoms) that have atomic weights that are approximately half the weight of the original atom. Fusion is the process that brings together two smaller atoms into a new larger atom, releasing energy in the process. Fusion is the primary energy process in stars, and it may be a future energy process if harnessed by humans. Spontaneous fission occurs naturally in terrestrial systems at a low rate, and induced fission is the current process exploited in commercial nuclear reactors. Energy released through fission, through fusion, or by accelerators (where ionizing radiation is created by using electricity) can also activate atoms to create radionuclides. This activation process is often the source of light radionuclide by-products [such as tritium (^{3}H) or ^{14}C] from nuclear reactors.

The science of environmental radioactivity is part of large international research efforts, the results of which are compiled and evaluated by international agencies and committees. The International Committee on Radiological Protection (ICRP) publishes recommendations related to radionuclides and radiological doses, usually coordinated

TABLE I Common Environmental Radionuclides Classed According to Primary Source[a]

Cosmogenic				
^{3}H (12)	^{14}C (5730)	^{7}Be (0.145)	^{10}Be (1.6 × 10^{6})	
^{26}Al (740,000)	^{32}P (0.039)	^{32}Si (450)	^{39}Ar (270)	^{81}Kr (210,000)
Primordial				
^{40}K (1.3 × 10^{9})	^{87}Rb (4.7 × 10^{10})	^{187}Re (5 × 10^{10})		
Thorium-232 (4*n*) decay chain (only radionuclides with half-lives >1 d shown): ^{232}Th (1.4 × 10^{10}) → ^{228}Ra (5.8)→ ^{228}Th (1.9) → ^{224}Ra (0.01) → ^{208}Pb (stable)				
Uranium-238 decay chain (4*n* + 2) (only radionuclides with half-lives >1 d shown): ^{238}U (4.5 × 10^{9}) → ^{234}Th (0.07) → ^{234}U (244,500) → ^{230}Th (77,000) → ^{226}Ra (1,600)→ ^{222}Rn (0.01) → ^{210}Pb (22.3) → ^{210}Bi (0.014) → ^{210}Po (0.38) → ^{206}Pb (stable)				
Uranium-235 decay chain (4*n* + 3) (only radionuclides with half-lives >1 d shown): ^{235}U (7 × 10^{8}) → ^{231}Th (0.0029) → ^{231}Pa (32,800) → ^{227}Ac (21.8) → ^{227}Th (0.051) → ^{223}Ra (0.031) → ^{207}Pb (stable)				
Anthropogenic				
^{41}Ca (140,000)	^{59}Ni (75,000)	^{60}Co (5.3)	^{63}Ni (96)	^{75}Se (0.33)
^{79}Se (65,000)	^{85}Kr (11)	^{86}Rb (0.05)	^{90}Sr (29)	^{90}Y (0.0073)
^{93}Zr (1.5 × 10^{6})	^{93}Mo (3,500)	^{93m}Nb (14)	^{94}Nb (20,300)	^{95}Nb (0.096)
^{95}Zr (0.18)	^{99}Mo (0.0075)	^{99}Tc (213,000)	^{99m}Tc (0.0007)	^{106}Ru (1)
^{107}Pd (6.5 × 10^{6})	^{113m}Cd (14)	^{125m}Te (0.16)	^{125}Sb (2.8)	^{126}Sb (0.034)
^{126}Sn (100,000)	^{125}I (0.16)	^{129}I (1.6 × 10^{7})	^{131}I (0.022)	^{133}Xe (0.014)
^{134}Cs (2.1)	^{135}Cs (2.3 × 10^{6})	^{137}Cs (30)	^{144}Ce (0.78)	^{182}Hf (9 × 10^{6})
^{182}Ta (0.32)	^{205}Pb (1.4 × 10^{7})	^{208}Bi (37,000)	^{232}U (72)	^{236}U (2.3 × 10^{7})
^{238}Pu (88)	^{239}Pu (24,000)	^{240}Pu (6,500)	^{242}Pu (376,300)	
Neptunium decay chain (4*n* + 1) (only longer-lived radionuclides shown): ^{241}Pu (14) → ^{241}Am (430) → ^{237}Np (2.1 × 10^{6}) → ^{233}Pa (0.074) → ^{233}U (1.6 × 10^{5}) → ^{229}Th (7,340) → ^{225}Ra (0.04) → ^{225}Ac (0.027) → ^{209}Bi (stable)				

[a] The half-lives in years are given in parentheses.

through international task groups. The United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) addresses many scientific issues related to environmental radioactivity and its effects and generates periodic reports for the United Nations General Assembly. The International Atomic Energy Agency (IAEA) is an intergovernmental forum of about 130 states for scientific and technical cooperation in the nuclear field, operating as a specialized agency within the United Nations system. The Nuclear Energy Agency (NEA) is a semiautonomous body within the Organization for Economic Cooperation and Development (OECD), contributing to the development of nuclear energy in support of about 27 member countries. In addition, many countries have their own government agencies dealing with the safety of ionizing radiation and regulation of nuclear energy.

II. SOURCES OF ENVIRONMENTAL RADIOACTIVITY

A. Natural Sources

1. Cosmic Radiation

Cosmic radiation originates in outer space and continuously irradiates the earth. It consists of high-energy electrons, protons, and complex nuclei that are produced by astronomical events. Some of this energy is absorbed by the earth's atmosphere, which causes a gradient in radiation flux with altitude. For example, the dose rate to humans is about five times greater at an altitude of 3200 m than at sea level. The greater flux of cosmic radiation outside the earth's atmosphere is an important consideration for the design of space vehicles and satellites, and for radiation protection of humans traveling in space. The proton is the primary type of particle at these altitudes, whereas at sea level, cosmic radiation is composed mainly of muons, with some gamma rays, neutrons, and electrons.

2. Cosmogenically Produced Radionuclides

Radionuclides produced from cosmic-energy interactions are usually formed high in the atmosphere where cosmic radiation is most intense. The spatial variability of these radionuclides is partly due to variation in the flux of cosmic radiation, which varies spatially with altitude and latitude and temporally with changes in cosmic-ray production rates. For example, the latitudinal production rate between the poles and the equator varies by about a factor of 4, and production rates of some radionuclides have varied by about 10% over the past 10 million yr. There are only a few radionuclides that are formed in sufficient quantities and are long-lived enough to contribute to the general radiation environment. Tritium (half-life = 12 yr) is formed as an unoxidized gas in the upper atmosphere by cosmic-energy interactions, mostly with nitrogen. This eventually enters the hydrologic cycle as tritiated water. Measurements prior to atmospheric weapons testing estimate surface water concentrations of about 10^{-18} atoms of ^{3}H per atom of ^{1}H.

Carbon-14 (^{14}C, half-life = 5730 yr) is also formed cosmogenically by interactions with atmospheric nitrogen. This quickly gets mixed with the stable carbon pool and background concentrations are typically about 227 Bq kg^{-1} stable carbon. Increasing fossil-fuel emissions, which are depleted in ^{14}C are diluting this carbon isotope so that the ratio of ^{14}C to ^{12}C is decreasing over time in the atmosphere.

Lesser amounts of ^{7}Be (half-life = 53 days), ^{10}Be (half-life = 1.6 million yr), ^{26}Al (half-life = 740,000 yr), ^{32}P (half-life = 14.3 days), and ^{32}Si (half-life = 450 yr) are also formed cosmogenically. Noble gases such as ^{39}Ar (half-life = 269 yr) and ^{81}Kr (half-life = 210,000 yr) are also formed in the atmosphere, tending to mostly remain there because of their chemical inertness.

3. Primordial Radionuclides

Primordial radionuclides are those that were present when the earth was formed about 5 billion yr ago. Typically, they have half-lives of the order of a billion years; otherwise they would have decayed away by the present time. These radionuclides have always been a part of the earth and have been decaying over time such that they pose a lesser radiation environment than in the past. Some of these primordial radionuclides are part of long decay chains, often associated with uranium. The decay chains described in the following sections are listed starting with a parent radionuclide that is still common on Earth. However, in some cases, these chains have shorter-lived precursors that are no longer common since they have decayed away over the earth's lifetime. Some of these are described in Section II.B. Also, there are a few abundant radionuclides, such as ^{40}K, that are not part of these long decay chains.

a. Thorium series (4n) decay chain. The long-lived parent of the thorium series is ^{232}Th with a half-life of 14 billion yr. Concentrations in soil range from 7 to 50 Bq kg^{-1} and in seawater are about 4 mBq m^{-3}. Most of the progeny of ^{232}Th have relatively short half-lives, with only ^{228}Ra and ^{228}Th having a half-life of greater than 1 yr. The decay chain and the type of emissions are given in Table I. The activities of ^{228}Ra, ^{228}Th, and ^{224}Ra are similar to that of ^{232}Th in the earth's crust, and this relationship generally holds in other parts of the environment. The stable end product of the chain is ^{208}Pb.

b. Neptunium series ($4n + 1$) *decay chain.* The neptunium series no longer exists naturally since all its members have relatively short half-lives compared with the age of the earth. The longer-lived members of the chain, in decay order (with half-lives), are ^{241}Pu (14.4 yr), ^{241}Am (430 yr), ^{237}Np (2.1 million yr), ^{233}Pa (27 days), ^{233}U (0.16 million yr), ^{229}Th (7340 yr), ^{225}Ra (14.8 days), and ^{225}Ac (10 days). This decay chain is included here for completeness, since it did occur naturally for the first few million years of the earth's existence. Humans have now re-created this decay series through nuclear reactions, such that some quantities of these radionuclides now exist. For example, soil concentrations of ^{241}Pu are on the order of 2000 Bq kg^{-1}, and the ratio of ^{241}Pu to ^{239}Pu from nuclear fallout is estimated at about 10. However, this ratio is quickly decreasing with the reduction in atmospheric nuclear testing and because ^{241}Pu has a half-life of only 14 yr. But some of the long-lived progeny, such as ^{237}Np and ^{233}U will remain for hundreds of thousands of years.

c. Uranium series ($4n + 2$) *decay chain.* The uranium series is the best known of the primordial decay chains, with the long-lived parent, ^{238}U, representing 99.2% of the total uranium (Table I). Typical soil concentrations of ^{238}U are about 10 Bq kg^{-1} and seawater concentrations are about 40 Bq m^{-3}. The half-life of ^{238}U (4.5 billion yr) is of the same order as the age of the earth, so that about half the original amount of ^{238}U is still present since the earth's formation. This decay chain has two long-lived progeny: ^{234}U and ^{230}Th, with half-lives of 244,500 yr and 77,000 yr, respectively. A very important decay product of ^{230}Th is ^{226}Ra, commonly referred to as just radium. Radium-226 has immense historical significance as one of the radionuclides used in early radioactive products, such as luminous watch dials, and was a prime research focus for the early radiochemists. Radium is also known as a "bone seeker" since it has biological properties similar to those of calcium and tends to accumulate in human bone.

The decay progeny of ^{226}Ra is ^{222}Rn, commonly called radon. Radon-222 is an inert gas with a half-life of about 4 days. It is easily transported through the environment, and its progeny of ^{218}Po, ^{214}Pb, ^{218}At, ^{214}Bi, ^{214}Po, and ^{210}Pb coalesce as particles and cause about half the background radiation dose to humans, mostly through deposition in the lungs. This dose from radon progeny varies spatially, depending on soil and geosphere concentrations of ^{226}Ra. However, the dose can also increase because of indoor buildup of ^{222}Rn in homes that is caused by low ventilation rates and house walls that are permeable to the gas. Many governments have imposed air-quality standards aimed at reducing the dose from the progeny of ^{222}Rn. Although there is little doubt that ^{222}Rn progeny are the single greatest radiation source for most humans, there is much debate on whether this poses a significant health hazard. The earlier concerns over the effect of ^{222}Rn progeny were caused by an excessive number of cases of lung cancer in uranium mine workers, who were exposed to very high levels of ^{222}Rn gas. But it is difficult to determine the effects caused by very low doses, and some studies show even decreased incidence of lung cancer in areas of relatively higher ^{222}Rn concentrations. The stable end product of this chain is ^{206}Pb.

d. Actinium series ($4n + 3$) *decay chain.* The actinium series is also called the uranium-235 series since ^{235}U is the parent with a half-life of 0.7 billion yr (Table I). Uranium-235 accounts for about 0.7% of the total uranium and is the fissionable material that is often used as fuel for nuclear power. Protactinium-231 and ^{227}Ac are the only progeny with half-lives greater than 1 yr. The stable end product is ^{207}Pb.

e. Other primordial radionuclides (^{40}K, ^{87}Rb, ^{187}Re). Potassium-40 has a half-life of 1.3 billion yr, is ubiquitous in the environment, and represents about 0.0119% of the total potassium inventory on Earth. It is one of the major radiation-dose contributors from natural sources, with both electron and photon emissions. Since potassium is a basic nutrient required for life, ^{40}K also cycles and accumulates freely in the biosphere. The combined dose from internal and external (e.g., soil and rocks) exposures is estimated to be about 8% of the total dose from all sources. Soil concentrations are estimated at about 370 Bq kg^{-1}, freshwater concentrations at about 50 Bq m^{-3}, and seawater concentrations at about 13 kBq m^{-3}.

Rubidium-87 has a half-life of 47 billion yr and represents 28% of the total rubidium in nature. The soil concentration is about 55 Bq kg^{-1}, the freshwater concentration is about 1 Bq m^{-1}, and the seawater concentration is about 100 Bq m^{-3}. Rhenium-187 has a half-life of about 50 billion yr, representing 63% of the total rhenium inventory. Soil concentrations are about 0.001 Bq kg^{-1} and freshwater concentrations are about 0.01 Bq m^{-3}. Neither ^{87}Rb nor ^{187}Re is a large-dose contributor to biota.

B. Anthropogenic Sources

1. Military Sources

Although experimental quantities of radionuclides were generated in the early part of the twentieth century, it was not until World War II that anthropogenically produced radionuclides became widespread. The development of nuclear weapons resulted in atmospheric releases between 1945 and 1962, when most atmospheric testing was banned by international agreement, although some limited atmospheric testing continued until 1980. The radioactive fallout was dispersed globally, such that many fission products could be detected in most parts of the environment. Fortunately, most of these fission products were relatively short-lived, with the greatest persistent inventories being tritium, ^{14}C, ^{137}Cs, ^{95}Zr, and ^{90}Sr. Other radionuclides measured in the environment from weapons include ^{95}Nb, ^{106}Ru, ^{131}I, ^{144}Ce, ^{239,240}Pu, ^{241}Pu, and ^{241}Am. These inventories continue to decrease in the environment, such that the gross beta activity in current precipitation is about 0.15% of the maximum concentrations measured in the early 1960s.

Nuclear weapons are not currently dispersing radionuclides through the global environment. However, the inventories are large, with more than 30,000 warheads owned by five countries, and some additional countries are developing arsenals. Initiatives are under way to decrease this inventory, but permanent disposal methods have not been developed. One possibility is to use the weapons' fissionable material in some nuclear power reactors, where the material is then converted to a nonweapons form. Some additional power benefits can also result, but the main goal is to decrease access to weapons material. The difficulty is to get cooperation between military and civilian organizations, transport the material to sites for conversion, and involve all countries that have a weapons inventory. Even if much of the existing inventory is disposed, the remaining weapons have great potential to release a large quantity of radionuclides to the environment.

Nuclear-powered watercraft (ships, submarines) are another source of radionuclides from the military. The potential exists for accidental and routine releases of radionuclides to occur in much the same way as from conventional power reactors. In addition, there are concerns about disposal of aging watercraft, especially in the Barents and Kara Seas, where some sunken submarines still contain fuel. Although dissolution of these inventories is slow, the potential still exists for a breach of containment with subsequent dispersion in the ocean.

2. Nuclear Power Industry

The nuclear power industry generates electricity from fission whereby a heavy radionuclide is split into smaller ones, which releases energy in the process. Although this process destroys some quantity of the heavy radionuclide (e.g., ^{235}U), much greater sources of ionizing radiation are created through the decay of the fission products (e.g., ^{131}I). Currently, there are over 400 power reactors operating in more than 30 countries, plus about 270 research reactors. Nuclear energy accounts for about 7% of the world energy consumption from all fuel types, and 18% of the electricity generated.

Several designs of power reactors are used. Pressurized water reactors dominate, and about 220 units have other designs, including boiling water reactors, pressurized heavy water reactors, gas-cooled reactors, fast breeder reactors, and light-water graphite reactors. Each of these reactor types has a slightly different characteristic regarding potential releases of radioactivity to the environment. In addition, the mining and milling of uranium and the fabrication of fuel are part of the nuclear fuel cycle. Contributions from all these sources are discussed in Section V.

Although all current power reactors use the fission process to generate electricity, research is being conducted on nuclear fusion. In this process, hydrogen (or deuterium and lithium) atoms are fused, which creates energy and a helium atom. This process has the advantage of more readily available fuel, such as hydrogen from water, and less radioactive waste because fission products are not created. However, tritium would be a by-product, and many materials used in the reactor would be activated to create additional radioactivity.

3. Medical Sources

The usefulness of X-rays for imaging hard tissues, such as bone, was recognized very early following the discovery of radioactivity. Early X-ray use was often not strictly controlled, with dangerous doses accumulated by patients and practitioners. However, its use is now much more sophisticated with much improved shielding, control over stray emissions, and regulatory checks. Today, it is a basic tool for medical and dental imaging in both human and veterinary medicine. Full X-ray examinations can cause doses ranging between 0.05 and 10 mSv each.

Radiopharmaceuticals are used in medical treatments as tracers to produce diagnostic images. Technetium-99m (^{99m}Tc, half-life = 6 hr) is a gamma-emitting radionuclide generated as a progeny of ^{99}Mo (half-life = 66 hr). It is used for imaging many tissues, such as the lung, liver, brain, and bone. The specific target organ depends on the mode of tracer preparation, and the short half-life ensures no lengthy radiation concerns. These procedures are carried out for specific purposes only and tend to cause doses to only specific local sites, without concerns about risk to the whole body. Doses are usually less than 10% of those caused by X-ray examinations.

Radionuclides can also be used to irradiate tissues directly for specific purposes, such as cancer treatments. External radiation fields can be generated by using photon emitters such as ^{60}Co or through an accelerator that generates photons from electrical energy without the need of a radionuclide inventory. This radiation can be targeted at specific locations, where the energy attacks the tumor. In some instances, electron emitters can be implanted or circulated into organisms, where the radionuclide targets and irradiates a specific site. For example, ^{131}I is used as an internal irradiator that can specifically target the thyroid or seek out a tumor when it is labeled to an appropriate antibody.

About 45% of all single-use medical products are sterilized with gamma radiation. Much of this uses ^{60}Co as a gamma source, with about 7×10^{12} Bq installed in about 180 facilities. Accelerators are also used for sterilization, but these rely on electrical energy as the energy source, so they do not have a radionuclide inventory that can cause potential environmental concerns.

4. Other Industrial Sources

Radionuclides are created for many other industrial uses. Some scientific instruments use radioactive sources as part of the measurement process. For example, tritium is often used as an electron source to measure gases in airstreams, and ^{241}Am (half-life = 432 yr) is used as a neutron source to measure soil moisture. Luminescent signs, especially for aircraft runways, are made with tritium. These applications tend to use only small amounts of radionuclides, and their distribution is limited, so they do not contribute much to the overall level of environmental radioactivity. The use of coal and phosphates (e.g., agricultural fertilizers) also releases uranium- and thorium-series radionuclides and ^{40}K.

Domestic smoke detectors commonly use ^{241}Am as an ion source for the detection of particles in the air. Millions of these units are in circulation, each containing less than 37 kBq of ^{241}Am. Although the external radiation field is small (weak photon source), the potential exists for discarded units to enter the environment where a small radiation dose could be received from the electron- and alpha-particle decay if it is ingested or if it is inhaled following combustion.

Satellites and other spacecraft use radionuclides as a power source. Some spacecraft have power reactors, and there have been a few instances in which these reactors have returned to Earth, which caused some contamination to be spread. One notable example is when a COSMOS satellite disintegrated on reentry into the atmosphere, which distributed radioactivity in northern Canada in 1978. A more common energy source is ^{238}Pu (half-life = 88 yr), which generates energy as it decays to power some satellite systems. There are few examples of this radionuclide reentering the earth's atmosphere, and its relatively short half-life makes it only a small potential source relative to background levels.

III. ENVIRONMENTAL TRANSPORT OF RADIONUCLIDES

A. The Geosphere

1. Consolidated Materials

Primordial radionuclides originate in the geosphere, and anthropogenic radionuclides are increasingly entering the geosphere because of human disposal practices. Mining and milling of uranium ore is one of the largest intrusions into consolidated rock with regard to radioactivity. Although uranium solubilized and was transported to specific deposit locations during the genesis of rock formation, it tends to be relatively immobile once deposited. However, groundwater does flow through the consolidated geosphere in fractures in hard igneous and metamorphic rocks, or through a more general percolation within more permeable sedimentary formations. This groundwater movement can solubilize radionuclides from underground deposits, depending on the chemistry of the groundwater and chemical interactions with specific radionuclides. If underground deposits become soluble, radionuclides are mostly retained in the rock because of chemical exchange along the water transport path. Inert gaseous radionuclides (e.g., ^{222}Rn) move more freely, but transport from deep underground to the surface takes years to millions of years, and radioactive decay often allows only small amounts of radionuclides to reach the surface. The nature of the underground deposit is important. For example, surface concentrations of uranium do not reflect a 1.3-billion-yr-old uraninite deposit 450 m below the surface in northern Canada. With very slow transport of radionuclides from deep underground, surface weathering of rock is the greatest source of radionuclides to the biosphere, hydrosphere, and atmosphere.

2. Unconsolidated Materials

The unconsolidated part of the geosphere comprises the top surface layers, including soil. This includes both inorganic (e.g., sand, clay) and organic materials. In many parts of the world, this material has formed through weathering of parent geological materials, and in some areas, deposits are formed through glacial or fluvial actions. The inorganic materials often have radionuclide compositions that reflect the underlying rock. The lateral and vertical flow of water transports radionuclides through this material quickly in the case of coarse sands and gravel, but very slowly in the case of fine silts and clays. Generally water, and hence radionuclide, flow is channeled through areas of greater permeability, with transit times ranging from hours to thousands of years depending on the permeability and distance. As a result, shorter-lived radionuclides can decay in transit.

Many radionuclides are often retained in surface soils, especially those with organic matter that binds the ions. For example, following the accidental release of ^{137}Cs in the forested areas near Chernobyl, most of the radionuclide is still retained in the top soil layers. This retention can prevent downward leaching of radionuclides to the water table, but it may pose other difficulties because of greater doses to organisms exposed to the soil surface.

The ability of radionuclides to either flow or be retained in unconsolidated materials can be defined on the basis of the partitioning of the radionuclide between solid and liquid forms. This reflects the chemical exchange of an element, which depends on the chemical form of the radionuclide and on the material. For example, the partitioning coefficient for plutonium (e.g., ^{239}Pu) increases by about a factor of 10 between sand (lower sorption) and clay soils (higher sorption). In addition, inorganic forms of iodine (e.g., ^{129}I) are about 1000 times more soluble than plutonium.

B. The Hydrosphere

1. Oceans

Weathering of the geosphere releases many long-lived radionuclides that are eventually transported to the oceans. In addition, many anthropogenic radionuclides are deposited to oceans, either through atmospheric deposition or through transport in rivers from terrestrial areas. Many power reactors are situated on oceans or major water bodies to take advantage of cooling water. These reactors release low concentrations of some radionuclides because they have imperfect filtering systems, which do not contain all products. In addition, thermal plumes slightly raise water temperatures, which can create areas of enhanced biological productivity at cold ocean sites.

Oceans have fluid dynamics that, in many ways, parallel those in the atmosphere. Soluble radionuclides move laterally with ocean circulation patterns. These preferential circulations are more predictable than atmospheric circulation, so the likely flow paths through the ocean can be estimated. In addition to radionuclides' being deposited to the ocean surface, they can be volatilized as gases or ejected as particles with sea spray. Oceanic biota, notably algae, are an important part of radionuclide cycling since they may trap radionuclides, eventually carrying them to the ocean floor, or release them as gases through organic conversions. They are also the basis for the food chain.

The most notable aspect is that vertical mixing and gravitational settling can move radionuclides to the deep ocean, where they are largely removed from circulation and may be deposited to the ocean floor. In some areas, the deep ocean can be a net sink for radionuclides, or vessels containing radionuclides, with little chance of general environmental contamination over the life of many radionuclides. However, there are numerous international conventions preventing the dumping of radioactive wastes at sea and most nuclear countries have adopted measures to eliminate pollution of oceans by radionuclides that originate from land-based sources.

2. Freshwater

Radionuclides are transported through freshwater bodies in much the same way as in oceans. However, most lakes do not have their own lateral circulation patterns, but instead are connected by rivers, which accomplish lateral transport. Another important feature is that many lakes in regions with substantial seasonal climate fluctuations experience semiannual water turnover (i.e., vertical mixing). This turnover moves radionuclides suspended or dissolved in the lower layers to the water surface. As in oceans, sedimentation and volatilization can occur. Tritium is easily evaporated from the water surface, as is ^{14}C in an oxidized ($^{14}CO_2$) or a reduced ($^{14}CH_4$) form.

Nuclear reactors are also situated on lakes or rivers, which act as sources of cooling and industrial water. Therefore, freshwater receives some radionuclide effluents during routine and emergency releases. Unlike the ocean, freshwater bodies have a much lower capacity for dilution because of their relatively small size, with water discharge rates varying from essentially zero (stagnate ponds) to more than 1000 $m^3\ s^{-1}$ in the case of large rivers.

C. The Atmosphere

The atmosphere is dominated by relatively rapid fluid motions, which quickly transport radionuclides. For most applications, atmospheric processes can be divided into suspension, dispersion, and deposition. Direct emissions result from atmospheric detonation of nuclear weapons and routine and emergency venting of radionuclides by industry. These are normally point sources, where the emissions originate from a defined point and activity. However, there are more general suspension processes that cause radionuclides to enter the atmosphere from soil or water surfaces. A few radionuclides are readily volatilized as gases. The most notable of these are ^{3}H, ^{14}C, and ^{222}Rn. Tritium is commonly evaporated in the oxidized form as water, but some microbes can covert tritium to a reduced form and release it as a tritiated hydrogen gas. Similarly, ^{14}C can be released as $^{14}CO_2$ during chemical conversions of carbonates, or respired by biota as a basic metabolic function or when organic matter is decomposed. Radon is easily emitted as an inert gas as its precursor, ^{226}Ra, decays. Some halogens can also be released as gases through both inorganic and organic chemical reactions. The most notable halogens that may be released as a gas are iodine (e.g., ^{125}I, ^{129}I, ^{131}I) and chlorine (^{36}Cl). Selenium (^{79}Se) is also known to be volatile from soil and lake surfaces. Other inert gases, such as Ar (^{39}Ar), Kr (^{81}Kr, ^{85}Kr), and Xe (^{133}Xe, ^{135}Xe), are found as gases in the atmosphere, but these typically either are released directly from some industrial source or are a decay product.

Most other radionuclides become suspended in the atmosphere as particles. Soil erosion by wind is a continuous process and this can be enhanced by mechanical disturbances. The heavier particles ($>20\ \mu$m in diameter) often are not truly suspended and move laterally near the surface through a process know as saltation. However, smaller particles can become truly suspended and contribute to the general atmospheric dust load, with those less than about 10 μm in diameter being most easily inhaled by humans. Particles are also released from water surfaces through direct wind action (e.g., sea spray) or through bubbles bursting at the surface. Fires can release radionuclides in the smoke, thereby mobilizing radionuclides that were tied up in organic matter. In some cases, fires get hot enough to chemically convert radionuclides to a gaseous form; for example, such a conversion starts to occur with cesium (^{134}Cs, ^{135}Cs, ^{137}Cs) when temperatures increase above 400°C.

Once radionuclides are suspended into the atmosphere, they are largely at the mercy of wind and the general circulation of the atmosphere. Releases from atmospheric testing of nuclear weapons in the 1950s and early 1960s resulted in widespread dispersion of many radionuclides throughout the world. Similarly, the 1986 accidental release from the damaged reactor at Chernobyl caused radionuclides to be widely spread through the Northern Hemisphere. These sources reached great heights in the atmosphere (>1 km), which allowed the upper airstreams to disperse materials widely. More commonly, radionuclides are emitted at lower heights in the atmosphere and tend to mostly be contained in the surface boundary layer. For example, emissions from reactors are usually at heights of tens to hundreds of meters, which in daytime is contained within the atmospheric mixing layer, with the plume being in close contact with the underlying surface. Suspension of contaminated soil may reach heights of only a few meters, depending on particle size. In these cases, dispersion is controlled by local wind direction and atmospheric stability, and radionuclide concentrations become more dilute with distance from the source. This dilution is an important aspect of atmospheric dispersion, with dilution factors (i.e., the ratio of air concentration to release rate from a point source) in the range of 10^{-7} to 10^{-5} s m^{-3}, depending on meteorology, source height, and distance from source.

Radionuclides are deposited to surfaces, unless they have a sufficiently short half-life to decay while suspended in the atmosphere. Deposition can be an active process, such as when animals inhale particles and gases that are directly deposited to the lungs. Deposition to other surfaces occurs through either wet or dry physical processes. Wet deposition involves the scavenging of particles and gases by precipitation, either through direct coalescence as part of precipitation formation or through “washout” of radionuclides by rain or snow falling through the atmosphere. Dry deposition does not need precipitation. It occurs when radionuclides attached to particles settle onto surfaces because of gravity, are impacted on surfaces while carried by the wind, are caught on sticky surfaces, or become attached as molecular motions bring particles close to surfaces. Gaseous radionuclides become deposited through chemical reactions or absorption in water at the surface or inside plant stomata. Although it greatly depends on the environment (especially the amount of precipitation), dry and wet deposition processes are typically of similar magnitude over long periods.

D. The Biosphere

In this section, transport through and among the biota is discussed, since the abiotic part of the environment is treated in the other sections. Microbes (e.g., bacteria, fungi) play an important role in the cycling of all elements, including radionuclides, through the biosphere. For example, they can transform tritiated water into tritiated hydrogen, or they convert organic matter containing ^{14}C to $^{14}CO_2$ or $^{14}CH_4$. Fungi establish mycorrhizal relationships with many plant species and are known accumulators of cesium, such that ingestion of ^{137}Cs in mushrooms is an important potential dose pathway in contaminated areas.

Plants passively and actively sorb and cycle radionuclides. Gases that enter plant stomata can be absorbed into cells and the radionuclides can be translocated through the plant. Plant roots take up radionuclides from the soil, depending on the nature of the element. Some large elements that are not physiologically needed, such as uranium (e.g., ^{238}U), are not easily transferred past the root endodermis, so that the aboveground plant parts often have relatively low concentrations. Other radionuclides, such as ^{3}H and ^{40}K, are easily taken up from soil and move through the plant in the water stream and then subsequently through the translocation stream. Mechanistic models of radionuclide uptake by plants are still being developed, and often simple transfer factors or concentration ratios are used to estimate plant concentrations on the basis of concentrations in the underlying soil.

Some radionuclides are physiologically similar to their nonradioactive, stable isotopes; for example, tritium behaves like stable hydrogen, and ^{14}C follows ^{12}C pathways. Actually, both tritium and ^{14}C are used to study the physiological pathways of their stable isotopes. This isotopic substitution tends to work for all radionuclides, although there is often a slight discrimination because of differences in molecular weight. In addition, some elements with similar characteristics in the periodic table often behave similarly. For example, cesium (e.g., ^{137}Cs) and potassium have similar chemical characteristics, so that often cesium isotopes follow potassium pathways. This similarity also occurs within animals, most notably when radium (e.g., ^{226}Ra) follows calcium pathways and is deposited in bone.

Terrestrial animals acquire radionuclides through ingestion of food and water (and soil, in some cases), inhalation of air, and sorption through the epidermis (skin). Radionuclides that are analogues of stable nutrients (e.g., ^{14}C, ^{40}K) are most easily metabolized in food and water. Rarer radionuclides, such as isotopes of plutonium and americium, are about four orders of magnitude less easily absorbed from food. Amphibians are often more sensitive to epidermal sorption because of their porous epidermis and immersion in potentially contaminated aquatic environments. Fish are also immersed in their environments so that concentrations of many radionuclides in their bodies are about 10–100 Bq kg^{-1} of wet flesh for every Bq L^{-1} of water (carbon, phosphorus, and cesium are notable exceptions).

Most radionuclides do not accumulate up the food chain, such that the concentration in primary consumers (plant eaters) is similar to that in higher-order consumers (carnivores, scavengers). However, ^{137}Cs has been found to biomagnify in some systems where wolves have concentrations about six times greater than those of their prey. Essentially, radionuclides behave biologically like their stable isotopes and do not have any special accumulation characteristics that make them more toxic. This is because they are usually in an inorganic chemical form, so they do not biomagnify as many organic contaminants do.

IV. EFFECTS OF RADIOACTIVITY ON BIOTA

A. Concepts of Radiation Dose

Radiation dose is created by the energy of ionizing radiation being absorbed by some substance, such as biological tissue. The basic unit of dose is the gray (Gy), and dose rates are expressed per unit time (e.g., Gy hr^{-1}). A Gy is defined as 1 J of energy absorbed in 1 kg of material. Each radionuclide emits radiation (photon, electron, alpha particles, or combinations of these) during its radioactive decay, with the energies expressed as MeV (10^6 electron volts) Bq^{-1} s^{-1} or as J Bq^{-1} s^{-1} (1.6×10^{-13} J $= 1$ MeV). The dose is then calculated on the basis of how much of this is absorbed in 1 kg of material, with alpha particles being readily absorbed and photons traveling farther before being absorbed. Denser materials are more opaque to radionuclide emissions, so denser tissues tend to absorb a greater dose.

Dose estimates use a radiation weighting factor to account for differential effects of the types of energy emissions. For humans, the usual convention is that alpha particles are given a weighting of 20, whereas many photons and electrons have a weighting of 1. Neutrons and protons have weighting factors varying from 5 to 20, depending on their energies. There is ongoing debate over some of these weightings, especially for nonhuman biota, such that greater or lesser values are sometimes used. The product of the absorbed dose and the radiation weighting factor is called the equivalent dose, which in the case of humans is expressed in units of sieverts (Sv).

Absorbed radiation can originate outside or inside an organism. External radiation comes mostly from photon sources in the general environment, such as cosmic rays or primordial radionuclides in the soil. Electrons and alpha particles also impose a dose to epidermal tissues, but their energies do not travel far into organisms. In addition, radionuclides within organisms irradiate inner tissues and organs. This dose depends on the location of the accumulation of the radionuclides (e.g., ^{226}Ra accumulates in bone) but because the energy is released internally, emissions that do not travel far (e.g., alpha particles) can still create a dose. This combination of dose contributions can be shown by using the example of ^{137}Cs in a contaminated environment. Humans would receive a dose from external radiation caused by photons emitted by ^{137m}Ba, the progeny of ^{137}Cs, because ^{137}Cs itself emits only electrons. However, ^{137}Cs is easily incorporated into the food chain and can be ingested and deposited in soft tissues, so that most of the dose is caused by internal electron radiation if humans eat local foods. The internal dose to humans would be largely avoided if food sources came from uncontaminated areas. However, plants are stationary and would receive doses from internal and external sources, with most of the dose being internal in the case of ^{137}Cs.

The dosimetry of the large range of plant, animal, and microbial species is still being developed since it is difficult to define geometries and relative biological effects for every organism. Instead, specific geometries are often used as a test case, such as a fish of a certain size that would represent a broad range of fish species. The alternative is to assume that all energy from internal radionuclides is absorbed by the organism (i.e., no photons escape), which gives a slight overestimation of the actual absorbed dose, with the overestimation increasing with decreasing organism size.

In contrast, human dosimetry is well developed, with the broad background of research summarized in recommendations by the International Commission on Radiological Protection (ICRP). Two important concepts are applied to human dosimetry. First, an effective dose is defined as the sum of the weighted equivalent doses in all tissues and organs in the body, and when the radiation weighting factors are included, becomes the effective dose equivalent. Second, a committed dose is based on the integration of the dose received following an intake of radionuclides over time, typically for 50 yr in the case of adults and 70 yr in the case of the general population. Human dosimetry is simplified through the definition of Reference Man, which represents a non-gender-specific person with defined anatomical and physiological characteristics. This definition provides a common basis from which doses from a variety of radionuclides to specific organs and to the whole body can be estimated. Another important concept is that of the critical group, a hypothetical group of individuals that is most highly exposed to a given source of radiation. Assessment of impacts on the critical group ensures that all less-exposed individuals are also assessed and protected.

The dose rate to individual humans varies between 1.5 and 6 mSv yr^{-1} around the world, with an average of about 3 mSv yr^{-1}. Natural sources that include cosmic and primordial sources account for about three-fourths of the annual radiation dose (Fig. 1), with most of the remaining anthropogenic contribution coming from medical sources in industrialized countries. However, there are many areas of the world where specific sources are much higher. For example, thorium and uranium deposits in Brazil, India, and northern Canada contribute greater doses from primordial sources. Also, the cosmic radiation dose increases with altitude, so residents of higher elevations receive a greater dose. As an example of increased dose with altitude, a single 10-hr commercial aircraft flight results in a dose of about 0.02 mSv, or almost 1% of the annual average dose. In contrast, medical exposures are quite variable among individuals, and the contribution from fallout from weapons testing, mostly prior to 1963, is decreasing with time but is estimated to be $<0.03\%$ of the present total dose (the data in Fig. 1 are committed doses for 1945–1992).

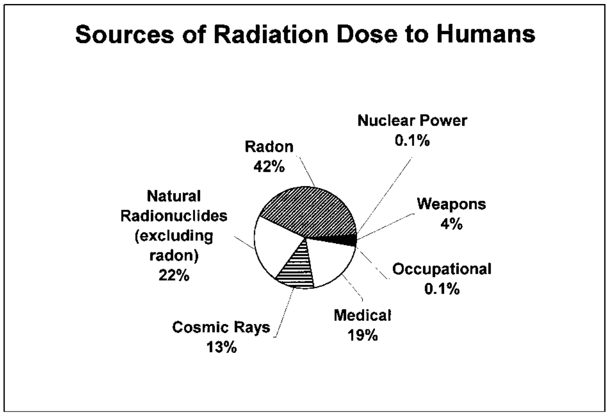


FIGURE 1 Sources of radiation dose to humans, based on 50-yr collective world population doses from practices occurring from 1945 to 1992. [Adapted from United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (1993). "Sources and Effects of Ionizing Radiation," United Nations Publication Sales No. E.94.IX.2, United Nations, New York.]

Another dose concept is that of collective, or population, dose. This is an integration of the dose over a whole population; for the world, this is simply the product of the average dose and the number of individuals. It is sometimes used to compare doses from wide-reaching global events, such as nuclear weapons fallout. However, aside from possible intercomparisons among dose sources, the absolute magnitude of the collective dose is not very useful, since it is unreasonable to extrapolate very low doses to a large population with the goal of inferring potential biological effects. This is because there is no evidence that minuscule doses are harmful, so that multiplication of these is not biologically meaningful. Certain practices are sometimes evaluated on the basis of their potential collective dose, where collective doses less than 1 person-Sv yr^{-1} are deemed trivial. However, many practices are excluded from collective dose estimates. For example, coal-fired power plants are estimated to cause, through the release of primordial radionuclides, collective doses of about 2 person-Sv per GW of energy produced annually.

Given that human activities have the potential to increase radiation dose, national regulatory agencies have developed criteria to limit exposure. Separate criteria are used for exposure to the general public and for occupational exposure to workers. These criteria vary with regulatory agency, but the 1990 recommendations of the ICRP limit occupational exposures to an effective dose of 100 mSv in a 5-yr period, giving an average annual value of 20 mSv, with a limit of 50 mSv in any given year. Annual limits for the public are generally in the range of 1 mSv yr^{-1}. Separate criteria are sometimes used for specific organs, such as the skin or the eye. The desired dose limit can be used to back-calculate the annual limit of intake for a given radionuclide, and recommendations are set for intake through ingestion or inhalation. In addition to these criteria, the nuclear industry generally subscribes to the principle of ALARA, as low as reasonably achievable, with economic and social factors taken into account. This principle optimizes radiation protection while recognizing that sometimes increased radiation dose has social benefits (e.g., medical X-rays) and that economics must be considered if huge expenditures do not appreciably lower doses that are already very low.

B. Effects on Humans

Effects on humans are separated into deterministic effects caused by severe exposures to high doses and stochastic effects (mostly cancer) caused by much lower exposures. Data on deterministic effects come from side effects of radiotherapy, exposure of the early radiobiologists, atomic bomb effects in Japan, and a few severe accidents. Data on stochastic effects are mostly based on epidemiological studies on the survivors of the atomic bomb detonations in Japan, on patients exposed to medical treatments, and on industrial exposures to workers. Animal studies are also used to evaluate human effects.

1. Deterministic Effects

Deterministic, or nonstochastic, effects are caused by direct energy damage to tissues and cells. This is usually caused by a high dose of more than 0.5 Sv delivered over a short time (seconds or minutes). The effects are much less, or negligible, if the exposure is at a much lower dose rate but extended over a longer period (years). This is because sublethal damage is reparable at the cellular level, and new cell proliferation can replace lethally damaged cells. The severe effects include suppression of bone marrow function, damage to the gastrointestinal tract, skin burns, cataracts, and reduced lymphocytes. Symptoms include nausea, pain, vomiting, and reddening of the skin. In some cases, peripheral blood stem-cell transplants have been successful at improving recovery following exposure. Acute doses greater than 10 Sv are often fatal. There are few cases in peacetime in which acute exposures have caused death, especially in recent years. However, in a few instances workers have been exposed to severe radiation, most notably during the Chernobyl accident in 1986 and more recently during a criticality accident at Tokaimura, Japan, in 1999.

2. Stochastic Effects

Stochastic effects include cancer and genetic damage. Cancer results when cells multiply out of control, and certain types of cancer are linked to exposure to ionizing radiation at high sublethal doses. An important feature is that cancer may not be expressed until decades following the exposure. The data from a study of the Japanese bomb survivors, which includes a study group of about 80,000 individuals, indicate a statistical increase in malignancies at doses above 0.2 Sv delivered at high dose rates.

There is a large uncertainty in the incidence of cancer caused by low dose rates in the range of background exposures. For example, no increase in the frequency of cancer has been documented in populations residing in areas of high natural background radioactivity. It is estimated that about 50 billion people would need to be studied for researchers to statistically identify serious health effects caused by exposures of about 0.1 mSv (i.e., about 5% of the annual natural background exposure). Despite these statistical challenges, a linear, nonthreshold hypothesis is used to estimate stochastic effects from radiation exposure. This hypothesis assumes a linear extrapolation of data collected at high doses to zero exposure, so that all doses are assumed to cause an effect that is linearly related to the amount of dose. The ICRP estimates fatal cancer risk at a rate of 0.04 per Sv for adult workers and 0.05 per Sv for the general population (Table II). The greater risk for the general population includes the risk to younger people. This means that a dose of the order of 20 μSv (i.e., 1% of the natural background exposure) is assumed to cause a risk of about one in one million that an individual member of the general public will receive a fatal cancer. An additional risk of nonfatal cancer is also estimated (see Table II).

TABLE II Nominal Probability Coefficients for Stochastic Effects[a]

	Detriment (10^{-2} Sv^{-1})			
Population	Fatal cancer	Nonfatal cancer	Severe hereditary efects	Total
Adult workers	4.0	0.8	0.8	5.6
Whole population	5.0	1.0	1.3	7.3

[a] From International Commission on Radiological Protection (ICRP) (1991). "1990 Recommendations of the ICRP," ICRP Publication 60, Pergamon Press, Oxford, UK.

Although experimental studies on plants and animals have shown stochastic effects on progeny, no hereditary effects have been demonstrated in humans. However, recessive mutations have the potential to affect subsequent generations, so a probability of such effects is assumed for protection from radiation (see Table II).

C. Effects on Nonhuman Biota

1. Plants

Plants actively grow in many areas of high radioactivity, but negative effects are observed only at very high doses. These effects can be expressed as morphological aberrations, such as abnormal leaf growth, but these tend to be minor effects that do not affect the plant population. High radiation can cause reduced growth and yield, loss of reproductive capacity, and eventually death. Plants tend to follow a hierarchy of radiation sensitivity related to the cell nuclear volume, and in particular, the interphase chromosome volume. The result is that simpler plants such as lichens are more resistant than most trees. Early experiments used high acute exposures of radiation, typically delivered in less than 1 day at doses greater than 0.5 Gy hr^{-1}. Acute lethal doses range from 10 to 1000 Gy depending on the species. However, many chronic experiments have also been performed to identify effects both on individual plants and on populations. Much of the evidence of ecological effects of radiation has come from field experiments, where a large gamma source was placed in the field to chronically irradiate the surrounding environment, sometimes for many years. These experiments demonstrated that plant death is related to both total dose and dose rate, with higher dose rates causing a greater impact for the same total dose. When plant populations are considered, dose rates less than 10 mGy day^{-1} have minimal impact on all ecosystems studied, and a safety level of 1 mGy day^{-1} should protect all plant populations. However, it is possible to find subtle effects on individual plants at these dose rates.

There are few examples in which high levels of radionuclides in the environment have been shown to affect plant growth. The severe accidental releases of radioactivity in the southern Ural mountains near Kyshtym in 1957 and near Chernobyl in 1986 caused effects on vegetation mostly from the radiation released in the plume during the accidents, or from short-lived radionuclides causing high dose rates immediately following the accidents. It has been much more difficult to detect chronic effects caused by longer-lived radionuclides, such as ^{137}Cs. Near Chernobyl, radioactivity in forest plant species is generally less than 5000 Bq kg^{-1} wet weight, giving dose rates of the order of 20 mGy hr^{-1}. This is much less than the dose rate where population effects have been established. Also, background dose rates to plants from "natural" radioactivity are estimated to be in the range of 0.7–0.8 μGy hr^{-1}, or 0.6–7 mGy yr^{-1}. It is also difficult to detect radiation effects in industrially contaminated areas, partly because chemical toxicity is often a confounding factor.

In addition to the negative effects of radiation at high doses, there is evidence that a hormetic effect occurs in some plants at lower doses. This tends to increase the yield in crop species when seeds are irradiated with gamma doses up to 10 Gy. However, the growth stimulation is often inconsistent among experiments, with other environmental variables exhibiting control. Therefore the phenomenon has not been widely used commercially.

2. Animals

a. Mammals. Effects on many mammals are similar to those on humans. Except for studies used as human analogues, many of the nonhuman mammal studies investigate impacts on populations, with less concern for impacts on individuals. Severe radiation effects include damage to the gastrointestinal tract and bone marrow failure, because stem cells are very radiosensitive. Acute doses in the range of 10–50 Gy usually cause death in mammals within the first 10 days following exposure because of damage to the gastrointestinal tract. Bone marrow failure usually occurs within weeks of receiving acute doses between 1.6 and 10 Gy. Protraction of the dose generally reduces the extent of nonstochastic effects because of the opportunity for cellular repair and replacement. If mammals survive acute exposure, there tends not to be any other life-shortening effect, except for radiation-induced cancers. Although the data are limited to only a small number of mammal species, there is little evidence that chronic dose rates less than 0.4 mGy hr^{-1} affect mortality from either deterministic or stochastic (cancer) effects.

Potential reproductive impacts are thought to be a more sensitive indicator of impacts on mammals since oocyte viability can affect a whole population. Experiments on continuous irradiation of multiple generations of rodents show no effects on fertility at dose rates between 0.5 and 1 mGy hr^{-1}, and it is unlikely that dose rates of less than 0.1 mGy hr^{-1} to the most exposed individuals of a wild mammal population will affect fecundity. The developing embryo and fetus are radiation-sensitive life stages. Most experiments have been done with rodents, and acute doses as low as 0.05 Gy have shown fetal mortality, with doses of 0.1 Gy impacting both the embryo and the fetus. Most experiments on chronic irradiation indicate that maximum dose rates of less than 0.1 mGy hr^{-1} are unlikely to induce damage to developing embryos that will affect the population as a whole.

Genetic damage causing mutations depends on whether the somatic tissue or germ cells are affected. The damage can cause numerical aberrations (e.g., polyploidy), structural aberrations (e.g., chromosomal deletions), or gene mutations (change in the DNA nucleotide sequence). Some chromosome aberrations in somatic cells also reduce fecundity and are implicitly included in the reproductive studies. Chromosome breakages have been detected in some mammal populations at dose rates between 1 and 10 μGy hr^{-1}, but the impact on the individual is uncertain. Genetic mutations, some of which can cause reproductive effects in subsequent generations, can occur in rodents at chronic exposures of about 5 mGy hr^{-1} or during acute, sublethal exposures. There is some indication of genetic sterility in chronically exposed male wild mice following the Chernobyl accident at dose rates of about 2 mGy hr^{-1}, but most individuals showed only temporary sterility.

b. Birds. Birds are about as equally sensitive as mammals to radiation, with mortality at acute doses in the

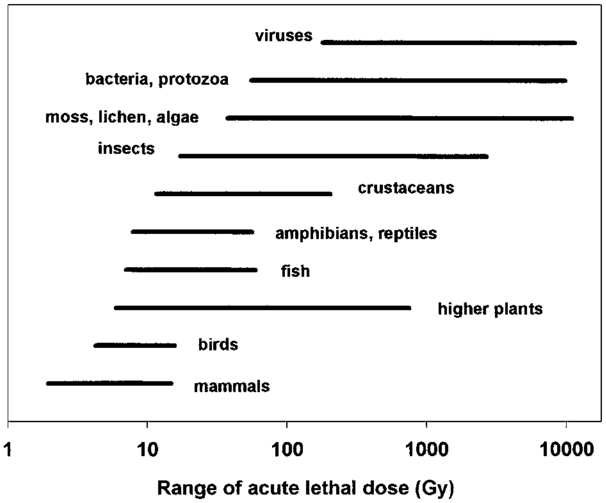


FIGURE 2 Range of acute lethal radiation dose to biota. [Adapted from United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (1996). "Sources and Effects of Ionizing Radiation," United Nations Publication Sales No. E.96.IX.3, United Nations, New York.]

range of 5–12 Gy. Significant depression in growth in tree swallows occurs at acute doses between 2 and 5 Gy, and radiation-exposed birds can have shorter lives. Chronic exposures of between 0.03 and 0.3 mGy hr^{-1} show no apparent effects. Some genetic germ-line impacts have been identified in barn swallows in contaminated areas following the Chernobyl accident, but it is difficult to estimate the dose received by these birds.

c. Reptiles and amphibians. Reptiles and amphibians also have radiation sensitivities close to those of mammals and birds, with mortality typically in the range of 2–22 Gy following acute exposure (Fig. 2). Reptiles tend to be slightly less sensitive because of a slower metabolism and shielding by thick skin or a carapace. Although there are few data on low chronic exposures, some effects have been observed at about 0.2 mGy hr^{-1}. In radionuclide-contaminated environments, amphibians may be sensitive indicators of potential radiation effects since they are in close contact with water, soil, and sediment and their skin often has a respiratory function allowing absorption of gases.

d. Fish. Fish tend to be the most sensitive group of nonmammalian aquatic organisms, with developing fish embryos being the most sensitive stage. Acute irradiation causes mortality at doses of about 10–25 Gy, although some fish can tolerate several hundred gray. Salmon embryos can be killed at doses as low as 0.16 Gy. Reproductive effects in fish from chronic radiation exposure are unlikely at dose rates less than 1 mGy hr^{-1}. Populations of aquatic organisms are not expected to be affected at chronic dose rates of less than 0.4 mGy hr^{-1}.

e. Invertebrates. Invertebrates are much less sensitive to radiation than vertebrates are (see Fig. 2). Acute doses in excess of 100 Gy are often needed to cause death in arthropods (e.g., insects), although some reduction in life span occurs at doses between 20 and 40 Gy. Doses in excess of 20 Gy can affect egg development in some species. Chronic exposures of 10–30 mGy hr^{-1} have no effect on the mortality of many of the larger marine invertebrates (snails, scallops), although some effects on egg capsule production has been observed at less than 1 mGy hr^{-1}. Exposure to 300–400 Gy causes sterilization in some moths; however, programs to release sterile individuals to reduce wild populations are seldom successful unless a sterile-to-wild male ratio exceeds 20 to 1. Radiosensitivity generally decreases from egg to adult stage, largely because of the relative development of tissues. Therefore for a given dose, fecundity and eggs are generally more affected than adults are.

3. Microbes

Microbes are much more radioresistant than higher plants and animals (see Fig. 2). Acute lethal doses to viruses are in the range of 200 Gy to 10 kGy, with protozoa and bacteria being killed at slightly lesser doses. Sterilization of medical products to kill microorganisms usually uses doses of about 25 kGy. In some countries, human food is irradiated to kill microbes and disease vectors, with doses ranging from 0.3 to 10 kGy, depending on the type of food. Background levels of environmental radiation are much lower than dose rates where microbes are expected to show effects, so there is little information on chronic dose effects. However, from the data on acute exposures, it is obvious that microbes can tolerate very high dose rates.

V. THE NUCLEAR FUEL CYCLE

A. Mining and Milling

The nuclear fuel cycle starts with the mining and milling of uranium ore. This activity is almost totally confined to extracting the ore from hard rock, although the oceans also contain large quantities of uranium and thorium in a more dilute form. The average geosphere concentration of uranium is about 2.7 parts per million, and uranium is mined economically when concentrations are between 0.03 and 10%. Known conventional uranium reserves are estimated at about 3 million tonnes, and the current demand is about

60,000 tonnes annually. Over half the uranium is mined in open-pit operations, where it is then crushed, ground, leached, extracted, concentrated, precipitated, and then dried to form "yellowcake." About one-quarter of the uranium is mined in underground shafts, with the remainder being leached from sandstone-based aquifers or gathered as a by-product of other mining operations. Mining and milling operations often result in enhanced local environmental concentrations of uranium, ^{226}Ra, ^{210}Pb, and ^{210}Po, mostly from dust emissions and effluent from the tailings.

B. Reactors

1. Routine Emissions

Nuclear reactors emit small amounts of radionuclides during their normal operations. Atmospheric releases result from noncontainment of all fission products, especially inert radioactive gases of argon, krypton, and xenon. Some reactors also release quantities of ^{3}H, ^{14}C, and ^{131}I to the atmosphere or to aquatic environments. These quantities are judged by comparison with a derived emission limit (DEL). The DEL is calculated on the basis of the annual dose limit for members of the public, such that no individual will receive an annual dose greater than a predetermined amount, typically less than 1 mSv yr^{-1} to the whole body, depending on the country's regulations. The actual DEL will vary with the location and construction of the reactor, since meteorological conditions and release height affect atmospheric dispersion, and dose rate. For example, a single reactor may be allowed to release about 10^{16} Bq of tritium oxide, 10^{12} Bq ^{131}I, and 10^{12} miscellaneous particulate matter annually to still meet the DEL. Typically, most reactors release less than 1% of the DEL, which helps them to satisfy the ALARA principle. At the global scale, only tritium, ^{14}C, ^{85}Kr, and ^{129}I are globally dispersed from routine reactor operations, although they contribute very small doses.

2. Emergency Releases

One of the greatest public fears concerns emergency releases of radionuclides during a severe accident at a nuclear facility. Most reactors have both passive and active safety systems whereby radionuclides are contained within the facility. Releases to aquatic environments, primarily through contaminated cooling water, are normally easily contained before the water reaches the general environment. However, releases to the atmosphere can be more difficult to contain if buildings and safety systems are damaged. Since the 1950s, there have been 14 reactor accidents in which there was damage to the reactor core. However, only the accidents at Windscale, England, in 1957 and at Chernobyl, Ukraine, in 1986 have released more than trace amounts of radioactivity.

The Chernobyl accident is the most severe case of an emergency release, where a reactor core was severely damaged and about 10^{19} Bq of various radionuclides were released to the atmosphere sporadically over a period of about 10 days. These radionuclides reached a sufficient height to enter the general atmospheric circulation of the Northern Hemisphere, such that the radionuclides could be detected throughout Europe, Asia, and North America. Close to the reactor, the radioactive plume caused extensive damage to vegetation in some wind sectors. Farther away, precipitation washed out radionuclides, which caused deposition patterns related to the position of the plume at the time of the precipitation. The initial radiation dose was caused by short-lived radionuclides, especially ^{131}I (half-life = 8 days), so that biota exposed to the plume were most greatly affected. Some radiation dose also resulted during the first few weeks following ^{131}I deposition, where this dose was primarily caused by ingestion of contaminated food and water. Approximately 140,000 residents were evacuated from the contaminated areas immediately following the accident. Following the decay of short-lived radionuclides, most of the contamination was caused by the fission products ^{137}Cs (half-life = 30 yr), ^{134}Cs (half-life = 2 yr), and ^{90}Sr (half-life = 29 yr), with a more localized region where ^{239}Pu (half-life = 24,065 yr) and ^{240}Pu (6537 yr) were deposited. About 25,000 km^{2} of territory in Belarus, Russia, and the Ukraine were contaminated with ^{137}Cs at levels in excess of 1.85×10^{5} Bq m^{-2}. Since this deposition, much of the ^{137}Cs is still retained in the top 10–20 cm of the soil, although it is slowly leaching downward. Remediation efforts in some areas have removed contaminated soil and vegetation. But for most areas, soil decontamination largely depends on radioactive decay.

Work related to biological effects is still ongoing in the Chernobyl area. Approximately 40 people died of direct causes related to the accident, with 28 of these caused by immediate severe radiation injuries. Another 140 people were diagnosed with acute radiation syndrome. There is a greater uncertainty associated with stochastic effects. More than 800 cases of thyroid cancer in children have been diagnosed, most of which are believed to be caused by doses received during the accident. A few of these cases have resulted in death, but this cancer is usually nonfatal if diagnosed and treated early. It is likely that there will be more cases diagnosed in the future, although children born after 1986 do not appear to be at risk. Other cancer effects are not statistically evident, even in the group of "liquidators," some of whom received doses of 1 Sv or more. There have been psychological and traumatic disorders, mostly related to the stress of the accident, relocation issues, the inability to eat locally grown foods, and fear of future radiation effects.

C. Waste Issues

1. High- and Intermediate-Level Wastes

The operation of over 400 power reactors, plus many smaller research and military reactors, has created a legacy of radioactive waste. Used fuel and highly contaminated parts of reactors are classed as high-level waste. The definition of intermediate waste varies among countries and activities, but this waste has a significant risk if it is allowed to freely enter the environment. In some countries, used fuel is reprocessed to recycle some elements, such as plutonium, for use in reactors. This practice reuses nuclear materials but still leaves high- and low-level wastes that require management. Many countries store their high-level wastes in water-filled pools or concrete bunkers to allow the shorter-lived radionuclides to decay. One year of operation of a typical power reactor of 1 GW capacity generates about 200 m^3 of low-level waste, 70 m^3 of intermediate-level waste, and 2.5 m^3 of vitrified high-level waste or 10 m^3 of spent fuel. All countries are searching for permanent solutions to disposal of their high-level wastes.

Disposal of high-level waste into space is deemed impractical because of the large volume of waste, potential high risks and environmental impacts, high energy needs, and high costs. Such disposal would become feasible only with new technologies or separation techniques that would allow a substantial decrease in waste volume.

Transmutation is a nuclear process that transforms one radionuclide into another nuclide, perhaps a stable one or a less toxic one with a different half-life. This requires reprocessing of fuel to separate target radionuclides for transmutation. It also involves an enhanced nuclear industry with available power for the process and potentially creates secondary waste streams. It is also impractical to treat a wide variety of radionuclides by using this method, so at present transmutation will not significantly reduce the risk from high-level wastes.

Geological disposal of high-level waste is being explored by almost every country with nuclear power. Possible geological media include ice sheets, the ocean bed, and terrestrial sites. Deep terrestrial disposal is favored, and countries are exploring crystalline bedrock, salt, shale, volcanic tuff, and clay formations. The concept involves placing a stable form of waste (e.g., ceramic) in sealed containers that resist corrosion and groundwater, surrounding these containers with an impermeable material such as clay, and placing them in underground excavations that are hundreds of meters below the surface. Some concepts suggest continuous monitoring of the waste, whereas others plan to totally seal the underground vaults to minimize transport pathways to the surface or intrusion by humans. Computer simulations show that most of the fission products decay away relatively quickly, and that often the waste will have about the same toxicity as a high-grade uranium ore body by about 10,000 yr following disposal. The geological formation selected should be stable, should minimize radionuclide transport to the surface, and should be available for use by a country. Geological disposal poses the lowest risk compared with that of all other current management options for high-level waste. However, there are many issues still to be resolved related to the location of waste repositories and transportation of highly radioactive materials to central locations. Despite many debates, regulations, siting activities, and government decisions, no country has yet placed high-level waste in a geological repository.

One natural analogue for geological disposal of nuclear waste is at a site near Oklo, Gabon, West Africa. About 2 billion yr ago, a uranium ore body sustained fission for a period of about 100,000 yr. Most of the fission and activation products from this natural reactor migrated only a short distance from the fission sites, which gives natural supporting evidence for the potential of radionuclide transport from underground waste disposal facilities.

2. Low-Level Waste

Low-level wastes originate from many sources, such as the nuclear power industry, medical and academic institutions, and a host of industries that use small amounts of radionuclides in their manufacturing processes. Examples are medical tracers that have been used, biological materials that were radioactively labeled, and plastic and paper products used in radioactive laboratories. Reprocessing of nuclear fuel also creates additional low-level wastes. These wastes generally do not pose severe risks from their radiation fields, but there is a risk associated with radionuclides migrating through the environment and entering the food chain. These wastes also tend to have a relatively large volume but are dilute in radioactivity. Tritium, ^{14}C, and ^{129}I often pose some of the greater risks because they are very mobile, but there are a large number of other radionuclides in these wastes.

Historical management of low-level wastes usually involved shallow burial in unconsolidated materials (e.g., sand), often in unlined trenches. In many cases, groundwater infiltration has transported some of the more mobile radionuclides through the overburden, so there has been some breaching of containment. Also, at some sites, gaseous evasion of ^{14}C can be measured. Current management has much improved containment, with wastes encapsulated in concrete bunkers or silos or placed in lined trenches with impervious caps and leachate collectors to minimize waste movement to the general environment.

D. Other Nuclear Industry Sources

The production of nuclear power requires other operations, such as fuel fabrication. This has caused a few accidental releases of radioactivity, most notably when errors were made that allowed chemical mixtures to reach nuclear criticality. Such situations create a chain reaction for only a short time because of the small amounts of fissile material available. However, workers and anyone else in the vicinity can be exposed to the radiation field or to the fission and activation products, such as ^{131}I or ^{133}Xe (half-life = 5 days). Most of these accidents occurred prior to 1965, but even in 1999 there was a criticality accident at Tokaimura in Japan where three workers were severely injured. One of the most severe environmental releases of radioactivity followed a chemical explosion at the Mayak complex near Kyshtym, Russia, in 1957, where about 74 PBq of fission products were released, which contaminated about 15,000 km^2.

In addition to accidental releases of radioactivity, some processing practices have resulted in chronic contamination of the environment. For example, the Hanford nuclear processing plant in Washington released ^{129}I from its operations prior to 1972, and elevated levels were still found in the air for several years following the plant closure. A processing plant in Karlsruhe, Germany, has also created elevated environmental levels of ^{129}I.

E. Transportation

Radioactive materials are moved throughout the world each day by road, rail, aircraft, and ships. Raw bulk materials such as uranium ore have low toxicity, even in large quantities, and travel from mining and milling sites to fuel fabrication sites. Raw fuel is transported to reactor sites with little risk, since most of the risk comes from fission and activation products created in the reactor. Spent fuel and military inventories are also transported, but these have greater associated risks. Since permanent disposal systems are not yet in place, most of these wastes are retained at the generation sites. Special transportation casks are used for these high-level packages to ensure that radionuclides will not be released in the event of collision or fire.

Small amounts of radioactive isotopes used as scientific and medical tracers are easily shipped, even through the public mail systems. Most of these tracers have small radiation fields associated with them, which can be contained with minor shielding.

VI. CONCLUDING REMARKS

The perception of environmental radioactivity has evolved over the past century, starting with an extended period of scientific exploration until the detonation of the first atomic bomb. During much of the second half of the twentieth century, bomb fallout and the threat of nuclear weapon detonations have dominated discussions of radioactivity. Some countries are decreasing their weapons arsenal, but this still requires a safe method to dispose of the radioactive materials, especially to stop nuclear proliferation. In addition, other countries have been developing nuclear programs and the potential risk of detonations is still real.

The development of commercial nuclear reactors has provided a substantial energy source for economic growth in many countries. However, there continues to be public concern about low amounts of radionuclides released during routine operations of the reactors, despite little evidence of adverse impacts to human health or the environment. Accidental releases are of greater concern because of the potentially larger doses, even if the probability of an accident is low. Further, the Chernobyl accident in 1986 has demonstrated the complexity of identifying adverse effects, even from such a severe accident. Many reactors constructed in the 1960s and 1970s are nearing the end of their planned lives and need to be decommissioned. In addition, alternative means of generating energy need to be developed to replace these aging reactors. The largest present issue deals with the disposal of used nuclear fuel and other high-level radioactive wastes. Society is focusing on closing the cycle of all industrial processes, and the need to resolve the waste issue will likely determine the future of energy generation from fission.

The use of fossil fuels has increased atmospheric carbon dioxide levels to the point where the climate is being affected. Nuclear energy offsets some of these fossil fuels, but it is unlikely that fission technologies will increase rapidly enough to offset the increasing demand for energy, especially in developing countries. A major breakthrough in fusion technology would help fill these energy needs. However, fusion will create radioactive by-products and there will be associated safety and waste issues to address. The general public needs to understand the risks posed by environmental radioactivity and to place these risks in a comparable framework with other risks, such as chemical toxicity, transportation accidents, and fossil-fuel use.

SEE ALSO THE FOLLOWING ARTICLES

COSMIC RADIATION • ENVIRONMENTAL GEOCHEMISTRY • ENVIRONMENTAL MEASUREMENTS • ENVIRONMENTAL

OBSERVATION AND FORECASTING SYSTEMS • ENVIRONMENTAL TOXICOLOGY • NUCLEAR ENERGY, RISK ANALYSIS • NUCLEAR FACILITIES EMERGENCY PLANNING • NUCLEAR SAFEGUARDS • RADIATION, ATMOSPHERIC • RADIATION SHIELDING AND PROTECTION • RADIOACTIVE WASTES • RADIOACTIVITY

BIBLIOGRAPHY

Desmet, G. *et al.* (1997). Freshwater and estuarine radioecology. *In* "Proceedings of the International Seminar on Freshwater and Estuarine Radioecology, Lisbon, Portugal," Elsevier Science, New York.

Draganic, I. G., Draganic, Z. D., and Adloff, J.-P. (1993). "Radiation and Radioactivity on Earth and Beyond," 2nd ed., CRC Press, Boca Raton, FL.

Eisenbud, M. (1987). "Environmental Radioactivity," Academic Press, Orlando, FL.

International Commission on Radiological Protection (ICRP) (1983). "Radionuclide Transformations, Energy and Intensity of Emissions," ICRP Publication 38, Pergamon Press, Oxford, UK.

International Commission on Radiological Protection (ICRP) (1991). "1990 Recommendations of the ICRP," ICRP Publication 60, Pergamon Press, Oxford, UK.

International Commission on Radiological Protection (ICRP) (1995). "Dose Coefficients for Intakes of Radionuclides by Workers," ICRP Publication 68, Pergamon Press/Elsevier Science Ltd., Oxford, UK.

International Commission on Radiological Protection (ICRP) (1996). "Radiological Protection and Safety in Medicine," ICRP Publication 73, Pergamon Press/Elsevier Science Ltd., Oxford, UK.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (1988). "Sources, Effects and Risks of Ionizing Radiation," United Nations Publication Sales No. E.88.IX.7, United Nations, New York.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (1993). "Sources and Effects of Ionizing Radiation," United Nations Publication Sales No. E.94.IX.2, United Nations, New York.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (1996). "Sources and Effects of Ionizing Radiation," United Nations Publication Sales No. E.96.IX.3, United Nations, New York.

Wilkening, M. (1990). "Radon in the Environment," Elsevier Science, New York.

Yamashita, S., and Shibata, Y. (1997). Chernobyl: A decade. *In* "Proceedings of the Fifth Chernobyl Sasakawa Medical Cooperation Symposium," Elsevier Science, New York.

Environmental Toxicology

Edward A. Laws
University of Hawaii at Manoa

GLOSSARY

Acceptable daily intake (ADI) The amount of a food additive that can be ingested daily in the diet without appreciable risk, determined on the basis of all facts known at the time. "Without appreciable risk" refers to the practical certainty that injury will not result, even after a lifetime of experience. The ADI is a practical approach to determining the safety of food additives and is a means of achieving some uniformity of approach in regulatory control. It serves to ensure that the actual human intake of a substance is well below toxic levels. In the United States, ADIs are determined by the U.S. Food and Drug Administration (FDA).

Action level The concentration of a contaminant in a product at or above which the FDA will take legal action to remove the product from the market. When no established action level exists, the FDA may take legal action against the product at the minimal detectable level of the contaminant. Action levels are established and revised according to criteria specified in Title 21, *Code of Federal Regulations,* Parts 109 and 509, and are revoked when a regulation establishing a tolerance for the same substance and use becomes effective.

Acute/chronic ratio The ratio of the concentration or level of a toxic substance or stress that produces toxic effects after a short period of exposure to the concentration or level of the same substance or stress that produces toxic effects after a long period of exposure.

Acute toxicity Toxicity resulting from exposure to a toxic substance or stress for a relatively brief period, typically no more than 48–96 hr, but never more than 10% of the natural lifetime of an organism.

Acute toxicity unit For a given species and a single toxic substance, the 96-hr median tolerance limit (TLm). For a mixture of toxicants, any combination of concentrations that would be expected to kill half the individuals of the same species in 96 hr.

Adverse effect Any effect that results in functional impairment and/or pathological lesions that may affect the performance of a whole organism or that reduces an organism's ability to respond to an additional challenge.

Bio concentration factor (BCF) The biological accumulation factor associated with direct uptake of a substance from the water in the absence of any possible intake through the food chain.

Biological accumulation factor The ratio of the concentration of a substance in one or more tissues of an aquatic organism to the concentration of the same substance in the water in which the organism has been living.

Chronic toxicity Toxicity resulting from exposure to a toxic substance or stress for a relatively long period, at least 10% of the natural lifetime of an organism.

Chronic toxicity unit For a given species and a single toxic substance or stress, the smallest concentration of the toxic substance or level of stress that would be expected to have an adverse effect on the species after a long period of exposure. For a mixture of toxicants or stresses, any combination of concentrations or stresses that would be expected to produce a similar adverse effect after a long period of exposure.

Cold-blooded Of or relating to an organism whose body temperature is determined largely by a passive exchange of heat between the organism and its surroundings.

Conformer An organism that does not regulate its internal state with respect to an environmental variable such as temperature or salinity. An organism that is a conformer with respect to one environmental variable may be a regulator with respect to another.

Criterion continuous concentration The 4-day average concentration of a toxicant not to be exceeded more than once every 3 yr and defined by the EPA to be the minimum of the final chronic value and the final plant value.

Criterion maximum concentration The 1-hr average concentration of a toxicant not to be exceeded more than once every 3 yr and defined by the EPA to be half the final acute value.

Dry weight The weight of an organism, tissue, or sediment sample after removal of water that can be evaporated off at a temperature of 105°C for 24 hr.

Final acute value The concentration of a toxic substance expected to equal or exceed the 96-hr TLm of no more than 5% of the genera in an aquatic ecosystem.

Final chronic value The concentration of a toxic substance expected to exert a chronic stress on no more than 5% of the genera in an aquatic ecosystem.

Final plant value The lowest concentration of a toxic substance that has been shown to reduce the growth of any aquatic plant.

Incipient lethal level The concentration of a toxic substance or level or stress expected to kill 50% of a group of organisms after a theoretically infinite period of time.

Lowest observed adverse effect level (LOAEL) The lowest concentration of a toxic substance or level of stress observed to produce an adverse effect on humans or animals.

Lowest observed effect level (LOEL) The lowest concentration of a toxic substance or level of stress observed to produce on humans or animals an effect that does not fit the definition of an adverse effect.

Maximum tolerated dose (MTD) The highest dose of a toxic substance that does not literally kill experimental animals when it is administered in their diet for a period of time.

Median survival time The time within which 50% of a group of organisms survive exposure to a given concentration of a toxic substance or level of stress.

Median tolerance limit (TLM) The concentration of a toxic substance or level of stress that is tolerated by (does not kill) 50% of a group of organisms after a specified exposure period.

Q_{10} The rate of a reaction or metabolic process at a given temperature divided by the rate of the same reaction or process at a temperature 10 degrees cooler.

Regulator An organism that regulates its internal state with respect to an environmental variable such as temperature or salinity. An organism that is a regulator with respect to one environmental variable may be a conformer with respect to another.

Toxicology Quantitative study of the effects of harmful chemicals or stressful conditions on organisms.

Wet weight The weight of a freshly collected organism, tissue, or sediment sample before any water has had a chance to evaporate.

TOXICOLOGY is the quantitative study of the effects of harmful substances or stressful conditions on organisms. This rather broad field is broken down into three major divisions: economic, forensic, and environmental toxicology. Economic toxicology is concerned with the deliberate use of toxic chemicals to produce harmful effects on target organisms such as bacteria, parasites, and insects. The obvious applications of economic toxicology are in medicine, agriculture, and forestry management, where it is frequently desirable to eliminate or at least control the numbers of various infectious organisms, parasites, and pests. Forensic toxicology is concerned with the medical and legal aspects of the adverse effects of harmful chemicals and stressful conditions on humans. The medical aspects of forensic toxicology concern the diagnosis and treatment of the adverse effects of toxic chemicals and stressful conditions. The legal aspects concern the cause-and-effect relationships between exposure to harmful chemicals or conditions and the effect of this exposure on human health. Environmental toxicology is concerned with the incidental exposure of plants and animals, including humans, to pollutant chemicals and unnatural environmental stresses. Environmental toxicology is related directly to the subject of water pollution, because it is environmental toxicological studies that reveal the quantitative relationships between, for example, the concentration of chemicals found in the water or in aquatic organisms and the effect of these chemicals on aquatic organisms and on persons who drink the water or consume the organisms.

I. THE ROLE OF TOXICOLOGY IN WATER QUALITY MANAGEMENT

The role of toxicology in water quality management is controversial because not everyone agrees, and indeed many people disagree, about what sorts of information are needed to manage water quality and how that information should be applied in setting water quality standards. Much information is available, for example, on short-term exposures that produced fatal effects. While the results of such studies are certainly relevant to water quality management, environmentalists and public health authorities must also be concerned about sublethal effects resulting from long-term exposure to much lower concentrations of a toxic substance. How does one extrapolate from the results of short-term studies concerned with fatal effects when one is trying to set water quality standards that will protect organisms, including humans, from subtle, sublethal effects resulting from long-term exposure?

A second consideration is the question of who or what the standards are intended to protect. Attitudes generally differ depending on whether one is considering humans or other organisms. In the latter case authorities may not be concerned about protecting every individual organism in an ecosystem, but where should they draw the line? Should policies try to ensure the long-term survival of every species, or is it acceptable for some species to become extinct? In the case of humans, should standards protect every individual in the population, or is it acceptable to sacrifice a few subsistence fishermen who eat a great many fish from a contaminated stream or lake or a few individuals who happen to live close to a nuclear power plant?

A third consideration is whether or not there is a threshold level associated with exposure to a particular toxic substance or stressful condition. In other words, is there a level of exposure below which there is no adverse effect? For many toxic substances and stressful conditions it is believed that there is a threshold or no-adverse-effect level, but determining whether such a threshold exists and where it lies may be a difficult task, since subtle adverse effects are inherently difficult to detect. In other cases it is assumed that there is no threshold. One must then decide what level of adverse effects is tolerable. It is often assumed, for example, that there is no threshold associated with the probability of developing cancer as a result of exposure to carcinogenic chemicals. Is it acceptable then for the concentration of certain chemicals in drinking water to cause 100 cases of cancer per year in a nation with a population of 275 million people? This is the sort of question that public health authorities must decide in the case of any substance or condition for which there is presumed to be no threshold level.

A major concern in toxicology today is the relevance of studies with animals, in particular rats and mice, to human health. It has been argued, for example, that the physiology of rats and mice is suffciently different from that of humans that chemicals which are harmful to such animals when administered in large doses over the lifetime of the organism may pose no threat to humans. This argument may have merit. The fact that authorities often rely on studies carried out with experimental animals when they are setting environmental standards for the protection of human health reflects a more general problem, namely that of setting environmental standards in the absence of directly relevant information. The fact is that information on the adverse effects of substances or conditions on humans is often lacking, and because it is unacceptable to carry out controlled experiments on human guinea pigs, authorities must rely instead on other types of information, usually the results of experimental studies with animals. Whether rats and mice are the best human analogue for such purposes is certainly debatable. Rats and mice are easy and relatively inexpensive to maintain in captivity, but monkeys, for example, are physiologically more closely related to humans. In some cases authorities do have some knowledge of the effects on human exposure, the ongoing study of 120,321 persons resident in Hiroshima and Nagasaki when atomic bombs were dropped in August 1945 being a case in point. Of those persons, 91,228 were exposed to radiation from the bombs. Incidents such as this are the exception rather than the rule, however, and in all such cases the degree of exposure is uncontrolled and in some cases difficult to estimate.

A final concern in water quality management is the problem of estimating effects when more than one toxic substance is present in the water. If there is no interaction between the toxic substances, then the toxic substance present in the greatest relative amount will determine the toxicity of the mixture, the presence of the other toxic substances being of no consequence. It is certainly possible, however, that there will be some interaction with respect to toxic effects. Since relatively little information exists on the toxicity of mixtures of toxic substances, establishing water quality standards is difficult in situations when two or more toxic substances are likely to be present at the same time. This situation may often exist in the vicinity of discharges of industrial wastewater. How are the government agencies charged with protecting the environment and human health to decide under what conditions such water is safe?

The foregoing discussion gives some idea of the complex problems that surround the use of toxicological information in setting water quality standards. Certainly not everyone agrees on how these issues should be resolved, but practical considerations have forced at least working resolution, subject to change and further debate. In the following sections the toxicological considerations that come into play when water quality standards are being set are reviewed. How the U.S. Environmental Protection Agency (EPA) has used toxicological information to establish water quality criteria for the United States is examined when appropriate. The use of EPA water quality criteria for illustrative purposes is not to be construed as an endorsement of EPA policy but merely as an example of how a major industrialized nation has dealt with the problem of protecting the quality of its waters.

II. KINDS OF TOXICITY

In a general sense, studies of toxic effects may be assigned to one of two categories. The first category involves studies of the general overall effects of a compound or stress on an organism. Such overall effects are often described as the result of either acute or chronic exposure. Acutely toxic effects are the result of exposure for a relatively long interval, usually 10% of an organism's life span or longer, and generally involve a lingering or continuous stimulus.

The effects of acute and chronic exposure may be either lethal or sublethal. Falling off a building is an example of an acute stress. It may be fatal, but it need not be. It may produce sublethal damage, such as broken bones, which the body repairs within a short time. It may also produce sublethal adverse effects, such as damage to the brain or spinal cord, which the body is unable to repair. The effect of eating fish with a high concentration of mercury for many years is an example of a chronic stress. The effect may be fatal; indeed some people have died from such exposure. The effect may also be sublethal and manifest in the form of brain damage or lack of coordination. The Mad Hatter in Lewis Carroll's *Alice's Adventures in Wonderland* is an example of a person suffering from chronic exposure to mercury. Mercury was used for years in making felt for hats, a practice banned in the United States since 1941.

The second category of toxicity studies involves experiments designed to evaluate specific kinds of toxicity in detail. The tests may be designed, for example, to study the tendency of a toxicant to cause abnormal fetal development (teratogenic tests), to affect the reproductive capacity of an organism, to cause mutations, to produce tumors, to cause cancer, to affect the photosynthetic rates of plants, and so forth. These specific tests are necessary because many of the important effects of toxicants on organisms, particularly at the sublethal level, do not become apparent in standard tests to evaluate overall effects.

A. Sublethal Effects

Concentrations of a toxic substance that are insufficient to kill organisms outright may nevertheless have a devastating effect on the population of the same organism over a sufficient period as a result of sublethal effects. Most sublethal effects may be classified as modifications of or interferences with reproduction, development or growth, and behavior. The aim of setting water quality standards in the United States is to eliminate or at least to minimize the sublethal effects of pollutants. Unfortunately the task of determining the level of stress that causes no significant sublethal effects has proven much more difficult to determine than was once imagined, in part because of the large number of potentially toxic substances and because of the large number of possible sublethal effects. In some cases simply designing an experiment to study certain types of sublethal effects is no trivial matter. A few examples will illustrate the variety and complexity of sublethal effects.

1. Reproduction

Obviously any pollutant that interferes with or blocks the reproduction of a species may completely eliminate that species without having any apparent effect on the adult members of the population. For example, DDE, which is a metabolite of DDT, was implicated as the cause of the reproductive failure of certain fish-eating birds during the 1960s and 1970s. Certain toxicological tests designed to examine the specific effects of DDE on the reproduction of various birds showed convincingly that high levels of DDE in the feed of female birds lowered the estrogen level in the birds and caused them to lay thin-shelled eggs by inhibiting the calcium pump in the oviduct membrane, which

TABLE I Reproduction and Eggshell Thickness Data from Penned Mallard Ducks Maintained for Two Seasons on Feed Containing DDE[a]

DDE added to feed (ppm)	No. of 14-day ducklings per hen (% of control results)		Shell thickness (% of control results)	
	Year 1	Year 2	Year 1	Year 2
10	51	24	92	89
40	35	20	87	86

[a] From Heath, R. G., Spann, J. W., and Kreitzer, J. F. (1969). "Marked DDE impairment on mallard reproduction in controlled studies." Nature (London) **224,** 47–48.

prevented adequate calcium from reaching the inside of the oviduct where the shell is formed. Table I illustrates the results of one such study.

In this particular experiment, which lasted for 2 test ducks began receiving feed containing either 10 or 40 parts per million (ppm) DDE several weeks before the onset of the first laying season. A control group of ducks received feed containing no added DDE. The ducks fed 10 or 40 ppm DDE produced 49–80% fewer 14-day ducklings per hen than the control group did and produced eggshells that were about 10% thinner than the control eggshells.

That fish-eating birds might have encountered food containing as high as 10 ppm DDE during the 1960s is confirmed by the data in Table II. In 1969 northern anchovies, which account for about 92% of the food consumed by brown pelicans in the southern California bight, were found to contain an average of 3.24 ppm DDE on a fresh weight basis. On a dry weight basis this concentration is equivalent to about 9.7 ppm and is therefore virtually identical to the concentration of 10 ppm dry weight that Table I indicates reduced the reproduction of mallard ducks by 50–80%. The sharp drop in anchovy DDE content between 1969 and 1970 coincides with the time when the Montrose Chemical Company, located in Los Angeles County and the sole manufacturer of DDT in the United States, began disposing of its liquid wastes in a sanitary landfill rather than discharging the wastes into the municipal sewer system. The DDE content of the anchovies then remained constant for about 3 yr, but dropped sharply between 1972 and 1973 when the EPA restriction on DDT use in the United States went into effect. The EPA restriction resulted in no small part from the examination of toxicological results such as those shown in Table I, which convincingly demonstrated the adverse effects of DDE on the reproduction of aquatic birds.

TABLE II Geometric Mean Residues of DDE in Anchovies Off the Southern California Coast[a]

Year	Anchovy whole bodies (ppm, fresh weight)
1969	3.24
1970	0.84
1971	0.87
1972	0.74
1973	0.18
1974	0.12
1975[b]	—

[a] Data from Anderson, D. W. (*et al.* (1975). "Brown pelicans: Improved reproduction off the southern California coast." *Science* **190,** 806–808; Anderson, D. W., Jurek, R. M., and Keith, J. O. (1977). "The status of brown pelicans at Anacapa Island in 1975." *Calif. Fish Game* **63,** 4–10.

[b] Data for 1975 from Anacapa and Santa Cruz Islands only.

The correlation between DDT use and the reproductive success of the southern California brown pelican is illustrated in Fig. 1. In 1969 and 1970 only 9 fledglings were produced from 1852 nesting attempts, and the thickness of intact eggshells averaged about 0.40 mm. In 1971 and 1972, following the diversion of Montrose's liquid waste to a sanitary landfill, reproduction improved to 249 fledglings out of 1161 nesting attempts, and eggshell thickness increased to about 0.45 mm. In 1973–1975, following the restriction on DDT use in the United States, reproductive success improved further to 1575 fledglings out of 2175 nesting attempts, and intact eggshell thickness averaged 0.50 mm. The most recent data have shown almost no evidence of crushed eggs due to thin eggshells, and by 1986–1987, eggshell thickness was averaging about 0.54 mm, within 5% of the pre-1947 norm of 0.57 mm.

The number of brown pelicans in the southern California bight colonies increased dramatically during the 1970s, in large part due to the increase in reproductive success of the adult birds. Since 1985 the pelicans have built an average of 5000 nests per year, up from an average of ∼1200 per year during the previous decade. However, since 1984 the fledgling rates of these colonies have averaged only about 0.63 per nest compared with an average of 1.0 per nest at similar colonies in the Gulf of California. It has been estimated that perhaps 25% of the adult southern California population in 1980 was the result of immigratiom from colonies in southern Baja California or the Gulf of California. Had the disastrous reproductive failures in the years immediately preceding the EPA's restriction on DDT been allowed to continue for much longer, the southern California brown pelican would have very likely become extinct. Although some of the population decline during the late 1960s may have been due to adult mortality from direct poisoning, the major cause of the decline was reproductive failure. Fortunately, careful toxicological studies of the sublethal effects of DDE on bird reproduction had been carried out and were presented

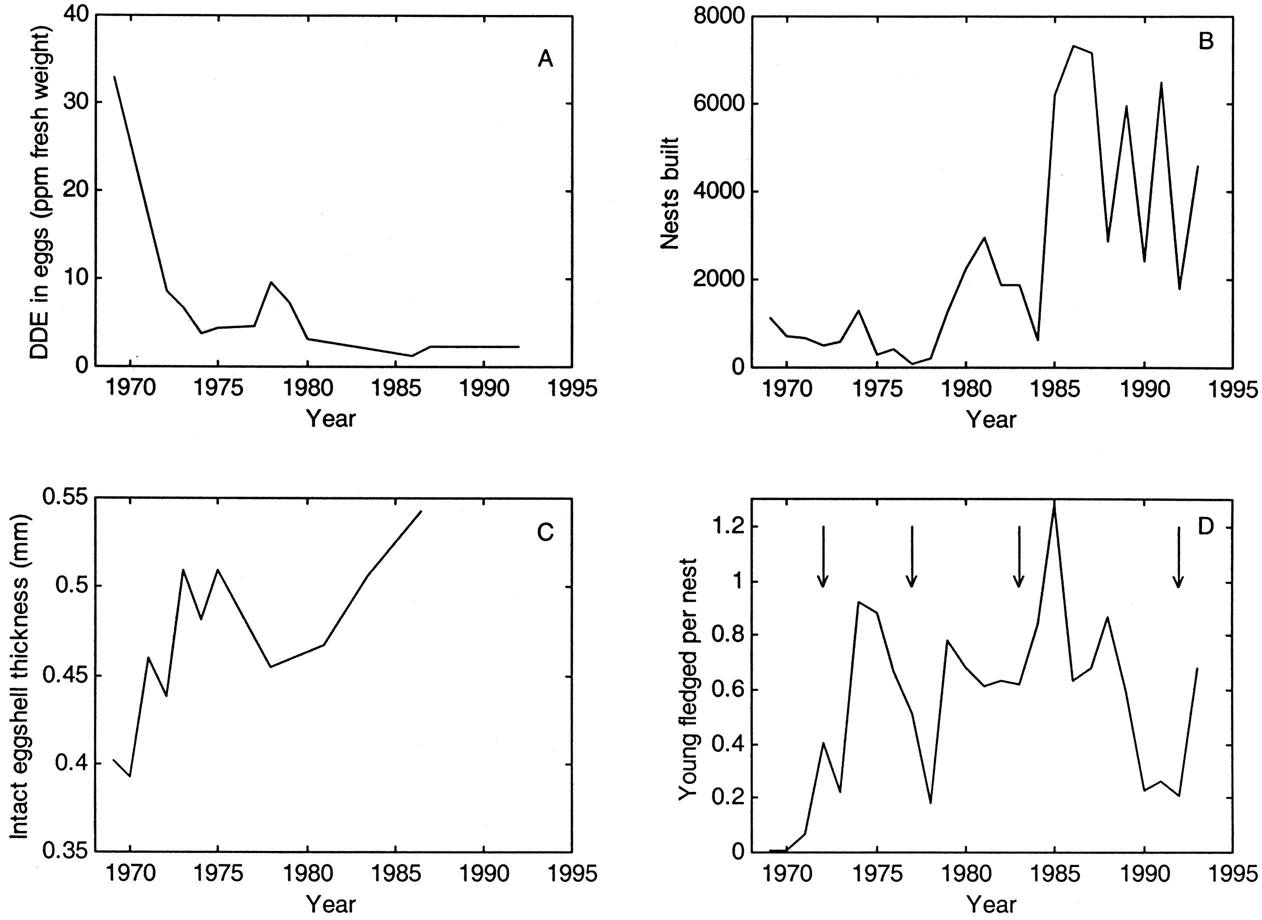


FIGURE 1 Recent history of DDE concentrations in southern California brown pelican eggs,eggshell thickness, and pelican reproductive success as reported by Anderson, D. W. *et al.* (1975). "Brown pelicans: Improved reproduction off the southern California coast." Science **190,** 806–808; Anderson, D. W., Jurek, R. M., and Keith, J. O. (1977), "The status of brown pelicans at Anacapa Island in 1975." *Calif. Fish Game* **63,** 4–10; Gustafson, J. R. (1990). "Five-Year Status Report for California Brown Pelican," Calif. Dept. Fish and Game, Nongame Bird and Mammal Sect., Rept. 90-4 (draft); and Gress, F. (1995)." Organochlorines, Eggshell Thinning, and Productivity Relationships is Brown Pelicans Breeding in the Southern California Bight," Ph.D. dissertation, Univ. of California, Davis. In panel d, vertical arrows indicate the years during which El Niño events produced unusually warm water conditions off the southern California coast.

to the EPA at the time of the 1972 DDT hearings. These studies did much to influence the EPA decision to restrict DDT use in the United States.

In recent years reproductive rates of the southern California bight brown pelicans have been strongly influenced by fluctuations in the local availability of anchovies, whose abundance typically declines during El Niño years (Fig. 1d). However, the fact that fledgling rates in these colonies are below average rates observed at colonies outside the southern California bight suggests that the former colonies may still be under some unnatural stress. The most recent analyses have shown that southern California brown pelican eggs still contain measurable concentrations of both DDE and polychlorinated biphenyls (PCBs), about 2 ppm and 1 ppm, respectively, on a fresh weight basis. The latter are believed to be at least as potent as DDE with respect to their adverse effects on reproduction.

2. Development and Growth

Interference with the normal physiological processes of an organism may adversely affect its ability to grow and develop without directly killing the organism. Nevertheless, because of the pressure of competition and predation, a population of organisms whose development or growth has been retarded may be rapidly eliminated from an ecosystem. For example, studies have shown that exposure of certain species of marine phytoplankton to mercury concentrations as low as 0.5 parts per billion (ppb) can cause more than a 50% reduction in photosynthesis. Figure 2 illustrates some of the experimental

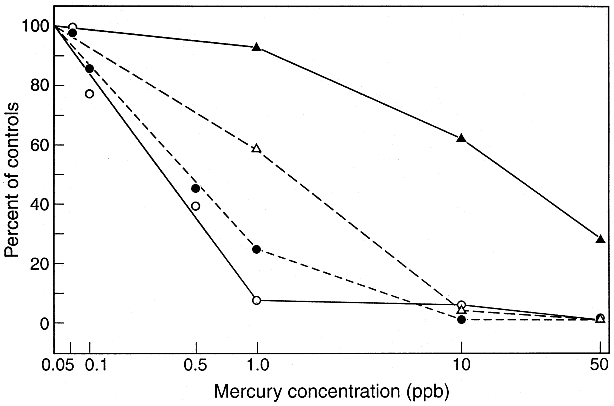


FIGURE 2 Photosynthesis by the marine diatom *Nitzschia delicatissima* after 24 hr of exposure to the following mercurials: ▲, diphenyl mercury; △, phenylmercuric acetate; ●, methylmercury dicyandiamide; ○, MEMMI (a compound containing methylmercury). [Reprinted with permission from from Harriss, R. C., White, D. B., and MacFarlane, R. B. (1970). Mercury compounds reduce photosynthesis by plankton. *Science* **170:** 736–737. © 1970 AAAS.]

results obtained with the diatom *Nitzschia delicatissima*. If the presence of a toxin in the water significantly reduces the photosynthetic rate of a particular species, it is possible that the species will be grazed to extinction by herbivores without being directly killed by the toxin. Similarly, PCB concentrations as low as 0.1–10 ppb have been found to reduce phytoplankton division rates and/or photosynthetic rates. The effect of PCBs on phytoplankton differs markedly between species, some species being much more sensitive than others. As a result, the species composition of a phytoplankton community may be greatly altered due to the presence of PCBs in the water at concentrations as low as 0.1 ppb. Studies conducted in a natural estuarine marsh showed that exposure to as little as 1.0 ppb PCBs reduced phytoplankton production and caused a shift in the size structure of the phytoplankton community toward smaller cells. This shift in the mean cell size of the phytoplankton could effectively increase the number of trophic levels between the primary producers and commercially useful fish and/or could divert more production toward various gelatinous predators such as jellyfish. Thus, the sublethal effects of a toxin on organisms at the base of the food chain can have a profound effect on the structure of the entire food chain and may significantly affect the productivity of the system with respect to commercially useful organisms

One of the best-documented effects of a stress on the development and growth of aquatic organisms is that of temperature. All organisms grow most efficiently over a limited temperature range. Temperatures above or below the optimum range lead to reduced growth rates or to reduced growth efficiency. Figure 3 illustrates in a qualitative way how temperature can influence the growth rate and efficiency of an organism. In the figure, A_c is the amount of food consumed by the organism, and A_g is the amount of that food converted into biomass (growth). A_w is the amount of food that is excreted (waste), while A_a, A_d, and A_s represent the amounts of food that are respired to support the activity of the organism (A_a), the degradation and metabolism of consumed food (A_d), and standard or basal metabolism (A_s). It is clear from Fig. 3 that there is some temperature at which growth, A_g, is a maximum, because the rate of food consumption, A_c, peaks at a certain temperature, whereas the basal respiration rate, A_s, rises exponentially with temperature. This dependence of A_c and A_s on temperature is characteristic of the behavior that might be expected from a poikilothermic (cold-blooded) organism. The amount of food converted to biomass is positively correlated with temperature only as long as the difference between A_c and A_w increases more rapidly than A_s.

Figure 3 illustrates one of the difficult problems encountered in converting toxicological information into water quality standards. In all cases shown, it is clear that a small deviation in temperature from the optimal value will result in some decrease in growth, but this effect by itself will not be lethal. However, the effect of a change in temperature on growth is obviously dependent on the availability of food to the organism. A change of temperature that would cause only a modest reduction in growth when the food

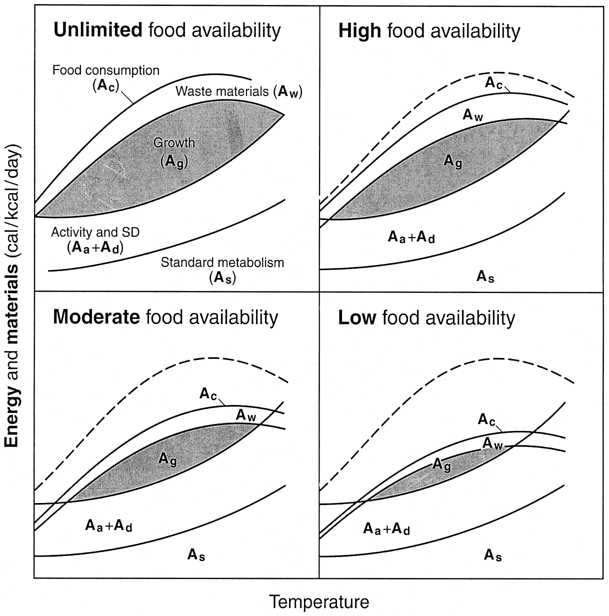


FIGURE 3 Effect of temperature and food availability on food consumption (A_c) and the partitioning of consumed food for various purposes by a poikilothermic animal. The curves are qualitative and purely theoretical. [Reprinted with permission from Warren, C. E. (1971). "Biology and Water Pollution." Saunders, Philadelphia. © 1971 by W. B. Saunders Co.]

supply is unlimited could be lethal when food availability is low. Furthermore, although a modest reduction in growth might not by itself be lethal, in a highly competitive system, even a small reduction in growth could lead to the elimination of a species through the combined effects of competition and predation. Similarly, the ecological impact of a given temperature stress might be far greater in a system in which the organisms were already subjected to other forms of stress, such as the presence of toxic chemicals, than in a system in which no other stresses existed. Thus, translating apparently straightforward toxicological information into water quality standards can be quite difficult.

3. Behavior

Modification in behavior patterns involves a wide variety of effects, such as alterations in migratory behavior, learning ability, feeding behavior, predator avoidance, and so forth. A few examples will illustrate some of the kinds of problems that pollutants may cause as a result of behavioral modification.

Salmon migrate from the ocean back to the freshwater stream where they were born in order to spawn. Although the mechanism that these fish use to find their way from the open ocean back to the general vicinity of their home stream is not well understood, experiments have shown clearly that salmon use their sense of smell to guide their migration upstream once they have entered freshwater. The chemical substances that guide this migration have been shown to include certain volatile organic compounds. Unfortunately, a number of streams once characterized by large salmon runs have become seriously polluted with industrial wastes, particularly in the northeastern United States. In many of these streams salmon runs are now no longer observed or they involve insignificant numbers of fish. There is good reason to believe that the presence of industrial wastes, including a variety of organic substances, in these streams has altered the characteristic smell of the water. As a result the fish find it difficult to identify their home stream. Thus, the presence of chemicals in the water at concentrations well below lethal levels may completely disrupt the migratory patterns of the salmon and, through this behavioral modification, effectively eliminate the species.

Seasonal changes in water temperature can lead to a variety of natural behavioral responses in aquatic organisms. These responses include the onset of reproductive cycles and migratory behavior. The discharge of heated wastewater from large electric power plants can completely disrupt these temperature cues and therefore seriously interfere with natural behavior patterns. For example, menhaden found in the coastal waters of New Jersey during the summer months normally migrate to warmer waters off the North Carolina coast during the winter. At Barnegat Bay, New Jersey, the Jersey Central Power and Light Company operates a nuclear power plant that discharges cooling water at a temperature about 14°C above ambient. During the winter, large numbers of menhaden are often found in the waters near the Oyster Creek outfall, either because they are attracted naturally to the warm water in the winter or because, finding themselves in the warmer waters near the outfall, they do not receive the proper temperature cue to migrate south. In this case the migratory behavior of the menhaden is altered completely by the presence of the heated effluent water, which by itself has no adverse physiological effect on the fish. In fact, during the winter months the temperature of the effluent water is undoubtedly much closer to the optimum temperature for the menhaden than is the temperature of the surrounding water. Unfortunately, the water temperature near the outfall can drop quickly back to ambient levels if the power plant is shut down. Such a shutdown occurred on January 28, 1972, and as a result 100,000–200,000 menhaden died of thermal shock. In a similar incident in Long Island Sound, approximately 10,000 bluefish that had been attracted to the heated plume from the Northport power plant died from thermal shock when winds and tidal currents casued the plume to suddenly shift position on January 17, 1972. Similar incidents can be anticipated wherever the heated wastewater from power plants attracts aquatic organisms during the colder months of the year.

III. DETERMINATION OF TOXICITY

One can perhaps imagine a set of water quality criteria based on essentially two types of standards, the first a concentration or stress level that should rarely or never be exceeded and the second a concentration or stress level whose long-term average value should not be exceeded. In both cases one would presumably be interested in protecting organisms from any form of stress, whether lethal or sublethal. To examine the adverse effects of long-term exposure, one must obviously keep the test organisms alive for a long time. For practical reasons then, the study of chronic toxicity has almost always involved the study of sublethal effects, and ideally the period of exposure has been the natural lifetime of the organism. Studies of shorter duration are acceptable when a long-term study is impractical, but in the United States the duration of exposure must be $\geq$10% of an organism's natural lifetime to qualify as a chronic toxicity study. Acute toxicity studies in the United States generally involve exposure to a stress for 48–96 hr. The length of exposure for purposes of determining acute toxicity is arbitrary, but for practical purposes the line needs to be drawn somewhere in order to facilitate comparison between studies. In practice acute toxicity studies have generally involved the observation of lethal effects (i.e., the organisms are dead after 48–96 hr). There is no compelling reason why this should be the case, for, as noted, the real concern should be sublethal effects resulting from acute exposure. Therefore some allowance must be made for the fact that acute toxicity studies have generally involved lethal stresses and that the goal of water quality management is to protect organisms from both lethal and sublethal effects.

Regardless of the type of toxicity test being carried out, some allowance must be made for the natural variability within a species to the toxic substance or stress being studied. Because of the natural variability among members of a given species, any sample of that population can be expected to contain individuals that differ in their tolerance to a given stress. How then can the response of the species to the stress be quantified? Should one use the response of the most sensitive individual or the least sensitive individual, or should one use the mean or median response? In the United States, the tendency has been to use the median response of a sample of individuals of a given species as a measure of the tolerance of that species to a stress. Whether one considers this policy conservative or liberal depends to a certain extent on one's frame of mind. For example, suppose one were interested in knowing how long a particular species could withstand a given stress before the stress became fatal. A polluter might argue that the time required to kill all the members of the species should be termed the survival time and that the median survival time was unduly conservative because half the members of the species were still alive at the end of the experiment. An environmentalist might argue that the time during which all members of the species survived should be termed the survival time and that the median survival time was unduly liberal because half the members of the species were dead at the end of the experiment. The argument in favor of using the median response is in part statistical. Median tolerance levels are far more reproducible from one subgroup of a species to the next than are the tolerance levels of the most sensitive and least sensitive members of a subgroup. Furthermore, in the United States, water quality criteria are ultimately based on the tolerance levels of the 5% of the genera that are most sensitive to a stress. In practice establishing water quality criteria in this way protects virtually all members of most genera, not just the 50% who happen to be most tolerant.

Exactly how acute and chronic toxicity studies are carried out and how the information obtained from these studies is incorporated into water quality criteria are examined in some detail in the following paragraphs.

A. Acute Toxicity Determination

As previously noted, acute toxicity studies have generally involved studies of lethal toxicity. Typically organisms are placed in a series of aquariums, each containing water with a known concentration of a toxic substance. The range of concentrations is chosen, on the basis of experience and/or educated guesswork, to vary from concentrations too low to have any effect on survival to concentrations expected to kill all the organisms during the exposure period. The survival of the organisms is noted after either 48 or 96 hr. The percentage survival of the test organisms is then plotted against the concentration of the toxic substance. By graphical or other means, the toxicant concentration corresponding to 50% survival is determined. This toxicant concentration is called the median tolerance limit, or TLm. In some cases the abbreviations TL50 (tolerance limit corresponding to 50% survival) or LC50 (lethal concentration corresponding to 50% mortality) are used to designate the same parameter. In a more general sense the abbreviation EC50 is often used to designate the concentration that produces a specified effect on 50% of the organisms after a certain exposure period. In all cases the symbol should be preceded by a designation of the associated time period. Thus, a 96-hr TLm would be the concentration of a toxicant associated with 50% survival after 96 hr.

Figure 4 illustrates the calculations of a 96-hr TLm. The data are taken from a study of the toxicity of copper

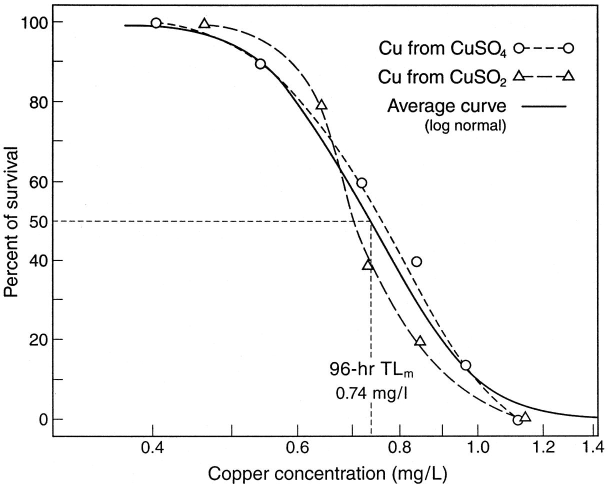


FIGURE 4 Percentage survival of bluegill fish after 96 hr versus copper concentration. The smooth curve is a log-normal curve adjusted to give the best fit to the combined $CuSO_4$ and $CuCl_2$ data. The intersection of this curve with the horizontal line at 50% survival gives the median tolerance limit (TLm) after 96 hr: 0.74 mg/L. [Reprinted with permission from Trama, F. B. (1954). The acute toxicity of copper to the common bluegill (*Lepomis macrochirus* Rafinesque). *Acad. Natl. Sci. Phila., Notulae Naturae* 257. ©Academy of Natural Sciences of Philadelphia.]

to bluegill fish. The fish were divided into two groups and exposed to various concentrations of either CuSO$_4$ or CuCl$_2$ for 96 hr. Percentage survival was then plotted against time. The two data sets were very similar, and a smooth curve was therefore drawn through the combined results. The intersection of this curve with the horizontal line at 50% survival corresponds to the TLm, which in this case was 0.74 mg/L.

B. Chronic Toxicity Determination

In some respects chronic toxicity studies are conducted in the same way as acute toxicity studies, but the former require much longer time periods. Ideally studies would last throughout the life cycle of the organism, and when such studies are feasible, this approach is the one recommended by the EPA. Partial life cycle studies are acceptable for fish that require more than a year to reach sexual maturity, and early life stage toxicity tests may be used to establish water quality criteria if complete or partial life cycle tests with a given species are unavailable. In the latter case exposure should begin shortly after fertilization and should last 28–32 days through early juvenile development, or 60 days posthatch for salmonids. According to the EPA guidelines these chronic toxicity studies should be conducted in flow-through systems, and toxicant concentrations must be measured in the test solutions. In the case of daphnids (commonly referred to as water fleas), renewal systems may be substituted for flow-through systems. Obviously such tests involve considerably more time and effort than that required by the 48- or 96-hr bioassays conducted for acute toxicity determination.

In the case of lethal effects, the goal of chronic toxicity studies is to estimate the stress level that would kill half the test organisms after an exposure period approximately equal to their natural lifetime. Since in practice it is often impractical to monitor the test organisms for such a long time, survival may be monitored for shorter time periods and the results extrapolated to time infinity. This approach leads to two important toxicological concepts: median survival times and incipient lethal levels.

C. Median Survival Times

The median survival time is the time that half the organisms in a random sample from a test population are able to survive a given level of stress. Figure 5 illustrates how median survival times are calculated. The data are taken from a study of the toxicity of low oxygen levels to juvenile brook trout (Shepard, 1955). The trout were conditioned for a period of time in water containing 10.5 mg/L dissolved oxygen and were then placed in experimental aquariums having lower concentrations of dissolved oxygen. The survival of the trout was then monitored over 2000 min. As each fish died, it was removed from its aquarium and its time of death was recorded. Ten fish were studied at each oxygen level. The results were plotted on semilog paper, and the data were fit with straight lines. At the two highest oxygen levels, over half the fish survived for the entire 2000 min, and no effort was therefore made to estimate median survival times at these oxygen concentrations. However, at the eight lowest oxygen levels, over half the fish died during the study period, and estimates of median survival time were made from the intercept of the regression lines fit to the data with the horizontal line at 50% mortality.

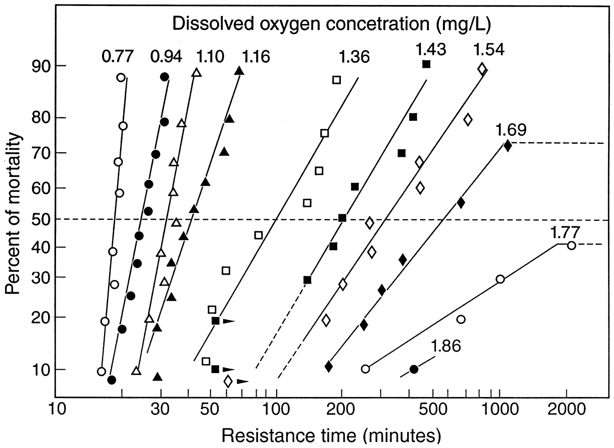


FIGURE 5 Percentage mortality versus resistance time for juvenile brook trout exposed to the indicated concentrations of dissolved oxygen after being conditioned to 10.5 mg/L oxygen. [Reprinted with permission from Shepard, M. P. (1955). Resistance and tolerance of young speckled trout (*Salvelinus fontinalis*) to oxygen lack, with special reference to low oxygen acclimation. *J. Fish. Res. Bd. Canada*, **12,** 387–446. ©1955 Fisheries Research Board of Canada.]

D. Incipient Lethal Levels

Once median survival times have been determined at a series of stress levels, the median survival times are plotted against the stress level and the results extrapolated to an infinite survival time, assuming of course that the data permit such an extrapolation. The hypothetical concentration of the toxicant corresponding to an infinite median survival time is called the incipient lethal level. Figure 6 illustrates how incipient lethal levels are calculated. The data are taken from a study of the toxicity of zinc to rainbow trout. Median survival times were calculated as a function of zinc concentration at three levels of water hardness. Note that both axes are plotted on a logarithmic scale. At water hardness levels of 50 and 320 mg/L, the data show distinct curvature, and a function of the form $(C - C_\mathrm{i})^n(T - T_\mathrm{i}) = K$

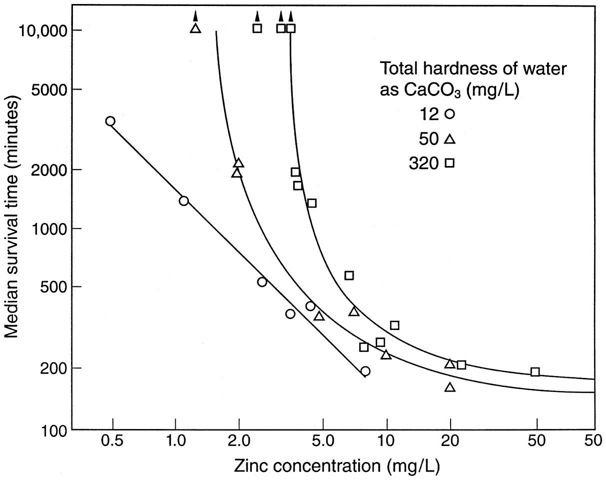


FIGURE 6 Median survival time of rainbow trout as a function of zinc concentration at three levels of water hardness. [Reprinted with permission from Lloyd, R. (1960). The toxicity of zinc sulfate to rainbow trout. *Annals Appl. Biol.*, **48,** 84–94. © 1960 Biochemical Society, London.]

was fit to the data. In this equation C is the concentration of zinc corresponding to median survival time T. As T approaches infinity, C approaches C_i, and C_i is therefore interpreted as the incipient lethal level. The values of C_i estimated from this curve fitting were 3.5 and 1.5 mg/L at water hardness levels of 320 and 50 mg/L, respectively. At a water hardness of 12 mg/L, it was impossible to estimate an incipient lethal level because the data were linear over the range of concentrations studied. All that may be said is that the incipient lethal level at 12 mg/L hardness is less than 0.5 mg/L zinc.

It should be obvious that the incipient lethal level is a hypothetical concept because no organism survives for an infinite time, regardless of the stress level. However, the incipient lethal level does provide a standard measure of the concentration or stress level that produces no adverse effects on survival.

E. Sublethal Effects

As already noted, a wide variety of sublethal effects could be examined in chronic toxicity studies. Ideally the procedure is to expose the test organisms to a given stress level for their natural lifetime and to monitor their behavior, development and growth, and reproduction during the exposure period. The results are compared with those from a control group, which is exposed to no stress. The treatment (i.e., exposure to the given concentration of toxicant or stress level) is considered to have adversely affected the organisms if a standard statistical test indicates that the performance of the stressed organisms was significantly worse than that of the control group. “Worse” in this context means, for example, that the treatment group produced fewer viable offspring, grew more slowly, or developed or behaved in an aberrant way.

TABLE III Chronic Toxicity of Four Compounds to *Daphnia magna*[a]

	Chronic values (ppb)		
Compound	Survival of adults	Reproduction	Mean chronic value
1,1,2-Trichloroethene	32,000	18,000	24,000
Dieldrin	100	32	57
Pentachlorophenol	180	320	240
3,4-Dichloroaniline	56	5.6	18

[a] Data from Adema, D. M. M. (1978). “*Daphnia magna* as a test animal in acute and chronic toxicity tests.” *Hydrobiol.* **59,** 125–134.

Results for a given species or genus are generally reported as mean chronic values, which means that they are the average of the stress levels that produced adverse effects on the organisms in chronic toxicity bioassays. Since there are many effects that might be examined and since species will invariably be more sensitive with respect to some kinds of effects than with respect to others, the particular effects that were studied can influence to some extent the mean chronic value for the species. For example, Table III lists the values of four pollutants which have been shown to adversely affect reproduction and the survival of adult *Daphnia magna* in chronic toxicity studies. In the case of reproduction, the reported values are the smallest concentrations that had a statistically significant adverse effect on reproduction in paired comparisons with a control group subject to no stress. The difference in the chronic values that affected reproduction and survival is as much as a factor of 10. Obviously one would reach rather different conclusions if one estimated chronic levels of these pollutants for *D. magna* solely on the basis of reproduction or solely on the basis of adult survival. Standard EPA practice is to average the reported chronic values by calculating their geometric mean, which is the nth root of n numbers. In this case, since there are two chronic values reported for each toxicant, the geometric mean is the square root of the product of the two numbers. The geometric means so calculated are listed in the last column of the table. It is noteworthy that the mean chronic value of 18 ppb for 3,4-dichloroaniline is over three times the concentration reported to adversely affect reproduction.

IV. WATER QUALITY STANDARDS

Once acute and chronic toxicity information is available for a sufficient number and variety of aquatic organisms, the process of establishing water quality criteria can begin. The rationale and procedures for determining these water quality guidelines in the United States are described in guidelines promulgated by the EPA. In establishing these guidelines, the EPA has not attempted to suggest standards that would make the water safe for every organism. The rationale behind the guidelines is summarized as follows:

> Aquatic communities can tolerate some stress and occasional adverse effects on a few species, and so total protection of all the species all of the time is not necessary. Rather, the Guidelines attempt to provide a reasonable and adequate amount of protection with only a small possibility of considerable overprotection or underprotection. Within these constraints, it seems appropriate to err on the side of overprotection. [EPA (1980c), p. 79342]

A. Acute Effects

Within the context of the EPA's statement, it is now useful to examine the minimum database that the EPA believes is needed to determine acutely toxic concentrations. The rationale behind establishing a minimum database is as follows:

> Results of acute and chronic toxicity tests with a reasonable number and variety of aquatic animals are necessary so that data available for tested species can be considered a useful indication of the sensitivities of the numerous untested species. [EPA (1980c), p. 79343]

The minimum database differs for freshwater and marine organisms and is stipulated as follows in the EPA guidelines (EPA, 1985b). The criterion in freshwater must be based on acute tests with freshwater animals in at least eight different families including all the following categories: (1) salmonid fish, (2) nonsalmonid fish, (3) a third vertebrate family, (4) planktonic crustaceans, (5) benthic crustaceans, (6) insects, (7) a family not included among vertebrates or insects, and (8) a family in any order of insect or any phylum not already represented. For marine organisms, the criterion should be based on acute tests with saltwater animals in at least eight different families subject to the following five constraints: (1) at least two different vertebrate families are included, (2) at least one species is from a family not included among the vertebrates and insects, (3) either the Mysidae family (crustaceans commonly referred to as opposum shrimps and resembling miniature crayfish) or the Penaeidae family (certain shrimp) or both are included, (4) there are representatives from at least three other families not included among the vertebrates, and (5) at least one other family is represented.

If studies have been carried out on an appropriate number and variety of organisms, it may be possible to establish what the EPA (1985b) refers to as the final acute value of the given toxicant, as long as the duration of exposure to the toxicant was appropriate. The guidelines require that results be based on 48- to 96-hr exposures, the recommended duration depending on the type of organism and the nature of the physiological response. In some cases acute values are based on TLm's, but EC50 values are used for effects such as decreased shell deposition in oysters. For each species for which satisfactory studies with a particular toxicant have been performed, the acute concentration is calculated as the geometric mean of the reported values. When available, the results of flow-through tests in which the toxicant concentrations were measured are used to determine the geometric mean. When no such data exist, the geometric mean is calculated from all available acute values, including flow-through tests in which the toxicant concentration was not measured and static and renewal tests in which the initial total toxicant concentration was specified.

Once mean acute concentrations have been determined for the appropriate number and variety of species, the species are grouped by genera, and genus mean acute values are calculated as the geometric mean of all the species mean acute values for each genus. The genus mean acute values are then ranked from lowest to highest, and a graphical technique is used to estimate the concentration of the toxicant that would exert an acutely toxic effect on no more than 5% of the species. The genus ranks are converted to cumulative probabilities by dividing each rank by $N+1$, where N is the number of genera in the list. The four genera with cumulative probabilities closest to 0.05 are then used to determine the final acute value. A plot is made of the logarithm of the genus mean acute values for the four genera against the square root of the cumulative probability, and a model II geometric mean least-squares regression line is fit to the data. The concentration of the toxicant corresponding to a cumulative probability of 0.05 on the regression line is taken to be the final acute value.

Table IV and Fig. 7 illustrate how this procedure would be used to establish a final acute value for the pesticide dieldrin in salt water at the time the EPA water quality criteria for dieldrin were promulgated. In this case acute values were available for 21 species, the range of values being about a factor of 70. The least sensitive species was the grass shrimp *Palaemonetes vulgaris*, with a TLm of 50.0 ppb, and the most sensitive species was the pink shrimp *Penaeus duorarum*, with a TLm of 0.7 ppb. Table IV includes species from 19 genera, and the cumulative probabilities associated with the four most sensitive genera are therefore 0.05, 0.10, 0.15, and 0.20. The genus mean acute values and the least-squares regression line are shown in Fig. 7. The regression line actually gives an excellent fit to the entire data set, although it was fit to only the four lowest genus mean acute values. According to the

TABLE IV Dieldrin Genus Mean Acute Values in Salt Water[a]

Species	Species mean acute value (ppb)	Genus mean acute value (ppb)
Sphaeroide maculatus, northern puffer	34.0	34.0
Crassostrea virginica, eastern oyster	31.2	31.2
Mugil cephalus, striped mullet	23.0	23.0
Palaemonetes vulgaris, grass shrimp	50.0	20.7
Palaemonetes vulgaris, grass shrimp	8.6	
Morone saxatilis, striped bass	19.7	19.7
Pagurus longicarpus, hermit crab	18.0	18.0
Gasterosteus aculeatus, threespine stickleback	14.2	14.2
Palaemon macrodactylus, Korean shrimp	10.8	10.8
Cyprinodon variegatus, sheepshead minnow	10.0	10.0
Crangon septemspinosa, sand shrimp	7.0	7.0
Fundulus heteroclitus, mummichog	8.9	6.7
Fundulus majalis, striped killifish	5.0	
Thalassoma bifasciatum, bluehead	6.0	6.0
Menidia menidia, Atlantic silverside	5.0	5.0
Mysidopsis bahia, mysid shrimp	4.5	4.5
Micrometrus minimus, dwarf perch	3.5	3.5
Cymatogaster aggregata, shiner perch	2.3	2.3
Oncorhynchus tshawytscha, chinook salmon	1.5	1.5
Anguilla rostrata, American eel	0.9	0.9
Penaeus duorarum, pink shrimp	0.7	0.7

[a] From EPA (1980b).

regression line, the dieldrin concentration corresponding to a cumulative probability of 5% is 0.63 ppb, and according to the 1986 EPA guidelines this concentration would therefore become the final acute value for dieldrin in salt water. On the basis of the data available for dieldrin, it is expected that if a large random sample of saltwater genera were taken, only 5% would have TLm values less than 0.63 ppb. In 1980, when the dieldrin criteria were being determined (EPA, 1980a), the EPA was using a somewhat different procedure for estimating the final acute value, and the final acute value determined at that time was 0.71 ppb. According to the regression line in Fig. 7, this concentration would be associated with a cumulative probability of 6%.

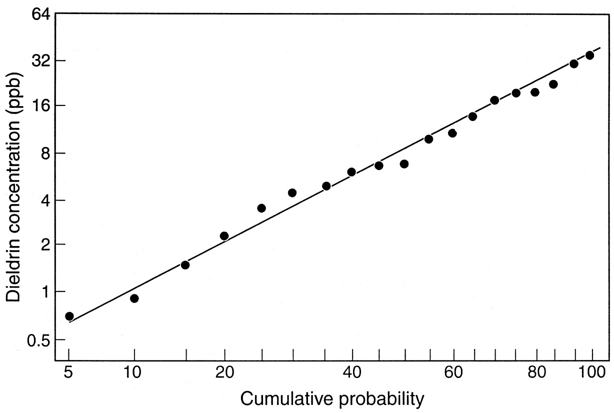


FIGURE 7 Cumulative probability distribution of dieldrin genus mean acute values based on the data in Table IV. The straight line is a least-squares fit to the four lowest data points.

B. Chronic Toxicity

If data on a sufficient number and diversity of organisms are available, a final chronic value for a particular toxicant may be calculated in the same way that final acute values are determined. In practice, however, there are seldom sufficient data to allow a direct graphical estimation of the toxicant concentration that would exert a chronic stress on no more than 5% of the species in the system. In such cases an acute toxicity standard is established on the basis of an adequate amount of short-term toxicity tests, and an average acute/chronic toxicity ratio is then calculated on the basis of a smaller amount of information. The rationale for this procedure is that for a given pollutant the acute/chronic ratio is likely to be more constant between species than is the chronic or sublethal stress level itself. Hence less information is required to estimate the acute/chronic ratio. The chronic toxicity standard is established by dividing the acute toxicity standard by the so-called final acute/chronic ratio. The EPA considers this procedure acceptable if acute/chronic ratios are available for at least three species and (a) at least one of the species is a fish, (b) at least one is an invertebrate, and (c) at least one is an acutely sensitive freshwater species or saltwater species when the ratio is being used to establish freshwater or marine criteria, respectively.

1. Acute/Chronic Ratios

The final acute/chronic ratio is considered to be a measure of the ratio of the concentration of the toxic substance associated with acutely toxic effects to the concentration associated with chronic toxicity effects. According to the methodology recommended by the EPA (1985a), the final acute/chronic ratio may be calculated in one of four ways. First, if there is no major trend apparent among the acute/chronic ratios for the different species and the species mean acute/chronic ratios lie within a factor of 10 for a number of species, the final acute/chronic ratio is the geometric mean of all the species acute/chronic ratios available for both freshwater and saltwater species. Second, if the species mean acute/chronic ratios seem to be correlated with the species mean acute values, the final acute/chronic ratio should be taken to be the acute/chronic ratio of species whose acute values lie close to the final acute value. Third, in the case of acute tests conducted on metals and possibly other substances with embryos and larvae of barnacles, bivalve mollusks, sea urchins, lobsters, crabs, shrimp, and abalones, the acute/chronic ratio is assumed to be 2. The rationale is that chronic tests are very difficult to conduct with such species, and the sensitivities of embryos and larvae would likely determine the results of life-cycle tests. Assuming the acute/chronic ratio to be 2 causes the final chronic value to equal the criterion maximum concentration (see Section V). Finally, if the most appropriate species mean acute/chronic ratios are less than 2, acclimation to the stress probably occurred during the chronic test (see Section VI.B). In such cases, the final acute/chronic ratio is assumed to be 2. The final chronic value for the toxicant is then calculated by dividing the final acute value by the final acute/chronic ratio.

An example of the calculation of a final acute/chronic ratio is shown in Table V for the pesticide dieldrin. Chronic values for this pesticide were available for only four species in 1980 when the guidelines for dieldrin were established (EPA, 1980a), and therefore it was necessary to use acute/chronic ratios to establish the final chronic value. Acute toxicity values were available in only three of the four cases where chronic effects were studied, but the three species satisfied the criteria for calculating acute/chronic ratios in both freshwater and salt water. Since the acute/chronic ratios for the three species differed by less than a factor of 2, it was appropriate to calculate the final acute/chronic ratio for dieldrin by taking the geometric mean of the three ratios, which is $[(11)(9.1)(6.2)]^{1/3} = 8.5$. Final acute values for dieldrin in freshwater (EPA, 1996b) and salt water are 0.48 and 0.71 ppb, respectively. Hence the final chronic values for dieldrin in freshwater and salt water are $0.48/8.5 = 0.056$ ppb and $0.71/8.5 = 0.084$ ppb, respectively.

C. Toxicity to Plants

The EPA guidelines for calculating acute and chronic concentrations of a toxicant clearly pertain to aquatic animals. Nevertheless, aquatic plants may be affected adversely by the presence of toxicants in the water, and sometimes at exceedingly low concentrations (e.g., Fig. 2). According to EPA guidelines (EPA, 1986), a substance is toxic to a plant at a given concentration if the growth of the plant is decreased in a 96-hr or longer experiment with an alga or in a chronic test with a vascular plant. Growth may be measured in a veriety of ways, including photosynthetic rate, change in dry weight, or change in chlorophyll concentration. Table VI, for example, provides a summary of data on the toxicity of dieldrin to aquatic plants. In this particular case, the concentrations that reduced plant growth were substantially higher than both the final acute and final chronic values for animals and therefore had no influence on the final water quality guidelines for dieldrin. However, because plants are the chief producers of organic matter in most aquatic food chains, it is critical that the effects of toxicants on aquatic plants be assessed. Reductions in photosynthetic rates and/or changes in the composition of the plant community may cause repercussions throughout the entire food web. The final plant value is the lowest concentration of the toxicant that reduces plant growth in an appropriate experiment. On the basis of the data in Table VI, the final plant value for dieldrin was set at 100 ppb in freshwater and 950 ppb in salt water (EPA, 1980a).

TABLE V Acute/Chronic Ratios for Dieldrin Toxicity to Three Species of Aquatic Animals[a]

Species	Acute value (ppb)	Chronic value (ppb)	Ratio
Salmo gairdneri, rainbow trout	2.5	0.22	11
Poecilia reticulata, guppy	4.1	0.45	9.1
Mysidopsis bahia, mysid shrimp	4.5	0.73	6.2

[a] From EPA (1980a).

TABLE VI Concentrations of Dieldrin That Reduce Growth in Aquatic Plants[a]

Species	Effects	Concentration (ppb)
Scenedesmus quadricaudata	22% reduction in biomass in 10 days	100
Navicula seminulum	50% reduction in growth in 5 days	12,800
Wolffia papulifera	Reduced population growth in 12 days	10,000
Agmenellum quadruplicatum	Reduced growth rate	950

[a] From EPA (1980a).

V. THE TWO-NUMBER CRITERION

In the United States, EPA water quality guidelines with respect to toxic substances are based on a two-number criterion (EPA, 1986). The first number is the 1-hr average concentration that is not to be exceeded more than once every 3 yr. This number is referred to as the criterion maximum concentration. The second number is the 4-day average concentration that is not be exceeded more than once every 3 yr. This number is referred to as the criterion continuous concentration. The rationale for using a

TABLE VII The Three Toxicant Concentrations Used to Establish Water Quality Guidelines for Dieldrin in Freshwater and Salt Water

	Value in freshwater (ppb)	Value in salt water (ppb)
Final acute value	0.48	0.71
Final chronic value	0.056	0.084
Final plant value	100	950
Criterion maximum concentration	0.24	0.355
Criterion continuous concentration	0.056	0.084

two-number criterion is that organisms can tolerate brief exposures to toxicant concentrations substantially higher than can be tolerated over a longer period. This fact is evident in Fig. 6. There is invariably a negative correlation between the concentration of a toxicant that causes a particular adverse effect and the duration of exposure. The recommended exceedence frequency of 3 yr is the EPA's best scientific judgment of the average amount of time it would take an unstressed system to recover from a pollution event in which exposure to a toxicant exceeded the criterion.

Under the present system (EPA, 1996b), the two numbers in the criterion are calculated from the final acute value, the final chronic value, and the final plant value. The three values for dieldrin in freshwater and salt water are shown in Table VII. The criterion maximum concentration is equated to half the final acute value. Division by 2 in this case to some extent corrects for the fact that much of the acute toxicity information is based on observations of lethal effects, whereas the real concern is protection of organisms from sublethal stresses. The criterion continuous concentration is the smaller of the final chronic value and the final plant value.

VI. COMPLICATING FACTORS

A. Interactions with Harmless Substances or Conditions

The toxicity of a particular substance to aquatic organisms may sometimes depend strongly on the concentrations of other substances or on environmental conditions that by themselves are harmless. For example, in aquatic systems a continuous interconversion takes place between ammonium ions (NH_4^+) and un-ionized ammonia (NH_3). The equation describing this interchange can be written $NH_3 + H^+ \longleftrightarrow NH_4^+$. Although NH_4^+ is nontoxic, NH_3 is highly toxic. The NH_3 96-hr median tolerance limits for a variety of freshwater fish lie in the range 0.08–4.6 ppm. The equilibrium distribution of NH_3 and NH_4^+ is a function of the hydrogen ion concentration (H^+) in the water, and the equilibrium shifts to NH_3 as the concentration or activity of H^+ is reduced. In water quality studies the H^+ activity commonly is reported in terms of pH, where pH is the negative of the common logarithm of the H^+ activity. (Most natural freshwaters have a pH in the range 6–8. The pH of seawater is about 8.1.) At steady state the concentrations of NH_3 and NH_4^+ are equal at a pH of 9.3, and the $[NH_3]/[NH_4^+]$ ratio increases by a factor of 10 whenever the pH increases by one unit. Thus water that is virtually harmless at a pH of 7 might become extremely toxic if the pH is raised to 9. Because of this fact, it is of relatively little use in water quality work to assay merely for $NH_3 + NH_4^+$ (by standard procedures) without simultaneously measuring the pH.

Temperature is another factor that can greatly influence the toxicity of a substance without itself casuing any significant stress. In general terms the performance of an organism is positively correlated with temperature below the optimum temperature for the organism, and then rapidly drops as the temperature increases further (Fig. 8). The rate of increase of the respiration rate or metabolic rate of an organism is usually expressed in terms of the organism's Q_{10}, which is the change in respiration rate or metabolic rate per 10°C increase in temperature. For temperature conformers (i.e., cold-blooded organisms) metabolic rates typically increase by a factor of 2 or 3 for every 10°C rise

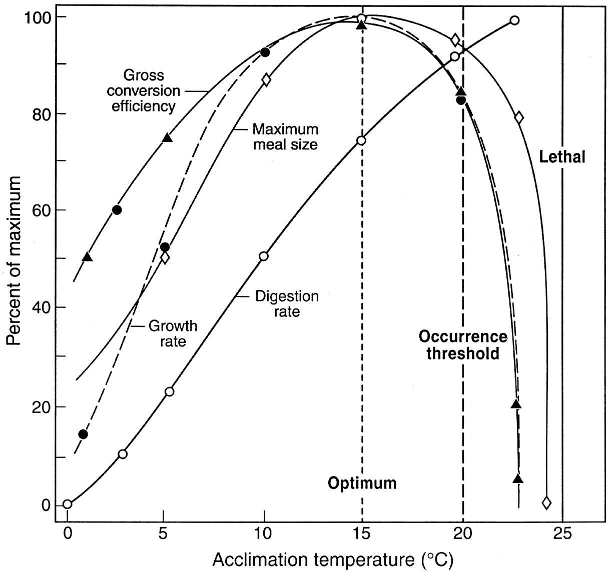


FIGURE 8 Performance of sockeye salmon as a function of acclimation temperature. [Reprinted from Brett, J. R. (1971). Energetic responses of salmon to temperature. A study of some thermal relations in the physiology and fresh water ecology of sockeye salmon (*Oncorhynchus nerka*). *Amer. Zool.* 11: 99–113. © American Society of Zoologists.]

in temperature as long as the temperature is not too close to the lethal temperature for the species. In other words, typical Q_{10} values for cold-blooded organisms are 2–3. Obviously any increase in respiration rate implies and increased demand for oxygen, and as a result, the TLm for oxygen is likely to be positively correlated with temperature. This fact in part underlies the tendency of fish kills caused by indequate oxygen concentrations to occur during the summer when water temperatures are highest. An increase in the respiration rate of a fish obviously implies an increase in the rate at which water is pumped over the gills. As a result, the intensity of exposure of the fish to any toxic substance in the water is increased. The effects of a given toxicant concentration on the organism may therefore become apparent more rapidly in warm water than in cold water.

Figure 6 provides another clear example of the interaction between otherwise harmless water quality characteristics and toxicant effects. The toxicity of zinc to rainbow trout is clearly correlated in a negative way with water hardness. Similar correlations between metal toxicity and water hardness have been noted for a variety of toxic metals. The mechanism responsible for the correlation is not well understood, and considering the wide range of metals for which the effect has been observed, it is unlikely that exactly the same mechanism is operative in all cases. Water hardness, which is a measure of the concentration of calcium and magnesium ions in the water, tends to be correlated with a number of other water quality characteristics such as pH and alkalinity. Therefore it is possible that at least in some cases the observed negative correlation between metal toxicity and water hardness actually reflects a cause-and-effect relationship with one of these other covariates.

1. Incorporation of Interactions into Water Quality Guidelines

If there is enough relevant information, it is possible to incorporate these interactions between toxicant effects and water quality characteristics into water quality criteria. In the United States, for example, the criterion maximum concentration and the criterion continuous concentration are sometimes expressed as explicit functions of certain water quality characteristics. In the case of NH_3, for example, toxicity has been shown to be negatively correlated with pH and, below 20–25°C, with temperature. The EPA water quality criteria for NH_3 therefore are expressed in terms of mathematical equations that relate the criterion maximum and criterion continuous concentrations of NH_3 to temperature and pH (EPA, 1986).

In the case of heavy metals, the criteria are often expressed as functions of water hardness. Figure 9 shows some of the data that were used to determine the freshwater criterion maximum concentration in the case of cadmium (EPA, 1984, 1999). Pertinent data relating cadmium toxicity to water hardness were available for five species, and the logarithms of the cadmium 96-hr TLm values were plotted against water hardness. The slopes of the five regression lines fit to the data were similar, the average being 1.128. The implication is that cadmium 96-hr TLm values can be expected to be proportional to water hardness raised to the 1.128 power. Freshwater 96-hr TLm data reported in the literature were then normalized to the same water hardness by assuming this dependence on water hardness, and the normalized 96-hr TLm values were then plotted as in Fig. 7. The normarlized final acute value was then calculated in the manner previously described. However, this procedure resulted in a normalized final acute value greater than the normalized species mean acute values of four salmonids, which were the most sensitive species among the 50 species included in the study. Since the salmonids are a commercially and recreationally important group of fish, the normalized final acute value was arbitrarily set to the normalized species mean acute value of rainbow trout, the most sensitive salmonid for which results of flow-through tests were available. The normalized freshwater criterion maximum concentration was then set equal to half this normalized final

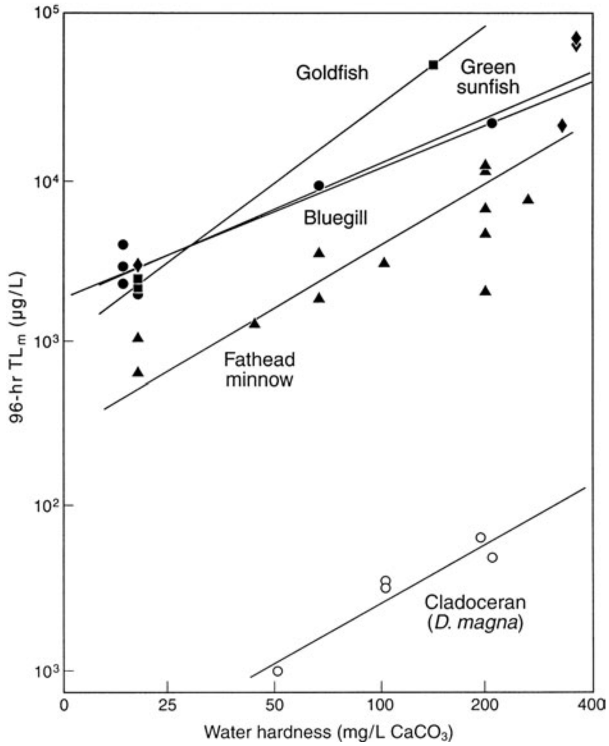


FIGURE 9 Cadmium freshwater TLm values as a function of water hardness for five species. [Data from EPA (1984).]

acute value. The criterion maximum concentration was then determined from the equation

$$\text{Criterion maximum concentration (ppb)} = 0.025\ \text{Hd}^{1.128},$$

where Hd is water hardness in mg/L. The value so calculated is the concentration of the so-called total recoverable fraction of the cadmium (EPA, 1997) and is operationally determined by measuring the cadmium concentration in an acidified sample of water. For purposes of establishing water quality criteria, this concentration is now multiplied by a conversion factor (cf) to determine the concentration of the dissolved metal (EPA, 1997). In the case of cadmium, the conversion factor is a function of the water hardness and is calculated from the equation $\text{cf} = 1.137 - 0.0418 \ln(\text{Hd})$. The complete equation for the cadmium criterion maximum concentration expressed as dissolved cadmium is then (EPA, 1999)

$$\text{Criterion maximum concentration (ppb)} = (0.025)(\text{cf})\text{Hd}^{1.128}.$$

An analogous equation for the final chronic value may be calculated by simply dividing the equation for the final acute value by the final acute/chronic ratio. However, if there is evidence that there is a difference in the functional dependence of chronic toxicity and acute toxicity on water quality characteristics such as temperature and hardness, then the final chronic equation may be determined independently of the final acute equation. In the case of cadmium, for example, chronic toxicity appears to be less sensitive to water hardness than acute toxicity appears to be, thus a final freshwater chronic equation was developed solely from chronic toxicity studies performed with 16 freshwater species. The final chronic equation for dissolved cadmium in freshwater is

$$\text{Final chronic value (ppb)} = (0.066)(\text{cf})\text{Hd}^{0.785},$$

where cf in this case equals $1.1017 - 0.0418 \ln(\text{Hd})$.

B. Conditioning and Acclimation

The toxicity of a particular stress to an aquatic organism can depend very much on the previous life history of the organism. This fact reflects the ability of many organisms to acclimate or adapt to changes in the environment and in particular to the presence of toxic substances in the water. Figure 10 provides a good example of the effects of adaptation on the toxic effects of temperature stress. The data are taken from a study of both lethal and sublethal temperature stresses on young sockeye salmon. Two features of the graph are of particular interest. First, sublethal effects on spawning and growth become apparent outside a much narrower temperature range than is the case for lethal effects. For example, the fish do not begin to die outright until the temperature exceeds 21–25°C or falls below 5–7°C. These observations simply illustrate and reemphasize the importance of distinguishing sublethal effects from lethal effects. Second, the range of temperatures within which the fish can effectively carry out certain activities is shifted to higher temperatures as the acclimation temperature is increased. For example, a young sockeye salmon acclimated to a temperature of 20°C would evidently die if placed in water at 4°C, but a similar fish acclimated at 10°C would not die if placed in the same water. There is, however, a limit to the adaptive capabilities of aquatic organisms. For example, it is apparent from Fig. 10 that young sockeye salmon cannot spawn outside the approximate temperature range 5–14°C, regardless of the acclimation temperature.

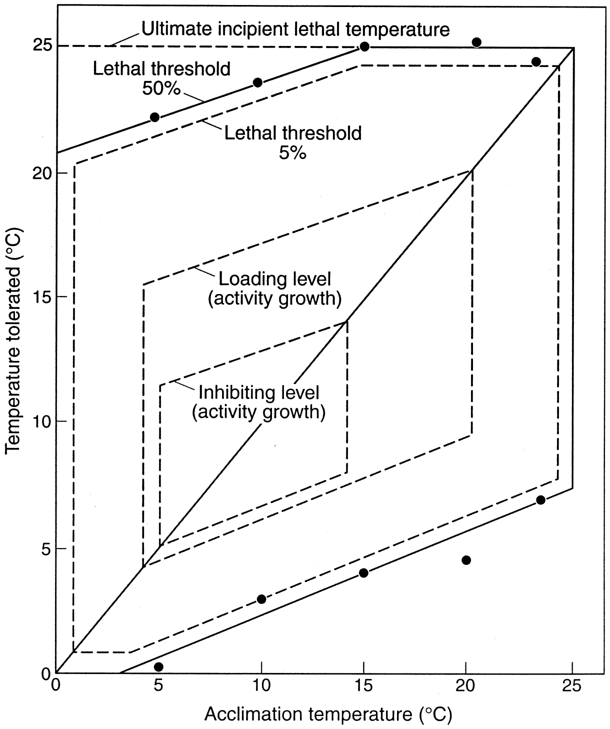


FIGURE 10 Temperature tolerance zones for young sockeye salmon as a function of acclimation temperature. [Reprinted with permission from Brett, J. R. (1960). Thermal requirements of fish—Three decades of study. *In* "Biological Problems of Water Pollution." (C. M. Tarzwell, ed.), pp. 110–117. U.S. Dept. of Health, Education, and Welfare.]

In some cases exposure of an organism to a stress may actually increasee the sensitivity of the organism to that stress, but in many cases prolonged exposure to a sublethal level of stress may actually desensitize or acclimate the organism to the stress. For example, many fish are capable of acclimation to low oxygen levels as well as to high and low temperatures. Some fish have been shown to acclimate to ammonia, cyanide, pH, phenol, synthetic detergents, and zinc. In organisms such as bacteria that can multiply rapidly, adaptation may actually involve genetic changes. Mutant strains may by chance be more resistant to a particular stress than the other members of the population are and time will gradually dominate the population. Genetic selection of this sort is believed to have been responsible for the emergence of pesticide-resistant strains of about 200 different insect pests. Among organisms that reproduce more slowly, genetic selection would take much longer, and acclimation in the short term must therefore involve other mechanisms such as biochemical, physiological, or behavioral adjustments. The processes involved in these adjustments are in general not well understood but probably involve changes in the functioning of hormones and/or enzymes and the response characteristics of the organism's nervous system.

In a general sense organisms that adapt to a stress may be classified as either conformers or regulators. For example, cold-blooded animals are conformers with respect to temperature because their internal body temperature is correlated strongly with the temperature of the environment. Warm-blooded animals are regulators with respect to temperature because their internal temperature is relatively independent of the temperature of the environment. Of course, for both conformers and regulators there is an optimum internal state at which the organism functions most effectively. Acclimation among regulators usually produces a change in the range of the external variable within which the organism can effectively control its internal state but does not change the optimum state of the organism. Acclimation among conformers may involve a change in the organism's optimal internal state and usually involves a change in the range of internal variability within which the organism can function efficiently.

Should the possible effects of acclimation be considered when water quality guidelins are being set? The attitude of the EPA is that acclimation effects should in general not be considered and that toxicity data obtained with acclimated organisms should not be used in deriving water quality guidelines. The rationale for this attitude is this: "Acclimated organisms are the exception rather than the norm. Rarely, if ever, can acclimation be depended on to protect organisms in a field situation because concentration often fluctuate and motile organisms do not stay in one location very long" (EPA, 1980c, p. 79364). This attitude seems a reasonable one to the extent that acclimation in toxicological studies is avoidable and tends to desensitize organisms to a stress. However, conditioning may sometimes sensitize rather than desensitize an organism to a stress, and in the case of temperature stress, it is impossible to aviod the issue of acclimation in experimental studies. Certainly one should be aware of the potentially confounding effects of acclimation in toxicological work.

C. Interactions between Toxic Substances

In the discussion to this point, a toxicant has been considered as if it were the only stress-producing substance in the water. What effect can be anticipated, however, if two or more toxic substances are present in the water at the same time? For example, according to the current EPA guidelines, the 4-day average concentrations of lead and zinc in salt water should not exceed 5.6 and 86 ppm, respectively, more than once every 3 yr. Now suppose that a particular body of salt water is consistently found to contain 4 ppm of lead and 60 ppm of zinc. Would this water be likely to exert a chronic stress that would in some sense violate the intentions of the EPA guidelines?

The answer to this question is that the effect of two or more toxic substances on the organisms in the water will depend on the manner in which the toxicants interact. The following example illustrates the various ways in which two toxicants might interact. Suppose that the 96-hr TLm for a particular organism is 1.0 and 10 ppb for toxicants A and B, respectively. Now suppose that toxicants A and B are added to the water in a flow-through acute toxicity experiment so that the concentrations of A and B are in fact 1.0 and 10 ppb, respectively. The test organisms are placed in the water and their survival is monitored for 96 hr. Any of the following outcomes is possible:

1. Exactly half the organisms are dead after 96 hr. This result is exactly what one would have expected if one had added 1.0 ppb A but no B to the water, or 10 ppb B but no A. In this case one says that there has been no interaction between A and B.
2. Fewer than half the organisms are dead after 96 hr. In this case the toxicity of the mixture is apparently less than the toxicity of a solution containing only 1.0 ppb A or only 10 ppb B, and the interaction between A and B is said to be antagonistic.
3. If more than half the organisms are dead after 96 hr, there are several possible interpretations:
    a. If the percentage of dead organisms is exactly what would have been expected if the water contained only 2.0 ppb A or only 20 ppb B, then the interaction of A and B is said to be strictly additive. In other words, 1 ppb A produces the same toxicity as 10 ppb B, and when the two substances are present together the toxicity of the mixture can be calculated by assuming that every 10 ppb of B are equivalent to 1.0 ppb of A.

b. If the fraction of organisms that are dead after 96 hr is greater than in case 3a, then the interaction is said to be supra-additive.
c. If the percentage of organisms dead after 96 hr is greater than 50% but less than in case 3a, then the interaction is said to be infra-additive.

It is useful in discussing these interactions to speak in terms of toxicity units rather than actual concentrations. In this hypothetical example, one acute toxicity unit of A would be 1.0 ppb of A, and one acute toxicity unit of B would be 10 ppb of B. The possible interactions between A and B may then be indicated graphically as in Fig. 11.

How should one take account of these interactions in setting water quality guidelines? One obvious practical problem is that there is not a great deal known about the interactions of toxic substances. In fact, a review of current EPA water quality criteria (EPA, 1999) shows that in many cases insufficient information is available to establish numerical guidelines for toxicants considered separately. The present guidelines make no allowance for interactions between toxicants. In response to criticisms concerning this policy, the EPA has stated, "Synergism and antagonism are possible between numerous combination [sic] of two or more pollutants, and some data indicate that such interactions are not only species specific, but also vary with the ratios and absolute concentrations of the pollutants and the life stage of the species" (EPA, 1980c, p. 79358).

In other words, not much is known, the interactions may be very complicated, and given these facts it is very difficult to estimate the toxicity of a solution containing more than one toxic substance. To the extent that the effects of two or more toxicants can be normalized by expressing concentrations in terms of toxicity units, Fig. 11 shows that in two cases calculation of the toxicity of a mixture of toxicants is straightforward. First, if there is no interaction between the toxicants, then the toxicity of the mixture would be determined by the toxicant present in the greatest number of toxicity units. For example, if a solution contains 0.75 toxic units of A and 0.5 toxic units of B, then if A and B do not interact, the toxicity of the solution would be equal to that of a solution containing only 0.75 toxic units of A. Second, if the interaction is strictly additive, the toxicity of the mixture would be equal to that of a solution containing 1.25 toxic units. In other words, if one assumes that there is no interaction or that the interactions are strictly additive, then the toxicity of a mixture of toxicants can easily be calculated. However, if the interactions are supra-additive, infra-additive, or antagonistic, then one must have detailed information about the nature of the interaction to calculate the toxicity of the mixture.

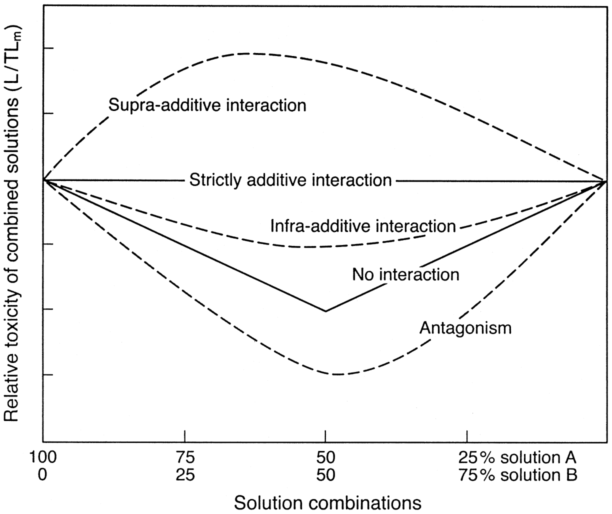


FIGURE 11 Possible interactions between two hypothetical toxicants, A and B. [Redrawn with permission from Warren, C. E. (1971). "Biology and Water Pollution," © W. B. Saunders, Philadelphia.]

It is apparent from Fig. 11 that the assumption of no interaction is conservative only if the true interactions are antagonistic. The assumption of no interaction would underestimate the true interactions if the true interactions were in any sense additive. The assumption of strictly additive interactions would appear to be conservative in all cases except the case in which the true interactions were supra-additive. This latter conclusion, however, is valid only to the extent that a given toxic effect can be associated uniquely with a certain number of toxicity units. Figure 11, for example, implicitly assumes that X toxicity units of A have the same effect as X toxicity units of B for all values of X. In other words, there is a universal curve that relates toxic effect to toxicity units. The normalization procedure ensures that the effects are the same when $X = 1$, but it does not follow automatically that the effects are the same for all values of X. For example, suppose that two acute toxicity units of A kill 75% of a given species in 96 hr and that two acute toxicity units of B kill 95% of the same species in 96 hr. Now suppose that 85% of the test organisms die after 96 hr of exposure in water containing one toxicity unit of both A and B. The interactions are clearly additive in some sense, since more than 50% of the test organisms died after 96 hr. However, the interactions appear supra-additive relative to toxicant A and infra-additive relative to toxicant B. Strictly speaking then, one cannot apply the assumption of strictly additive interactions unless one feels confident that the curves relating toxic effects and toxicity units are the same or very similar for all toxicants of concern.

There is no doubt that some toxicants do not interact in a strictly additive manner. For example, the toxicity of cadmium is decreased by other metal ions, particularly low concentrations of copper and zinc, and it is known that the cholinesterase-inhibiting pesticides EPN (*O*-ethyl-*O*-paranitrophenyl phenylphosphonothioate) and malathion interact in a supra-additive manner. The TLm of a mixture containing equal numbers of toxic units of EPN and malathion is only one-twelfth the TLm of the individual compounds. The effect is due to the ability of EPN to inhibit the enzyme system that is responsible for detoxifying malathion after it has been converted to malaoxon. In contrast, chlorinated hydrocarbons such as DDT that induce some of the liver microsomial enzyme systems may interact antagonistically with chemicals that are detoxified by the same enzyme systems (e.g., hexobarbital). Although admittedly it is illogical to expect the toxicity of poisons with different toxicological properties and different concentration–response curves to be strictly additive,

> Nevertheless, the method has been found to work empirically . . . . If, for the present, the poisons can be regarded as agents producing stress . . . , each of which produces a degree of shock with resulting non-specific effects, it might be considered reasonable that summation of the overall stress is possible. [From Brown, V. M. (1968). "The calculation of the acute toxicity of mixtures of poisons to rainbow trout." *Water Res.* **2,** 723–733.]

Application of the strictly additive interaction assumption to water quality guidelines was made in the 1973 EPA water quality criteria, where the following was noted:

> The system of adding different toxicants in this way is based on the premise that their lethal actions are additive. Unlikely as it seems, this simple rule has been found to govern the combined lethal action of many pairs and mixtures of quite dissimilar toxicants, such as copper and ammonia, and zinc and phenol in the laboratory . . . . The rule holds true in field studies . . . . The method of addition is useful and reasonably accurate for predicting thresholds of lethal effects in mixtures. (EPA, 1973, p. 122)

This statement notwithstanding, the assumption of strictly additive interactions was abandoned in 1980 (EPA, 1980c, pp. 79318–79379) and replaced with the assumption of no interactions. Since the latter assumption causes one to underestimate the toxicity of a mixture whenever the interactions are in any sense additive, adoption of this policy seems inconsistent with the expressed intentions of the EPA "to err on the side of overprotection" (EPA, 1980c, p. 79342). If a general guideline in establishing water quality criteria is to err on the side of overprotection and if detailed information on the interactions between combinations of toxicants is not available, then the assumption of strictly additive interactions would seem preferable to the assumption of no interactions.

VII. PUBLIC HEALTH

In the previous discussion, the use of toxicological studies was considered for purposes of setting water quality criteria for the protection of marine and freshwater organisms. An equally if not more important use of toxicology is to establish water quality criteria for the protection of the general public. In the United States, such criteria are established by different methods depending on whether the toxicants are believed to be carcinogenic or not.

A. Noncarcinogenic Effects

In the case of noncarcinogenic substances, it is generally assumed that there is a threshold dose below which the substance exerts no adverse effects on humans. It is obvious that such a threshold must exist in the case of essential trace metals, which are required in small amounts by all organisms but become toxic if administered in large doses. In all cases, however, the existence and value of the threshold must be established on the basis of experiments with animals or experience with humans. Water quality criteria are then established on the basis of the estimated thresholds.

One can imagine several types of experimental or observational data that might be used to estimate the threshold doses for noncarcinogenic substances. For instance, there may have been unfortunate incidents in which humans were exposed to toxic concentrations of a particular substance. If the dose received by the people is known or can be estimated and the effect on the persons documented, then the data are useful for establishing an upper bound on the threshold dose. In some cases there may have been experimental feeding studies with human volunteers, and in such cases it may be possible to obtain an even better estimate of the threshold dose. In the absence of information directly related to humans, there may be data obtained from experimental feeding studies with animals. These data may be used to estimate human effects, but the uncertainty factor is greater in the absence of any direct observations on humans.

The duration of exposure must also be considered. Water quality criteria are based on an assumed lifetime exposure, but experimental observations, whether on humans or on animals, do not always cover the lifetime of the organisms. If the only data available are from acute rather than chronic observational or experimental studies, then the uncertainty factor used in establishing water quality criteria must be increased.

Finally, one can imagine several general types of effects or responses associated with exposure to a toxicant. The first type of response would be no observed adverse effect. The dose associated with no observed adverse effect is referred to as the no observed adverse effect level (NOAEL) and is theoretically a lower bound to the threshold dose that causes some adverse effect. The second type of response would be an observed adverse effect, and the lowest observed adverse effect level (LOAEL) is theoretically an upper bound to the threshold dose. Adverse effects are defined by the EPA as "any effects which result in functional impairment and/or pathological lesions which may affect the performance of the whole organism, or which reduce an organism's ability to respond to an additional challenge" (EPA, 1980c, p. 79353). It is also possible that exposure to a toxicant will produce a response that is not an adverse effect according to the above definition. For example, rather than passing through the digestive tract and being excreted, the toxicant may be assimilated and then detoxified or broken down by the organism. Therefore the organism has been affected by the toxicant to the extent that it has diverted some of its metabolic energy to detoxifying or breaking down the toxicant, but should this response be regarded as an adverse effect? The job of classifying such responses is obviously somewhat subjective and has become increasingly difficult as more sophisticated testing protocols are developed and more subtle responses are identified. One result has been that some responses are classified simply as observed effects. One therefore finds in the EPA literature, for example, reference to no observed effect levels (NOELs) and lowest observed effect levels (LOELs). It is up to public health administrators, given information on these various types of responses or effects and after applying appropriate uncertainty factors, to determine what is called the reference dose (RfD), which is the maximum acceptable rate of consumption of the toxic substance per unit body weight. Table VIII lists the uncertainty factors used in the United States.

Once a RfD has been established, the human health criterion (HHC) for the toxicant is calculated from the equation

$$\mathrm{HHC} = \frac{(\mathrm{RfD})(\mathrm{BW})}{\mathrm{WC} + (\mathrm{FC})(\mathrm{PBCF})}, \qquad (1)$$

where HHC is the concentration of the toxic substance in the water ($\mu\mathrm{g\,kg}^{-1}$), RfD is the reference dose ($\mu\mathrm{g\,kg}^{-1}\,\mathrm{d}^{-1}$), BW is body weight (kg), WC is the water consumption rate ($\mathrm{kg\,d}^{-1}$), FC is the fish and shellfish consumption rate ($\mathrm{kg\,d}^{-1}$), and PBCF is the practical bioconcentration factor (dimensionless). For adults, BW is assumed to be 70 kg, and WC is assumed to be 2 kg (2 L) per day. The product (RfD)(BW) is often referred to as the acceptable daily intake, or ADI. Average fish and shellfish consumption rates in the United States are as follows (EPA, 1996a):

Fresh and estuarine fish and shellfish	$0.0056\ \mathrm{kg\,d}^{-1}$
Marine fish and shellfish	$0.0121\ \mathrm{kg\,d}^{-1}$
Total	$0.0177\ \mathrm{kg\,d}^{-1}$

1. Application to Cadmium

Application of the foregoing methodology is illustrated in the following example of cadmium. In this case there are a number of examples of humans who were exposed to levels that clearly caused adverse effects. The best-documented example concerns people living in the Jintsu River basin of Japan. The health effects caused by their consumption of cadmium, commonly referred to as itai-itai disease, included severe pain in the back, joints, and lower abdomen; development of a waddling or ducklike gait; kidney lesions; proteinuria; glycosuria; and loss of calcium from the bones leading in some cases to multiple bone fractures. Approximately 200 persons were afflicted with itai-itai disease, and of these nearly 100 died. In the area of Japan where itai-itai disease occurred, about 85% of the cadmium intake was due to the consumption of

TABLE VIII Uncertainty Factors for Deriving Criteria for Threshold Effects of Toxicants from NOEL, NOAEL, LOEL, and/or LOAEL Data[a,b]

Nature of experimental results	Uncertainty factor
Valid experimental results from studies on prolonged ingestion by humans with no indication of carcinogenicity	10
Experimental result of studies of human ingestion not available or scanty (e.g., acute exposure only) with valid results of long-term feeding studies on experimental animals, or in the absence of human studies, valid animal studies on one or more species; no indication of carcinogenicity	100
No long-term or acute human data; scanty results on experimental animals with no indication of carcinogenicity	1000
Additional judgmental uncertainty factor to be applied to LOAEL data	1–10

[a] From National Academy of Sciences (1977). "Drinking Water and Health," National Academy of Sciences, Washington DC.

[b] NOEL, no observed effect level; NOAEL, no observed adverse effect level; LOEL, lowest observed effect level; LOAEL, lowest observed adverse effect level.

cadmium-contaminated rice. A sensitive indicator of the disease was tubular proteinuria, the incidence of which rose above that observed in control populations when the cadmium concentration in the rice exceeded about 0.45 ppm. Japanese in the area were known to consume about 430 g of rice per day. Therefore an adverse effect level for Japanese equals $(430)(0.45)/(0.85) = 230 \mu$g/day. For Americans and western Europeans, who are about 32% larger than Japanese, the corresponding adverse effect level would be about 300 μg/day. Since only about 5% of ingested Cd is actually absorbed by the body (EPA, 1980b), the figure of 300 μg/day ingested corresponds to only about 15 μg/day absorbed.

Similar conclusions have been reached from other studies. For example, from a study of workers who inhaled Cd dust, researchers concluded that effects began to appear when the concentration of Cd in the air equaled 21 μg m^{-3}. If one assumes that the average person inhales about 10 m^3 of air per day, that lung retention of Cd is about 25% efficient (EPA, 1980b), and that the workers were exposed to the contaminated air 5 days per week, the calculated level of absorption corresponding to the appearance of effects was $(21)(10)(0.25)(5/7) = 38 \mu$g/day. Similarly, the Working Group of Experts for the Commission of European Communities has estimated that the threshold level of Cd exposure by ingestion is about 200 μg/day, which would correspond to absorption of 10 μg/day.

Applying the uncertainty factor of 10 in Table VIII appropriate for studies on prolonged ingestion by humans, one would conclude that an ADI for adults (assuming 5% absorption efficiency) would be $10/((10)(0.05)) = 20 \ \mu\text{g d}^{-1}$ to $38/((10)(0.05)) = 76 \ \mu\text{g d}^{-1}$ for Cd. It is apparent from Table IX that for the average American the rate of Cd intake in food exceeds the lower of these two figures and equals about 40% of the higher figure. It is also apparent that drinking water contributes a minor fraction of the total Cd absorbed. In fact, the average drinking water in the United States contains only 1.3 ppb Cd. Only 3 of 969 community water supplies studied by the EPA contained Cd in excess of 10 ppb (EPA, 1980b).

TABLE IX Average Daily Intake and Absorption of Cadmium by Americans[a]

Exposure source	Exposure	Cd intake (μg)	% Absorbed	Absorption (μg d^{-1})
Air, ambient	0.03 μg m^{-3}	0.6	25	0.15
Air, smoking (one pack)	3.0 μg/pack	3.0	25	0.75
Food	—	30.0	5	1.50
Drinking water	1.3 μg/L	2.6	5	0.13

[a] From EPA (1980b).

TABLE X Cadmium Geometric Mean Bioconcentration Factors (BCFs) for Edible parts of Consumed Aquatic Species[a]

Species	Mean BCF
Bivalve mollusks	
Asiatic clam, *Corbicula fluminea*	2570
Blue mussel, *Mytilus edulis*	186
Bay scallop, *Argopecten irradians*	2040
Eastern oyster, *Crassostrea virginica*	1660
Soft-shell calm, *Mya arenaria*	160
Others fish and shellfish	
Brook trout, *Salmo gairdneri*	21.5
Green crab, *Carcinus maenas*	5.9

[a] From EPA (1984).

How much Cd would an average person absorb from drinking water and consuming fish and shellfish if all water contained the existing drinking water standard of 5 ppb Cd (EPA, 1991)? Bioconcentration factors (BCFs) for Cd in the edible parts of commercially important fish and shellfish are shown in Table X. The BCFs are consistently higher for the five bivalve mollusks than for the other two species, the geometric mean BCFs being 763 and 11, respectively. Bivalve mollusks account for only 0.6 g d^{-1} of the 17.7 g d^{-1} of fish and shellfish consumed by the average American. It therefore seems appropriate to weight the PBCF according to the relative amounts of the different kinds of fish and shellfish. The weighted PBCF is therefore $[(0.6)(763) + (17.1)(11)]/17.7 = 36.5$. The daily intake of Cd from drinking water and consuming fish and shellfish if all water contained 5 ppb (i.e., 5 μg kg^{-1}) Cd would therefore be $5[2 + (0.0177)(36.5)] = 13.2 \mu$g/day. Of this figure, only 24% comes from the consumption of fish and shellfish. For the average American then, the consumption of fish and shellfish makes a minor contribution to Cd intake. This conclusion may not apply, however, to persons who consume large amounts of bivalve mollusks.

In considering the effects of Cd on human health, the EPA (1980b) commented, “It appears that a water criterion needs to be no more stringent than the existing primary drinking water standard (10 μg/L) to provide ample protection of human health” (p. C-66). This conclusion is thought provoking. If a drinking water supply actually contained 10 μg/L of Cd, the intake of Cd from drinking water alone would be 20 μg/day, which equals the lower bound of the ADIs for Cd estimated from the observed adverse effects of Cd on human health. If the estimated intake of 30 μg/day from food (Table IX) is added to this figure, the total is 2.5 times the lower bound. The EPA’s conclusion seems to have been based largely on the fact that “water constitutes only a relatively minor portion of man’s daily cadmium intake” (EPA, 1980b, p. C-66). This fact arises, however, not because average water contains

10 μg/L cadmium, but because average water contains almost eight times less cadmium. The decision to lower the drinking water standard to 5 μg/L Cd (EPA, 1991) was a step in the right direction.

B. Carcinogenic Effects

The whole issue of carcinogenic effects has become controversial because of the suggestion that faulty experimental and analytical methods may have been used to determine whether a substance is carcinogenic and, if so, the relationship between dose and response. Although in some cases evidence of carcinogenicity has been based on unintentional human exposure, in many cases toxicological studies relative to cancer have been conducted with rats or mice. In the United States, the standard procedure has been to conduct initial feeding experiments with the test chemical, and on the basis of these preliminary experiments to determine the so-called maximum tolerated dose (MTD). The MTD is the highest dose of the chemical that does not literally kill the animals, although if repeated over a 2-wk period, the MTD usually leads to a noticeable but tolerated reduction in weight. Chronic feeding studies are then conducted at doses equal to or less than the MTD over the animal's lifetime. The experimental doses vastly exceed those to which humans are likely to be exposed.

If an abnormal incidence of cancer occurs in the experimental animals compared with that of a control group that is not exposed to the chemical, the excess incidence of cancer is plotted as a function of dose and an analytical function is fit to the data. The analytical function has the form

$$A(d) = 1 - \exp\left(-q_1 d - q_2 d^2 - \cdots - q_k d^k\right), \tag{2}$$

where $A(d)$ is the extra risk over background associated with a dose d, and the q_i are constants to be determined by a least-squares curve-fitting procedure. An important characteristic of Eq. (2) is the fact that at low doses it reduces to the form

$$A(d) = q_1 d \tag{3}$$

and hence conforms to the assumption that there is no threshold dose below which the chemical exerts no carcinogenic effects. The levels of exposure for most humans are likely to fall in the range where, according to the model, $A(d)$ can for all intents and purposes be assumed to be directly proportional to d. The cancer risks at these low dose levels represent considerable extrapolations of the experimental data.

What is wrong with these studies? Actually this sort of approach was probably a reasonable one in the 1970s because medical understanding of the etiology of cancer was relatively crude compared with that of the present time and it undoubtedly seemed wise to err on the side of safety. As understanding of the mechanisms that cause cancer has improved, however, this approach has drawn much criticism. Abelson (1990a), for example, has commented, "The standard carcinogen tests that use rodents are an obsolescent relic of the ignorance of past decades" (p. 1497).

The problems with the rodent tests are severalfold. First, the rodents in question are pure strains that have been inbred for numerous generations. This procedure of inbreeding was intended to produce rats and mice with a more-or-less well-defined genotype, but such inbreeding often leads to genetic impairments. In the experimental mice, for example, there is a high natural incidence of liver tumors, and, as Abelson (1987) points out, "The usual response of these animals to massive doses of a chemical is to develop an even higher incidence of liver tumors" (p. 473). This increase in liver tumors among naturally tumorigenic mice is of doubtful relevance to humans because primary liver cancer is rare in humans with the exception of persons with alcoholism and persons who have had hepatitis (Abelson, 1987).

Second, recent evidence indicates that the practice of inbreeding has led to substantial genetic drift in the rodents used in risk-assessment experiments. In most tests conducted by the EPA one or more of three strains of rodents have been used: Sprague-Dawley (SD) rats, Fischer (F-344) rats, and B6C3F1 mice. Abelson (1995) points out, for example, that at the Merck Research Laboratory in the 1970s the survival rate at age 2 yr of SD rats used as controls was 58%. In the 1980s the corresponding survival rate was 44%, and in the 1990s it had dropped to 24%. The incidence of tumors in control rodents has changed also. The number of liver tumors, for example, in control B6C3F1 mice increased from an average of 32% in 1980 to roughly 50% in 1984 (Abelson, 1995).

Third, it has been customary for the companies responsible for breeding the rodents to give them as much food as they want to eat. This practice is referred to as feeding the animals *ad libitum (ad lib)*. The result is that the animals overeat, and their health is impaired. A recent experiment involving SD rats compared the longevity of control rats fed *ad lib* with diet-restricted rats fed 65% of the *ad-lib* amounts. At maturity the *ad-lib* male rats weighed 60% more than the diet-restricted males. Only 7% of the *ad-lib* males survived to age 2 yr, while 72% of the diet-restricted males survived longer than 2 yr (Abelson, 1995).

Finally, a major criticism of the rodent studies is the fact that prolonged feeding at the MTD tends to cause mitogenesis (induced cell division), and it is now recognized that nongenotoxic chemicals such as saccharin can be carcinogenic at high doses simply because they cause mitogenesis. The reason is that a dividing cell is much more at risk of mutating than is a quiescent cell (Ames and Gold, 1990). Since there is believed to be a threshold dose below which such nonmutagenic toxins exert no effect on cell division, according to Ames and Gold (1990), "At the low doses of most human exposures (where cell-killing and mitogenesis do not occur), the hazards [of nonmutagenic toxins] may be much lower than is commonly assumed and often will be zero" (p. 971).

Given the above information, one should not be surprised to learn that about 58% of all chemicals tested chronically at the MTD are judged to be carcinogens. Although much more emphasis has been given to testing synthetic chemicals than to testing naturally occurring compounds, the percentage of synthetic chemicals judged to be carcinogenic on the basis of these tests is not much different than the corresponding percentage of naturally occurring compounds, 61% versus 48%, respectively. Of the natural chemicals that have been tested, about two-thirds are natural pesticides produced by plants to defend themselves. These natural pesticides account for over 99.99% of the pesticides consumed by persons in the United States, and on the basis of the rodent bioassays, 27 of 52, or about 52%, of the natural pesticides tested are carcinogens. These 27 naturally occurring pesticides are found in 57 different foods, including apples, bananas, carrots, celery, coffee, lettuce, orange juice, peas, potatoes, and tomatoes (Abelson, 1990a). A cup of coffee, for example, contains about 10 mg of rodent carcinogens, primarily caffeic acid, catechol, furfural, hydrogen peroxide, and hydroquinone. The implication of these revelations is that the rodent bioassays probably have been very misleading concerning the carcinogenic threat associated with numerous chemicals, both synthetic and natural. Diets rich in fruits and vegetables, for example, tend to reduce the incidence of cancer in humans. This observation seems inconsistent with the results of the rodent bioassays.

Although not all scientists agree that the rodent bioassays have been misleading, there does seem to be a general consensus that a better understanding of the mechanisms that cause cancer is needed at this time. For example, we need to know the hormonal determinants of breast cancer, the viral determinants of cervical cancer, and the dietary determinants of stomach and colon cancer. Unleaded gasoline has been implicated as a carcinogen because it causes kidney tumors in male rats, but we now know that branched-chain hydrocarbons, which are key components of unleaded gas, interfere with the mechanism for excreting a low-molecular-weight protein by the male rat. Research conducted at the Chemical Industry Institute of Toxicology indicates that this interference may be the cause of the kidney cancer in male rats exposed to gasoline. A similar mechanism does not exist in female rats, in male or female mice, or in humans (Abelson, 1990b). Similarly, we now know that "saccharin's ability to induce bladder tumors in male rats is solely due to the proliferative effects that high doses have on the bladder lining" (Marx, 1990, p. 744). Clearly it will be impossible to make an informed judgment about the carcinogenic threat of various chemicals to humans until the mechanisms responsible for causing cancer in humans are more thoroughly understood.

VIII. PROTECTION OF WILDLIFE

The EPA has published procedures for the determination of water quality criteria for the protection of wildlife that depend on fish for food either directly or indirectly (EPA, 1995). The criteria were developed as part of an effort to establish various water quality criteria for the Great Lakes. The rationale behind the so-called wildlife criteria (WC) is straightforward. The criteria are intended to protect fish-eating animals and animals farther up the food chain from being adversely affected by the presence of toxic substances in the prey that they consume and the water that they drink. The approach used in establishing the criteria is similar to the approach used to determine the human health criteria, but as noted by the EPA (1995), the approach "focuses on endpoints related to reproduction and population survival rather than the survival of individual members of a species" (p. 19). At least in the case of the Great Lakes, the analysis focused on biocumulative chemicals (BCCs) because "these are the chemicals of greatest concern to the higher trophic level wildlife species feeding from the aquatic food web" (EPA, 1995, p. 19). The food chains leading to three representative avian species (bald eagle, kingfisher, and herring gull) and two mammal species (mink and otter) were chosen for analysis. A wildlife value was determined for each species, and an avian value and a mammal value were determined from the geometric means of the three avian species and the two mammal species, respectively. The lower of the mammalian and avian values became the Great Lakes Wildlife Criterion.

WC are not listed in national EPA water quality criteria documents, in part because the development of these criteria is relatively new and in part because the criteria are specific to particular aquatic ecosystems. The WC for the Great Lakes are thought provoking. In the case of mercury, for example, the WC is 1.3 ng/L. This compares with the national HHC for mercury of 50 ng/L. Why is the WC almost 40 times lower than the HHC? The answer is that only a fraction of the diet of most Americans comes from fish. Animals that prey on fish almost exclusively or are predators of such animals are obviously at much greater risk from the standpoint of exposure to toxic substances.

IX. COMMENTARY

At the present time there are approximately 75,000 chemicals in use or being distributed through the environment, and 500–1000 are added each year. Although not all these chemicals are toxic, merely keeping up-to-date information on the possible toxicity of so many compounds is a prodigious task. The discussion in this article has indicated how involved the process of setting water quality guidelines is for even a single chemical. Large amounts of the right kind of information are needed. In its 1999 update to *Quality Criteria for Water*, the EPA listed water quality criteria for 120 priority toxic chemicals. In only 20 of those cases was there sufficient information to determine a criterion maximum concentration and a criterion continuous concentration in both salt water and freshwater. This fact dramatically illustrates the nature of the problem that confronts governmental agencies charged with the responsibility of controlling water pollution. On the one hand, there is an increasingly urgent demand to establish water quality guidelines for the large number of toxic chemicals that pollute both groundwater and surface water. On the other hand, there is a need to develop intelligent and legally defensible techniques for establishing those guidelines. Ultimately government agencies must decide in a somewhat arbitrary manner how much of what kind of toxicological information is necessary and sufficient to establish water quality criteria. Establishing the rules of the game is obviously a critical first step, and in the case of carcinogenic substances, there is disagreement over what the rules should be. Even when the rules have been established, there often remain information gaps that must be filled before criteria can be recommended. Invariably there will remain disagreement over whether the methodology required to establish the guidelines is too simple or too cumbersome, too strict or too lax, and too detailed or too ambiguous. There will remain differences of opinion as to what types of toxicological data are useful and acceptable and how best to utilize those data in developing guidelines. The wisdom and collaboration of many conscientious scientists and administrators will be needed if the legal efforts to control the quality of water are to keep pace with the growing threat of pollution.

SEE ALSO THE FOLLOWING ARTICLES

Drinking Water Quality and Treatment • Environmental Measurements • Environmental Radioactivity • Pollution, Air • Pollution, Environmental • Soil and Groundwater Pollution • Transport and Fate of Chemicals in the Environment • Wastewater Treatment and Water Reclamation • Water Pollution

BIBLIOGRAPHY

Abelson, P. (1995). “Flaws in risk assessments,” *Science* **270,** 215.

Abelson, P. H. (1987). “Cancer phobia,” *Science* **237,** 473.

Abelson, P. H. (1990a). “Incorporation of new science into risk assessment,” *Science* **250,** 1497.

Abelson, P. H. (1990b). “Testing for carcinogens with rodents,” *Science* **249,** 1357.

Ames, B. N., and Gold, L. S. (1990). “Too many rodent carcinogens: Mitogenesis increases mutagenesis,” *Science* **249,** 970–971.

Environmental Protection Agency (1973). “Water Quality Criteria 1972,” EPA/R3/73/033, Environmental Protection Agency, Washington, DC.

Environmental Protection Agency (1980a). “Ambient Water Quality Criteria for Aldrin/Dieldrin,” EPA 440/5-80-019, Environmental Protection Agency, Washington, DC.

Environmental Protection Agency (1980b). “Ambient Water Quality Criteria for Cadmium,” EPA 440/5-80-025, Environmental Protection Agency, Washington, DC.

Environmental Protection Agency (1980c). “Water quality criteria documents: Availability,” Fed. Reg. **45**(231), 79318–79379.

Environmental Protection Agency (1984). “Ambient Water Quality Criteria for Cadmium—1984,” EPA 440/5-84-032, Environmental Protection Agency, Washington, DC.

Environmental Protection Agency (1985a). “Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses,” PB85-227049, Environmental Protection Agency, Washington, DC.

Environmental Protection Agency (1985b). “Methods for Measuring the Acute Toxicity of Effluents to Freshwater and Marine Organisms,” EPA/600/4-85/013, Environmental Protection Agency, Washington, DC.

Environmental Protection Agency (1986). “Quality Criteria for Water,” EPA 440/5-86-001, Environmental Protection Agency, Washington, DC.

Environmental Protection Agency (1991). “National primary drinking water standards,” *Fed. Reg.* **56,** 3526.

Environmental Protection Agency (1995). “Final water quality guidance for the Great Lakes system: Final rule,” *Fed. Reg.* **60**(56), 15365–15425.

Environmental Protection Agency (1996a). “Exposure Factors Handbook. Volume II. Food Ingestion Factors,” EPA 600/P/95/002 Bb, Office of Research and Development, National Center for Environmental Assessment, Washington, DC.

Environmental Protection Agency (1996b). “1995 Updates: Water Quality Criteria Documents for the Protection of Aquatic Life in Ambient Water,” EPA-820-B-96-001, Environmental Protection Agency, Washington, DC.

Environmental Protection Agency (1997). “Water quality standards: Establishment of numeric criteria for priority toxic pollutants for the state of California; Proposed rule,” 40 CFR, Part 131. *Fed. Reg.* **62,** 42159–42208.

Environmental Protection Agency (1999). “National Recommended Water Quality Criteria—Correction,” EPA 822-Z-99-001, Environmental Protection Agency, Washington, DC.

Laws, E. A. (2000). “Aquatic Pollution,” 3rd ed., Wiley (Interscience), New York.

Marx, J. (1990). “Animal carcinogen testing challenged,” *Science* **250,** 743–745.

Shepard, M. P. (1955). “Resistance and tolerance of young speckled trout (*Salvelinus fontinalis*) to oxygen lack, with special reference to low oxygen acclimation,” *J. Fish. Res. Bd. Canada* **12,** 387–444.

Geoenvironmental Engineering

Charles D. Shackelford
Colorado State University

GLOSSARY

Containment Control of migration of gaseous, liquid, or solid contaminated media from a site by use of containment measure(s) such as engineered liners, covers, or *in situ* vertical barriers.

Contaminant Substance that can cause harm to human health or the environment.

Cover (or cap) system One or more layers of material, such as soils, suitable mineral wastes, and geosynthetics, constructed on the surface of a site or waste impoundment and designed to control egress of contaminants from and infiltration of precipitation into a source of contamination.

Hazardous solid waste (HSW) A solid waste that meets any one of the following four criteria: (1) the waste is specifically listed as a hazardous waste by regulation, (2) the waste is a mixture containing a hazardous waste, (3) the waste is derived from the treatment, storage, or disposal of a hazardous waste (e.g., the ash residue resulting from incineration of a hazardous waste is also a hazardous waste), or (4) the waste exhibits the characteristics of ignitability, corrosivity, reactivity, or toxicity.

Hydrocarbons (HCs) Compounds consisting of hydrogen and carbon.

Leachate A chemical solution containing contaminants derived from water percolating through a solid waste.

Liner A low-permeability barrier used to prevent or reduce the migration of potentially harmful chemicals, or contaminants, into the surrounding environment.

Liner system An engineered barrier system that is composed of combination of one or more liners and drainage layers. The drainage layers are placed immediately above the liner(s) to drain the leachate that has percolated through the waste to a collection system for removal and treatment.

Municipal solid waste (MSW) “Nonhazardous” wastes from residential, commercial, institutional, and some industrial sources.

Permeability A measure of the relative ease with which a fluid (liquid or gas) will flow through a porous medium; the higher the permeability with respect to the fluid, the easier the fluid will flow through the medium.

Remediation The process of restoring land that has been polluted with contaminants.

Vadose zone The zone between the ground surface and the water table within which the water pressure is negative. The zone is divided into (1) a saturated portion, or capillary fringe, located just above the water table due to capillary rise, and (2) an unsaturated zone.

Water table The surface of water existing in the ground on which the fluid pressure is atmospheric.

GEOENVIRONMENTAL ENGINEERING can be described as the engineering of geologic (earthen) and geosynthetic (polymer) materials for problems related to the protection of human health and the environment. The primary environmental problems addressed by geoenvironmental engineers are related to either the potential for or the existence of subsurface (below the ground surface) pollution, also referred to as "contaminated land." In this context, contaminated land includes the solid portion of the ground (soil, rock) as well as the water and gas or vapor in the ground. Thus, geoenvironmental engineering is concerned with the protection of uncontaminated land as well as the remediation, or clean up, of contaminated or polluted land.

Subsurface pollution can result from a multiplicity of activities. For example, subsurface pollution can result from relatively localized or "point"-source activities, such as industrial chemical spills (e.g., jet fuel spills), leaking waste containment facilities (e.g., landfills), and leaking above-ground and underground storage tanks (e.g., oil storage tank farms, buried gasoline reservoirs), as well as from relatively dispersed or "non-point"-source activities, such as infiltration of pesticides spread over large agricultural areas. In many cases, subsurface pollution has resulted from controlled activities in the distant past (e.g., more than 30–40 years ago in the United States) not perceived to be a potential threat to the environment. In other, more recent cases, subsurface pollution typically has resulted from uncontrolled activities, such as accidental spills, or from negligent or illegal disposal practices.

With respect to prevention of subsurface pollution, geoenvironmental engineering pertains to the engineering aspects related to design, construction, and monitoring of containment facilities for a wide variety of waste disposal activities. Examples of such facilities include landfills for disposal of municipal and hazardous solid waste, impoundments for disposal of waste and by-products resulting from mining activities, lagoons for disposal of animal waste, evaporation ponds for disposal of contaminated process waters, and isolation barriers for disposal of high-level nuclear waste. In all of these applications, the principal goal of the containment facility is to prevent the escape of contaminants associated with the waste thereby preventing pollution of the surrounding environment.

With respect to the existence of subsurface pollution, geoenvironmental engineering pertains to the engineering aspects related to the remediation of the contaminated land. Remediation, also commonly referred to as reclamation, of contaminated land may be defined as the process of restoring land that has been polluted with contaminants by some activity. Although the words "remediation" and "reclamation" often are used interchangeably in terms of environmental contamination, the two words arguably have slightly different meanings. For example, reclamation of contaminated land refers to restoring the land for reuse, whereas remediation of contaminated land refers to restoration that will prevent or minimize a real (e.g., medical) or perceived (e.g., legal) risk of harm to humans (i.e., regardless of whether or not the land will be reused). This distinction is relevant in the United Kingdom, for example, where the term reclamation often is associated with derelict land (e.g., former industrial land that is no longer used) and remediation for land that is contaminated. However, as derelict land also is often contaminated, there are many situations where reclamation involves both minimization and prevention of risk as well as reuse of the land. In the United States, the two terms tend to be used interchangeably, with remediation probably being the more common term.

I. TECHNICAL CONSIDERATIONS

A. Materials

1. Geologic Materials

Geologic or earthen materials include both rock and soil. Rock exists as geologic formations, also referred to as intact or parent rock, and as fragments. Soil is the product of the effect of chemical and physical weathering or degradation of preexisting or parent material (rock or soil).

In general, earthen materials comprise three phases: (1) a gas phase, (2) a liquid phase, and (3) a solid or mineral phase, where the gas and liquid phases collectively comprise the fluid phase of the soil. In the case of rock fragments and soils, the earthen materials comprise a multitude of individual solid (or mineral) particles in contact with each other and pores (or voids) that exist between the particles. As a result, rock fragments and soils also are referred to as particulate materials.

Since the possible range of particle sizes associated with particulate earthen materials is tremendous, various classification systems based on particle size have been devised for the purpose of categorizing the behavior of soil and rock. For example, as shown in Table I, the types of earthen materials based on particle size include rock fragments consisting of boulders and cobbles, coarse-grained

TABLE I Types of Earthen Materials Based on Particle Size

Category	Type	Range of particle (or grain) sizes
Rock fragments	Boulders	>300 mm (>12 in.)
	Cobbles	76–300 mm (3–12 in.)
Coarse-grained soils	Gravel	4.75–76 mm (0.75–3 in.)
	Sand	0.075–4.75 mm (0.003–0.75 in.)
Fine-grained soils	Silt	0.002–0.075 mm (2–75 μm)
	Clay	<0.002 mm (<2 μm)

soils consisting of gravels and sands, and fine-grained soils consisting of silts and clays. Coarse-grained soils contain particles that can be seen with the unaided eye, whereas fine-grained soils require a microscope to observe individual particles.

In general, naturally occurring soils (as opposed to mined or specially processed soils) may comprise various percentages of gravel, sand, silt, and/or clay based on particle size. In this case, a fine-grained soil typically is assumed to be a soil that contains greater than 50% (w/w) of silt and/or clay, whereas a coarse-grained soil must contain 50% (w/w) or more of gravel and/or sand. Such soils are further classified as either a silt or a clay in the case of fine-grained soils, or as either a gravel or sand in the case of coarse-grained soils, based on which soil type dominates the fine-grained or coarse-grained fractions, respectively, of the soil. Coarse-grained soils that contain small amounts of fines (e.g., $<5\%$ silts and/or clays) typically are referred to as "clean" soils.

As noted in Table I, the term "clay" refers to that part of the soil that is composed of particle sizes less than 2 μm. However, the term "clay" also can refer to the clay mineral content of a soil regardless of the particle size. Finally, the term "clay" often is used to denote a soil mass that possesses properties similar to those of a pure clay even though only a relative small percentage of the overall soil mass consists of clay-sized particles and/or clay minerals. For example, mixtures of sand and bentonite, a high-swelling clay, containing 4–10% (w/w) of bentonite typically possess some properties that are characteristic of clay soils, even though 90% (w/w) or more of these mixtures is made up of sand.

The pore spaces in earthen materials may be occupied by gases, liquids, or both. For example, with respect to the liquid phase in a soil, a "dry soil" results when the pores are completely occupied by a gas, typically air (Fig. 1a), whereas a "saturated soil" results when the pores are completely filled with a liquid, typically water (Fig. 1b). An "unsaturated" soil results when the pores comprise both a gas and a liquid (Fig. 1c). In the case of unsaturated soils containing both water and air (Fig. 1c), the water is the "wetting fluid" because the water is preferentially absorbed (i.e., relative to the air, the "nonwetting fluid) to the surface of the solid soil particles and therefore tends to wet the surface.

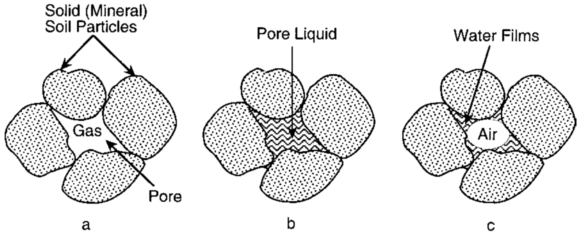


FIGURE 1 Particulate nature of soils: (a) dry soil; (b) saturated soil; (c) unsaturated soil.

When the liquid phase in a soil is water, the liquid phase is more specifically referred to as the "aqueous phase." In reality, the aqueous phase in soils is made up of the compound water (H_2O) and other chemical compounds or species that are dissolved in the water. In this case, the water serves as the "solvent" and the dissolved compounds are referred to as "solutes." Solutes in aqueous solutions may exist in a variety of forms or species. For example, solutes may exist as uncharged chemical species, or neutral compounds, as well as charged chemical species, or ions. Positively charged ions are called cations, and negatively charged ions are called anions. For example, ionic species common in natural pore waters of soils include the monovalent cations of potassium (K^+) and sodium (Na^+), the divalent cations of calcium (Ca^{2+}) and magnesium (Mg^{2+}), the monovalent anion of chlorine, or chloride (Cl^-), and the divalent sulfate anion (SO_4^{2-}).

The concentrations of the solutes in the aqueous phase serve as a basis for evaluating the quality of the water. For example, water that contains an abundance of dissolved calcium and magnesium commonly is referred to as "hard water." Contamination of soils results when specific solutes, referred to as contaminants of concern (COCs), exist in the subsurface, and pollution of the subsurface results when these COCs exist at concentrations or levels that are considered to be harmful to human health and the environment.

Because of the porous nature of earthen materials, fluids can flow through earthen materials when the pores are interconnected. As a result, earthen materials also are referred to as porous media. In general, the ease with which a fluid will flow through a porous medium increases with an increase in the size of the pores, and the size of the pores increases with an increase in the size of the particles making up the porous medium. Thus, according to Table I, the ease with which a fluid will flow through soils decreases in the following order: gravel > sand > silt > clay. The ability of contaminants to migrate through the interconnected pores of soils and rocks is the most important technical consideration with respect to either preventing the migration of contaminants into the subsurface in the case of waste disposal, or removing contaminants from the subsurface in the case of remediation. For example, only soils that tend to restrict fluid flow (e.g., clays) are considered for use as containment barriers in waste disposal applications, whereas remediation of subsurface pollution typically is effective only in soils that allow rapid movement of fluids (e.g., sands and gravels).

As a liquid-saturated soil begins to dry, some of the liquid (e.g., water) in the soil pores is displaced by gas (e.g., air). In this initial stage of drying, the liquid phase is continuous throughout the soil and therefore the liquid can flow freely through the soil. However, the gas exists as isolated, or occluded, bubbles. In this condition, the gas phase is discontinuous, restricting free flow of the gas. As the soil is dried further, the gas phase continues to occupy a greater percentage of the pore space until the gas phase eventually becomes continuous and can flow freely through the soil. However, at the same time, the liquid phase is continuously occupying a lesser percentage of the pore space, such that the liquid phase eventually becomes discontinuous, restricting the free flow of the liquid. Upon rewetting an unsaturated soil, the invading liquid displaces the remnant gas, and the process reverses. Thus, in general, the greater the degree of liquid saturation (i.e., percentage of the total pore space occupied by the liquid), the greater the ease with which a liquid will flow through the soil, whereas the lesser the degree of liquid saturation, the greater the ease with which a gas will flow through the soil.

2. Geosynthetics

Geosynthetics are manufactured synthetic materials (i.e., made from polymers or hydrocarbon chains) that are used for a wide range of engineering applications. Seven categories of geosynthetic materials are manufactured (Koerner 1998): geocomposites, geogrids, geomembranes, geonets, geopipes, geotextiles, and geosynthetic clay liners. A geocomposite consists of a combination of one or more geosynthetics, specifically a geogrid, a geotextile, a geomembrane, and/or a geonet, with another material. Geotextiles are used primarily for applications requiring separation, filtration, reinforcement, and drainage. Geogrids are used as reinforcement to improve the strength of soil or other materials. Geonets are relatively thin, planar geosynthetics used for drainage applications, whereas geopipes are buried plastic pipes used for drainage. Geomembranes and geosynthetic clay liners are thin ($\leq$10 mm), relatively impervious geosynthetics that are used as barriers in containment applications for waste disposal and *in situ* remediation. Geosynthetics have gained widespread use only since about 1990, but their use in geoenvironmental engineering applications involving liquid and waste containment has grown substantially in the last decade. A detailed description of the properties, characteristics, and uses of each of these types of geosynthetic can be found in Koerner (1998).

B. Fluids

1. Liquids

The properties of the liquid phase of the porous medium (i.e., pore liquid) will depend on the nature and chemical composition of the pore liquid. As shown in Fig. 2, pore liquids may be categorized as aqueous liquids, i.e., solutions containing contaminants that are miscible in water (also known as hydrophilic or "water-loving"), or nonaqueous liquids consisting of organic compounds that are immiscible in water (also known as hydrophobic or "water-hating"). In addition, mixtures of both aqueous and nonaqueous liquids resulting in the formation of two separate liquid phases may occur. Aqueous liquids contain inorganic chemicals (acids, bases, salts) and/or hydrophilic organic compounds. In the case of applications involving nuclear wastes, aqueous solutes also may be radioactive.

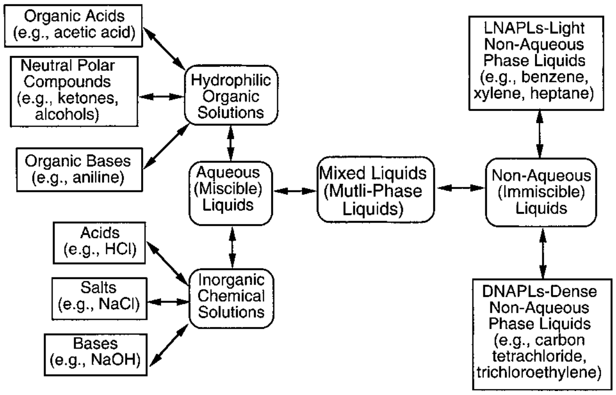


FIGURE 2 Categories of liquids. [Adapted from Shackelford and Jefferis (2000).]

Hydrophilic organic compounds are distinguished from hydrophobic organic compounds based on the concept of "like dissolves like," i.e., polar organic compounds usually will readily dissolve in water, a polar molecule, whereas nonpolar organic compounds are repelled by water. A hydrophobic compound also is typically further categorized as either an LNAPL (light nonaqueous-phase liquid) or DNAPL (dense nonaqueous-phase liquid) based on whether the density of the compound is lower or greater than that of water, respectively. Thus, LNAPLs will tend to float on water and DNAPLs will tend to sink in water.

In general, even though NAPLs are considered to be aqueous immiscible, most NAPLs have some aqueous solubility. Although the aqueous solubility of NAPLs is small relative to hydrophilic compounds, the solubilities of most NAPLs typically result in aqueous-phase concentrations that far exceed regulatory limits. Thus, NAPLs existing in the subsurface represent sources of continuous aqueous-phase contamination when the NAPLs come into contact with flowing ground water.

Subsurface pollution can result from any category of liquid shown in Fig. 2. For example, toxic heavy or trace metals are derived from inorganic chemical solutions resulting from industrial activities as well as from the leaching of metals from solid waste during passage of infiltrating water. Some heavy metals commonly of environmental concern include arsenic, barium, cadmium, chromium, copper, lead, mercury, and nickel.

Organic solvents frequently are used as cleaning (e.g., degreasing) agents in industrial manufacturing and processing activities. Among the many types of organic solvents used in industry, a group of DNAPLs known as the chlorinated hydrocarbons are particularly prevalent. In particular, trichloroethethylene (TCE) and tetrachloroethylene, also known as perchloroethylene (PCE), are two of the most ubiquitous groundwater pollutants in industrialized countries.

Another major group of chemicals commonly encountered in conjunction with groundwater pollution are a group of LNAPLs referred to as the BTEX (benzene, toluene, ethyl benzene, and xylene) compounds. The BTEX compounds are derived from petroleum, are major constituents in gasoline and other petroleum products, and represent a significant environmental concern with respect to pollution of the subsurface.

2. Gases

The gas phase in liquid unsaturated porous media can contain a wide variety of important chemical constituents. For example, the amount of carbon dioxide gas (CO_2) in the soil can affect the chemistry of aqueous solutions, and the tendency for aerobic and anaerobic conditions is reflected by the existence and amount of oxygen gas (O_2).

Gas-phase contamination can result from the direct production of a gas, such as the production of methane gas (CH_4) resulting from the decomposition of municipal solid waste, or from the volatility of chemical compounds. Volatility of hydrocarbons is related to the vapor pressure of the compound (i.e., the pressure in the vapor above the liquid at equilibrium). In general, the greater the vapor pressure, the more volatile the compound. For example, volatile organic compounds (VOCs) are hydrocarbons with vapor pressures at 20°C that are greater than 1 mm Hg (1.3×10^{-3} atm). Semivolatile organic compounds (SVOCs) are hydrocarbons with vapor pressures between 10^{-10} mm Hg (1.3×10^{-13} atm) and 1 mm Hg (1.3×10^{-3} atm). Although the lower limit of vapor pressures for SVOCs is 10^{-10} mm Hg (1.3×10^{-13} atm), hydrocarbons with vapor pressures $<10^{-7}$ mm Hg ($<1.3 \times 10^{-10}$ atm) are not expected to volatilize significantly. Volatility of chemical compounds also is indicated by the boiling point (i.e., the lower the boiling point, the greater the susceptibility to volatilization via heating).

Five VOCs that are commonly associated with pollution of drinking water are tetrachloroethylene (PCE), trichloroethylene (TCE), vinyl chloride (VC), 1,2-dichloroethane (DCE), and carbon tetrachloride (CCl_4). Examples of SVOCs include cresol, phenol, pyrene, 1,2- dichlorobenzene, and 1,4-dichlorobenzene.

C. Fluid Flow

1. Liquid Flow

a. Hydraulic conductivity. Hydraulic conductivity refers to the coefficient of proportionality in Darcy's law describing liquid flow through porous media. For example, flow of liquids in and through soil may be described in accordance with Darcy's law as follows:

$$q = \frac{Q}{A} = -ki, \tag{1}$$

where q is the liquid flux or flow rate (length/time, L T^{-1}), Q is the volumetric flow rate of the liquid (L^3 T^{-1}), A is the total cross-sectional area of the soil (solids plus voids) perpendicular to the direction of flow (L^2), k is the hydraulic conductivity or coefficient of permeability (L T^{-1}), and i is the hydraulic gradient (dimensionless). The hydraulic gradient represents the energy or driving force for liquid flow. The hydraulic conductivity is a material property that reflects the relative ease of liquid flow through porous media, and is among the most variable material properties in engineering. Approximate ranges in the hydraulic conductivities under saturated conditions of various materials commonly encountered in geoenvironmental engineering are provided in Table II.

For a given hydraulic gradient, the fluid flow rate q is directly proportional to the hydraulic conductivity of the porous medium in accordance with Eq. (1). Thus, in accordance with Eq. (1), the objective of most low-permeability

TABLE II Approximate Ranges of Saturated Hydraulic Conductivities Expected for Materials Commonly Encountered in Geoenvironmental Engineering Based on Permeation with Water

Material	Saturated hydraulic conductivity k_{sat} (m/sec)	Comments
Gravel	10^{-2}–10^{-3}	Values based on "clean" soils; variation in k_{sat} based on particle size distribution
Sand	10^{-3}–10^{-5}	
Silt	10^{-5}–10^{-8}	Variation in k_{sat} based on mineralogical composition of silt particles
Clay	10^{-8}–10^{-12}	Variation in k_{sat} based on mineralogical composition of clay particles
Geosynthetic clay liner	10^{-10}–10^{-11}	Values based on sodium bentonite sandwiched between two geotextiles
Sand–bentonite mixture	10^{-9}–10^{-10}	Values based on a mixture of clean sand (w/o fines) and 4–10% (w/w) sodium bentonite

barriers used for waste containment systems prescribed by regulation is to minimize the amount of liquid percolation by (1) achieving a suitably low saturated hydraulic conductivity for the barrier material and (2) restricting the amount of liquid buildup on top of the barrier to maintain a suitably low hydraulic gradient. For example, in the United States, the regulatory maximum value for the saturated hydraulic conductivity k_{sat} prescribed for municipal and solid waste landfill liners is 10^{-9} m/sec and the maximum height of liquid allowed to accumulate on top of the liner is 300 mm. This regulatory limit on k_{sat} may vary depending on the country, the specific component of the containment facility (e.g., liner versus cover), and/or the specific application (e.g., solid waste disposal versus nuclear waste disposal).

b. Compatibility. The hydraulic conductivity k is a function of the properties of both the solid matrix of the porous medium and the liquid in accordance with the following equation:

$$k = K\left(\frac{\gamma}{\mu}\right) = K\left(\frac{\rho g}{\mu}\right) = K\left(\frac{g}{\upsilon}\right), \tag{2}$$

where K is the intrinsic, absolute, or specific permeability of the soil (L^2), γ is the unit weight of the liquid (mass/length^2/time, $M\ L^{-2}\ T^{-1}$), μ is the absolute or dynamic viscosity of the liquid ($M\ L^{-1}\ T^{-1}$), ρ is the mass density of the liquid ($M\ L^{-3}$), g is acceleration due to gravity ($L\ T^{-2}$), and υ is the kinematic viscosity of the liquid ($L^2\ T^{-1}$). The intrinsic permeability K represents the influence of the soil structure on the hydraulic conductivity, whereas the ratio γ/μ reflects the influence of the permeant liquid on the hydraulic conductivity.

Compatibility refers to the effect of potential interactions between the permeating liquid and the solid particles on the hydraulic conductivity of the porous medium. For example, changes in the pore spaces available for flow can occur when a clay soil is permeated with liquids with significantly different chemical properties relative to the naturally occurring pore water, such as in the case of clay soils used for waste containment barriers. These changes result from incompatibility between the permeant liquid and the solid soil particles.

With respect to Eq. (2), when the permeant liquid has a value of γ/μ that is significantly different than that of water, incompatibility between the permeating liquid and the porous material is indicated by an increase in the intrinsic permeability of the porous material (i.e., $\Delta K > 0$). In many cases, the values of γ/μ for the permeant liquid and water are sufficiently close such that incompatibility also will be reflected by an increase in hydraulic conductivity (i.e., $\Delta k > 0$). From a practical viewpoint, an increase in hydraulic conductivity may be considered significant if the final (steady-state) value exceeds the regulatory maximum value of the hydraulic conductivity. When this occurs, the soil being considered for use as the barrier layer in the containment system and the waste liquid are considered "incompatible," and a different soil and/or material must be evaluated for use as the barrier.

c. Hydraulic conductivity of unsaturated media. The concavity of the interface toward the nonwetting fluid in a liquid unsaturated soil (e.g., see Fig. 1c) reflects a higher pressure in the nonwetting fluid relative to the wetting fluid. The resulting difference between the pressure in the nonwetting fluid and the pressure in the wetting fluid is called the capillary pressure. In the case where the nonwetting fluid is at atmospheric pressure, or zero reference pressure, the pressure in the wetting fluid is negative, reflecting that the liquid is in tension. As the pores desaturate further, the concavity of the interface increases, reflecting an increase in the capillary pressure and an increase in the magnitude of the negative pressure in the wetting fluid. In the case where the liquid is water, the capillary pressure also often is referred to as the soil–water suction. As a result, the capillary pressure, or soil–water suction, increases in magnitude as the porous medium desaturates.

The hydraulic conductivity of a porous medium with respect to the wetting fluid (e.g., water) decreases as the medium desaturates, all other factors being equal. Thus, the maximum value of the hydraulic conductivity for a given porous medium is the saturated hydraulic conductivity. The decrease in hydraulic conductivity with decreasing saturation results from two effects: (1) the cross-sectional area available for flow of the liquid via the liquid phase decreases as the porous medium desaturates (e.g., compare Figs. 1b and 1c), and (2) the pathways for liquid migration through the liquid phase in porous medium become more tortuous. The increased tortuosity results from two effects. First, as the medium desaturates, the liquid is forced to migrate closer to the particle surface (see Fig. 1c) resulting in an increase in the effective length of the migration pathway through the medium. Second, since the most conductive pores are the largest pores, these pores drain faster than the smaller pores resulting in some of the larger pores being excluded from liquid flow.

Since larger pores drain faster than smaller pores, the rate of decrease in hydraulic conductivity for porous media that are made up primarily of larger pores, such as coarse-grained soils, is greater than that associated with porous media made up primarily of smaller pores, such as fine-grained soils. Thus, although the saturated hydraulic conductivity of coarse-grained soils is significantly higher than that of fine-grained soils (Table II), the reverse situation occurs as the degree of liquid saturation of the soil

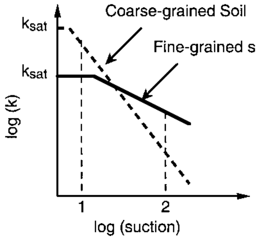


FIGURE 3 Relative trends in hydraulic conductivity of two soil types as a function of soil suction.

decreases (i.e., as soil–water suction increases). These trends are reflected schematically in Fig. 3.

2. Gas Flow

The migration of gas becomes an important consideration in certain geoenvironmental engineering applications when the gas content in a porous medium is sufficiently high such that the gas phase is continuous. Examples of such applications include the removal of VOCs and SVOCs from the unsaturated or vadose zone above the water table in the subsurface via the gas phase and the minimization of oxygen influx or radon efflux from engineered covers for tailings disposal applications. In the former case, removal efficiency is improved as gas permeability increases, whereas in the latter case, the objective is to minimize the gas permeability and therefore minimize the gas flow.

When the gas phase is continuous, gas may migrate through the porous medium under a pressure gradient in accordance with an analogous form of Darcy's law for liquid-phase flow as follows:

$$q_{\mathrm{g}} = \frac{Q_{\mathrm{g}}}{A} = -\frac{K_{\mathrm{g}}}{\mu_{\mathrm{g}}} i_{\mathrm{P}}, \qquad (3)$$

where q_{g} is the gas flux or flow rate, Q_{g} is the volumetric flow rate of the gas, K_{g} is the intrinsic permeability of the porous medium with respect to gas flow, μ_{g} is the absolute or dynamic viscosity of the gas, and i_{P} is the pressure gradient.

As with liquid flow, the intrinsic permeability K_{g} with respect to gas flow is a function of only the porous medium. However, the ability of a gas to migrate through a porous medium will depend on the liquid content in the porous medium. When the porous medium is completely dry, K_{g} represents the continuity of flow through the pores attributable only to the porous solids (i.e., $K_{\mathrm{g}} = K$). However, as the pores are filled with liquid (e.g., water) such that the percentage of gas-filled pores decreases, the gas may become occluded in some pores such that some of the pathways for gas-phase migration are cut off. When the porous medium is sufficiently saturated with liquid such that gas in the pores exists as occluded bubbles (i.e., bubbles completely surrounded by the liquid phase), the gas phase becomes discontinuous and gas-phase migration via Eq. (3) effectively stops. Thus, the gas-phase permeability as reflected by K_{g} also is a function of the degree of liquid saturation of the porous medium.

D. Contaminant Migration

Most geoenvironmental engineering applications involve some consideration of the rate and extent of migration of contaminants through a porous medium, often referred to as contaminant transport. In the case of waste containment, the objective is to minimize the rate of contaminant migration from the containment facility, whereas in the case of remediation, the objective may be either to enhance or to minimize the rate of contaminant migration.

1. Aqueous-Phase Transport

Aqueous-phase, or aqueous miscible, contaminant transport in porous media is controlled by a variety of physical, chemical, and biological processes. As indicated in Table III, the primary physical processes governing aqueous miscible contaminant transport are advection, diffusion, and dispersion. Diffusion tends to be the dominant transport process under relatively low flow-rate conditions, such as occur through clay barriers used for containment in waste disposal and remediation applications. However, advection, or hydraulically driven transport, and dispersion dominate under relatively high flow-rate conditions, such as occur for contaminant migration through coarse-grained aquifer materials.

The chemical and biological processes commonly considered to potentially affect contaminants during aqueous-phase migration through porous media are described in Table IV. Some of the processes, such as adsorption, radioactive decay, precipitation, hydrolysis, and biodegradation, typically are considered as "attenuation processes" in that these processes generally immobilize,

TABLE III Physical Processes Affecting Aqueous Miscible Contaminant Transport[a]

Process	Definition	Significance
Advection	Mass transport due to bulk water flow	Most dominant process in high-flow-rate media
Diffusion	Mass spreading due to concentration gradients	Most dominant process in low-flow-rate media
Dispersion	Mass spreading due to heterogeneities in the flow field	Results in greater mass spreading than predicted by advection

[a] National Research Council (1990), Shackelford and Rowe (1998).

TABLE IV Chemical and Biological Processes Affecting Aqueous Miscible Contaminant Transport[a]

Process	Definition	Significance
Sorption	Partitioning of contaminant between pore water and porous medium	Limiting case of ion exchange; adsorption reduces the rate of apparent contaminant migration and makes contaminant removal difficult by desorption
Radioactive decay	Irreversible decline in the activity of a radionuclide	Important attenuation mechanism when the half-life for decay is $\leq$ residence time in flow system; results in by-products
Dissolution/precipitation	Reactions resulting in release of contaminants from solids or removal of contaminants as solids	Dissolution is mainly significant at the source or at the migration front; precipitation is an important attenuation mechanism, particularly in high-pH system (pH >7)
Acid/base	Reactions involving a transfer of protons (H^+)	Important in controlling other reactions (e.g., dissolution/precipitation)
Complexation	Combination of anions and cations into a more complex form	Affects speciation that can affect sorption, solubility, etc.
Hydrolysis/substitution	Reaction of a halogenated organic compound with water or a component ion of water (hydrolysis) or with another anion (substitution)	Typically make an organic compound more susceptible to biodegradation and more soluble
Oxidation/reduction (redox)	Reactions involving a transfer of electrons	Important attenuation mechanism in terms of controlling precipitation of metals
Biodegradation	Reactions controlled by microorganisms	Important attenuation mechanism for organic compounds; may result in undesirable by-products

[a] National Research Council (1990), Shackelford and Rowe (1998).

retard, or otherwise degrade chemical constituents that exist in the aqueous phase of the porous media. However, in some cases, these attenuation processes may not be effective in actually reducing the potential impact of the contaminants. For example, the radioactive decay of the initial contaminant results in by-products that also may represent a potential adverse environmental impact. Similarly, subsequent desorption of a previously adsorbed contaminant or dissolution of a previous precipitated contaminant also may result in negative environmental impacts.

2. Aqueous Immiscible Flow

Aqueous immiscible flow, or immiscible displacement, involves the simultaneous flow of two or more immiscible fluids in the porous medium. Since the fluids are immiscible, the interfacial tension between the two fluids is not zero, and a distinct fluid–fluid interface separates the fluids.

The migration of two separate fluid phases in fluid-saturated soil, referred to as two-phase flow, requires special consideration for the interaction between the two fluid phases. Due to surface and interfacial tension effects, one of the two fluids preferentially wets the surface of the pores resulting in a characteristic concave curvature of the fluid–fluid interface toward the nonwetting fluid and a discontinuity in pressure (i.e., capillary pressure) exists across the interface separating the two fluids, as illustrated in Fig. 4. The nonwetting fluid can displace the wetting fluid in the pore space of the soil only after a minimum pressure, known as the entry pressure, for the nonwetting fluid is reached. For example, in the case of airflow through initially water-saturated soil, this minimum pressure is known as the air-entry pressure.

For the two-fluid flow scenario, such as in the case of immiscible displacement of a water-saturated soil by a nonaqueous-phase liquid (NAPL) that is nonwetting, the migrations of both the wetting and the nonwetting fluids are described by a modified form of Darcy's law [Eq. (1)] written for each fluid. In this case, fluid flow is complicated by the fact that the permeabilities for the wetting and nonwetting fluids are functions of the degree of saturation of the wetting and nonwetting phases, respectively, and therefore the permeabilities change with location and time. Thus, the governing system of flow equations

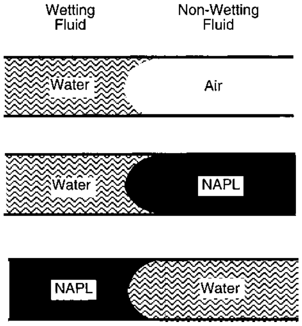


FIGURE 4 Capillary tube schematics of wetting versus nonwetting fluids. NAPL, Nonaqueous-phase liquid. [Adapted from Shackelford and Jefferis (2000).]

typically is highly nonlinear requiring the use of numerical methods for the solution. Further details regarding the application of this approach for describing DNAPL migration can be found in Pankow and Cherry (1996) and Charbeneau (2000).

3. Gas-Phase Transport

In addition to a gas migrating in response to a pressure gradient [Eq. (3)], gas also can migrate in the absence of a pressure difference via diffusion in response to a difference in the gas-phase concentration of the gas. All other factors being equivalent, the diffusive migration of chemical compounds through the gas phase typically ranges from 10^4 to 10^5 times higher than the liquid-phase diffusive migration of the same compounds. Thus, diffusion of gas is a potentially significant consideration in liquid unsaturated porous media. Further details regarding gas-phase diffusion can be found in Charbeneau and Daniel (1993) and Shackelford and Rowe (1998).

E. Mechanical Considerations

In geoenvironmental engineering applications involving design and construction of waste containment facilities, knowledge of the strength and stress–strain properties of the engineering materials being used in the containment facility as well as the waste being disposed is required. For example, consider a liquid waste containment facility that is constructed by excavating a disposal pit followed by placing a geomembrane liner (GML) that is subsequently covered with a layer of soil, as shown in Fig. 5. The cover soil is typically placed to protect the GML from damage resulting from exposure to ultraviolet light prior to placement of the waste. Design of this facility requires several considerations that require a knowledge of the strength and stress–strain properties of the component materials.

First, the stability of the side slopes (Fig. 5a) with respect to sliding within the foundation soil must be determined by calculating the shear stress τ along the assumed failure plane resulting from the self-weight of the failure wedge and the resisting, shear strength s provided by the soil along the same shear plane. In the case where $\tau > s$, either the side slopes must be cut back (i.e., reduce the side slope angle β), possibility resulting in a reduction in storage capacity, or the strength of the soil must be increased by engineering a stabilization scheme.

Second, for a given β, the potential for sliding of the soil cover along the interface of the soil cover and the GML (Figs. 5b and 5c) must be evaluated by calculating the shear stress τ along the assumed failure plane resulting from the self-weight of the soil cover and the shear strength s offered by the frictional resistance along the soil cover–GML interface. If $\tau > s$ in this case, the engineer may select a different geomembrane product for the GML, select a different soil cover material, and/or reduce β.

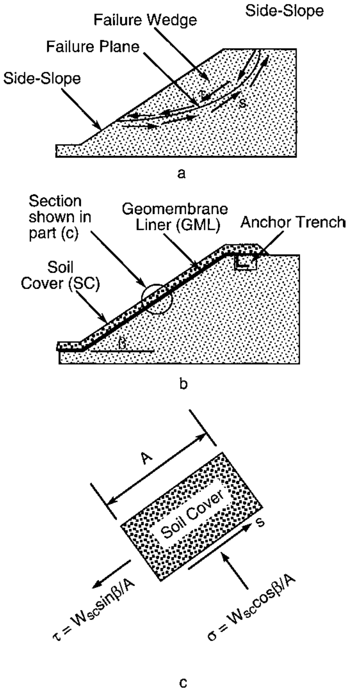


FIGURE 5 Design considerations for installation of a liquid waste containment facility: (a) side slope stability from disposal pit excavation; (b) placement of liner and soil cover; (c) free-body diagram for sliding along soil cover–GML interface (τ is the applied shear stress, σ is the normal stress, s is the shear strength, and W_{SC} is the weight of the soil cover).

Third, in the case where the interface friction between the GML and side slope is sufficient to prevent the soil cover from sliding along the surface of the GML, the GML and the soil cover may slide together down the slopes if the GML is not properly anchored or the GML has inadequate tensile strength and fails by rupture. Thus, the engineer must ensure that suitable tensile capacity and suitable anchorage capacity, for example, through design of an anchor trench (Fig. 5b), are provided.

In addition to mechanical considerations involving the strength of component materials used in waste containment facilities, consideration also must be given to the compressibility of the foundation soil and the compressibility of waste being disposed. For example, municipal solid waste (MSW) typically is highly compressible and degrades with time due to biological-mediated decomposition. Thus, as illustrated schematically in Fig. 6, a significant potential for large, nonuniform settlement of the surface of the waste pile exists in MSW landfills. In the case where final covers are placed on top of the waste pile to keep precipitation from coming into contact with the waste after closure of the facility, such large

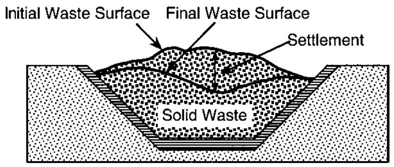


FIGURE 6 Nonuniform settlement in a solid waste disposal facility resulting from compressibility and time-dependent degradation of the waste.

settlements can have deleterious effects in terms of the integrity of the cover system. Thus, the engineer must design the cover system to withstand these settlements, and special precautions, such as compacting the waste during disposal, must be employed to minimize the potential for such settlements.

II. WASTE CONTAINMENT

Subsurface pollution resulting from waste disposal activities occurs when potentially harmful liquids reach the environment beneath the waste disposal site. The liquids usually originate from precipitation, disposal activities, and/or compression of the waste. The amount and quality of liquids generated depend primarily on the climatic conditions, the physical and chemical properties of the waste, and the disposal operations. This potential for groundwater pollution typically is controlled through encapsulation of the liquid or waste material using engineered liner and cover systems.

A. Liners

1. Types of Liners

The primary objective of a liner, or barrier layer, is to prevent or reduce the migration of potentially harmful chemicals, or contaminants, into the surrounding environment. With respect to this objective, several different types of liners are used for the containment of waste. As outlined in Table V, the different types of liners may be separated into three broad classifications: (1) earthen (soil) liners, (2) geosynthetic (polymer) liners, and (3) composite liners. Earthen liners may be further divided into naturally occurring low-permeability soil liners and manmade liners (e.g., compacted or mechanically stabilized low-permeability soils).

Natural (soil) liners (NLs) are formed by low-permeability geologic formations, such as aquitards or aquicludes. An *aquitard* is a geologic formation (e.g., interbedded sand and clay lenses) that transmits water at a very slow rate relative to an aquifer, and an *aquiclude* is a geologic formation so impervious that it completely obstructs the flow of ground water (e.g., shale). Several disposal scenarios based on natural soil liners for liquid waste disposal are shown in Fig. 7.

TABLE V Types of Liners Used for Waste Containment

Liner type	Origin	Examples
Earthen or soil	Geologic formation	Aquicludes and aquitards (NL)
	Borrow source/quarry	Compacted clay liner (CCL)
Geosynthetic	Manufactured	Geomembrane liner (GML), geosynthetic clay liner (GCL)
Composite	Manufactured-single	GCL made w/geomembrane
	Manufactured-combined	GML/GCL
	Borrow source/quarry + geologic formation	CCL/NL
	Manufactured + geologic formation	GML/NL or GCL/NL
	Manufactured + borrow source/quarry	GML/CCL or GCL/CCL

Compacted soil liners also are commonly referred to as compacted clay liners (CCLs) because only soils that contain a sufficient amount of clay can provide the suitably low saturated hydraulic conductivity typically required for containment applications (e.g., $k_{\text{sat}} \leq 10^{-9}$ m/sec). The primary purposes of compacting a source clay for use as a liner material are to reduce the saturated hydraulic conductivity by densifying the clay (i.e., reducing the pore sizes) and to destroy any small secondary defects (e.g., fissures or cracks) that may act as conduits for flow in the natural soil formation.

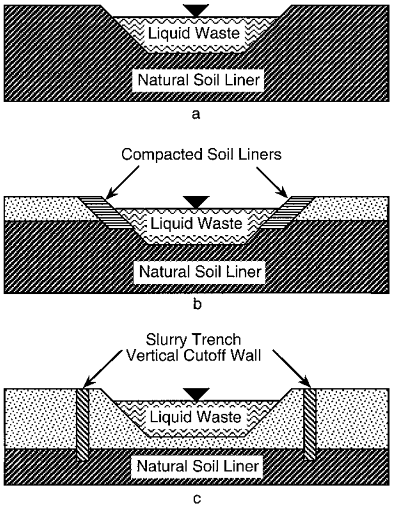


FIGURE 7 Examples of waste disposal cells (a) entirely within, (b) partially within, and (c) entirely above a natural soil liner.

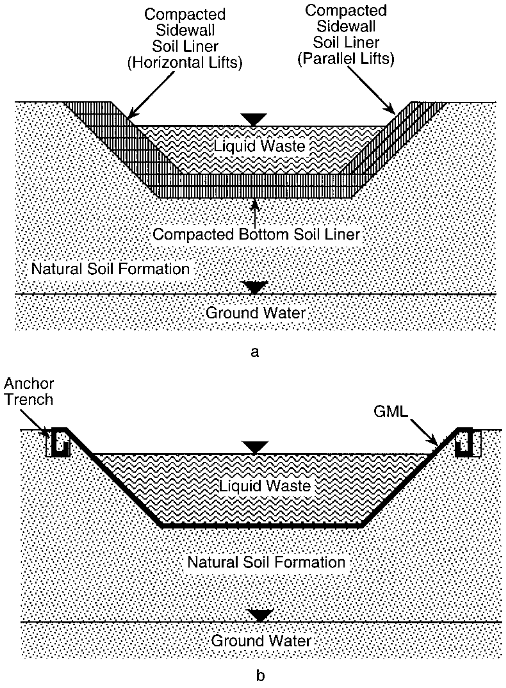


FIGURE 8 Liquid waste disposal cells lined with engineered liners: (a) compacted clay (soil) liner and (b) geomembrane liner (GML).

As schematically illustrated in Fig. 8a, CCLs typically are constructed in 150-mm-thick layers, or lifts, with the total thickness of the CCL being equal to the number of compacted lifts times the lift thickness. Thus, a 0.6-m-thick CCL would comprise four compacted lifts of clay. The saturated hydraulic conductivity of a CCL is particularly sensitive to the amount of water in the pores at compaction, with significant decreases (typically 10–1000 times) in the saturated hydraulic conductivity occurring as the water saturation increases at compaction. Thus, CCLs typically must be compacted at high water saturations (e.g., 85–95%) in order to achieve suitably low saturated hydraulic conductivities.

Geomembranes liners (GMLs) are flexible, thin [0.25–5.0 mm (10–200 mils)] sheets of rubber or plastic material that are manufactured at a factory, transported to the site, rolled out, and welded together to form the liner containment system. Geomembrane liners can be utilized in the same manner as CCLs, as illustrated in Fig. 8b. The minimum thickness of GMLs for waste disposal applications usually is 0.75 mm (30 mils). Unless punctured, GMLs generally are impervious to liquid migration, but VOCs typically can migrate readily through GMLs via molecular diffusion.

Geosynthetic clay liners (GCLs) are thin ($\leq$10 mm), flexible sheets of sodium bentonite ($\sim$4.9 kg/m^2 of bentonite) either sandwiched between two geotextiles (Figs. 9a and 9b) or glued to a geomembrane (Fig. 9c). In some cases, an adhesive is mixed with the bentonite to hold the GCL together (Figs. 9a and 9c). In other cases, the GCL is held together by needle-punched fibers or stitching (Fig. 9b). The GCLs are manufactured in panels (4–5 m by 25–60 m), placed on rolls at the factory, shipped to the site, and rolled out to form a thin barrier. Sheets of panels can be stacked to form thicker barriers. An advantage of GCLs relative to GMLs is that the GCL panels do not have to be seamed or otherwise fused together as do GMLs, since the swelling bentonite in the presence of water provides a seam by simply overlapping the panels. In some cases, loose bentonite is placed between the overlapped portions of the panels to assist in sealing the panels.

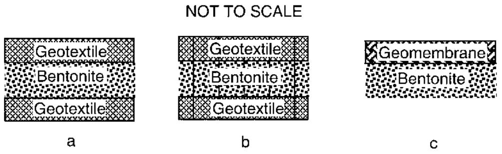


FIGURE 9 Typical GCL cross sections: (a) sodium bentonite with adhesive between two geotextiles, (b) sodium bentonite stitched between two geotextiles, and (c) sodium bentonite with adhesive attached to geomembrane.

Although GMLs and GCLs typically are much less permeable than compacted clay liners (CCLs), GMLs and GCLs also are very thin. Thus, a puncture or tear in a GML or GCL can have deleterious results with respect to the integrity of a liner system. For example, a small puncture in the bottom of the GML shown in Fig. 8b would result in a situation analogous to pulling the plug in a bathtub filled with water. As a result, the concept of the composite liner system, in which a natural and/or man-made soil liner serves as a backup to a GML or GCL, has emerged.

Composite liners are combinations of soil liners and/or geosynthetic materials. The properties of composite liners are essentially the same as the component materials for the composite liner. However, an important aspect for the effective use of composite liners is that the composite materials must be in intimate contact (i.e., without separation between materials). In reality, the type of GCL shown in Fig. 9c is a manufactured composite liner. Composite liner systems typically are used in conjunction with municipal and hazardous solid waste disposal.

2. Liner Systems for Solid Waste Disposal

The disposal of municipal solid waste (MSW) and, to a lesser extent, hazardous solid waste (HSW) represents a significant activity associated with geoenvironmental engineering. In most countries, the minimum design requirements for liner systems designed and constructed to contain either MSW or HSW are prescribed by regulation and therefore are referred to as prescriptive liner systems. In general, liner systems for MSW and HSW waste disposal consist of a combination of one or more drainage layers and liners. In these systems, the liner is a low-permeability barrier layer that is placed to impede the migration of leachate and gas from the solid waste, and the drainage layer is placed immediately above the liner to drain off the leachate that has percolated through the waste, thereby limiting the hydraulic gradient for contaminant migration through the liner. The collected leachate typically is drained to a collection basin (sump) where the leachate can be removed for treatment by pumping.

The basic components of solid waste liner systems typically consist, from top to bottom, of a leachate collection and removal system (LCRS), a top liner, a leakage detection system (LDS), and a bottom liner. The LCRS typically is made up of either a highly permeable soil (e.g., gravel) or a geosynthetic drainage material (e.g., geonet), with possibly a network of perforated pipes to drain the collected leachate to the sump (e.g., geopipes). The top liner consists of either a geomembrane liner (GML) or a composite liner. The primary purpose of the LDS, which is made up of the same materials as the LCRS, is to allow the detection of contaminants that may have migrated through the top liner (e.g., due to a leak). However, in the case of a leak, the LDS also can serve to collect and drain the leakage. An LDS is required only when a bottom liner is required. The bottom liner, if required, typically consists of a composite liner made up of a GML in intimate contact with 0.9 m of a low-permeability (typically $k_{\text{sat}} < 10^{-9}$ m/sec) CCL.

The minimum prescriptive (regulated) cross sections for liner systems for MSW and HSW disposal in the United States are shown in Fig. 10. As shown in Fig. 10a, the minimum requirements for MSW include a single composite liner overlain by an LCRS. Thus, an LDS and bottom liner are not required for MSW disposal in the United States. However, a double liner system consisting of an LDS sandwiched between a bottom composite liner system and a GML top liner is required in the case of HSW disposal (Fig. 10b). The LDS and bottom liner in the HSW requirements provide a level of redundancy in the containment that is required due to the particularly hazardous nature of HSW.

B. Covers

Once the disposal capacity of the facility or containment cell is reached, an engineered final cover is placed over the waste to (1) prevent the generation of leachate by minimizing the amount of precipitation percolating through the waste during the inactive (postclosure) period, (2) minimize the escape of noxious and/or potentially harmful gas into the surrounding atmosphere, and (3) prevent physical dispersion of the waste by wind and water. Covers typically are separated into two categories: traditional covers and alternative covers.

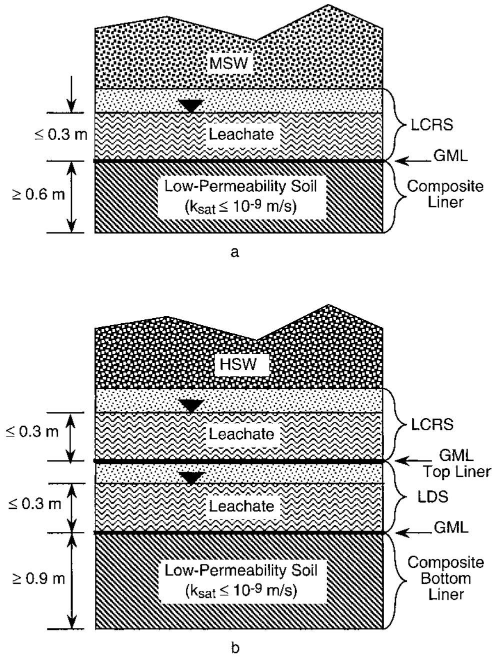


FIGURE 10 Minimum regulated (prescriptive) cross sections required in the United States for engineered liners used for disposal of (a) municipal solid waste (MSW) and (b) hazardous solid waste (HSW).

1. Traditional Covers

Traditional covers or caps often are referred to as "resistive barriers" because the design typically is based on minimizing infiltration of water by minimizing the saturated hydraulic conductivity of the barrier component of the cover. In the case of the use of compacted clay as the sole or primary material in the barrier, minimizing the saturated hydraulic conductivity follows the same approach as for a CCL.

a. Layering in traditional covers. As shown in Fig. 11, traditional covers include the following six possible components, from top to bottom: (1) surface layer, (2) protection layer, (3) drainage layer, (4) barrier layer, (5) gas collection layer, and (6) foundation layer. The primary purposes of the surface layer are to support vegetation, resist erosion, and reduce temperature and moisture extremes in underlying layers. The primary purposes of the protection layer are to store and release water through evaporation and transpiration (ET), separate underlying

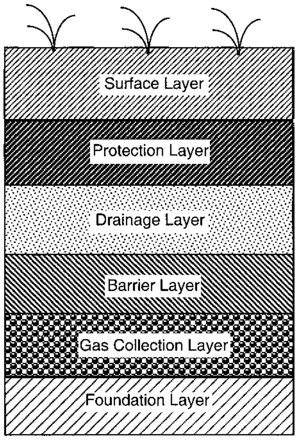


FIGURE 11 Cross section of layering in traditional caps (see Rumer and Mitchell, 1997).

TABLE VI Materials for Covers[a]

Layer	Materials
Surface	Topsoil (vegetated)
	Geosynthetic erosion control layer over topsoil (vegetated)
	Cobbles
	Paving material
Protection	Soil
	Cobbles
	Recycled or reused waste (e.g., fly ash, bottom ash, and paper mill sludge)
Drainage	Sand or gravel
	Geonet or geocomposite
	Recycled or reused waste (e.g., tire chips)
Barrier	Compacted clay
	Geomembrane
	Geosynthetic clay liner
	Recycled or reused waste (low permeability)
	Asphalt
	Sand or gravel capillary barrier
Gas collection/Removal	Sand or gravel
	Geonet or geocomposite
	Geotextile
	Recycled or reused waste (e.g., tire chips)
Foundation	Sand or gravel
	Geonet or geocomposite
	Recycled or reused waste select waste

[a] Modified from Rumer and Mitchell (1997).

waste from humans, burrowing animals, and plant roots (biointrusion), and protect underlying layers from "environmental" stress resulting from climatic conditions such as alternating wetting and drying cycles and/or freezing and thawing cycles. The drainage layer is used to allow drainage of the water from overlying layers to reduce the buildup of infiltrating water, thereby reducing the hydraulic gradient for flow through the barrier layer, and increase stability by reducing pore water pressures. The barrier layer functions to impede infiltration of water into the underlying soil and to restrict the outward migration of gases from the underlying waste or contamination. The gas collection layer serves to collect and remove gases to reduce the potential for detrimental effects (e.g., explosions) and to provide for energy recovery and/or vapor treatment. Finally, the foundation layer provides a surface on which to construct the other layers of the cap.

The potential materials for each layer are summarized in Table VI. Some layers, such as the hydraulic barrier layer, may consist of more than one material, such as a composite barrier comprising a geomembrane overlying a compacted clay.

All covers require a surface layer, but all of the six layers are not required at all sites. For example, a drainage layer may not be needed at an arid site. However, all of the layers that are used in a cover must have adequate shear strength to ensure stability against slope failure, and must be sufficiently durable to function adequately over the design life of the cap. More detail regarding the design and performance of covers can be found in Rumer and Mitchell (1997).

b. Design considerations for traditional covers. Two approaches can be used to design traditional covers: (1) the water balance approach and (2) the unsaturated flow approach. In the water balance approach, the amount of percolation through the cover P_r is determined by subtracting all of the components contributing to removal of water from the cover system from the precipitation as follows:

$$P_r = P - R_s - ET - L_f - \Delta S_w, \qquad (4)$$

where P is precipitation, R_s is surface runoff, ΔS_w is change in soil–water storage, ET is evaporation plus transpiration (evapotranspiration), and L_f is intralayer lateral flow. This water balance approach forms the basis for the HELP (hydrologic evaluation of landfill performance) model that commonly is used in the design of cover systems. A schematic cross section of a cover illustrating the water balance approach in the HELP model is shown in Fig. 12.

In the HELP model, each layer of the cover must be specified as a vertical percolation layer, barrier soil liner, lateral drainage layer, or geomembrane liner depending on the function and hydraulic properties of the layer. Unsaturated flow of water occurs only in a vertical percolation layer. In contrast, a barrier layer (soil liner), with low

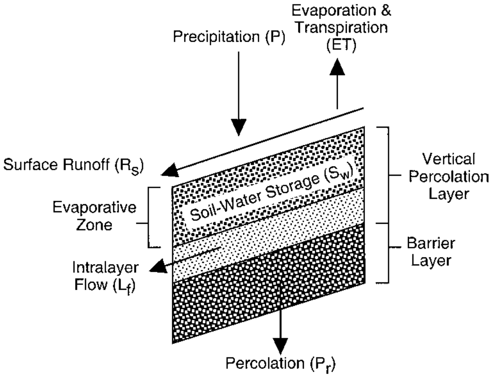


FIGURE 12 Schematic cover cross section used in the water balance approach of the HELP model. [Modified after Khire *et al.* (1997).]

saturated hydraulic conductivity, is assumed always to be saturated and percolation through the barrier layer can occur only when water ponds on the surface of the layer. As a result, unsaturated flow through the liner is not considered by the HELP model.

Precipitation in the HELP model is separated into surface runoff and infiltration based on an empirical approach using a modification of the Soil Conservation Service (SCS) runoff curve number method. Evaporation and transpiration removes water only from the evaporative zone of the cover in the HELP model. The procedure for determining evaporation and transpiration also is empirical. Further details on the procedures and assumptions used in the HELP model are provided by Khire *et al.* (1997).

The water balance of landfill covers also has been simulated using the unsaturated flow approach in the form of a one-dimensional, finite-difference computer program, UNSAT-H, that solves a modified version of Richards' equation that governs unsaturated flow through porous media (Khire *et al.* 1997). In this case, surface runoff is determined as the difference between precipitation and infiltration, where infiltration is based on the saturated and unsaturated hydraulic conductivities of the soils constituting the cover. Thus, unlike the empirical approach in the HELP model, the determination of surface runoff in UNSAT-H is based directly on the physical properties of the soil profile.

c. Traditional covers for solid waste disposal facilities. The minimum requirements for cover systems for MSW and HSW disposal prescribed by regulation in the United States are shown in Fig. 13. Both cross sections include a composite barrier layer consisting of a GML in intimate contact with a low-permeability soil, but the thickness required for this soil in the MSW cover is less than that in the HSW cover, and the maximum saturated

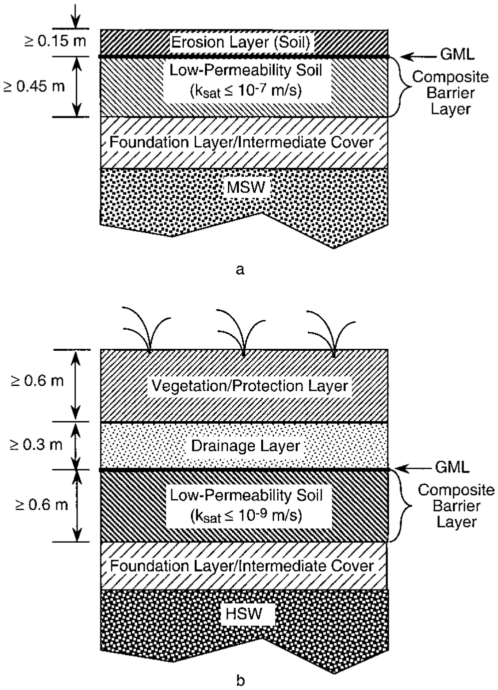


FIGURE 13 Minimum regulated (prescriptive) cross sections required in the United States for engineered covers used in disposal of (a) municipal solid waste (MSW) and (b) hazardous solid waste (HSW).

hydraulic conductivity required by the low-permeability soil in the MSW cover is 100 times higher than that in the HSW cover. Also, the MSW cover only requires a relatively thin erosion layer, whereas the HSW cover requires both a thick vegetation layer and an underlying drainage layer.

2. Alternative Covers

Interest in the use of alternative earthen final covers (AEFCs) has increased in recent years due in part to the relatively high costs associated with the more traditional covers and to other problems (e.g., desiccation cracking) associated with resistive covers containing low-permeability clay. Alternative earthen final covers are earthen covers designed on water storage principles that perform as well as more traditional, prescriptive covers but have greater durability and/or lower cost. The use of AEFCs typically is suitable only in drier regions where potential evaporation and transpiration exceeds precipitation, so that the AEFC can be designed to have sufficient storage capacity to retain water during wet periods without transmitting appreciable percolation. For example, AEFCs currently are receiving significant consideration for use in the western regions of the United States where semiarid and arid climates prevail. Although a variety of designs are being considered for AEFCs, most designs can be classified as either monolithic barriers or capillary barriers.

Monolithic covers (MCs) consist of a relatively thick vegetated layer of finer textured soil that has high water storage capacity such that water contents near the base of the cover remain fairly low. In addition, unlike resistive barrier layers (e.g., CCLs) that are compacted at high water saturations to minimize the saturated hydraulic conductivity, MCs are compacted at relatively low water saturations to increase the water-holding capacity and to reduce the (unsaturated) hydraulic conductivity. This combination of a relatively high water storage capacity together with the relatively low unsaturated hydraulic conductivity results in percolation rates from the base of the cover that can meet target percolation rates. Additional advantages of MCs relative to traditional, resistive covers include (a) less vulnerability to desiccation and cracking during and after installation, (b) greater simplicity of construction, and (c) lower maintenance. Also, MCs are economical to implement because MCs can be constructed of a reasonably broad range of soils and are typically constructed using soils from a nearby area.

A capillary barrier results when unsaturated flow occurs through a relatively fine layer overlying a relatively coarse layer (e.g., clay over sand, silt over gravel, and sand over gravel). In a capillary barrier, only a small fraction of the liquid flux associated with the wetting front through the finer layer is transmitted into the underlying coarser layer. A capillary barrier effect occurs because the unsaturated hydraulic conductivity of the coarser (underlying) layer decreases to a lower value than the unsaturated hydraulic conductivity of the finer (overlying) layer as the soil–water suction increases (see Fig. 3), and the residual suction in the finer layer after first passage of the wetting front is larger than the gravitational force on the wetting front. In general, the capillary barrier effect increases with an increase in the contrast in soil properties (e.g., unsaturated hydraulic conductivity) between the finer and coarser layers.

3. Gas-Phase Migration Through Covers

Control of gas-phase migration represents another important objective of engineered covers. In the case of municipal solid waste, failure to adequately control and vent the methane gas (CH_4) generated by decomposition of the waste represents a potential explosion hazard. Thus, the final cover serves both to prevent the escape of large quantities of methane gas and to assist in the collection of the gas for controlled venting through an engineered collection network.

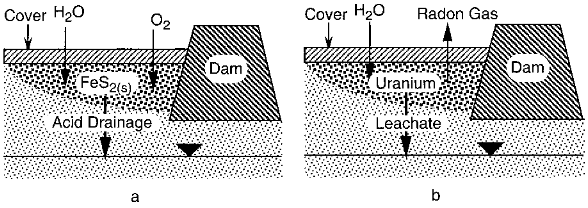


FIGURE 14 Tailings disposal scenarios requiring consideration of gas-phase migration through covers: (a) acid drainage, (b) radon gas emission.

As illustrated in Fig. 14, two problems associated with the environmentally safe disposal of mine tailings concern gas-phase migration through the final cover. First, the problem of acid drainage can occur when sulfidic tailings, such as pyrite, $FeS_{2(s)}$, are oxidized by the influx of oxygen into the tailings (Fig. 14a). The result of this oxidation is the production of a low-pH (acidic) solution (e.g., $pH < 6$) containing relatively high concentrations of potentially toxic heavy metals associated with the tailings, such as Fe^{2+}. Second, the disposal of uranium tailings results in a source of radon gas that can be environmentally harmful if not controlled properly (Fig. 14b). In both of these cases, the objective in the cover design must include steps taken to minimize the influx (O_2) or efflux (radon) of gas through the cover as well as minimize water infiltration.

The currently recommended design approach for soil covers in terms of the acid drainage scenario has been to compact a fine-grained soil at a water saturation sufficiently high to minimize both the saturated hydraulic conductivity and the extent of a continuous air phase available for oxygen diffusion. Minimization of continuous air phase is important in this case since oxygen is relatively insoluble in water and the oxygen diffusion coefficient in air is 10^5 times larger than that in water. An underlying coarse-grained layer also is usually included in the design to provide for lateral drainage of percolating water.

III. REMEDIATION

A. Remediation Procedure

The process of remediation of subsurface contamination commonly involves several steps, as outlined in Fig. 15. After contamination is discovered at a site, an emergency response and abatement phase typically is initiated to provide an initial assessment of the nature of the contamination, and rapid measures to abate or prevent further contamination are taken. After this initial assessment, a more detailed site characterization is undertaken to provide more accurate assessment of the problem (e.g., soil stratigraphy, concentrations of the contaminants of concern), followed by a detailed exposure assessment and a

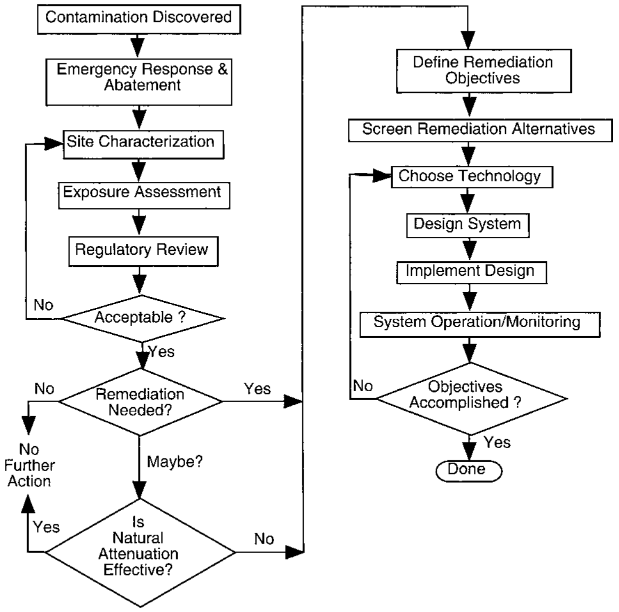


FIGURE 15 Flow chart of the remediation process. [From Shackelford and Jefferis (2000).]

review of the outcomes of these assessments by the governing regulatory authority, a process that is ongoing and interactive.

After the site and exposure assessment is approved, a decision regarding the necessity of remediation is made. This decision typically results in one of three outcomes: (1) there is no need for further action, (2) remediation must be implemented, or (3) the site should be monitored to determine whether or not natural attenuation processes (e.g., biodegradation) will remediate the site, with the potential for implementation of a more aggressive remediation technology some time in the future.

If remediation is required, the next phases of the remediation process involve defining the remediation objectives (e.g., concentration levels, containment versus treatment, etc.), screening possible remediation technologies that can achieve the objectives, choosing the most appropriate technology, and designing and installing the chosen remediation technology and system. Finally, monitoring of the performance of the system is required to ensure that the chosen technology is working to achieve the stated objectives. If the desired objectives cannot be achieved with the chosen technology, either the design of the chosen technology must be reevaluated and modified and/or an additional technology must be used. This process is repeated until the desired remediation objectives are achieved.

B. Remediation Technologies

Remediation technologies can be classified in numerous ways, such as according to the process involved in the remediation (i.e., physical, chemical, biological, or thermal), the objective of the technology (containment versus treatment), or the location of the remediation process (*in situ* versus *ex situ*).

1. Remediation Processes

Remediation processes commonly are categorized as physical, chemical, biological, or thermal processes. Physical processes include interphase mass transfer reactions, such as sorption and ion exchange, as well as binding processes, such as cement fixation of chemicals to soil.

Chemical processes involve chemical reactions. Two chemical processes of relatively recent interest in remediation are the oxidative degradation of halogenated hydrocarbons (e.g., PCE, TCE) in the presence of oxidants and reductive dehalogenation of halogenated hydrocarbons in the presence of solid, zero-valent iron metal, such as the iron filings found in metal machine shops. For example, oxidation of trichloroethylene, or TCE (C_2HCl_3), by potassium permanganate ($KMnO_4$) is described by the following chemical reaction:

$$2KMnO_4 + C_2HCl_3 \rightarrow 2CO_2 + 2MnO_2 + 2KCl + HCl, \tag{5}$$

whereas dechlorination of TCE to dichloroethylene, or DCE ($C_2H_2Cl_2$), in the presence of solid, zero-valent iron metal ($Fe^0_{(s)}$) is described by the following chemical reaction:

$$2Fe^0_{(s)} + C_2HCl_3 + 3H_2O \rightarrow C_2H_2Cl_2 + H_2 + Cl^- + 3OH^- + 2Fe^{2+}. \tag{6}$$

The potential products of these reactions include other chlorinated hydrocarbons (e.g., DCE) that subsequently degrade to vinyl chloride (C_2H_3Cl, or VC), ethene (C_2H_4), or ethane (C_2H_6), hydrogen gas (H_2), chloride (Cl^-), and iron oxide (FeO) and hydroxide [$Fe(OH)_2$] precipitates.

In biologically mediated processes, the contaminant, typically a hydrocarbon compound, acts as a carbon source and a source of energy for the bacteria. For example, consider the following reactions for bacteria-mediated degradation (i.e., biodegradation) of benzene (C_6H_6):

$$C_6H_6 + 2.5O_2 + NH_3 \xrightarrow{Pseudonomas} C_5H_7O_2N + CO_2 + H_2O \tag{7}$$

and

$$C_5H_7O_2N + 5O_2 \xrightarrow{Pseudonomas} 5CO_2 + 2H_2O + NH_3. \tag{8}$$

The first (top) reaction represents oxidation of benzene (C_6H_6) to cell mass and the second (bottom) reaction represents complete mineralization to carbon dioxide (CO_2), water (H_2O), and ammonia (NH_3). These reactions also illustrate the need for large quantities of oxygen (O_2) for aerobic oxidation of hydrocarbons. For example, in this case, 3.1 kg of oxygen is required per kilogram of contaminant.

Thermal processes involve heating the contaminated soil to induce volatilization and removal of contaminants (e.g., radiofrequency heating) or to melt and solidify the contaminated soil mass (e.g., vitrification).

2. *In Situ* versus *ex Situ* Remediation

In situ versus *ex situ* remediation refers to the location of the remediation technology. In this context, an important distinction arises relative to the location of the application of the technology versus the location of the treatment. For example, in the pump-and-treat approach to remediation, the pumping occurs *in situ*, but the treatment of the pumped, contaminated water occurs *ex situ*.

3. Containment versus Treatment

Containment refers to technologies used to prevent the spread of contamination without necessarily resulting in degradation or transformation of the contaminants. Treatment refers to technologies used to degrade or otherwise transform contaminants into less toxic or nontoxic concentrations. Although *in situ* containment technologies have been available for decades, interest in the use of *in situ* containment as the primary technology for remediation applications has gained momentum in recent years because containment (1) typically is cheaper than treatment, (2) can be used until a more efficient treatment technology is developed, (3) can provide a means for evaluating the potential for natural attenuation processes to degrade the contaminants, and (4) can present a lower overall risk since major excavation, contaminant exposure, etc., can be avoided.

Many remediation scenarios will involve both containment and treatment. However, in some countries where economic resources are not plentiful, using a remediation technology that requires *ex situ* treatment may not be a viable option, particularly if treatment of wastewater, for example, is not already practiced routinely.

4. Active versus Passive Considerations

In the case of cleanup, or remediation, of existing subsurface pollution, both active and passive remediation activities are possible. Active remediation involves application of a conventional technology, such as a pump-and-treat system. However, active remediation is relatively expensive due to the high costs associated with both construction (e.g., installation of wells) and operation (e.g., pumping and treating). Thus, more passive technologies, such as low-permeability ($k_{sat} < 10^{-9}$ m/sec) vertical cutoff walls, also known as slurry walls, frequently are installed to prevent further spreading of subsurface pollution, thereby reducing the cleanup duration and cost.

In response to the increasing costs associated with active remediation of contaminated sites, an increasing emphasis is being placed on more cost-effective, passive treatment technologies. For example, the use of permeable reactive walls (PRWs) for *in situ* treatment of contaminated ground water has gained momentum in recent years. A PRW (also known as a passive treatment system or *in situ* treatment zone or curtain) incorporates chemical and/or biological reagents or catalysts into the porous medium used for the wall to degrade or otherwise reduce the concentrations of pollutants during passage of the contaminant. A major difference between the more traditional low-permeability containment barriers such as slurry walls, and PRWs is that the hydraulic conductivity of a PRW must be sufficiently large to allow timely and efficient processing of the contaminated ground water.

5. Classification of *in Situ* Remediation Technologies

Based on the aforementioned considerations, a general classification system of possible remediation technologies is given in Table VII. With respect to the *in situ* technologies, brief descriptions of some of the more commonly used technologies follow. More detailed descriptions can be found in Shackelford and Jefferis (2000).

6. Descriptions of Remediation Technologies

a. Vertical containment technologies. Pump and treat refers to the process of pumping contaminated ground water to a surface collection system through wells screened in the saturated zone and then treating the contaminated water with one or more *ex situ* treatment technologies. In this context, pump and treat may be considered to be a treatment rather than containment option. However, in terms of the *in situ* aspects of pump and treat, which is pumping, not treatment, the pump-and-treat technology typically is categorized as a containment option.

For example, pumping wells may be located within the contaminated area to provide drawdown of the groundwater table that directs flow into the site, thus providing active containment by minimizing or preventing further migration of the plume away from source or site. Alternatively, pumping wells can be located downgradient from the contamination to provide a drawdown barrier.

TABLE VII Classification of Remediation Technologies Based on Soil as the Contaminated Medium[a]

Soil removal?	Technology category	Technique/process	Example(s)	Comment(s)
Yes (*ex situ*)	Containment	Disposal	Landfills	On-site vs. off-site; new vs. existing
	Treatment	Chemical	Neutralization, solvent extraction	Treated soil may require disposal in a landfill or may be returned to the site
		Physical	Soil washing, stabilization/solidification, vitrification	
		Biological	Biopiles, bioreactors	
		Thermal	Incineration, vitrification	
No (*in situ*)	Containment	Pump and treat	Vertical wells, horizontal wells	Both passive and active containment are possible; in pump and treat, pumping is used to control hydraulic gradient and collect contaminated water; treatment is *ex situ*
		Capping	Traditional covers, alternative covers, geochemical covers	
		Vertical (low-permeable) barriers	Slurry walls, grout curtains, sheet piling, biobarriers, reactive barriers	
		Horizontal (low-permeable) barriers	Grout injected liners	
	Treatment	Chemical	Oxidation, chemical reduction	Technologies with asterisk require removal of gas and/or liquid phases and *ex situ* treatment; both passive and active treatment are possible
		Physical	Soil flushing, stabilization/solidification, vitrification, air sparging (AS), soil vapor extraction (SVE), electrokinetics (EK)	
		Biological	Monitored natural attenuation (MNA), bioventing, bioslurping, biosparging	
		Thermal	Steam injection, radio frequency heating (RF), vitrification	

[a] From Shackelford and Jefferis (2000).

Slurry walls, also known as slurry cutoff walls or vertical cutoff walls (see Fig. 7c), are vertical walls constructed by excavating a trench with a back-hoe or clamshell and simultaneously filling the trench with a stabilizing slurry which is typically prepared from a mixture of bentonite clay and water, or bentonite clay, cement, and water. The slurry forms a thin filter cake (typically <3 mm for bentonite slurry and perhaps a few tens of millimeters for cement–bentonite slurry) with a low hydraulic conductivity (typically $<10^{-10}$ m/sec) on the sides of the trench. The filter cake minimizes slurry loss from the trench, stabilizes native soil on the sidewalls of the trench, and provides a plane for slurry stabilization of the excavated trench. Both bentonite and cement–bentonite slurries typically contain 4–7% (w/w) sodium bentonite mixed with water. The hydraulic conductivity of slurry walls is affected significantly by the quality of the backfill material and typically ranges from $\sim 10^{-7}$ to $\sim 10^{-11}$ m/sec, with lower values associated with backfill materials that contain sufficient quantities of clay.

Other passive vertical barriers include walls constructed using deep soil mixing or jet grouting using chemical grouts (e.g., silicates, resins, and polymers), grout curtains, and sheet pile walls. However, the application of these technologies for long-term containment has been limited due to concern with the integrity of the containment system and potential leakage of contaminants through “windows” in the barriers, such as high-permeability zones between the grout in grout curtains, or through the interlocks in sheet piles, or corrosion with respect to the use of sheet-pile walls. Nonetheless, these technologies can be used in conjunction with other remediation technologies to aid in temporary, partial containment.

b. Horizontal containment technologies. Caps or covers are used as horizontal surface barriers to minimize the leaching of contaminants into the ground water. Two possible scenarios involving applications of caps for *in situ* remediation applications are illustrated in Fig. 16.

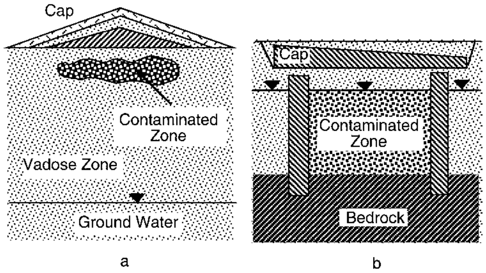


FIGURE 16 Schematic example applications for caps for *in situ* remediation: (a) long-term containment of vadose zone contamination in arid climates, (b) capping of slurry wall containment system for saturated zone contamination. [From Shackelford and Jefferis (2000).]

As shown in Fig. 16a, a cap may be considered as the only technology needed in cases where the climate is arid, the water table is deep, and the site is relatively isolated.

The schematic diagram shown in Fig. 16b depicts the conceptual use of an *in situ* cap designed and placed to prevent the potential for overtopping of the contaminated ground water within a vertically contained zone due to infiltration, an occurrence known as the "bathtub effect." The cap is constructed by excavating the soil above the containment area, placing or compacting the low-permeability barrier, and backfilling with the native soil to provide a protective surface layer. The backfill is shown to original grade to allow for subsequent use of the site. Special design features include sloping the surface of the barrier layer to promote subsurface runoff down gradient of the containment zone. Also, periodic separations between the horizontal (cap) barrier layer and the vertical containment barrier may be required to allow for venting of volatile organics and/or minimizing the potential buildup of gas pressure beneath the cap and eventual uplift of the cap.

Subsurface horizontal barriers include indigenous barriers (aquicludes, aquitards) and artificially emplaced barriers. Artificially emplaced barriers include thin diaphragm walls created by directional and nondirectional drilling with jet grouting, conventional jet grouting, and soil freezing technologies.

Indigenous barriers rarely are relied upon for permanent containment because of the uncertainty associated with the integrity of the system. Application of the artificially emplaced barriers also is limited for the same reasons as the alternative passive vertical barriers, namely uncertainty regarding integrity and longevity of the placed system, and costs associated with placement of the technology.

c. Immobilization technologies. Stabilization and solidification (S/S) methods are intended to immobilize dissolved contaminants and, in certain cases, DNAPLs. Stabilization refers to techniques that reduce contaminant hazard potential by converting the contaminants to less soluble, mobile, or toxic forms. Solidification refers to techniques that encapsulate the contaminant in a monolithic solid of high structural integrity. Common S/S reagents include cement and other pozzolanic materials, such as lime, fly ash, and slag. These reagents typically result in stabilization (fixation) of the contaminants.

Vitrification is a thermal process that results in vaporization and mobilization of organic compounds and solidification and immobilization of inorganic compounds. However, the primary application of the technology historically has been in immobilizing inorganic, radioactive contaminants. The *in situ* vitrification process uses a subsurface electrode array to heat the soil to temperatures as high as 1800°C through the application of large currents, resulting in melting of the soil. The melting process initiates at the ground surface and propagates downward, resulting in vaporization of organic compounds. The vapors migrate to the ground surface and are collected using a vacuum hood. Upon cooling, the remaining contaminants are solidified within a rocklike mass resembling obsidian.

d. Enhanced removal technologies. Soil vapor extraction, also referred to as *in situ* soil venting, couples vapor extraction (recovery wells) with blowers or vacuum pumps to remove vapors from the vadose zone and therefore reduce the levels of residual soil contaminants.

Air sparging, also known as *in situ* air stripping or *in situ* volatilization, is effective in removing SVOCs and VOCs. Air sparging involves injecting air through wells into the contaminated aquifer below the groundwater table. The air sparging process cleans the contaminated groundwater zone through two processes. First, injection of the air creates a turbulent condition that enhances volatilization of the contaminants. Second, the injected, or sparged, air increases the oxygen content of the ground water, enhancing aerobic biodegradation of the contaminants.

Soil flushing refers to the enhanced mobilization of contaminants in contaminated soil for recovery and treatment by forcing a fluid through the contaminated soil to either displace the contaminants or induce physical, chemical, biological, and/or thermal conditions that favor removal. Although the terms *soil flushing* and *soil washing* often are used interchangeably to describe the same technology, the term *soil flushing* generally is used in connection with *in situ* remediation, whereas the term *soil washing* is reserved for *ex situ* remediation. Flushing fluids may be water, enhanced water, and gaseous mixtures. Typical soil flushing solutions include water, dilute acids and bases, complexing and chelating agents, reducing agents, solvents, and surfactants. The fluids are introduced through spraying, surface flooding, subsurface leach fields, and subsurface injection.

In some cases, steam has been used as the fluid in soil flushing applications. The steam is injected into the subsurface through a ring of injection wells configured to surround all or part of the subsurface plume to be treated. Liquid and vapor are extracted from one or more wells located near the center of the well pattern while steam is injected into the permeable layers through screened portions of wells constructed around the plume. The steam directly heats contaminated permeable layers in the process zone to lower the viscosity and increase volatility of the contaminants. The contaminants are swept toward the extraction wells as ground water is displaced in advance of the injected steam front.

Radio frequency heating involves the use of electromagnetic energy to vaporize hydrocarbons in soil. The method involves using a system of electrodes implanted in vertical and/or horizontal holes drilled into the soil. Electromagnetic energy in the radio frequency band is generated using a modified radio transmitter and directed from the electrodes into the ground to heat the soil to a predetermined temperature to volatilize trapped hydrocarbons.

Electrokinetics involves the use of an applied electrical gradient to speed the removal of contaminants from contaminated, low-permeability, fine-grained soils. In general, application of the electrical current results in generation of protons (H^+) at the anode, due to hydrolysis reactions, that migrate together with the metal cations to the negatively charged cathode for removal and processing. In addition, the migrating metal ions (cations) result in bulk water flow from the anode to the cathode, a process referred to as electroosmosis.

e. Bioremediation. Bioremediation may be defined as the process whereby a biological agent (e.g., bacteria, fungi, plants, enzymes) is used to reduce contaminant mass and toxicity in soil, ground water, and air. The goal of bioremediation is to reduce toxicity and target contaminant concentrations in water and/or soil through either biodegradation (of organics) or biotransformation (of metals).

Biotransformation is a general term representing any biologically catalyzed conversion of a metal or organic chemical. Biodegradation refers to a single biological reaction or sequence of reactions that result in the conversion of an organic substrate to a simpler molecule. Mineralization, or complete biodegradation, refers to conversion of the target compound to carbon dioxide and water under aerobic conditions, or methane (or ethane or ethene) under anaerobic conditions. Biotransformation and biodegradation of a contaminant do not necessarily result in a decrease in toxicity.

Biowalls, also referred to as permeable reactive biowalls, or in-line microbial filters, are essentially *in situ* bioreactors or permeable reactive walls that treat contaminants passively through either biostimulation (stimulation of indigenous bacteria or microbes) or bioaugmentation (addition of nonindigenous bacteria or microbes). Much of the effort has been directed toward the use of oxygen and/or nutrient sources as reagents in biowalls to stimulate the activity of indigenous microorganisms.

f. Monitored natural attenuation. Monitored natural attenuation is a passive approach to *in situ* remediation in that the approach relies on natural processes to attenuate the contaminants in the soil and ground water. Monitored natural attenuation can be used as a remediation approach when natural processes will reduce the mass, toxicity, mobility, and/or volume of the contaminants. Natural processes resulting in a decrease in or spreading of contaminants of concern (COCs) associated with the unsaturated, or vadose, zone include volatilization, sorption, leaching to ground water, and natural bioventing. Natural processes resulting in a decrease in or spreading of COCs associated with the saturated, or groundwater, zone include sorption, diffusion/dispersion, biodegradation, and dilution. Thus, MNA requires (1) a very thorough understanding of the fate and transport of the COCs and (2) close monitoring of soil and groundwater samples to verify continuous removal of the COCs.

SEE ALSO THE FOLLOWING ARTICLES

ENVIRONMENTAL TOXICOLOGY • HAZARDOUS WASTE INCINERATION • POLLUTION, ENVIRONMENTAL • POLLUTION CONTROL • ROCK MECHANICS • SOIL AND GROUNDWATER POLLUTION • SOIL MECHANICS • WASTE-TO-ENERGY SYSTEMS • WASTEWATER TREATMENT AND WATER RECLAMATION • WATER POLLUTION

BIBLIOGRAPHY

Acar, Y. B., and Daniel, D. E. (1995). “Geoenvironment 2000: Characterization, Containment, Remediation, and Performance in Environmental Geotechnics,” ASCE, Reston, VA.

Charbeneau, R. J. (2000). “Groundwater Hydraulics and Pollutant Transport,” Prentice-Hall, Inc. Upper Saddle River, NJ.

Charbeneau, R. J., and Daniel, D. E. (1993). “Contaminant transport in unsaturated flow,” *In* “Handbook of Hydrology” (D. R. Maidment, ed.), pp. 15-1–15-54, McGraw-Hill, New York.

Daniel, D. E. (1993). “Geotechnical Practice for Waste Disposal,” Chapman & Hall, London.

Daniel, D. E., and Koerner, R. M. (1995). “Waste Containment Facilities, Guidance for Construction Quality Assurance and Quality Control of Liner and Cover,” ASCE, Reston, VA.

Grubb, D. G., and Sitar, N. (1994). “Evaluation of Technologies for *In-situ* Cleanup of DNAPL Contaminated Sites,” EPA/600/R-94/120, U.S. EPA, Washington, DC.

Khire, M. V., Benson, C. H., and Bosscher (1997). "Water balance modeling of earthen final covers," *J. Geotechnical Geoenviron. Eng.* **123**(8), 744–754.

Koerner, R. M. (1998). "Designing with Geosynthetics," 4th ed., Prentice-Hall, Upper Saddle River, NJ.

Koerner, R. M., and Daniel, D. E. (1997). "Final Covers for Solid Waste Landfills and Abandoned Dumps," ASCE, Reston, VA.

Manassero, M., Benson, C. H., and Bouazza, A. (2000). "Solid waste containment systems," *In* "International Conference on Geotechnical and Geoenvironmental Engineering (GeoEng2000)," Vol. 1, pp. 520–642, Technomic, Lancaster, PA.

McBean, E. A., Rovers, F. A., and Farquhar, G. J. (1995). "Solid Waste Landfill Engineering and Design," Prentice-Hall, Upper Saddle River, NJ.

Mitchell, J. K. (1993). "Fundamentals of Soil Behavior," 2nd ed., Wiley, New York.

National Research Council (1990). "Ground WaterModels: Scientific and Regulatory Applications," National Academy Press, Washington, DC.

Oweis, I. S., and Khera, R. J. (1998). "Geotechnology of Waste Management," 2nd ed., PWS, Boston.

Pankow, J. F., and Cherry, J. A. (1996). "Dense Chlorinated Solvents and other DNAPLS in Groundwater," Waterloo Press, Portland, OR.

Rowe, R. K., Quigley, R. M., and Booker, J. R. (1995). "Clayey Barrier Systems for Waste Disposal Facilities," E & FN Spon, London.

Rumer, R. R., and Mitchell, J. K. (1997). "Assessment of Barrier Containment Technologies," National Technical Information Service (NTIS), Springfield, VA.

Shackelford, C. D. (1997). "Modeling and analysis in environmental geotechnics: An overview of practical applications," *In* "2nd International Congress on Environmental Geotechnics, IS-Osaka '96," Vol. 3, pp. 1375–1404, Balkema, Rotterdam.

Shackelford, C. D., and Rowe, R. K. (1998). "Contaminant transport modeling," *In* "3rd International Congress on Environmental Geotechnics," Vol. 3, pp. 939–956, Balkema, Rotterdam.

Shackelford, C. D., and Jefferis, S. A. (2000). "Geoenvironmental engineering for *in situ* remediation," *In* "International Conference on Geotechnical and Geoenvironmental Engineering (GeoEng2000)," Vol. 1, pp. 121–185, Technomic, Lancaster, PA.

Sharma, H. D., and Lewis, S. P. (1994). "Waste Containment Systems, Waste Stabilization, and Landfills: Design and Evaluation," Wiley, New York.

Hazardous Waste Incineration

Joseph J. Santoleri
RMT/Four Nines

GLOSSARY

Destruction and removal efficiency (DRE) Measure of the degree to which a given principal organic constituent is destroyed in an incineration system. Given as a percentage of the mass of the constituent fed.

Excess air Incremental quantity of air above the stoichiometric quantity required for complete combustion of a waste. Given as a percentage of the stoichiometric quantity.

Hazardous waste A solid waste or combination of solid wastes which, because of its quantity, concentration, or physical, chemical, or infectious characteristics, may (a) cause or significantly contribute to an increase in mortality or an increase in serious irreversible or incapacitated reversible illness, or (b) pose a substantial present or potential hazard to human health or the environment when improperly treated, stored, transported, disposed of, or otherwise managed.

Incineration system System used for the thermal destruction of hazardous wastes. Usually includes a high-temperature furnace, quench, and air pollution control device.

Principal organic hazardous constituent (POHC) Any of the toxic substances listed in Appendix VIII of the Code of Federal Regulations, Chapter 40, Section 261.

Pyrolysis Destructive distillation of a solid by heating in the absence of stoichiometric oxygen.

Residence time Calculated maximum retention period for waste species in an incinerator furnace, assuming all gases move with plug flow and are at uniform temperature and pressure.

Resource conservation and recovery act (RCRA) Set of legally enforceable requirements on generators of hazardous wastes and on owners and operators of facilities for treatment, storage, and disposal of such wastes.

Trial burn Test of an incineration facility while burning hazardous wastes at the operating conditions specified in the permit application. Used to demonstrate that the incinerator meets the performance standards given by RCRA.

INCINERATION can be adapted to the destruction of a wide variety of hazardous wastes, and, unlike many other methods, is often a permanent solution to the disposal problem. Incineration is not a new technology and has been used for treating organic hazardous waste for many years. The major benefit of incineration is that the process actually destroys most of the waste rather than just disposing of or storing it (Oppelt, 1987).

In spite of its acceptance by the Environmental Protection Agency (EPA) as the most reliable process for destruction of hazardous waste, incineration is no longer a popular method for hazardous waste disposal even though it has a proven track record. Incineration of chemicals now considered hazardous is a technology that is less than 40 years old. After World War II, the chemical industries in the United States began to grow rapidly and with that growth came a commensurate growth in chemical wastes to be disposed of. Solid materials, whether municipal wastes or by-products of process operations, were often burned in open dumps. In the early 1950s, the widespread installation of municipal incinerators for the treating of these solid wastes occurred, but most liquid process wastes were released directly to sewers or nearby streams and rivers.

The legal aspects of the hazardous waste problem (i.e., the standards and regulations imposed on industry by federal, state, and local governments), which have a significant impact on the methods industry uses to handle particular wastes, are covered in Section I.

The most developed and commonly used incinerators are liquid injection incinerators and rotary kiln incinerators. Although liquid injection incinerators have been the most popular in Europe and the United States, rotary kiln incinerators are better suited to handle all physical forms of hazardous wastes. High-temperature incinerators are capable of destroying highly toxic organic wastes (e.g., organochlorine, herbicide orange, and polychlorinated biphenyls, PCBs) with high efficiency. Fluidized-bed and other technologies such as cement kilns and lightweight aggregate kilns have proven they are economically, technically, and environmentally acceptable. The various types of incinerators are further discussed in Section II. The quench systems and air pollution controls necessary to meet today's emission standards are described in Section III.

I. LEGAL REQUIREMENTS

A. Resource Conservation and Recovery Act (RCRA)

It was not until the mid-1970s that Congress enacted the Resource Conservation and Recovery Act (RCRA), the first major hazardous waste regulatory document to address waste generation and disposal effectively[1]. With the enactment of RCRA in 1976, EPA brought about significant progress in source testing. Under RCRA, EPA promulgated an interim final regulation in July 1982. The regulation set performance standards for three process parameters (RCRA, 1981/1991):

1. *Standards to control particulate matter* mandate that the incinerator may not emit particulate matter in excess of 180 mg/dry standard cubic meter (dscm) or 0.08 grain/dry standard cubic foot after correction to 7% O_2 in the stack gas. The correction factor (CF) is defined as follows:

$$\mathrm{CF} = (21 - 7)/(21 - Y) = 14/(21 - Y),$$

where Y is the measured O_2 concentration in the stack gas on a dry basis (%).

When systems operate with excess air that results in stack O_2 above 7% d.v., the CF is less than 1. If the system operates with stack O_2 less than 7% d.v., the CF is greater than 1. This impacts systems that dilute the combustion gases with excess air. This CF is used when air (21% O_2 by volume) is used as the combustion air gas.

2. *Standards to control hydrogen chloride* require the control of hydrogen chloride emissions such that the rate of emission is no greater than the larger of either 1.8 kg/hr (4 lb/hr) or 1% of the hydrogen chloride in the stack gas prior to entering any pollution control equipment.

3. *Standards to control organic emissions* require a destruction and removal efficiency (DRE) of selected principal organic hazardous constituents (POHCs) of 99.99% (four-nines DRE). The DRE for an incinerator/air pollution control system is defined by the following formula:

$$\mathrm{DRE}\ (\%) = (W_{\mathrm{in}} - W_{\mathrm{out}})/W_{\mathrm{in}} \times 100,$$

where DRE is the destruction and removal efficiency (in percent), W_{in} is the mass feed rate of POHC fed to the incinerator (pounds per hour), and W_{out} is the mass emission rate of POHC to the atmosphere (pounds per hour).

Permitting of Hazardous Waste Incinerators

The permitting of hazardous waste incinerators is a comple, multifaceted program conducted simultaneously on the federal, state, and local levels. Hazardous waste incinerators require the following permits: federal RCRA, state RCRA, Toxic Substance Control Act (for PCBs), federal air quality (Title V), state air quality, state waste water discharge, and federal waste water discharge permits.

A *trial burn plan* must be included as part of the engineering description section of the RCRA Part B permit application. The trial burn plan describes the goals and procedures of the trial burn. The plan includes discussions on the waste characteristics and the system equipment, including controls and monitoring devices, descriptions of the conditions which will be demonstrated during the test, and descriptions of the sampling and analysis methods which will be used to collect the data.

The ASME Research Committee on Industrial & Municipal Wastes (ASME, 1984), while strongly supporting the reduction and/or recycle of hazardous wastes, also realistically recognizes that some wastes are inevitably generated. Therefore, proper treatment and disposal are required. Incineration is a full-scale, proven, effective technology for treatment of a broad range of organic wastes.

B. Toxic Substance Control Act (TSCA)

In 1976, the Toxic Substances and Control Act prohibited the further manufacture of polychlorinated biphenyls (PCBs) after July 2, 1979, established limits on PCB use in commerce, and established regulations for proper disposal of PCBs. In the disposal of wastes containing PCBs, both RCRA and TSCA standards apply. The TSCA standards for incineration of PCB wastes are generally more restrictive than the RCRA standards. The TSCA standards may be summarized as follows:

1. The incinerator furnace must operate with at least a 99.9% combustion efficiency (CE),

   $$\mathrm{CE} = (\%\mathrm{CO_2})/(\%\mathrm{CO_2} + \%\mathrm{CO}) \times 100,$$

   where $\%\mathrm{CO_2}$ is the stack carbon dioxide percent by volume, dry basis, and %CO is the stack carbon monoxide percent by volume, dry basis.
2. Scrubbers are required to remove hydrochloric acid from the furnace exit gases; scrubber efficiency is not specified in the regulations.
3. Furnace temperature, stack percent oxygen, stack percent carbon monoxide, and the PCB feed rate must be measured and recorded on a continuous basis.
4. The PCB feed to the incinerator must be automatically cut off on failure of the stack gas monitors or failure of the PCB measuring and recording equipment.
5. The standards for liquid PCB incinerators must meet one of the following conditions:
   a. Temperature, 1200°C; residence time, 2.0 sec; stack O_2, dry, 3.0%.
   b. Temperature, 1600°C, residence time, 1.5 sec, stack O_2, dry, 2.0%.

   The furnace for either operating condition is about the same. However, fuel consumption is considerably higher at 1600°C.
6. Incinerators burning solid PCB wastes must have a 99.9999% destruction efficiency (six-nines). For each 1 million g of PCBs fed, only 1 g is allowed to leave the stack. TSCA standards are much more restrictive than RCRA, which allows 1 g to be released for each 10,000 g of waste fed (99.99%, or four-nines destruction efficiency).

C. Comprehensive Environmental Response, Compensation and Liability Act (CERCLA)

In 1980, the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) established a national fund (Superfund) to assist in remedial actions of cleanup of the uncontrolled waste sites created by the poor disposal practices of the past.

D. Hazardous and Solid Waste Act (HSWA)

In 1984, the Hazardous and Solid Waste Act (HSWA) amended the RCRA and established a strict timeline for restricting untreated hazardous waste from land disposal. HSWA also further articulated the national policy to emphasize pollution prevention as the first and primary goal for the waste management program. Another factor that distinguishes the HSWA is the ambitious timetable it established; many of the HSWA provisions were already in place in 1984 and others went into effect shortly thereafter. The HSWA also established what are called *hammer* provisions—statutory requirements ultimately banning the landfilling of hazardous wastes that go into effect automatically if EPA fails to issue regulations by certain dates.

E. Superfund Amendments and Reauthorization Act (SARA)

The Superfund Amendments and Reauthorization Act (SARA) of 1986 not only reauthorized the Superfund Program, but greatly expanded the provision and funding of the initial act.

F. Clean Air Act Amendments (CAAA)

The Clean Air Act Amendments (CAAA) of 1990, together with numerous state and county regulations, amended the CAA of 1970 and regulated emissions from various incineration sources.

G. Boiler and Industrial Furnace Regulation (BIF)

In the mid-1980s, issues regarding toxic metal and organic emissions from combustion of hazardous wastes in boilers and other process furnaces as well as incinerators became of concern to the EPA. Boilers and industrial furnaces burning hazardous waste were initially exempt from the RCRA incinerator requirements. Under RCRA, these units are considered to be “resource recovery” units, which do not fall under the definition of incinerators. In 1991, the final rules governing the operation of boilers and industrial furnaces (BIF) burning hazardous waste were promulgated (Code of Federal Regulations, Chapter 40, Part 266). These rules cover the burning of hazardous wastes in boilers, industrial furnaces, and kilns for cement, lime, and aggregate. The rule requires that these facilities obtain RCRA Part B permits to burn hazardous waste in resource recovery units (CFR, 1991).

H. Hazardous Waste Combustor Maximum Achievable Control Technology Standards for Hazardous Waste Incinerators, Cement Kilns, and Lightweight Aggregate Kilns

The emissions of products of incomplete combustion (PICs) and especially the polychlorinated dibenzodioxins (PCDD) and polychlorinated dibenzofurans (PCDF) from all hazardous waste combustors (HWCs) caused EPA to issue a moratorium in 1993 on all new permits for incinerators. The Maximum Achievable Control Technology (MACT) rules were initially drafted and released on April 19, 1996 for HWCs—hazardous waste incinerators, cement kilns (CKs), and lightweight aggregate kilns (LWAKs) under joint authority of the Clean Air Act Amendments of 1990 (CAAA) and the Resource Conservation and Recovery Act (RCRA) (61 FR 17358). The proposed MACT standards limit emissions from both new and existing facilities in each equipment category. Pollutants regulated under the rule are dioxins and furans (D/F), mercury (Hg), total chlorine (HCl/Cl_2), semivolatile metals (SVM) including lead and cadmium, low-volatility metals (LVM) including antimony, arsenic, beryllium, and chromium, particulate matter (PM), carbon monoxide (CO), and hydrocarbons (HC). Boilers were not included at that time. The final MACT standards for hazardous waste combustors can be found in the *Federal Register*, Volume 64, No. 189, Thursday, September 30, 1999.

The standards also require continuous emissions monitoring systems (CEMS) and performance testing.

1. Dioxin/Furan Emission Levels

Dioxin is a term used to describe a large group of chemical compounds having a similar basic structure. The most common subgroup of dioxin compounds (of which there are 75 different varieties) comprises those which include chlorine atoms. A few are toxic; the most toxic is 2,3,7,8-tetrachlorodibenzo(para) dioxin (2,3,7,8-TCDD). There are also 135 polychlorinated dibenzofuran compounds, known as "furans."

Note that EPA believes that it is appropriate to control dioxin/furan emissions for HWCs based on toxicity equivalents (TEQs). Scientists developed procedures for quantitative evaluation based on available toxicological data. This formed the basis for expressing the toxic potential of each congener by a factor ("toxicity equivalency factor" or TEF). The total toxic equivalent (TEQ) of a given environmental sample is calculated by multiplying the amount of each congener present in the sample by its specific TEF and adding up the total. This approach assigns TEQs to the various dioxin and furan congeners to account for their TEQ relative to 2,3,7,8-TCDD.

2. MACT Standards for Hazardous Wast Incinerators

The MACT standards include emission limits for both new and existing hazardous waste incinerators. The standards require a minimum *destruction and removal efficiency* (DRE) of 99.99% for each principal organic hazardous constituent (POHC) and a minimum DRE of 99.9999% for dioxin-listed wastes (Table I).

3. MACT Standards for Cement Kilns

Cement kilns have the same DRE requirements as incinerators under the MACT standards (Table II). Other emission standards are similar to those for incinerators, with some exceptions due to differences in equipment and operations.

4. MACT Standards for Lightweight Aggregate Kilns

LWAKs have the same DRE requirements as incinerators under the MACT standards. Other emission standards are similar to those for incinerators with some exceptions due to differences in equipment and operations (Table III).

TABLE I Emissions Limits from Incinerators[a]

Pollutant	Units	Existing	New
D/F	ng TEQ/dscm	0.2 or 0.4[b]	0.2 or 0.4[b]
Mercury	μg/dscm	130	45
SVM	μg/dscm	240	24
LVM	μg/dscm	97	97
HCl/Cl_2	ppm_v (dry)	77	21
PM	mg/dscm (g/dscf)	34 (0.015)	34 (0.015)
CO	ppm_v (dry)	100	100
HC	ppm_v (dry)	10	10

[a] All emission rates are corrected to 7% oxygen.

[b] If the gas temperature at the inlet of the initial particulate control device is maintained below 400°F.

TABLE II Emissions Limits from Cement Kilns[a]

Pollutant	Units	Existing	New
D/F	ng TEQ/dscm	0.2 or 0.4[b]	0.2 or 0.4[b]
Mercury	μg/dscm	120	56
SVM	μg/dscm	240	180
LVM	μg/dscm	56	54
HCl/Cl_2	ppm_v (dry)	130	86
PM	kg/mg	0.15	0.15
Opacity	%	20	20
CO	ppm_v (dry)	100[c]	100[c]
HC	ppm_v (dry)	10[d] or 20[e]	10[d] or 20[e]

[a] All emission rates are corrected to 7% oxygen.

[b] If the gas temperature at the inlet of the dry particulate control device is maintained below 400°F.

[c] In the bypass duct or main stack.

[d] In the bypass duct.

[e] In the main stack.

II. INCINERATION SYSTEMS

This section discusses the various pieces of equipment that constitute a hazardous waste combustion facility. Sections II.A (liquids) and II.B (solids) discuss the major types of incinerators and waste combustors in use today and point out the types of wastes that each one is best suited to handle.

The major benefit of waste incineration is that the process actually destroys most of the waste rather than disposing of or storing it. As of January 1997, there were 156 permitted commercial and industrial hazardous waste incinerators in the United States at 110 locations, and 27 permitted commercial hazardous waste incinerators at 17 locations.

Hazardous waste incineration involves the application of combustion processes under controlled conditions to convert wastes containing hazardous material to inert mineral residues and gases. The required "three T's" of combustion must be present along with sufficient oxygen for the reaction. Four conditions are required:

1. Temperature of the combustion zone must be maintained; the burning mixture (air, wastes, and fuel) must be raised to a sufficient temperature to destroy all organic components.
2. Turbulence is required for the constant mixing of waste and oxygen.
3. Residence time, the time of exposure to combustion temperatures, should be sufficient to ensure that even the slowest combustion reaction has gone to completion.
4. Oxygen must always be available in the combustion zone.

Thus, four parameters influence the mechanisms of incineration: temperature, turbulence, residence time, and oxygen.

This section deals with the types of incinerators and waste combustors used to destroy hazardous waste. The liquid injection (LI) incinerator is the most common type, and is used for destroying a wide range of liquids generated in the process industries. The liquid burners are also adapted to the solids waste incinerators, the most common being the rotary kiln incinerator. LI incinerators are also applied to boilers, cement kilns, and lightweight aggregate kilns burning hazardous wastes as a source of energy for the industrial furnace or boiler.

TABLE III Emission Limits from LWAKs[a]

Pollutant	Units	Existing	New
D/F	Ng TEQ/dscm	0.2 or 0.4[b]	0.2 or 0.4[b]
Mercury	μg/dscm	47	43
SVM	μg/dscm	250	43
LVM	μg/dscm	110	110
HCl/Cl_2	ppm_v (dry)	230	41
PM	mg/dscm	57	57
CO	ppm_v (dry)	100	100
HC	ppm_v (dry)	20	20

[a] All emission rates are corrected to 7% oxygen.

[b] If the combustion gas is rapidly quenched to 400°F or lower.

A. Liquid Injection Incineration

In the field of hazardous waste incineration, there is more experience with LI incinerators than all other types combined. In 1981, approximately 64% of the total number of incinerators in service were of the LI type. Various types of incinerators are used to handle liquid waste as well as wastes in other forms, i.e., solids, sludges, slurries, and fumes. In the LI incinerator, which, as the name implies, is applicable almost exclusively to pumpable liquid waste, the waste is burned directly in a burner (combustor) or injected into the flame zone or combustion zone of the incinerator chamber (furnace) via nozzles. The heating value of the waste is the primary determining factor for nozzle location.

LI incinerators handle the various liquid streams that are being generated by the process industries. These units may be on site (at the location of the generator) or off site (at a sister facility or at a commercial disposal operation).

LI incinerators are usually refractory-lined chambers (horizontal, vertical orientation up or down), generally cylindrical, and equipped with a primary combustor (waste and auxiliary fuel fired) and often secondary combustors or injection nozzles for low-calorific-value materials (aqueous wastes containing either organic or inorganic compounds, or both).

These units operate at temperature levels from 950°C (1742°F) to 1700°C (3092°F). Residence time in the chamber may vary from milliseconds to 2.5 sec. The viscosity determines whether the material being incinerated is considered a liquid, slurry, or sludge. LI units today are capable of burning high-viscosity materials of 4500 SSU or less. Critical to the operation of the unit is be the atomizing nozzle used to convert the liquid stream into finely atomized droplets.

TABLE IV Waste Liquid Fuel Data

Chemical composition (including halides, sulfur, nitrogen, etc.)
Specific gravity
Heat of combustion
Viscosity
Corrsivity
Ignitability
Reactivity
Polymerization
Solids content (type, percentage, physical data)
Metals
Slagging properties (temperature, eutectic data)

1. Liquid Incinerator System Design

a. Storage. In order to design the storage system to permit a feed of uniform mixture and heating value, proper mixing in the storage tank is required. Layering often occurs in large tanks, especially where wastes are introduced from various sources. Venting of the tank is also necessary to prevent pressure buildup. Vent gases should be exhausted to the incinerator chamber.

b. Transport. High-viscosity materials (slurries and sludges) and liquids containing solids are a serious concern in the design of the feed system to the incinerator. Various pump types are available and their selection and use is predicated on the properties of the waste stream. Progressive-cavity pumps have been used successfully, provided the materials of the rotor and stator are selected properly. Gear pumps have been the workhorse for most lower viscosity materials (equivalent to No. 6 fuel oil). In systems where a variety of waste streams must be pumped into the incinerator, the operator must be instructed about the proper feed system to use with each waste. In some cases, pumps may not be practical because of the high maintenance costs and shutdowns that result. Pressurizing with an inert gas (nitrogen) may be the safest and most economical method. Diaphragm pumps are in use at many installations where other pumps have failed.

The atomizer design usually dictates the pressure requirements of the transport system.

Certain liquid wastes and sludges must be maintained at temperature levels sufficiently high to permit pumping. If cooled, the material will solidify in the pipeline.

c. Waste-liquid data. The physical, chemical, and thermodynamic properties of the waste must be considered in the basic design requirements of the total incinerator system. This includes storage tanks, mixers, pumps, control valves, piping, atomizers, combustors, refractory, heat recovery, quench system, and air pollution equipment. The types of data needed by the designer properly to engineer the total system and the individual components are listed in Table IV.

d. Atomizers. The method of injection of the liquid into the burner or the waste combustor is one of the most important features of a well-designed system. The main purposes for injecting the liquid as a spray are to break up the liquid into fine droplets, place the liquid droplets in a specific zone with a specific pattern with sufficient kinetic energy and penetration into the burning zone, and control the rate of flow of the liquid discharged to the nozzle.

Liquid waste atomization can be achieved by any of the following atomizers: rotary cup; single-fluid pressure; two-fluid, low-pressure air (internal mix); two-fluid, high-pressure air or steam (internal mix); and two-fluid, high-pressure air or steam (external mix).

e. Combustion air system. Proper mixing of the combustion air with the liquid droplets is necessary. As the liquid is vaporized and superheated to ignition temperature, oxygen reacts with the hydrocarbon vapor to allow release of the energy.

As the viscosity of the liquid increases, droplet sizes tend to get larger. Increased turbulence resulting from high-intensity mixing permits this reaction to be achieved rapidly. Energy is imparted to the combustion air of the combustor and this in turn permits more rapid mixing (turbulence) of the air with the fuel droplet.

As the superheated liquid reaches ignition temperature, oxygen must be available to complete the oxidation reaction. Many problems result if the vaporized hydrocarbon contacts a low- oxygen-level stream. Pyrolysis results with a cracking of the hydrocarbon into carbon, carbon monoxide, and hydrogen, and time for completion of combustion is increased.

f. Combustors. Primary and secondary combustion units are utilized in the LI incinerator systems. Primary units are utilized for those wastes with sufficient heating value to supply the necessary heat input to the system without need for auxiliary fuel. With nozzle-type burners, wastes with heating values from approximately 2500 kcal/kg (4500 Btu/lb) and above can be burned satisfactorily. The type of burner design, air mixing, and turbulence determine the minimum heating value for burning without auxiliary fuel.

Low-intensity or laminar flame burners (air pressure drop varies from 50–150 to 2–6 mm H_2O) utilize high excess-air levels (25–60%) to provide the mixing of air with the waste fuel. These usually demonstrate low turbulence. High-intensity burners operate with combustion air pressures of (200–500 to 8–20 mm H_2O). As a result, the increased turbulence resulting from the energy imparted by the air allows operation at much lower excess-air levels (0–20%).

Each burner, regardless of type, is usually mounted in a refractory block or ignition tile (see Fig. 1). This is necessary to confine the primary combustion air introduced through the burner, ensure proper air–fuel–waste mixing, and maintain ignition. Manufacturers each have their own geometric specifications, which have been developed through past experience.

Improved utilization of combustion space and higher release rates can be achieved with the utilization of swirl or vortex burners or designs involving tangential entry.

g. Halogenated hydrocarbons. In the plastics industry, the generation of by-products from vinyl chloride monomer (VCM) and polyvinyl chloride (PVC) processes has expanded rapidly. In many of the larger facilities where the majority of product is produced, an increase in toxic gaseous vents and viscous liquid by-product material results. Since many of these materials are listed as Appendix VIII compounds (CFR, 1991), it is necessary to recycle or dispose of them in a safe and proper manner to protect the health and the environment of local communities.

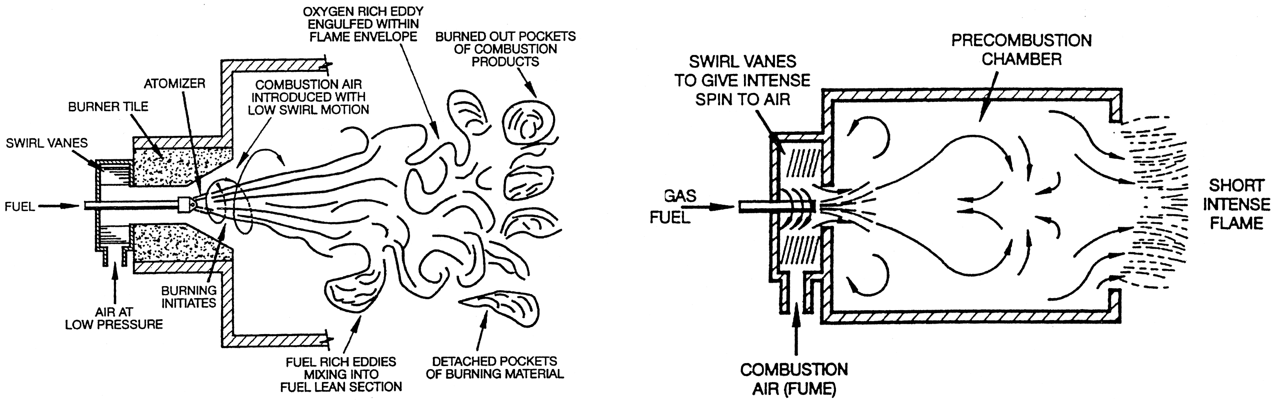
TABLE V Organic Waste Liquors

Organic (nonaqueous)	Aqueous with ash	Aqueous without ash
Solvents	Sodium Glutamate	Phthalic anhydride
Waste oils, sludges	Mlasses ferment	Phenol
Halogenated hydrocarbons	Synheic phenols	Oil water
Organic acids	Polyesters	Maleic anhydride
Organometal components	Caprolactum	
Aromatics	Agricultural-process wastes	

The objective in incinerating residues containing chlorine is to convert as much as possible of the chlorine content to hydrogen chloride. HCl can be absorbed in water from the reaction gas mixture. Chlorine requires a reagent (caustic) for proper scrubbing and neutralization.

h. Aqueous waste systems with ash. In reviewing the aqueous wastes in Table V, one should note the distinction between those wastes with and without ash. Non-ash-bearing wastes are easily handled in either a horizontal or vertical (up-oriented) LI incinerator system. The waste is usually injected downstream of the burner combustion chamber into the flame of the main burner.

Atomizer nozzles should be protected in the event of loss of waste flow by maintaining the flow of the atomizing



(a) (b)

FIGURE 1 (a) Low-air-pressure, long-flame combustor; (b) high-air pressure, short-flame vortex combustor. [From Santoleri, J. J. (1995). "Hazardous Waste Incineration Course." AWMA Course No. WST-258, Kansas City, MO (March 27, 1995).]

fluid. The design should provide sufficient air with the atomized liquid at the injection point to oxidize the organics contained in the waste stream. Incinerators have been known to produce soot as a result of a pyrolysis condition at the entry of the aqueous waste (lack of oxidizing air).

Incinerator designs for the oxidation of those wastes with high ash content are vertically-down oriented (see Fig. 2). This allows gravity discharge of the molten ash material into the quench zone [quench pot with sprays, submerged exhaust (sub-X^{TM}), or hot tap port]. Critical to the design and operation of incinerators for this waste type is the selection and design of the refractory. Severe spalling and excessive maintenance and down time have resulted from improper selection of refractories. (Santoleri, 1995).

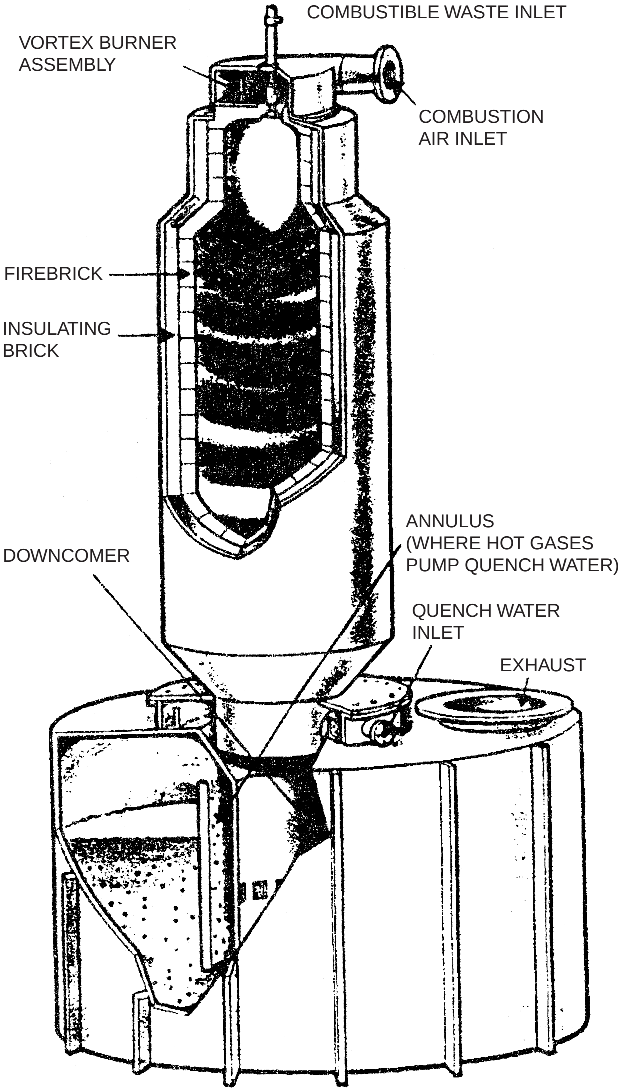


FIGURE 2 Sub-X^{TM} Incinerator. [Courtesy T-Thermal Co.]

B. Solids Incineration

1. Rotary Kiln Incinerator

The rotary kiln (RK) is a cylindrical refractory-lined shell that is mounted at a slight incline from the horizontal plane to facilitate mixing the waste materials and exposing the surface to the auxiliary burner flame and the waste fuel flames and flames generated over the burning surface of solids. The kiln accepts all types of solid and liquid waste materials with heating values between 555 and 8333 kcal/kg (1000–15,000 Btu/lb), and even higher. Solid wastes and drummed wastes are usually fed by a pack-and-drum feed system, which may consist of a bucket elevator for loose solids and a conveyor system for drummed wastes. Pumpable sludges and slurries are injected into the kiln through lances or nozzles. Temperatures for burning vary from 700°C to 1315°C (1300–2400°F) depending upon the waste types and shape.

RKs are classed as slagging ($>$1000°C, 1832°F) or nonslagging ($<$1000°C, 1832°F). The nonslagging rotary kiln for hazardous waste destruction does not attain high destruction and removal efficiencies (DRE) required of an incinerator due to the basic manner in which it operates. It will volatilize the organic content of the feed materials which as volatiles are carried in the combustion products exit gas stream. A secondary combustion chamber (SCC) is located immediately downstream of the kiln to perform the final destruction process on the kiln exit combustion products as shown in Fig. 3. The nonvolatile part of the material fed into the kiln such as metals, stones, sand, and other inorganics exit as ash. The ash and residue exiting the kiln should contain minimum to zero organics and may be treated and disposed of in a hazardous waste landfill. Due to the operating temperature some metals and other elements will vaporize within the kiln and be carried in the exhaust gas stream to the downstream equipment.

The rotary kiln combines three very important parameters of waste material processing into one piece of equipment: mixing, temperature, and time. Hazardous waste feed streams are generally not of uniform size, shape, or composition. The waste material that is placed into the kiln must have sufficient time for its organic components to volatilize. With proper time, the heat source within the kiln can raise the waste material to the desired temperature. The rotary kiln provides excellent mixing through its tumbling action to distribute the heat evenly to all of the material being processed and to expose all of the material to the heat source.

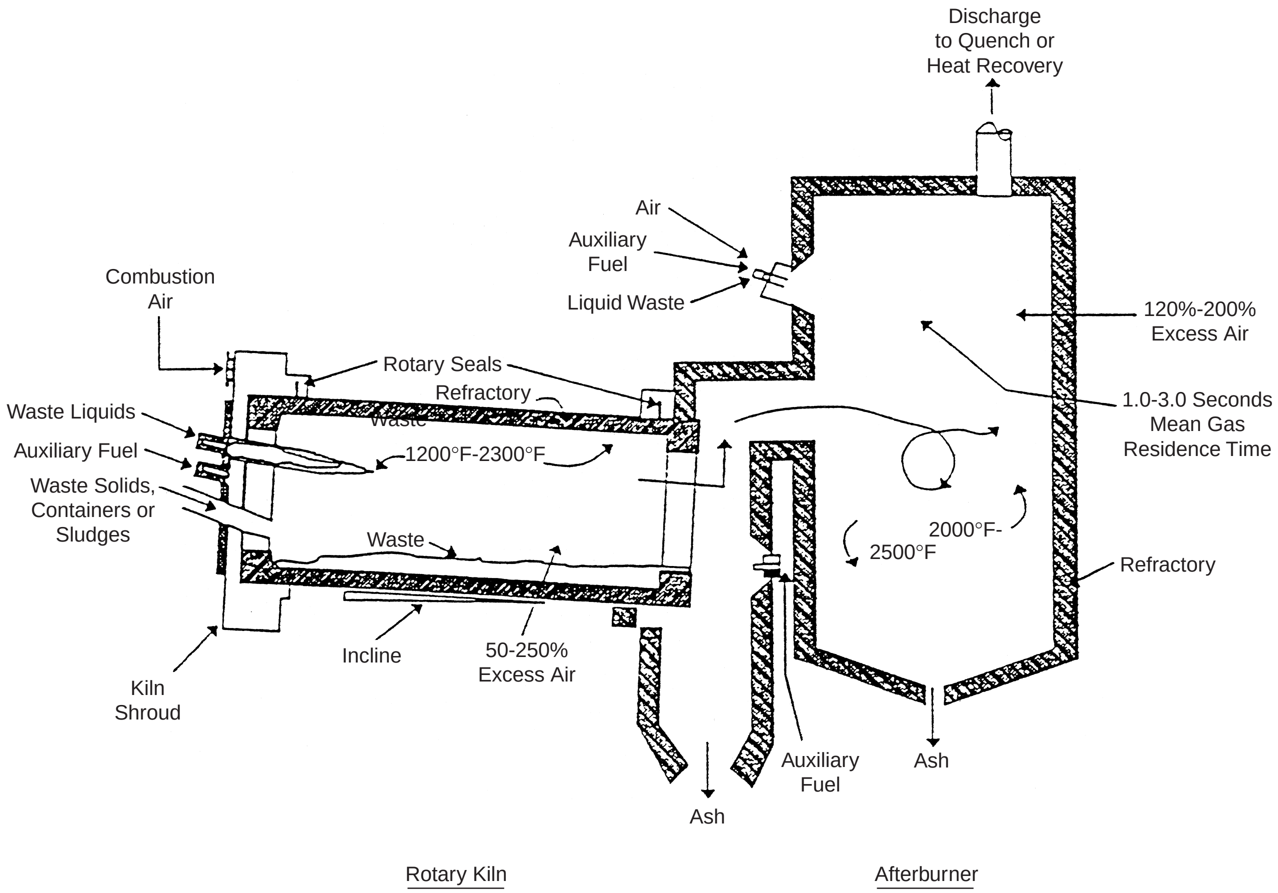


FIGURE 3 Rotary kiln/SCC incinerator system. [From Dempsey, C., and Oppelt, T. (1993). "Incineration of hazardous waste: A critical review." *Air Waste* **43**(January), 35.]

a. Kiln selection criteria. Two types of rotary kilns are being manufactured: cocurrent (burner at the front end with waste feed) and countercurrent (burner at the back end). For a waste having low combustibility (such as a high-water-volume sludge), the countercurrent design offers the advantage of controlling temperature at both ends, which all but eliminates problems such as overheating the refractory lining.

Another feature of the rotary kiln is that it can operate under substoichiometric (oxygen-deficient) conditions to pyrolyze the wastes. The combustible off-gases can then be incinerated with liquid wastes in the SCC. This mode of operation reduces the total mass flow of combustion gases from the SCC into the APC system. It also reduces the particulate matter carryover in the kiln gases. The rotating action allows better volatilization of the nonpumpable wastes than do stationary or fixed-hearth incinerators. Care must be exercised to prevent air leakage into the kiln or into the interconnecting duct between the kiln and SCC. This will create an explosive mixture with the rich gases, ignition temperatures, and oxygen availability.

b. Feed systems. As stated above, the kiln generally processes solid materials. The solid material can be either loose or packaged.

Containerized solids are fed into the kiln with either a chute or a ram system. Air locks are provided to permit control of entrained air. The method and control of the waste feed must be closely monitored. When a container is added to the kiln, it may release all of it thermal energy in the very short period of time that it takes to degrade the container. This is dependent upon the volatility of the material in the container, the heat content in Btu/lb, and the percentage of the total heat release that the container represents. The combustion air flow rate to the kiln is almost constant due to the operation and control of the induced draft (ID) fan. The oxygen available for combustion is also constant. Most kilns operate with excess air levels in the range of 50–100%. This establishes the range of heat release rate of the kiln. As higher Btu rates are fed to the kiln, the oxygen in the air will be depleted and the temperature will rise. A batch-fed system will have a continuous rise and fall of the oxygen at the exit. As oxygen falls, the temperature will rise. As burnout occurs, the oxygen will increase since air is entering at a fixed rate. This is followed by a drop in temperature. As a result, feed heating value and feed rate are critical to the proper operation and control of the kiln. Refractory damage often occurs due to high temperatures resulting from high heat input and substoichiometric oxygen levels.

c. Secondary combustion chamber. The RK acts as the primary combustion chamber (PCC) to volatilize and oxidize combustibles in the wastes. The volatilized combustibles exit the kiln and enter the SCC, where additional oxygen is available and ignitable liquid wastes or fuel can be introduced. Complete combustion of the waste and fuel occurs in the SCC. Both the SCC and RK are usually equipped with an auxiliary fuel firing system to bring the units up to the desired operating temperatures.

d. Design criteria. The flow of material through a rotary kiln is determined by the kiln slope, rotational speed, and characteristics of the material at the kiln operating temperatures. The kiln is installed with a slight slope (between 0.01 and 0.07 vertical foot per horizontal foot.) to allow the bed of solids to advance through the kiln by gravity. As the kiln is rotated, the solid material is carried up on the kiln walls and eventually tumbles or slides down the wall. Gravity forces the material to fall down at the point that it is released by the wall. The forward movement is directly related to the slope of the kiln.

Residence times for RK waste incinerators range from 15 min up to a maximum of 2 hr for the countercurrent kilns. The kiln revolution rate ranges from 0.15 to 2.5 rpm. The length normally ranges from 15 to 60 ft for the cocurrent kilns. The countercurrent kilns range up to 120 ft to obtain the additional residence times.

The combustion gases remain in the kiln for a very short time. Due to the short times and low gas velocities, there is very little gas mixing. Air that enters the system may not mix with all of the organics emitted by the waste. Hot and cold layers will exist within the kiln. The SCC must process all of the kiln exit gases to properly mix and oxidize all of the organics in the gas stream.

2. Fluidized-Bed Combustors

The primary application of fluidized-bed combustors (FBC) has been to municipal sewage sludge, but they have also been used for industrial and hazardous waste treatment. Among the industrial wastes burned since the early 1960s are chemical process wastes, petroleum tank bottoms, sludge from pharmaceutical, pulp and paper, and nylon manufacturing operations, waste plastics, carbon black, spent activated carbon, precoat filter cake, and waste oils and solvents. Industry began to apply fluid beds to the destruction of hazardous wastes in the 1980s.

The advantages of FBC technology include efficient combustion, ease of control, ability to handle variable feeds, ability to operate intermittently, and reasonable-to-low capital and operating costs. Additional advantages relative to hazardous wastes include fuel savings and lower emissions of nitrogen oxides (NO_x) and metals.

The FBC, shown schematically in Fig. 4, is a vertical, cylindrical, refractory-lined carbon steel vessel with four major components:

- A plenum, or windbox, into which the fluidizing/combustion air is received
- An air-distribution plate, which transmits the air from the windbox to fluidize the bed and prevents the bed material from sifting into the windbox
- The fluid bed of inert material, usually sand or product ash, into which the feed and/or fuel is injected
- The freeboard area above the bed

The flue gas exits from the bed through the freeboard, where oxidation (particularly of volatiles) is completed and the larger solid particles entrained in the flue gas fall back into the bed. Due to attrition or burnout, ash is usually reduced to a sufficiently small size that it exits the combustor with the flue gas. The bed material that accumulates is typically removed from the bed by an overflow drain, thus taking advantage of the bed's fluid characteristics.

In most instances, the FBC is of the forced-air type with a single fluidizing blower at the front end. A total system including the air pollution control system is shown in Fig. 4. If the downstream heat recovery or air pollution control devices are unable to withstand pressure, an induced draft fan may be added to provide a balanced draft. Balanced draft systems are also utilized to provide a negative pressure in the FBC so that difficult-to-move materials can be fed via gravity through a chute-and-airlock arrangement.

FBCs find their primary application to feeds that are nonatomizable liquids, slurries, sludges, tars, and granular solids, with the potential for combusting liquids or gases in combination with these. Thus, in terms of feed form, FBCs are intermediate between the simple liquid injection incinerator and the rotary kiln.

3. Fixed Hearth Incinerator

In the early 1960s, a new, intrinsically simple type of incinerator started gaining in popularity as public attention focused on air pollution from waste burning. The modular combustion unit (MCU) has become an economical and efficient system for on-site and central destruction

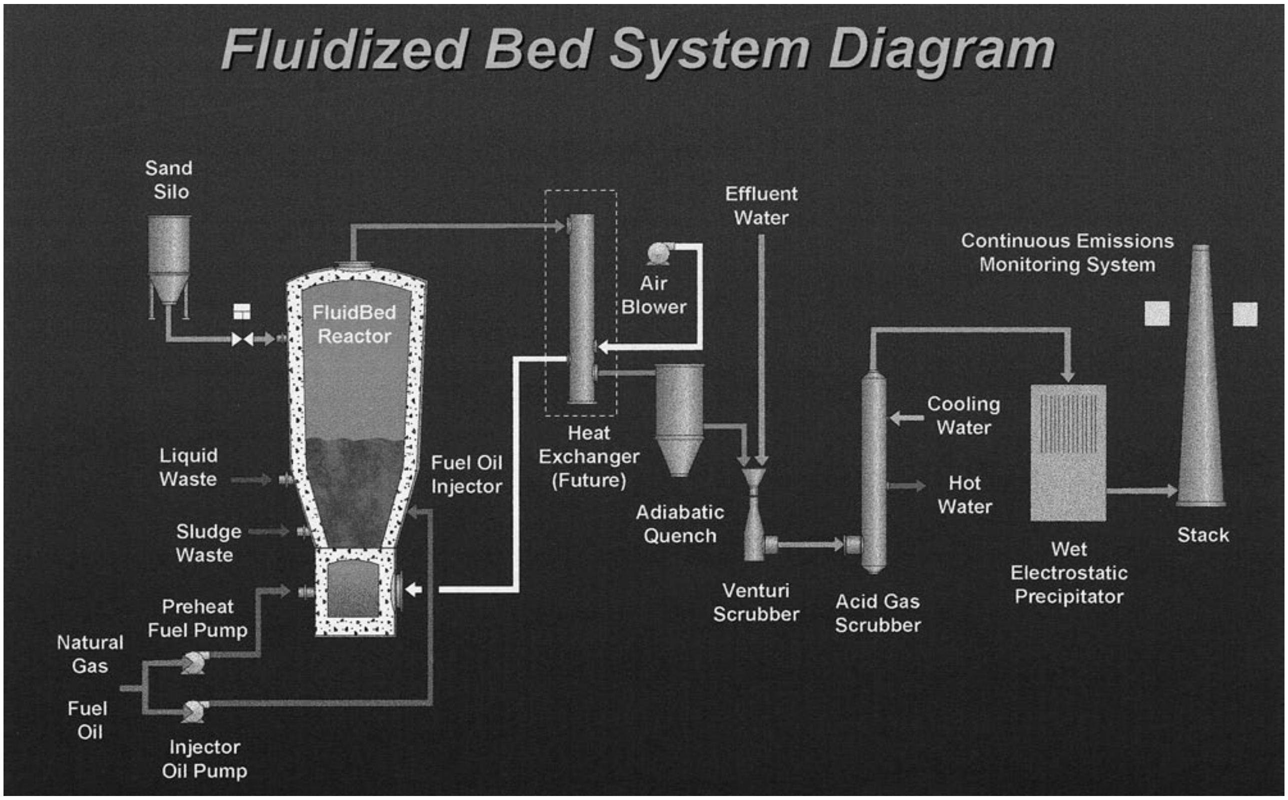


FIGURE 4 Fluidized-bed incinerator system.

of certain wastes. These incinerators are also known as starved air and controlled air units.

Fixed hearth incinerators, which are used extensively for medical waste incineration, are less commonly used to incinerate hazardous wastes. There are several units employed at commercial sites as well as plant sites for hazardous waste destruction. Fixed hearths can handle bulk solids and liquids, however, this design does not have the versatility of a rotary kiln.

The MCU consists of two major furnace components as shown in Fig. 5: a primary (PCC) and a secondary chamber (SCC). Waste is charged into the PCC and a carefully controlled flow of air is introduced. Ideally, enough air is provided to allow sufficient burning for pyrolysis to occur. Typically, 70–80% of the stoichiometric air requirement is introduced into the PCC.

The off-gas generated by the pyrolysis reaction contains combustibles; this gas is burned in the SCC, which is sized for sufficient residence time to destruct organics in the off-gas. A carefully controlled quantity of air is also introduced into the SCC, but in this case, excess air, 50–100% of the off-gas stoichiometric requirement, is maintained for complete combustion. In the past, gas cleaning devices such as bag houses or electrostatic precipitators were not required. The burnout of the off-gas in the SCC was usually sufficient to clean the gas to meet air emission standards. However, with today's CAAA, RCRA, and MACT standards, scrubbers are required to reduce the particulate and acid gas emissions to their proper and safe levels. Figure 6 shows a total system which includes the controlled air incinerator followed by a waste heat boiler, wet scrubber system, induced draft fan, and stack.

MCUs were developed for the destruction of trash, but they are applicable to other solid waste destruction, and their SCC can be used for destruction of gaseous or liquid waste in suspension. Units have been installed and permitted for incineration of hazardous waste as well as codisposal of treatment-plant sludge with general solid waste.

Wastes containing plastics will generate acid gases (e.g., HCl) as a product of combustion and will require acid gas scrubbers.

Small units, under 750 lb/hr (9.4×10^{-2} kg/sec), are normally batch-fed. Waste is charged over a period of hours, and after a full load has been placed in the chamber, the chamber is sealed and the waste fired. Larger units use a continuous waste charging system such as a ram feeder, screw feeder, or series of moving grates.

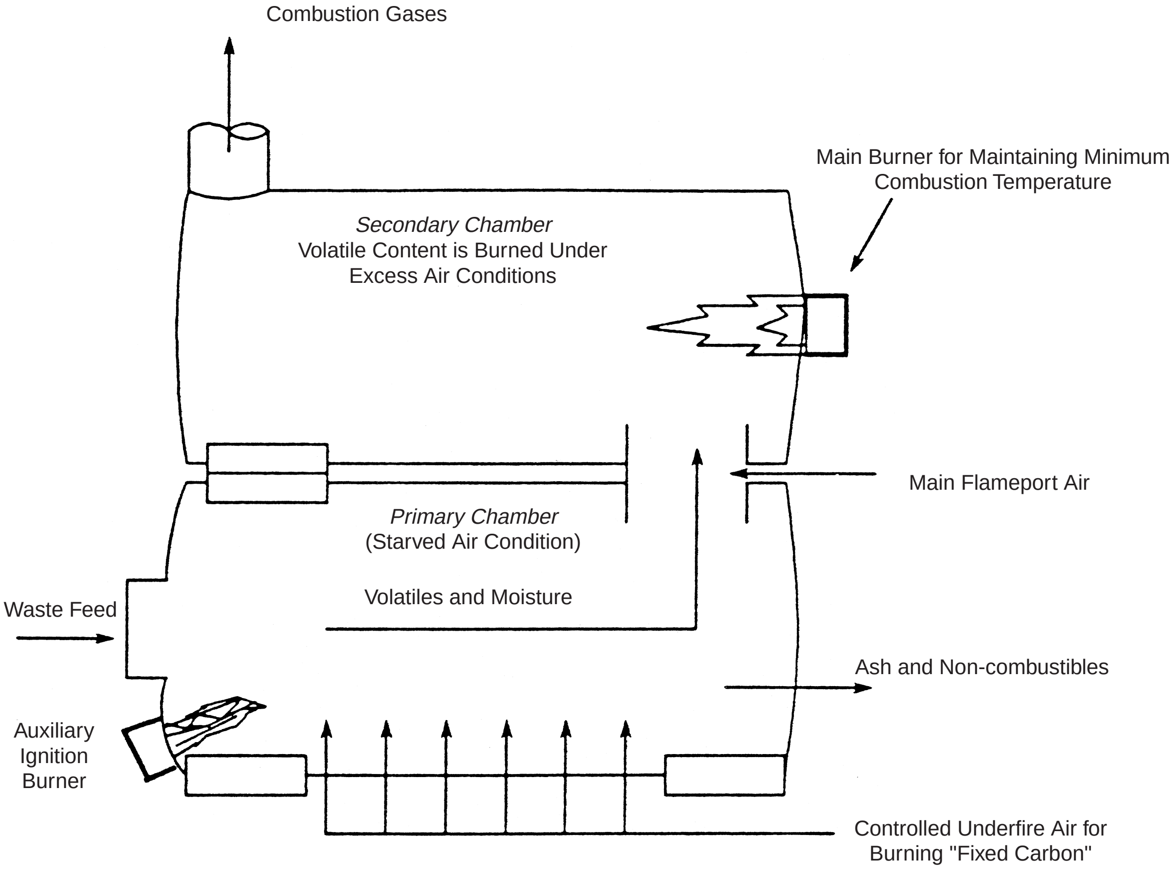


FIGURE 5 Controlled air incinerator.

As with waste charging, MCUs are provided with both manual and automatic discharge systems. After burnout, the chamber of smaller units is opened and ash residue is manually raked out. With continuous-operating units, ash is continually discharged, normally into a wet well, from which it is transferred to a container or truck by means of a drag conveyor.

4. Multiple Hearth Incinerator

The multiple hearth incinerator (also known as the Herreshoff furnace) was developed at the turn of the century for ore roasting. Since the 1930s, it has been used extensively for sludge burning and until the 1980s was the most prevalent incinerator for the disposal of treatment-plant sludge in the United States. It has been adapted to carbon regeneration, recalcining, and numerous industrial applications. There have been no installations for hazardous waste incineration.

Due to the more stringent requirements of RCRA and the CAAA of 1990, many multiple hearths are being phased out and replaced with fluid beds (Lundberg *et al.*, 1995). Fluid-bed incinerators operate at much lower excess air rates, resulting in higher combustion efficiency, lower CO emissions, and lower volumes of exhaust gas products. This reduces the size requirements for the APC system and horsepower for the ID fans.

C. Other Hazardous Waste Combustors

1. Waste Heat Boiler

A boiler is a closed vessel in which water under pressure is transformed into steam by the application of heat. More specifically, EPA has defined the following characteristics of a boiler for regulatory purposes under the RCRA:

- The combustion furnace and energy recovery section must be of integral design.
- Thermal recovery efficiency must be at least 60%, and at least 75% of the recovered energy must be "exported" (i.e., not for internal boiler uses).

Boilers come in a variety of sizes, configurations, and designs. The two common boiler design categories which

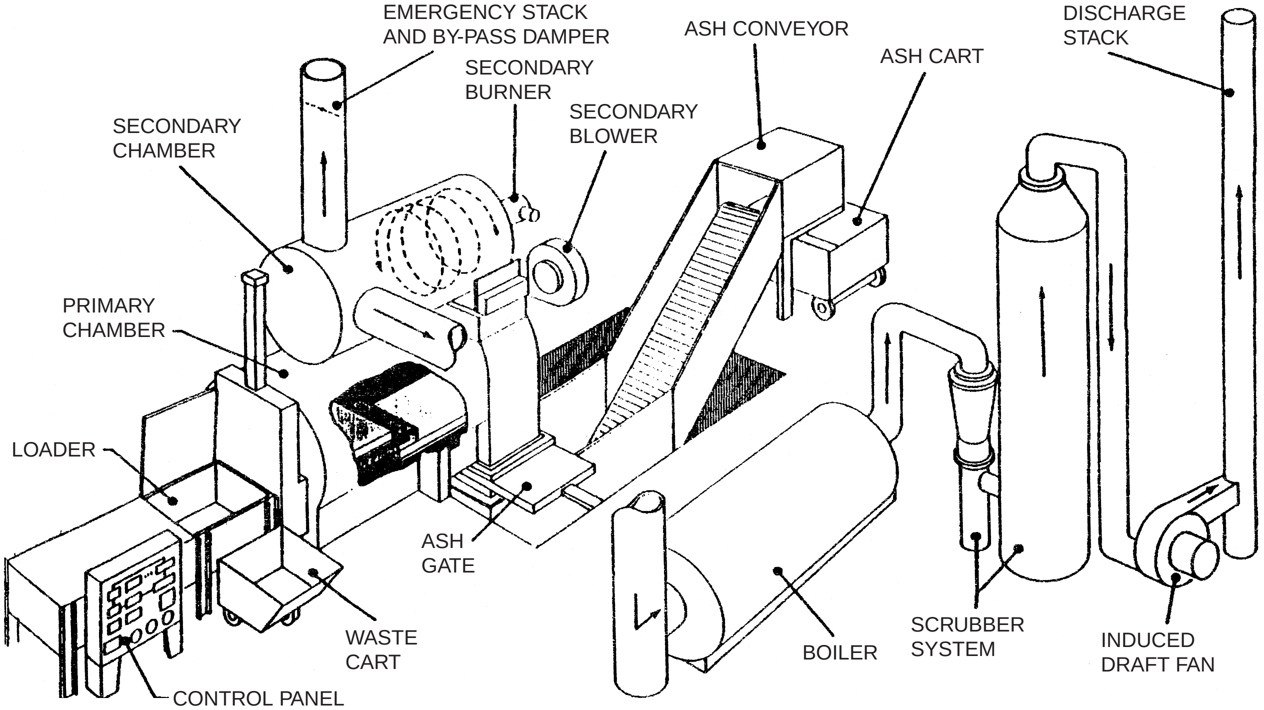


FIGURE 6 Controlled air incinerator air pollution control system.

burn hazardous waste are the firetube boiler and the watertube boiler.

a. Firetube boiler. The firetube boiler can be described in simple terms as a water-filled cylinder with tubes running through it which provide the escape path for the combustion gases or flue gas. As the flue gas passes through the tubes, the hot gases heat the tubes, which then heat the water to produce steam. The firetube boiler is primarily used in industrial applications.

Firetube boilers are compact, low in initial cost, and easy to modularize based on plant requirements. Firetube boilers usually range in size from less than 2 to 50 million Btu/hr. Most industrial firetube boilers are either of horizontal return tube (HRT), scotch marine, or firebox type.

Firetube boilers may be directly fired with waste fuels, either liquid or gaseous. They also serve to recover heat from incinerators fired on waste fuels.

b. Watertube boiler. The basic design of a watertube boiler circulates the water through tubes, with the hot combustion gases passing over the outside surfaces of the tubes. Generally, the boiler can be physically divided into two sections, the furnace and the convection pass, as shown in Fig. 7. Furnaces (fireboxes, combustion chambers) vary in configuration and size, but their function is to contain the flaming combustion gases and transfer the heat energy to the water-cooled walls. The convection pass contains the *superheaters*, *reheater*, *economizer*, and *air preheater* heat exchangers, where the heat of the combustion flue gases is used to increase the temperature of the steam, water, and combustion air.

Watertube boilers can withstand much higher operating pressures and temperatures than firetube boilers. In addition, the watertube boiler design is safer. They can also burn a wide variety of fuels as well as wastes and have the ability to expand and contract more easily than firetube boilers. The major drawback is that watertube boilers are more expensive to install. The also require more complicated furnaces and repair techniques.

2. Cement Kiln Incineration

Cement kiln incineration is a proven technology for treating hazardous wastes. This process is used commercially since reliable data from research, demonstration, and operating units have been provided to overcome public opposition.

Figure 8 is a schematic of a cement kiln process. It has been demonstrated that existing cement kilns, when properly operated, can destroy most toxic organic chemical wastes, including PCBs, which can be converted to less noxious forms during normal cement kiln operation.

Since many chemical wastes have significant heat content of 5555 kcal/kg (10,000 Btu/lb) or more, cement

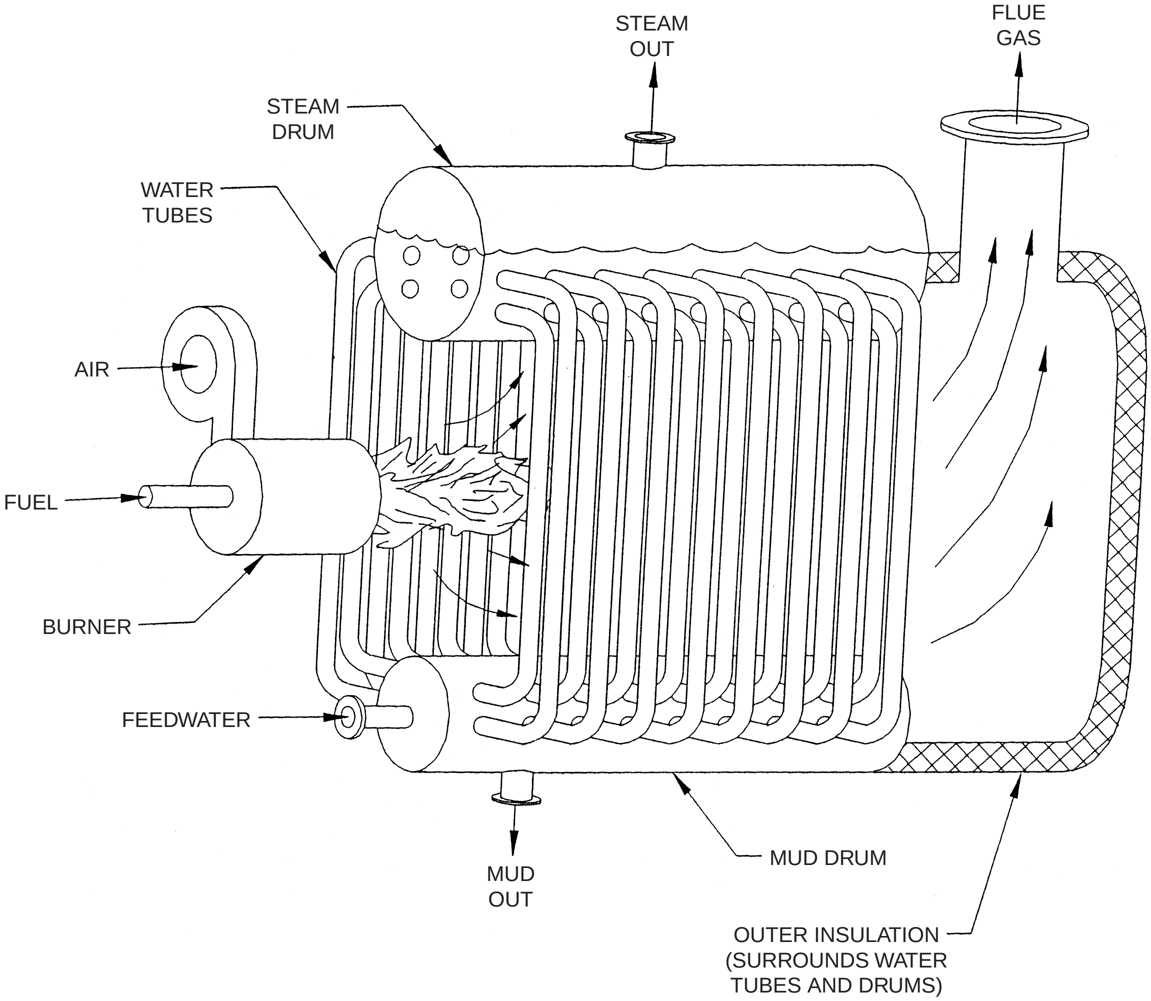


FIGURE 7 Watertube boiler cross section.

plants can save large quantities of energy by burning such liquid chemical wastes as alternate or synthetic fuels.

a. Cement kiln types. Cement can be produced in three different types of arrangements:

- Wet process kilns. A schematic of the wet cement making process is shown in Fig. 8.
- Dry process kilns. There are three different dry process configurations: long, preheater, or preheater/precalciner.
- Semi-dry process kilns.

3. Lightweight Aggregate Kiln

Lightweight aggregates include a wide variety of raw materials which when combined with cement form concrete products. Lightweight aggregate concrete is produced either for structural purposes or for thermal insulation purposes. This section presents an overview of the lightweight aggregate kiln process and a characterization of the lightweight aggregate kilns currently burning hazardous waste.

a. Process description. A lightweight aggregate plant is composed of a quarry, a raw material preparation area, a kiln, a cooler, and a product storage area. The material flows from the quarry to the raw material preparation area. From there, the material is fed into the rotary kiln. The rotary kiln consists of a long, steel cylinder lined internally with refractory bricks which is capable of rotating about its axis and is inclined at an angle of about 5 deg to the horizontal. The length of the kiln is usually 90–180 ft. The prepared raw material is fed into the kiln at the higher end, while firing takes place at the lower end. The dry raw material fed into the kiln is initially preheated

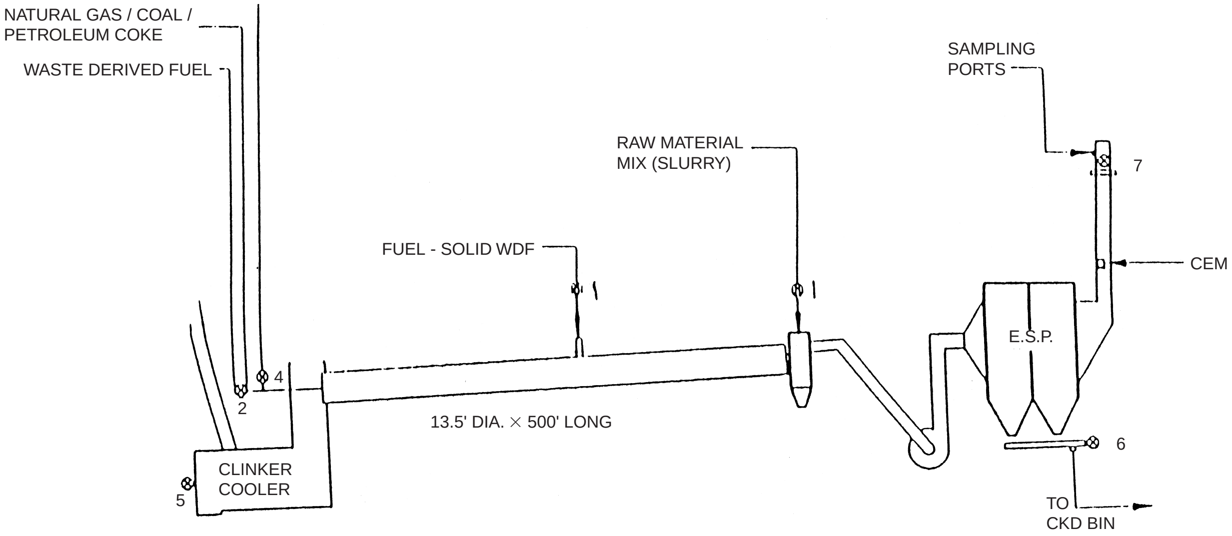


FIGURE 8 Wet process cement kiln.

by hot combustion gases. Once the material is preheated, it passes into a second furnace zone, where it melts to a semiplastic state and begins to generate gases which serve as the bloating or expanding agent. In this zone, specific compounds begin to decompose and form gases such as CO_2, SO_x, and O_2 that eventually trigger the desired bloating action within the material. As temperatures reach their maximum at approximately 1150°C (2100°F), the semiplastic raw material becomes viscous and entraps the expanding gases. This bloating action produces small, unconnected gas cells, which remain in the material after it cools and solidifies. The product exits the kiln and enters the cooler section of the process, where it is cooled with cold air and then conveyed to the discharge.

b. Details. A typical lightweight aggregate rotary kiln has a combustion gas residence time of 4 sec based on maximum temperature and a temperature range of 370–1150°C (700–2100°F). A typical rotary kiln processes approximately 450 Mg (500 tons) of lightweight aggregate per day. This kiln has an average furnace volume of 1000 m^3 (35,343 ft^3) and an average volumetric flow rate of 1920 Nm^3/min (73,170 scfm). Lightweight aggregate kilns that burm hazardous waste typically burm 100% liquid hazardous waste (i.e., hazardous waste is the sole fuel).

III. AIR POLLUTION CONTROLS

This section covers equipment that is always or often part of a waste combustion system. The hot gases exiting the combustion chamber contain gaseous and particulate pollutants that cannot be permitted to escape into the atmosphere. The greatest mass of air contaminants consists primarily of the following criteria pollutants: oxides of nitrogen, oxides of sulfur, and particulate matter (PM). Their removal from the gas stream requires the use of air pollution control devices. Trace levels of noncriteria pollutants such as chlorinated by-products, benzene, heavy metals, and acid gases are also of great concern. The emission rates are dependent on the waste incineration rate, chemical composition of the waste, incinerator type, air pollution control equipment, and incinerator system operating parameters Because of materials limitations (e.g., the fabric materials used in filter baghouses), most types of air pollution control equipment cannot handle gas temperatures typical of incinerator exhaust streams. Before reaching the air pollution control equipment, therefore, the hot gases must be cooled. This is usually accomplished either through the use of a waste heat boiler, in which a good portion of the high enthalpy of the combustion chamber exhaust gases is recovered, or by utilizing a quenching technique. The choice of method to cool the gases is dictated by the economics of the particular installation (Mycock *et al.*, 1995).

A. Quench System

A quench system is used downstream of the incinerator furnace to cool hot furnace gases to adiabatic saturation temperature, about 185°F (85°C). A quench system includes a refractory-lined carbon steel chamber with a watertight rubber film between the refractory and the carbon steel. Where acids are generated in the furnace, the refractory is acid resistant. Either water or air can be used as the quenching medium. When water is used, it is evaporated while removing about 1000 Btu (1.1×10^6 J) from the hot gas for each pound evaporated. Air quenching results in a much higher gas flow through a quench system and scrubbers located downstream because of the large quantities of air that must be used. If a heat recovery device is located downstream of a quench system, air quenching offers greater potential for energy recovery than water quenching.

Quenching reduces the volume of gas leaving the furnace, allowing smaller equipment to be used downstream. The temperature downstream of a quench tank is maintained below about 185°F (85°C) allowing less-expensive and more-corrosion-resistant materials such as fiber glass-reinforced plastic to be used in the scrubbing system and stack.

B. Classes of Air Pollution Control Equipment

There are four classes of air pollution equipment employed for acid gas and PM control: wet scrubbers, dry scrubbers, wet–dry scrubbers and combinations of wet–dry–wet scrubbers. The selection procedure for choosing one of these control devices depends on the inlet acid gas loading, potential acid removal capability, particulate loading, particle size distribution, and regulatory requirements. The technologies for these devices are well established.

Most waste combustion facilities today employ one of four possible process schemes for air pollution control, as described below.

1. Wet Scrubber Systems

a. Venturi scrubber (VS). A VS is used for particulates followed by a packed bed absorber scrubber (PBS) for gases. The VS could be a low ΔP (LVS) or high ΔP (HVS) scrubber based on the PM loading and size. If the VS is used only as a primary quench upstream of the PBS, an LVS is used.

A Venturi scrubber consists of a convergent section where the velocity of the gas is accelerated to as high as 400 ft/sec (122 m/sec). The high-velocity gas impacts atomized water droplets in the Venturi throat. The atomized droplets offer an extremely large surface for mass transfer. The outlet of the scrubber includes a divergent section that allows gradual expansion of the gas, thus reducing pressure losses across the scrubber (see Fig. 9). For a given removal efficiency, as particulate leaving a furnace gets finer, more pressure drop is required across a Venturi scrubber, especially for particulate in the submicrometer range.

Data indicate that Venturis at HWI facilities operate with pressure drops in the range 30–100 in. H_2O. Liquid-to-gas ratios for Venturi scrubbers are usually in the range 5–20 gal/1000 ft^3 of gas. At existing HWI facilities, liquid-to-gas ratios ranging from 7 to 45 gal/1000 ft^3 of gas have been reported. In many cases, a minimum ratio of 7.5 gal/1000 ft^3 is needed to ensure that adequate liquid is supplied to provide good gas sweeping. Gas velocities for Venturi scrubbers operating at HWI facilities are in the range 100–400 ft/sec, similar to other industrial applications. The low end of this range, 100–150 ft/sec, is typical of power plant applications, while the upper end of the range has been applied to lime kilns and blast furnaces.

b. Wet electrostatic precipitator (WESP). WESPs are used for PM combined with a PBS (for gases). The two major features in a WESP system are as follows:

1. A preconditioning step, where inlet sprays in the entry section are provided for cooling, gas absorption, and removal of coarse particles. A more recent application utilizes a vertical tube, water cooled on the outside causing condensation of water vapor carried in the flue gases.
2. A wetted collection surface, where the liquid (either sprayed into the gases or condensed from the gases) is used to continuously flush away collected materials.

Particle collection is achieved by the introduction of liquid to the gas stream through either sprays located above the electrostatic field sections or condensation of the water contained in the saturated gases, and migration of the charged particles and sprayed liquid droplets to the collection plates. In general, applications of the WESP fall into two areas: removal of fine particles and removal of condensed organic fumes. Outlet particulate concentrations are typically in the range 10^{-3}–10^{-2} g/ft^3. This device has been installed to control HF emissions. Using a liquid-to-gas ratio of 5 gal/1000 acf and a liquid pH between 8 and 9, fluoride removal efficiencies $>98\%$ have been measured; outlet concentrations of HF were found to be <1 ppm.

There are WESP installations at HWI facilities worldwide. A typical application is the use of the WESP in conjunction with a low-pressure-drop VS upstream, where a major portion of the gaseous contaminants and heavy particles is removed. The WESP then serves as a second-stage control device for removal of the submicrometer particles, in particular the metal fumes, and remaining gaseous pollutants. With the MACT standards issued in September 1999, many facilities require retrofit to meet the lower PM emission standards. Pilot testing is recommended prior to to full-scale design and installation.

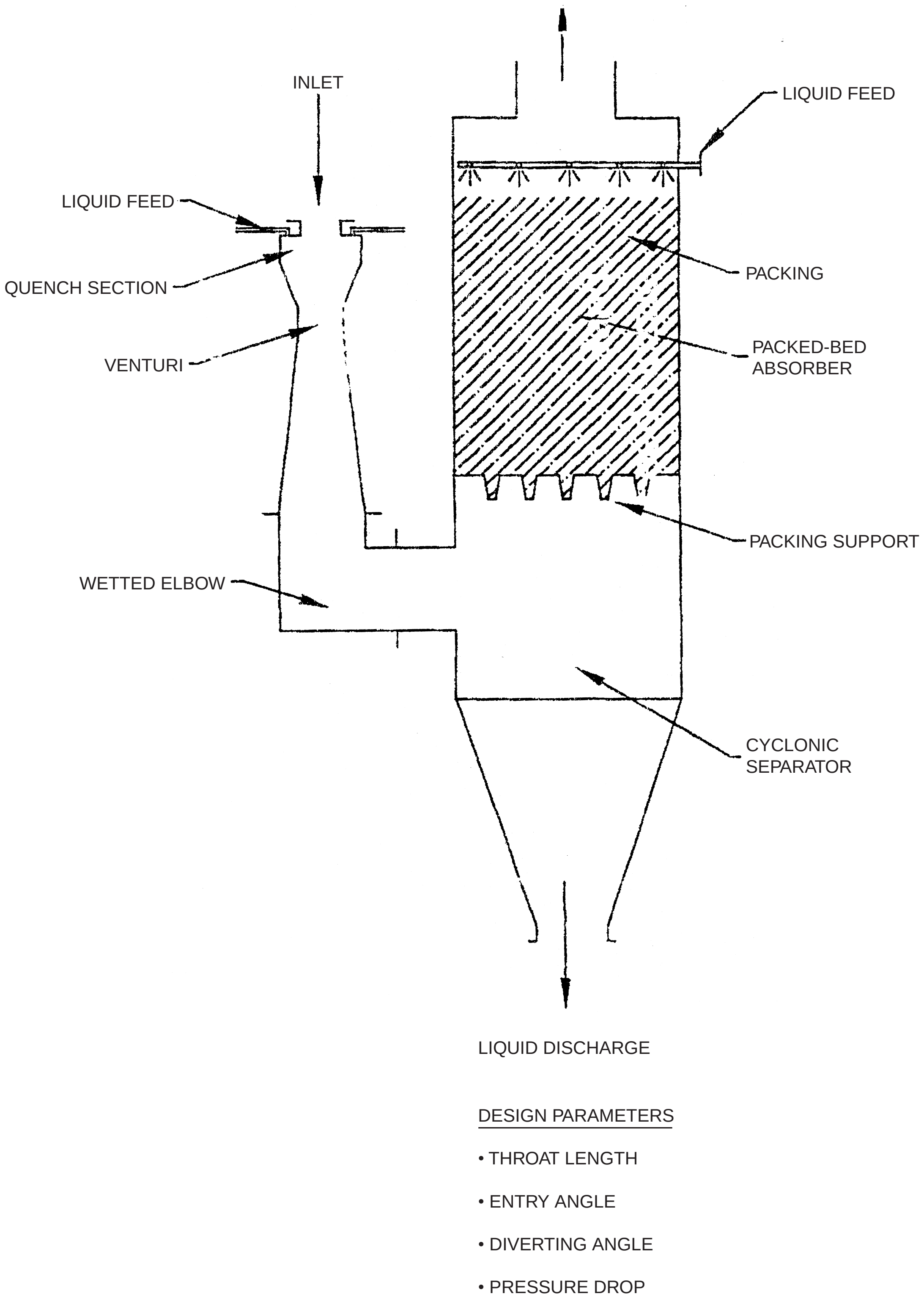


FIGURE 9 Venturi scrubber/packed-bed scrubber.

c. Ionizing wet scrubber (IWS). The IWS is used for PM followed by a PBS (for gases). The IWS was a relatively new development in the technology of removal of particulate matter from a gas stream in the early 1980s. These devices have been incorporated in on-site as well as off-site commercial incineration facilities.

In the IWS, high-voltage ionization in a charge section places a static electrical charge on the particles in the gas stream, which then passes through a cross-flow packed-bed scrubber. The packing is polypropylene in the form of circular-wound spirals and gearlike wheel configurations, providing a large surface area. Particles with sizes of 3 μm or larger are trapped by inertial impaction within the packed bed. Smaller charged particles pass close to the surface of either the packing material or a scrubbing water droplet. An opposite charge in that surface is induced by the charged particle, which is then attracted and attached to the surface. All collected particles are eventually washed out of the scrubber. The scrubbing water also functions to absorb gaseous pollutants.

The collection efficiency of a two-stage IWS is greater than that of a baghouse or a conventional ESP for particles in the range 0.2–0.6 μm. For 0.8 μm and above, the efficiency of bag collectors is greater; at 1.8 μm and above, the ESP is as effective as the IWS. Scrubbing water can include caustic soda or soda ash when needed for efficient absorption of acid gases. Corrosion resistance of the IWS is achieved by fabricating its shell and most internal parts from glass fiber-reinforced plastic (FRP) and thermoplastic materials. Pressure drop through a single-stage IWS is around 5 in. H_2O (primarily through the wet scrubber section). All internal areas of the ionizer section are periodically deluge flushed with recycled liquid from the scrubber recycle system.

d. Packed bed absorber towers. These are used for absorbing and neutralizing acid gases directly after an HVS which removes the PM. It is also used in a wet–dry–wet system downstream of a dry scrubber for more efficient removal of acid gases.

The packed-bed scrubber (PBS), a major type of wet scrubber applied as the APC in waste combustion systems, is not considered a PM control device, especially for submicrometer particulates. The major problems in using packed-bed scrubbers for particulate removal are plugging of the packing and the maintenance problems subsequently incurred. A general rule of thumb is to allow a maximum of 0.2 g/ft^3 of dust to enter a packed tower.

Packed columns are used for the continuous contact between liquid and gas. The countercurrent packed column (see Fig. 10) is the most common type of unit in gaseous pollutant control for the removal of the undesirable gas, vapor, or odor. The gas stream containing the pollutant moves upward through the packed bed against an absorbing or reacting liquid that is injected at the top of the packing. This results in the highest possible efficiency.

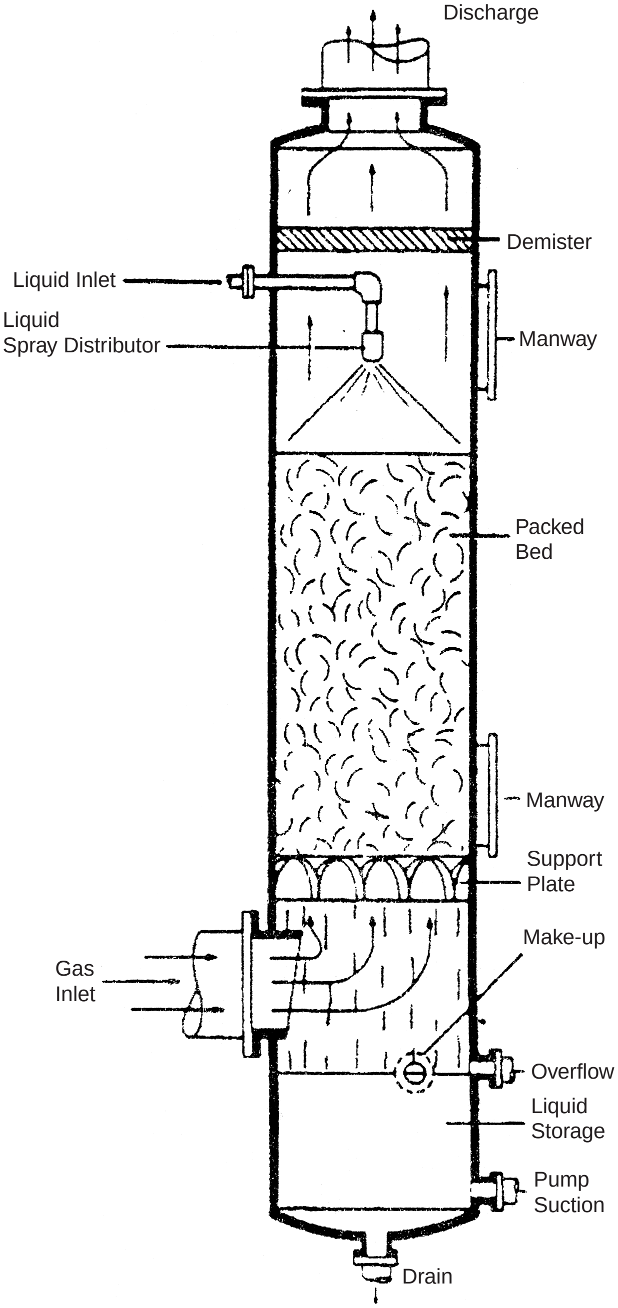


FIGURE 10 Packed-bed scrubber.

Water is often employed in absorbers to control emissions. When water is used as the scrubbing medium, an acidic stream is produced that must be neutralized prior to discharge. Hydrochloric acid concentration in the scrubbing liquor is normally limited to 1–2% by adjusting the makeup water and gas flow rates.

Caustic solution (typically 10–20 wt% caustic soda in water) is also commonly used in packed-bed absorbers to control HCl, Cl_2, and HF emissions. Because these compounds react with caustic, the driving force for mass transfer is increased and more efficient removal is achieved at the same liquid-to-gas ratio and packing depth.

e. Latest designs for wet scrubbers. Several specialty air pollution control devices, particularly in the scrubber area, are receiving more attention. Although on the market for many years, their unique collection mechanisms have recently attracted more interest. These include the Dyna-Wave^{TM} unit, the multi-micro-Venturi (MMV), the electro-dynamic Venturi (EDV), the condensing scrubber, and the collision scrubber. These are briefly described below.

The Dyna-Wave^{TM} unit creates a "froth" zone through which contaminated gas must pass. There is extreme turbulence in the zone, and efficient collection of submicrometer particles occurs (Goosens and Kint, 1995). The MMV scrubber consists of a "stacked" bank of staggered tubular elements arranged in a pattern similar to a bank of heat exchanger tubes. The thin spacing causes successive micro-Venturi flows at the "pinch" sections leading to high efficiencies. The EDV concept consists of a sequence of fundamental functions, namely saturation, condensation, ionization, and filtration. (Eagleson and Gilman, 1995). In condensing scrubbers, water vapor in the flue gas stream is condensed into particulates, causing the particle to grow to a size which enables easier collection by conventional means. At the same time, the temperature and volume of the contaminated gases are reduced providing a lower capital and operating cost for the downstream collection system and i.d. fan. Finally, in the collision scrubber two gas streams are made to collide head-on at the discharge of opposed Venturi throats containing water droplets. The collision action shreds the water droplets into finer ones that can more effectively collect submicrometer particles. The process also produces a larger liquid surface area for gas absorption (Bacon and Liang, 1997).

2. Dry Scrubber Systems

Dry scrubber systems are installed after the boiler or quench system. They comprise a dry lime or caustic injection system followed by either an electrostatic precipitator (ESP) for PM removal only or a fabric filter baghouse (FF) for PM removal with a percentage of acid gas neutralization on the bag surface.

a. Electrostatic precipitators. Electrostatic precipitator technology was developed in the United States by Dr. Fredrick Gardner at the turn of the century. Frederick Cottrell's precipitator was successfully applied in 1907 to the collection of sulfuric acid mist and shortly thereafter proven in a number of ore processing, chemical, and cement plants.

Electrostatic precipitators are satisfactory devices for removing small particles from moving gas streams at high collection efficiencies. They have been used almost universally in power plants for removing fly ash from the gases prior to discharge. Although electrostatic precipitators have found limited application in treating gases from incinerators in the Unied States because of their high cost and the varying nature of the electrical resistivity of the particulate matter, they do have the capability of fine particulate control. Resistivity plays an important role in determining whether a particle can be readily collected in this device.

b. Fabric filters (FF)—baghouses. One of the oldest, simplest, and most efficient methods for removing solid particulate contaminants from gas streams is by filtration through fabric media. The fabric filter is capable of providing high collection efficiencies for particles as small as 0.1 μm and can remove a substantial quantity of particles as small as 0.01 μm. In its simplest form the industrial fabric filter consists of a woven or felted fabric through which dust-laden gases are forced.

The fabric filters or bags are usually tubular or flat. The structure in which the bags hang is referred to as a *baghouse*; the number of bags in a baghouse may vary from a couple to several thousand. Quite often, when great numbers of bags are involved, the baghouse is compartmentalized so that one compartment may be cleaned while others are still in service.

Bottom-feed units are characterized by the introduction of dust-laden gas through the baghouse hopper and then to the interior of the filter tube. In top-feed units, dust-laden gas enters the top of the filters to the interior or clean-air side. When the gas flow is from inside the bag to the outside, by virtue of the pressure differential, the internal area of the filter element will be open and self-supporting; unsupported filter elements are tubular. When the filtration process is reversed, with the gas flow from outside the bag to inside, it is necessary to support the media against the developed pressures so that the degree of collapse is controlled. Supported filter elements are either of the tubular or envelope shape.

Baghouse collectors are available for either intermittent or continuous operation. For most air pollution control installations following an incinerator, it is desirable to use collectors that allow for continuous operation. This is accomplished by arranging several filter areas in a parallel flow system and cleaning one area at a time according to some preset mode of operation (see Fig. 11). Perhaps the outstanding characteristic of fabric filters is their inherent very high collection efficiency with even the finest particles. These units usually have the capability of achieving efficiencies of 99% almost automatically, provided they are properly constructed and maintained in satisfactory operating condition.

The size of a baghouse is primarily determined by the area of filter cloth required to filter the gases. The choice of a filtration velocity (or its equivalent, the *air-to-cloth ratio*, ACR, in cubic feet per minute of gas filtered per square foot of filter area) must take certain factors into consideration. The rule of thumb for ACR ratios for conventional fabric filter baghouses with woven cloth is 1.5–3.0 $\text{ft}^3/\text{min-ft}^2$. When dust is fine or loadings are high, it is recommended that the filtering velocity be <3 acfm/ft^2. Experience has also demonstrated that more cloth is required with increased temperature.

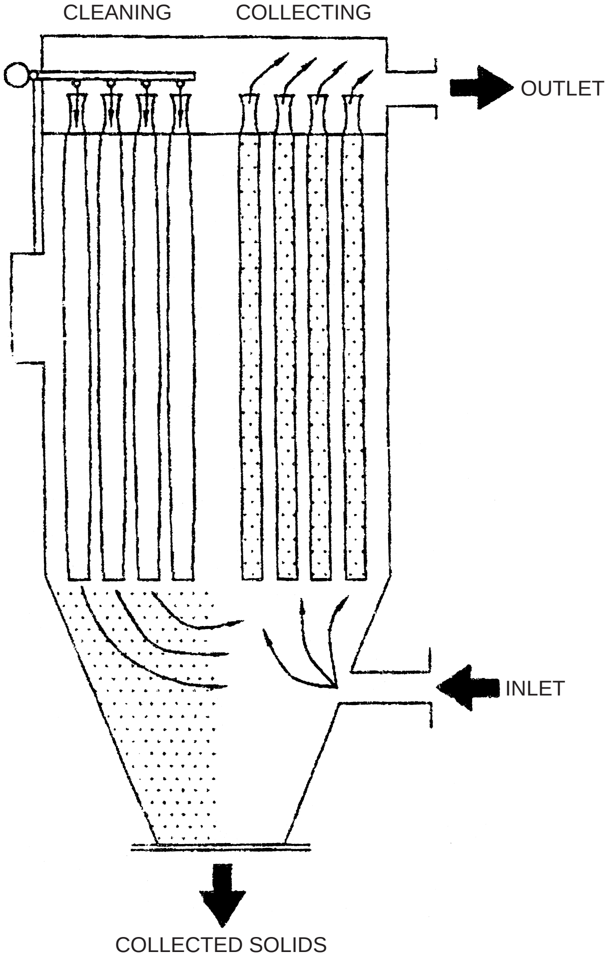


FIGURE 11 Baghouse. Simultaneous on-stream cleaning and collecting.

3. Wet–Dry Srubber System

a. Spray dryer absorber (SDA). This can be used for quench and neutralization of acid gases followed by an ESP or FF for PM removal. The method of operation of the SDA is relatively simple, requiring only two major equipment items—a spray dryer similar to those used in the chemical food-processing and mineral-preparation industries, and a fabric filter (baghouse) or ESP to collect the fly ash and entrained solids. In the SDA, the sorbent solution, or slurry, is atomized into the incoming flue gas stream to increase the liquid–gas interface and to promote the mass transfer of the HCl or SO_2 from the gas to the slurry droplets, where it is absorbed. The lime slurry solution may be mechanically atomized by either a rotary atomizer or by spray nozzles and injected into the combustion gas stream to be cleaned. The atomized lime droplets react with the HCl and SO_2 in the combustion gas stream, forming salts. The water droplet evaporates because of the 400–500°F temperature of the combustion gases, leaving a dry particulate, providing a sufficient drying time is allowed.

Optimum SDA performance is achieved through proper choice of L/D, droplet size, and residence time. The designed residence time for most SDAs is 10–12 sec at maximum design. Inspection and maintenance is critical to continued good operation of the scrubber.

After leaving the SDA, the solid-bearing gas passes through the fabric filter (or ESP), where the dry product is collected and a percentage of the unreacted alkali reacts with SO_2 for further removal. The cleaned gas is then discharged through the fabric-filter plenum to an induced draft (ID) fan and the stack. Figure 12 shows the SDA/FF system as part of the wet–dry–wet system described below.

Control of mercury compounds and other volatile metals, along with the trace organic PCDD/PCDF (dioxin/dibenzofuran) compounds, has been effectively accomplished in SD/FF systems by the addition of a relatively small quantity of activated carbon of approximately 0.003 lb/lb of . . . Results of tests conducted at Waste Technologies Industries, in East Liverpool, Ohio, have shown that activated carbon injection is extremely effective in reducing dioxin emissions in APC systems operating at higher temperatures than is typical in a rapid quench system (Roeck and Sigg, 1995). Comparisons have been made between HWIs that use rapid quench systems (wet scrubbers) versus those using waste heat boilers or SDA/FFs downstream of the high-temperature combustion zone.

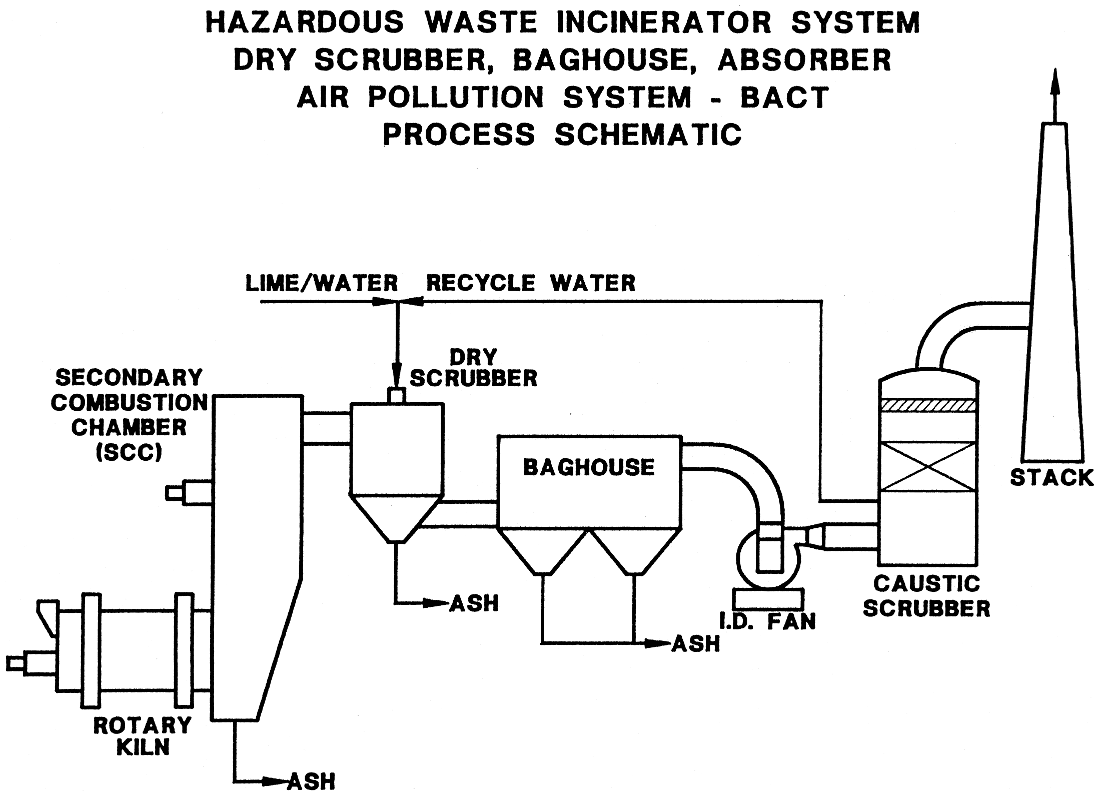


FIGURE 12 Wet–dry–wet scrubber system.

Systems that are able to quench the high-temperature combustion gases from 1000–1200°C (1832–2192°F) to adiabatic (75–90°C) in fractions of a second have emissions well below the levels established by the recent MACT standards (0.2 ng/dscm TEQ). This rapid quenching action freezes the reaction from the combustion zone preventing changes to the reaction products. Gradual cooling through the critical temperature zone of 300°C (572°F) with contact with PM, precursors, copper or iron, and Cl_2 results in the reformation of PCDD/PCDFs. Systems utilizing rapid quench have the lowest levels of PCDD/PCDF emissions. Most are well below the 0.2 ng TEQ/dscm-limit established by the MACT standards for new incinerators. This provides a major advantage to systems using rapid

TABLE VI Scrubber System Comparisons[a]

Parameter	Venturi	Packed bed	Wet ESP	IWS	SDP/ESP	SDA/FF
PM removal	Good	Poor	Excellent	Excellent	Excellent	Excellent
Metal removal	Good	Poor	Excellent	Excellent	Good	Excellent
Acid gas removal	Good	Excellent	Fair	Fair	Good	Excellent
Residue	Scrubber liquor	Scrubber liquor	Scrubber liquor	Scrubber liquor	Flyash	Flyash
Auxiliary equipment	Demister, liquid S&T	Demister, liquid S&T	Demister, liquid S&T	Liquid S&T	Lime/ash handling	Lime/ash handling
Turndown	2:1	5:1	2:1	2:1	5:1	10:1
Plume suppression	Difficult	Difficult	Fair	Fair	Easy	Easy
Pressure drop	High	Moderate	Low	Low	Low	Low
Capital Cost	Low	Low	Moderate	Moderate	High	High

[a] ESP, Electrostatic precipitator; IWS, ionizing wet scrubber; SDA, spray dryer absorber; FF, fabric filter.
[b] S&T, . . .

quench wet scrubbers immediately following the high-temperature combustion zones. This includes spray towers, Venturi scrubbers, and submerged exhaust quenchers (Santoleri, 1998).

b. Wet–dry–wet scrubber system. These systems are a combination of the wet–dry scrubber followed by a wet scrubber system (Fig. 12):

1. SDA for quench and neutralization, followed by PM control device as noted above, followed by a packed-bed scrubber tower (PBS) for quenching and final acid gas removal to meet emission standards for HCl and SO_2 not attained by the dry scrubber systems above.
2. SDA for quench and neutralization, followed by ESP or FF for PM removal, followed by low-pressure-drop quencher (ST or VS), followed by a PBS for final acid gas removal.

Table VI provides a comparison of the various scrubber systems described above.

SEE ALSO THE FOLLOWING ARTICLES

COMBUSTION • ENVIRONMENTAL TOXICOLOGY • POLLUTION, AIR • POLLUTION PREVENTION FROM CHEMICAL PROCESSES • RADIOACTIVE WASTE DISPOSAL • SOIL AND GROUNDWATER POLLUTION • WASTEWATER TREATMENT • WATER POLLUTION

BIBLIOGRAPHY

ASME. (1984). “Hazardous Waste Incineration: What Engineering Experts Say,” ASME CRTD, Vol. 32, American Society of Mechanical Engineers.

Bacon, G. H., Li, R., and Liang, K. Y. (1997). “Control of particulate and metal HAPs,” *Chem. Eng. Prog.*

CFR. (1991). “Boiling of hazardous wastes in boilers and industrial furnaces,” Code of Federal Regulations, Chapter 40, Parts 261, 265, 266 (August 21, 1991).

Eagleson, S. T., and Gilmand, K. (1995). “Complicance strategies for CETRED,” Presented at the 14th International Incineration Conference, Bellevue, WA.

Goosens, P. B., and Kint, W. (1995). “Incineration flue gas scrubbing using reverse jet scrubbers,” Presented at the 14th International Incineration Conference, Belleuve, WA.

Lundber, L. A., Meckel, B. D., and Mazrchese, N. J. (1995). “Converstion of an existing multiple hearth furnace to a fluidized bed,” Presented at the AWMA 88th Annual Meeting, San Antonio, TX, Paper No. 95-WP106.03.

Mycock, J., McKenna, J., and Theodore, L. (1995). “Handbook of Air Pollution Control Engineering and Technical Equipment,” CRC Press, Boca Raton, FL.

Oppelt, E. T. (1987). “Incineration of hazardous waste: A critical review,” *J. Air Pollution Control Assoc.* **37**(5).

RCRA. (1981/1991). Resource Conservation and Recovery Act, “Standards for Owners and Operators of Waste Facilities: Incinerators,” 40 CFR 264, RCRA 3004, Jan. 23, 1981, Rev. June 24, 1982; April 1983; Jan. 4, 1985; Nov. 24, 1985; Feb. 21, 1991.

Roeck, D. R., and Sigg, A. (1995). “Evaluation of carbon injection for controlling PCDD/PCDF emissions at WTI’s commercial hazardous waste incineration facility,” Presented at the 14th International Incineration Conference, Bellevue, WA.

Santoleri, J. J. (1995). “Hazardous Waste Incineration Course,” AWMA, Course No. WST-258, Kansas City, MO (March 27, 1995).

Santoleri, J. J. (1998). “The impact of thermal oxidation quench systems on dioxan formation,” Presented at the 17th Annual International Conference on Incineration and Thermal Technologies, Salt Lake City, UT.

Pollution Control

Arthur C. Stern
University of North Carolina at Chapel Hill

GLOSSARY

Acid rain (acidic deposition) Rain or snow or dry particulate matter with a high enough concentration of nitrogen and sulfur oxides to be acid (caused by absorption of these oxides from plumes from terrestrial sources and subsequent chemical reactions in the troposphere.

Ambient air Air at a receptor or a monitoring site; generally, the air "outdoors" as opposed to the air inside homes, buildings, or factories.

Depletion of the ozone layer by chlorofluorohydrocarbons and related chemicals Physicochemical reactions in the ozone layer of the earth's stratosphere, decreasing its effectiveness in screening out ultraviolet solar radiation reaching the earth.

Global warming (greenhouse effect) Hypothesis that buildup of carbon dioxide in the troposphere from combustion of fossil fuels and other gases with similar properties will allow solar radiation to reach the earth but impede radiation of infrared energy from the earth to the troposphere, thus increasing the temperature of the earth and its immediate atmosphere in this century.

Monitoring Continuous measurement of the pollutant concentration of ambient air or a source.

PM_{10} Particulate matter with an aerodynamic diameter less than or equal to a nominal 10 μm.

Primary and secondary pollutants Pollutants in the from in which they are emitted from sources. Primary pollutants may or may not react in the atmosphere with other pollutants or atmospheric constituents to form secondary or higher order pollutants. Primary pollutants are not necessarily chemically or physically simple. They can be emitted from their sources in quite complex chemical and physical form.

Rainout Absorption or adsorption of a gas or vapor by the formation of a raindrop or snowflake in a cloud; also, the inclusion of a particle as a nucleus for a raindrop or snowflake and subsequent deposition of the drop or flake onto a surface.

Receptors Humans, animals, aquatic life, vegetation, materials, or structures adversely affected by pollutants.

Sampling Extracting a quantity of ambient air or the effluent from a source for measurement and analysis.

Sedimentation Deposition of previously airborne suspended particulate matter onto surfaces.

Sources Places from which primary pollutants are emitted to the atmosphere. Sources can be categorized in several ways. One categorization is stationary sources, mobile sources, fugitive sources, and indirect sources. Others are primary and secondary sources; natural and anthropogenic sources; domestic, commercial, and industrial sources; and, finally, point, line, and area sources.

Suspended particulate matter Measured in terms of the number of particles per unit volume of air, their surface area, or their volume. When the density of the particles is known, volume can be converted to mass, and vice versa (see Fig. 1 and Table IV). Chemical and physical processes in the atmosphere cause fine particles to grow by condensation and coagulation mechanisms to become coarse particles, which leave the atmosphere by sedimentation, rainout, and washout.

Toxic air pollutants Airborne elements and compounds, which, on long enough exposure, are toxic to vegetable, animal, or human life.

Unpolluted air Concept of what the air would be like if there were no anthropogenic sources. This is no more than a concept, because the act of measuring air requires the presence of people and their equipment even at remote locations at sea, the poles, deserts, or mountains, where the air can be best described as dilute polluted air (see Tables I, II, and III). The suspended particulate matter concentration of unpolluted air is quite variable depending on whether it has been recently loaded by natural events such as volcanic eruptions or recently cleansed by rain- or snowfall.

Washout Removal of a particle or a gas or vapor molecule by impaction by a falling raindrop or snowflake and the subsequent deposition of the drop or flake onto a surface.

AIR POLLUTION control is the reduction or elimination of the emission of pollutants from anthropogenic sources into the atmosphere. (There are no means of reducing or eliminating the emission of natural pollutants into the atmosphere.) Control is treated here as engineering control and regulatory control. Engineering control is the application of engineering principles to the reduction or elimination of emission of pollutants from sources to the atmosphere. (Engineering control may be applied whether or not it is required by regulatory control.) Regulatory control is the imposition of governmental regulations on emissions from anthropogenic sources to the atmosphere.

I. AIR POLLUTANTS REQUIRING CONTROL

There are thousands of gases, vapors, and kinds of suspended particulate matter in the air—some from natural sources, others from anthropogenic sources. Some molecules are from local sources; others may have traveled around the world to arrive at a receptor or an air-sampling location. Most are present in extremely low concentration, so low that we do not have the analytical capability of measuring their concentration in air. Others can be measured only with extremely sophisticated and expensive technology, which is brought into use only for special studies, not for routine monitoring.

Tables I and II give the gaseous composition of unpolluted air on a dry and a wet (31,200 ppm water vapor)

TABLE I Gaseous Composition of Unpolluted Air (Dry Basis)[a]

Constituent	ppm (vol)	μg/m^3
Nitrogen	780,900	8.95×10^8
Oxygen	209,400	2.74×10^8
Water	—	—
Argon	9,300	1.52×10^7
Carbon dioxide	315	5.67×10^5
Neon	18	1.49×10^4
Helium	5.2	8.50×10^2
Methane	1.0–1.2	$6.56–7.87 \times 10^2$
Krypton	1.0	3.43×10^3
Nitrous oxide	0.5	9.00×10^2
Hydrogen	0.5	4.13×10^1
Xenon	0.08	4.29×10^2
Organic vapors	~0.02	—

[a] From Stern, A. C., Boubel, R. W., Turner, R. B., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.

TABLE II Gaseous Composition of Unpolluted Air (Wet Basis)[a]

Constituent	ppm (vol)	μg/m^3
Nitrogen	756,500	8.67×10^8
Oxygen	202,900	2.65×10^8
Water	31,200	2.30×10^7
Argon	9,000	1.47×10^7
Carbon dioxide	305	5.49×10^5
Neon	17.4	1.44×10^4
Helium	5.0	8.25×10^2
Methane	0.97–1.16	$6.35–7.63 \times 10^2$
Krypton	0.97	3.32×10^3
Nitrous oxide	0.49	8.73×10^2
Hydrogen	0.49	4.00×10^1
Xenon	0.08	4.17×10^2
Organic vapors	~0.02	—

[a] From Stern, A. C., Boubel, R. W., Turner, R. B., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.

basis, respectively. The important things to note are that (1) the "organic vapors" are of natural origin, and (2) the concentrations of pollutants we monitor in polluted air are too low in unpolluted air to appear in either table. More detailed information on them is given in Table III. Routine monitoring is limited to only a few pollutants, except where the possibility of accidental release of a specific pollutant requires monitoring for that pollutant. Damage to vegetation, a measure of air pollution, has been reported since the Middle Ages. However, except for a few published nineteenth century measurements of rainfall acidity, air pollution measurement and monitoring is of twentieth century initiation.

A. Particulate Matter

Particulate matter exists in the air in a wide range of particle sizes (Table IV) and concentrations (Fig. 1). The earliest monitoring was the routine gravimetric measurement of "sootfall" (i.e., the deposition of material by sedimentation, rainout, or washout) into open-top receptacles (buckets), sometimes also analyzed for combustible matter and ash. The next form of monitoring was the measurement of the discoloration of paper through which a known volume of air was filtered. This was known as "smoke shade," measured originally by visual comparison of the discoloration of the filter with an artificially prepared chart of a scale of gray from white to black. Still later, visual comparison was supplanted by the use of photocell meters to indicate electrically either the reflectance of the filter or its light transmittance, in coefficient of haze (COH) units. Even later, calibrations of reflectance units in gravimetric units (μg/m^3) were published. Although these calibrations came into widespread use, the reporting of the data in gravimetric units resulted in many suspect data because of the nonapplicability of the published "standard" calibrations to the actual color scale of the local pollution being reported.

True routine gravimetric monitoring of particulate matter began in the 1950s by weighing of previously tared glass fiber filters with measured 24-h air flow through them provided by vacuum-cleaner-type suction blowers. These samplers came to be known as high-volume samplers. After initial weighing, these filters were cut into sections for separate analysis of organic-solvent-soluble material by extraction in Soxhlet extractors, for analysis of several water-soluble ions, and in some cases, for analysis of metal ions. However, the metals in the glass-fiber filter paper interfered with the analysis of metallic compounds deposited on the filter. Therefore, for such analyses, organic filter media, including paper, had to be used. Another technology for the measurement of filter deposits is the use of β gauges.

Although filtration and sedimentation are the most common means of measuring and monitoring particulate pollutants, the absorption and scattering of light provide a means for routine monitoring of the number of particles in the air and their surface area by nephelometers, transmissiometers, and particle counting devices (see also Section II.B). These methods allow the measurement of particle size distribution of airborne particulate matter, but there are no routine methods for chemical analysis of particulate matter without their prior capture by filtration or sedimentation.

Airborne pollen, spores, and bacteria are measured by sedimentation onto either glass microscope slides or petri dishes coated with nutrient medium. The slides are placed in a microscope for indentification and counting of the deposited material. The same procedure can be used for identification, counting, and size frequency distribution of inorganic particles. Except for pollen counting, these are not routine procedures. Radioactive particles are usually measured for their radioactivity after collection by sedimentation (i.e., fallout) or filtration and, without prior collection, by specialized instruments. Such measurements are routinely performed by military and nuclear materials installations but are not commonly done by air pollution control agencies.

B. Gases and Vapors

The earliest monitoring of pollutant gases and vapors was the observation of vegetation damage and analysis of rain. Chemical analysis of airborne pollutants, a twentieth century development, is of two kinds: static and dynamic. In static sampling and analysis, a reactive chemical is exposed for a specific time. The extent of reaction is determined by a color change (e.g., of impregnated filter paper) or by chemical analysis (e.g., of lead sulfide). Since there is no measurement of air flow past or through the sampler, static samplers cannot measure in terms of mass of pollutant for unit mass or volume of air. However, static samplers can discriminate areas of relative intensity of exposure and are useful where there is no source of energy to operate dynamic samplers and recorders.

The simplest dynamic sampler consists of a suction device, usually an electric-motor-driven blower (but sometimes a hydraulic, compressed air, or gas suction device) to pull air from the atmosphere through a sampler, and a flowmeter to measure the rate of air flow through the sampler. The sampler can be an absorber to remove and retain the pollutant being measured from the air flowing through it, with subsequent chemical or physical measurement of the absorbed material (Table V). Although Table V lists tetrachloromercurate for absorbing sulfur dioxide, the absorbent most commonly used by the World

TABLE III Summary of Sources, Concentrations, and Important Scavenging Processes of Atmospheric Trace Gases[a]

Contaminant	Major pollutant sources	Important natural sources	Estimated annual emissions (Tg/year)[b]		Atmospheric background concentrations	Estimated atmospheric residence time	Removal reactions and sinks	Remarks
			Pollutants	Natural				
SO_2	Fossil fuel combustion	Volcanoes, reactions of biogenic S emissions	212[c]	20[d]	About 0.1 ppb[e]	1–4 days	Oxidation to sulfate by photochemical reactions or in liquid droplets	High reaction rates in summer due to photochemical processes
H_2S and organic sulfides[f]	Chemical processes, sewage treatment	Volcanoes, biogenic processes in soil and water	3 (as sulfur)	84[g] (as sulfur)	H_2S: 0.05–0.1 ppb COS: 0.5 ppb[h] CS_2:0.05 ppb[h]	H_2S: 1–2 days COS: 1–2 year[h]	Oxidation to SO_2 and SO_4	Atmospheric data are very incomplete; COS residence time may be 20 year[d]
CO	Auto exhaust, general combustion	Forest fires, photochemical reactions	700[i]	2100[i]	0.1–0.2 ppm (N. Hemisphere) 0.04–0.06 ppm (S. Hemisphere)	1–3 months	Photochemical reactions with CH_4 and OH	No long-term changes in the atmosphere have been detected
NO, NO_2	Combustion	Biogenic processes in soil, lightning	75[j] (as NO_2)	180[k] (as NO_2)	∼0.1 ppb[l]	2–5 days	Oxidation to nitrate	Natural processes are mostly estimated; background concentrations are in doubt, but may be as low as 0.01 ppb
NH_3	Waste treatment, combustion	Biogenic processes in soil	6[m]	260[k]	∼10 ppm[k]	1–7 days	Reaction with SO_2 to form $(NH_4)_2SO_4$ aerosol	Atmospheric measurements are very sparse
N_2O	Small amounts from combustion	Biogenic processes in soil	3[n]	340°	300 ppb	20–100 year	Photochemical in stratosphere	Some estimates place natural source at 100 Tg or less[n]
CH_4	Combustion, natural gas leakage	Biogenic processes in soil and water	160[p]	1050[p]	1.5 ppm	8 year[g]	Reaction with OH to form CO	Pollutant source includes 60 Tg/year from biomass burning.
Isoprene and terpenes	None	Biogenic plant emissions	None	830[i]	0	1–2 h	Photochemical reactions with OH and O_3	Not found in ambient atmosphere away from source regions

Total non-CH_4 hydrocarbons	Combustion	Biogenic processes in soil and vegetation	40^r	$2 \times 10^{4\,s}$	0–1 μg/m^3 for C_2's	Hours to a few days	Photochemical reactions with NO and O_3	Concentration given for C_2's in rural atmosphere
CO^2	Combustion	Biological processes	$22{,}000^s$	$10^{6\,o}$	345 ppm (1981)	2–4 year	Biogenic processes, photosynthesis, absorption in oceans	Forest destruction and changes in the earth's biomass may add 20–30 × 10^3 Tg CO_2/year to atmosphere^t
CH_3Cl	Combustion	Oceanic biological processes	2^u	$4–6^{s,u}$	600 ppt^{u,v}	1–2 year^u	Stratospheric reactions	Photochemical reactions in stratosphere may impact on O_3 layer
HCl, Cl_2	Combustion, Cl manufacturing	Atmospheric reactions of NaCl, volcanoes	4^w	$100–200^x$	~0.5 ppb^w	About 1 week	Precipitation	Volcanoes may release 10–20 Tg Cl/year^y

[a] From Elmer Robinson, National Oceanic and Atmospheric Administration, Mauna Loa GMCC Observatory, Hilo, Hawaii.

[b] Tg/year is 10^{12} g/year, or 10^6 metric tons/year.

[c] Based on 1978 global fuel usage and estimated sulfur contents.

[d] Major reference is Cadle, R. D. (1980). *Rev. Geophys. Space Phys.* **18,** 746–752.

[e] Maroulis, P. J., Torres, A. L., Goldberg, A. B., and Bandy, A. R. (1980). *J. Geophys. Res.* **85,** 7345–7349.

[f] Includes COS, CS_2, $(CH_3)_2S$, $(CH_3)_2S_2$, CH_3, SH, etc.

[g] Adapted from Adams, D. F., Farwell, S. O., Robinson, E., and Pack, M. R. (1980). "Biogenic Sulfur Emissions in the SURE Region," Final report by Washington State University for Electric Power Research Institute, EPRI Report No. EA-1516.

[h] Torres, A. L., Maroulis, P. J., Goldberg, A. B., and Bandy, A. R. (1980). *J. Geophys. Res.* **85,** 7357–7360.

[i] Zimmerman, P. R., Chatfield, R. B., Fishman, J., Crutzen, P. J., and Hanst, P. L. (1978). *Geophys. Res. Lett.* **5,** 679–682.

[j] Based on 1978 global combustion estimates.

[k] Galbally, I. E. (1975). *Tellus* **27,** 67–70.

[l] Approximate value combining values given in several references.

[m] Söderlund, R., and Svensson, B. H. (1976). *In* SCOPE Report 7, Swedish National Science Research Council, Stockholm.

[n] 1978 fuel usage figures applied to the following references: Weiss, R. F., and Craig, H. (1976). *Greophys. Res. Lett.* **3,** 751–753; and Pierotti, D., and Rasmussen, R. A. (1976). *Geophys. Res. Lett.* **3,** 265–267.

[o] Robinson, E., and Robbins, R. C. (1972). *In* "Air Pollution Control" (W. Strauss, ed.), Pt. 2, pp. 1–93. Wiley, New York.

[p] Sheppard, J. C., Westberg, H., Hopper, J. F., and Ganesan, K. (1982). *J. Geophys. Res.* **87,** 1305–1312.

[q] Heidt, L. E., Krasnec, J. P., Lueb, R. A., Pollock, W. H., Henry, B. E., and Crutzen, P. J. (1980). *J. Geophys. Res.* **85,** 7329–7336.

[r] Reference in footnote *o* tabulation updated to approximate 1978 emissions.

[s] Graedel, R. E. (1979). *J. Geophys. Res.* **84,** 273–286.

[t] Woodwell, G. M., Whittaker, R. H., Reiners, W. A., Likens, G. E., Delwiche, C. C., and Botkin, D. B. (1978). *Science* **199,** 141–146.

[u] Rasmussen, R. A., Rasmussen, L. E., Khalil, M. A. K., and Dalluge, R. W. (1980). *J. Geophys. Res.* **85,** 7350–7356.

[v] Robinson, E., Rasmussen, R. A., Krasnec, J., Pierotti, D., and Jakubovic, M. (1997).*Atmos. Environ.* **11,** 213–215.

[w] Ryan, J. A., and Mukherjee, N. R. (1975). *Rev. Geophys. Space Phys.* **13,** 650–658.

[x] Based on estimated reaction of NaCl to form Cl_2.

[y] Cadle, R. D. (1980). *Rev. Geophys. Space Phys.* **18,** 746–752.

TABLE IV Particle Size Ranges and Methods of Measurement[a]

Particle size range (μm)	Ions	Nuclei	Visibility	Suspended or settleable; nonairborne	Dispersion aerosol	Condensation aerosol	Pollen and spores	Sedimentation, diffusion, and settling
10^{-4} to 10^{-3}	Small	—	—	Suspended	—	Gas molecules	—	Diffusion
10^{-3} to 10^{-2}	Intermediate and large	Aitken nuclei	Electron microscope	Suspended	—	Vapor molecules	—	Diffusion
10^{-2} to 10^{-1}	Large	Aitken and condensation nuclei	Electron microscope	Suspended	—	Fume–mist	—	Diffusion
Air pollution 10^{-1} to 10^{-0}	—	Condensation nuclei	Microscope: electron and optical	Suspended	Dust–mist	Fume–mist	—	Diffusion and sedimentation
10^{0} to 10^{1}	—	—	Microscope: optical	Suspended and settleable	Dust–mist	Fume–mist	—	Sedimentation
10^{1} to 10^{2}	—	—	Eye; sieves	Settleable	Dust–mist	Mist–fog	Pollen and spores	
10^{2} to 10^{3}	—	—	Eye; sieves	Nonairborne	Dust–spray	Drizzle–rain	—	Sedimentation
10^{3} to 10^{4}	—	—	Eye; sieves	Nonairborne	Sand–rocks	Rain	—	Sedimentation

[a] From Stern, A. C., Boubel, R. W., Turner, B. D., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.

Health Organization–United Nations Environmental Program worldwide air-sampling network is hydrogen peroxide with subsequent analysis of the absorbed sulfur dioxide. A sampler that absorbs a pollutant for a specific time period gives the average pollutant concentration for that time period (e.g., 2, 8, or 24 h). This is called the averaging time of the measurement. For measurements with very low averaging times (i.e., 1 min or less), there are continuous monitors and recorders. These are commercially available for only a limited number of pollutant gases and vapors: carbon monoxide, ozone, nitrogen dioxide, sulfur dioxide, and nonmethane volatile hydrocarbons. Air-monitoring stations that continuously monitor the atmosphere are therefore limited to the operation of one or more of these instruments plus one of the means of monitoring atmospheric particulate matter discussed in Section I.A. Because of their initial and operating costs, stations that continuously monitor four or five pollutants are relatively few in number and are located mainly in big cities.

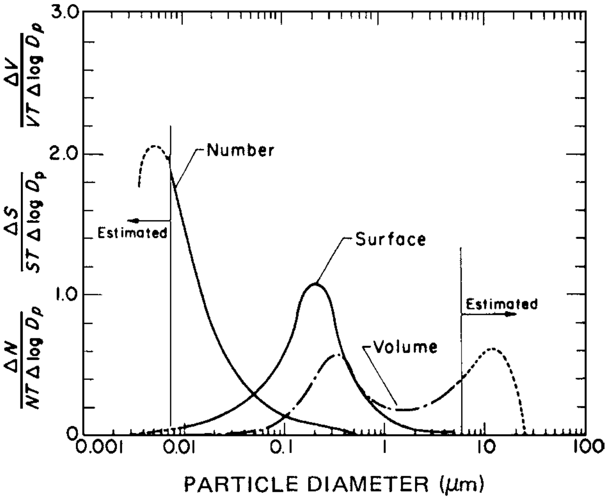


FIGURE 1 Grand average number N, surface area S, and volume V distributions of Los Angeles smog. The linear ordinate normalized by total number NT, area ST, or volume VT is used so that the apparent area under the curves is proportional to the quantity in that size range. [From Corn, M. (1976). *In* "Air Pollution," 3rd ed. (A. C. Stern, ed.), Vol. 1, p. 123, Academic Press, New York.]

II. REASONS FOR AIR POLLUTION CONTROL

Air pollution is controlled because of a number of adverse effects on materials and structures; the atmosphere, particularly visibility reduction; vegetation, animals, and aquatic systems; and human health (Fig. 2).

A. Adverse Effects on Materials and Structures

Air pollution causes corrosion and deterioration of ferrous and nonferrous metals, stone, concrete, ceramics and glass, and a host of organic materials: paint, paper, photographic emulsions, textiles, textile dyes, and leather. It also causes deterioration of cultural property: architectural monuments, museums, libraries and archives, and stained

TABLE V Collection of Gases by Absorption[a]

Gas	Sampler	Sorption medium	Air flow (liters/m)	Minimum sample (liters)	Collection efficiency	Analysis	Interferences
Ammonia	Midget impinger	25 ml 0.1 *N* sulfuric acid	1–3	10		Nessler reagent	—
Hydrochloric acid	Fritted bubbler	0.005 *N* Sodium hydroxide	10	100	95	Titration with 0.01 *N* silver nitrate	Other chlorides
Hydrogen sulfide	Midget impinger	15 ml 5% cadmium sulfate	1–2	20	95	Add 0.05 *N* iodine, 6 *N* sulfuric acid, backtitrate 0.01 *N* sodium thiosulfate	Mercaptans, carbon disulfide, organic sulfur compounds
Nitrogen dioxide	Fritted bubbler (60–70 μm pore size)	20–30 ml Slatzman reagent[b]	0.4	Sample until color appears; probably 10 ml of air	94–99	Reacts with absorbing solution	Ozone in fivefold excess peroxyacyl nitrate
Ozone	Midget impinger	1% potassium iodide in 1 *N* potassium hydroxide	1	25	95	Measures color of iodine liberated	Other oxidizing agents
Sulfur dioxide	Midget impinger, fritted rubber	10 ml sodium tetrachloromercurate	2–3	2	99	Reaction of dichloro sulfitomercurate and formaldehyde-*p*-rosaniline	Nitrogen dioxide[c] hydrogen sulfide[d]

[a] From Pagnotto, L. D., and Keenan, R. G. (1973). *In* "The Industrial Environment—Its Evaluation and Control," pp. 167–179. United States Department of Health, Education, and Welfare, U.S. Government Printing Office, Washington, D.C.

[b] Five-grams sulfanilic: 140-ml glacial acetic acid: 20-ml 0.1% aqueous *N*-(1-naphthyl)ethylenediamine.

[c] Add sulfamic acid after sampling.

[d] Filter or centrifuge any precipitate.

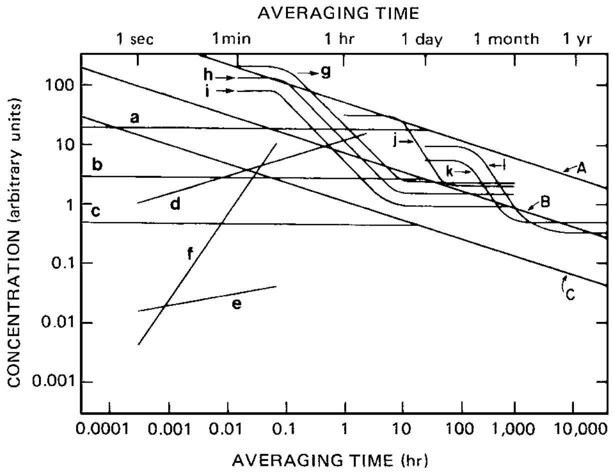


FIGURE 2 Adverse responses to various pollution levels. A, Pollution level high: B, pollution level moderate: C, pollution level low. a, One-mile visibility: b, 10-mile visibility; c, 100-mile visibility: d, nasopharyngeal and eye irritation; e, material response to acid droplets: f, odor response; g, 100% crop loss: h, rubber cracking i, zero crop loss; j, acute respiratory disease; k, livestock flurosis; l, human—chronic respiratory disease, vegetation—decreased yield of fruit and forest, material—corrosion, deterioration, soiling. [From Stern, A. C., Boubel, R. W., Turner, D. B., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.]

glass. By attacking metals, stone, and concrete, it threatens the structural stability of buildings and bridges built of these materials and requires the increased cost of applying paints and protective coatings to preserve them. Other costs are those of cleaning and laundering to remove soil and that of the deterioration of these materials by repeated cleaning.

B. Visibility Reduction

Diminution of visibility is valued differently under different circumstances and is related to the ready availability of a means of amelioration. The reduction of visibility on a highway due to an occasional vehicle with a smoking exhaust represents a failure to enforce laws and regulations prohibiting such emissions and requires that such vehicles be taken off the road until the engine deficiency has been rectified. The coal smoke that made Pittsburgh a smoky city with reduced visibility from point to point in the city was ameliorated when the domestic heating plants in the city switched to the burning of natural gas. The pea-soup fogs of London disappeared with the advent of smokeless zones, in which the burning of coal in the fireplaces of each room in a building was made illegal.

The visibility problem that has resisted solution is that caused by photochemical smog, particularly in the national parks of the western United States. Environmentalists have called this phenomenon “the graying of the West.” Even though the absolute diminution of visibility in the Grand Canyon may be less than that experienced in eastern U.S. cities, a much higher aesthetic premium is placed on maintaining the scenic value of such important landmarks.

C. Adverse Effects on Vegetation and on Animal and Marine Life

For commerical crops a distinction is made between injury and damage. Even though injury can be shown to have occurred, a crop may still be harvestable and salable. Injuries that cause the crop to have a lower sales value, to have required excess costs to grow, or to have a lower yield are all costs related to air pollution. If the injury is so great that it is uneconomical to harvest the crop and it has to be ploughed under, the loss is complete. The principal mitigative measures are to (1) decrease pollution levels and (2) breed pollutant-resistant cultivars.

The pollutants of concern to vegetation damage are, in decreasing order of importance, ozone, sulfur dioxide, nitrogen dioxide, fluoride, peroxyacyl nitrate (PAN), hydrogen chloride, and hydrogen sulfide. In specific areas and for specific crops, this order of importance can vary. The pollutant of principal concern to animals is fluoride. The pollutants of principal concern to aquatic life are the acidic compounds of sulfur and nitrogen, which can change the pH of lakes and other water bodies. It is still an open question as to whether these same sources of acidic deposition are the principal cause of damage to forests, or whether they are secondary to other causative agents. Some trees are particularly susceptible to sulfur dioxide damage, others to ozone and PAN damage, and still others to fluroide damage. The annual vegetation loss due to air pollution damage has been estimated to be many billions of dollars in the United States alone.

D. Adverse Effects on Human Health

Air pollution effects on human health are of two types: (1) those caused by diminution of solar energy and (2) those caused by inhalation of pollutants. Humans have adapted genetically to living in an environment with the amount of ultraviolet (UV) energy received at the earth’s surface with a minimum of its absorption by tropospheric pollutants and natural regulation by the ozone layer above the troposphere of the amount of stratospheric UV reaching the troposphere. Too little UV energy reaching the earth causes rickets in children, which we attempt to compensate for by adding artificial irradiated ergesterol (vitamin D) to milk or other foods. Too much UV light can cause increased erethema (sunburn) and, potentially, skin cancer.

Pollution in the troposphere also causes a diminution in the amount of light we use for vision that reaches ground level. Decreased solar illumination at ground level, on the one hand, is psychologically depressing and, on the other, requires increased artificial illumination, which is costly and is a cause of still more pollution, if it is provided by means requiring the combustion of fuels to produce light or electrical power.

Some inhaled pollutants, particularly tobacco smoke inhaled by cigarette smokers, can cause lung cancer. Nonsmokers are at risk for lung cancer only in special situations, such as those involving exposure to asbestos fibers, working in mines with high radon levels, or working where there is occupational exposure to high levels of airborne carcinogens and chemical compounds, such as benzo[*a*]pyrene. There is no evidence that persons who breathe the ambient air of urban or rural areas are at risk for lung cancer. If a vast increase in the use of diesel-engine-driven vehicles were to be permitted without adequate exhaust emission control, there might be some grounds for further concern about lung cancer. However, this does not seem likely to occur in the industrialized nations of the world, which presumably will require adequate controls of potentially hazardous emissions from new diesel vehicles.

The other concerns to human health due to respiration are (1) the inhalation of elevated levels of carbon monoxide or other potentially lethal gases or vapors and (2) respiratory disease. Most cases of inhalation of elevated levels of potentially lethal gases or vapors are accidental. One of the worst of these was the accidental release in Bhopal, India, in 1984 of methyl isocyanate, which killed more than 2000 people. Other, less devastating releases of a variety of toxic gases and vapors occur annually due to accidents at manufacturing plants, spills due to accidents to tank cars in railroad collisions and derailments, highway accidents to tank trucks, and accidents at marine vessels and aircraft. Occasionally releases occur at stationary storage facilities, pipelines, and waste-disposal sites. Because such releases are accidental, the only recourse consists of employing regular operating and maintenance schedules, training employees for response to accidental situations, and preplanning for response to such situations. The greater the potential hazard, the greater is the need for preplanning, for example, the elaborate precautions taken during rocket blastoffs at launching sites such as Cape Canaveral.

Exposure to elevated carbon monoxide levels is less dramatic but more frequent. It can occur wherever there is combustion of fuel. The adverse effects are proportional to the elevation of the carbon monoxide level and the duration of exposure. Exposure to carbon monoxide from vehicular exhaust can occur at street level and highways with considerable traffic (Table VI). At relatively low carbon monoxide exposures, there is concern for the psychomotor response of people and the increased risk of angina pectoris patients to attack.

Our concerns with the induction and exacerbation of respiratory diseases encompass bronchitis, emphysema, asthma, and other nonoccupational diseases. (Pneumoconiosis and silicosis are considered diseases due to occupational exposure, rather than to inhalation of ambient air.) The pollutants principally indicted are ozone, sulfur dioxide, nitrogen dioxide, suspended particulate matter, and mixtures of these pollutants. Concern over the effects of suspended particulate matter on human health is particularly focused on what are called “respirable” or “inhalable” particles, usually defined as particles less than 10 nm in size. Lead dust is the exception to the rule that particles must be inhaled to be hazardous to health. Lead is toxic whether it is inhaled or ingested. In fact, the average person ingests much more lead than he or she inhales. This has made the setting of limits for airborne lead particulate matter quite complex (Table VII). With increasingly stringent limits being set on the amount of lead in gasoline, the amount of airborne lead particulate matter should decrease in the future.

TABLE VI U.S. Ambient Air Quality Criteria for Carbon Monoxide[a]

Level of carboxyhemoglobin in blood (%)	Human symptoms associated with this level
80	Death
60	Loss of consciousness; death if exposure is continued
40	Collapse on exercise; confusion
30	Headache, fatigue: judgment disturbed
20	Cardiovascular damage; electrocardiographic abnormalities
5	Decline (linear with increasing carboxyhemoglobin level) in maximal oxygen uptake of healthy young men undergoing strenuous exercise; decrements in visual perception, manual dexterity, and performance of complex sensorimotor tasks
4	Decrements in vigilance (i.e., ability to detect small changes in one’s environment that occur at unpredictable times); decreased exercise performance in both healthy persons and those with chronic obstructive pulmonary disease
2.5	Aggravation of cardiovascular disease (i.e., decreased exercise capacity in patients with angina pectoris, intermittent claudication, or peripheral arteriosclerosis)

[a] Data from Henderson, Y., and Haggard, H. W. (1927). “Noxious Gases,” Chemical Catalog Co., New York; and U.S. Environmental Protection Agency (1979). “Air Quality Criteria for Carbon Monoxide” (preprint), EPA-600/8-79-022, special series, Research Triangle Park, North Carolina.

TABLE VII U.S. Ambient Air Quality Criteria for Lead[a]

Level of lead in blood (μg/dl)	Human symptoms associated with this level
300	Coma; convulsions; death
100	Acute lead intoxication; deleterious irreversible effects; gastrointestinal and encephalopathic symptoms
80	Permanent severe neurological damage in children
50	Peripheral neuropathies; anemia in adults; central nervous system defects in children
40	Anemia in children: coproporphyrin elevation; increased urinary aminolevulinic acid excretion
30	Maximum safe level for an individual child
15[b]	Geometric mean maximum safe level for entire population of children: erythrocyte protoporphyrin elevation.

[a] Data from “Air Quality Criteria for Lead,” EPA 600/8-77-017, Dec. 1977, special series. United States Environmental Protection Agency, Research Triangle Park; North carolina; National ambient air quality standards for lead, final rules, *Fed. Reg.* (Part IV), Oct. 5, 1978, pp. 46246–46277.

[b] After attributing a blood lead level of 12 μg/dl to nonairborne ingested sources (i.e., food, water, soil, paint), this leaves 3 μg as the maximum safe contribution from inhaled air. Assuming airborne lead (μg Pb/m air) as 0.5 blood lead, this yields a safe air quality level of 1.5 μg Pb/μg air, which was promulgated as the U.S. national ambient air quality standard for lead, as the maximum arithmetic mean averaged over a calendar quarter.

III. AIR POLLUTION CONTROL

A. History

Little was written about air pollution before the nineteenth century, but there is pictorial evidence that some of the processes that produce pollution now have done so since ancient times. In times past there were fewer people and these processes took place on a smaller scale, but the people most likely lived nearer the sources. Ancient nomadic people periodically moved away from the stench of the animal, vegetable, and human wastes they generated. Indoor pollution levels were high when cooking and heating took place inside dwellings without chimneys.

The principal industries associated with the production of air pollution in the centuries preceding the Industrial

TABLE VIII Categories of the Air Pollution Problem[a]

Category	Vertical scale	Temporal scale	Scale of organization required for resolution
Local	Height of stacks	Hours	Municipal
Urban	Lowest mile	Days	County or multicounty
Regional	Troposphere	Months	State, provincial, or national
Continental	Stratosphere	Years	National or international
Global	Atmosphere	Decades	International

[a] From Stern, A. C., Boubel, R. W., Turner, B. D., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.

Revolution were metallurgy, ceramics, and the preservation of animal products.

The nineteenth century was associated with an increase in the scale of industrial sources, the advent of coal-fired industrial boiler furnaces, the development of pollution-producing metallurgical and chemical processes, and, in the home, the increasing substitution of coal for wood as fuel.

Before the nineteenth century, air pollution control consisted of regulations prohibiting "noxious trades" (i.e., those that produced offensive odors), and the burning of soft coal, within municipal boundaries. In the nineteenth century, it consisted of placing limitations on the emission of dense smoke from factories and locomotives and on the emission of acid gases from specific industrial sources.

In the twentieth century, dense smoke emission was gradually brought under control, and emphasis shifted to control of photochemical "smog" (i.e., control of hydrocarbon and oxide of nitrogen emissions from motor vehicles and stationary sources), of particulate matter and sulfur oxide emissions, of "toxic" and hazardous emissions (e.g., lead, mercury, beryllium, vinyl chloride, and asbestos), and "acidic deposition."

The categories of the air pollution problem that are recognized today are listed in Table VIII.

B. Philosophy

Air pollution control involves long-term strategies and short-term tactics. Long-term strategies are aimed at achieving improvement of air quality 5, 10, or 15 years hence. Short-term tactics are aimed at preventing the buildup of dangerous pollution levels, say, next week, when there are meteorological forecasts of an impending stagnation.

Air pollution control has two major thrusts: engineering control and regulatory control.

IV. ENGINEERING CONTROL

A. Engineering Control Methods

Engineering methods of control (Table IX) are best applied in the original design of a process or plant, but they can also be applied to an ongoing process or an existing plant.

TABLE IX Engineering Control Methods[a]

I. Applicable to all emissions
  A. Decrease or eliminate production of emission
    1. Change specification of product
    2. Change design of product
    3. Change process temperature, pressure, or cycle
    4. Change specification of materials
    5. Change the product
  B. Confine the emissions
    1. Enclose the source of emissions
    2. Capture the emissions in an industrial exhaust system
    3. Prevent drafts
  C. Separate the contaminant from effluent gas stream
    Scrub with liquid
II. Applicable specifically to particulate matter emissions
  A. Decrease or eliminate particulate matter production
    1. Change to process that does not require blasting, blending, buffing, calcining, chipping, crushing, drilling, drying, grinding, milling, polishing, pulverizing, sanding, sawing, spraying, tumbling
    2. Change from solid to liquid or gaseous material
    3. Change from dry to wet solid material
    4. Change particle size of solid material
    5. Change to process that does not require particulate material
  B. Separate the contaminant from effluent gas stream
    1. Gravity separator
    2. Centrifugal separator
    3. Filter
    4. Electrostatic precipitator
III. Applicable specifically to gaseous emissions
  A. Decrease or eliminate gas or vapor production
    1. Change to process that does not require annealing, baking, boiling, burning, casting, coating, cooking, dehydrating, dipping, distilling, expelling, galvanizing, melting, pickling, plating, quenching, reducing, rendering, roasting, smelting
    2. Change from liquid or gaseous to solid material
    3. Change to process that does not require gaseous or liquid material
  B. Burn the contaminant to CO_2 and H_2O
    1. Incinerator
    2. Catalytic burner
  C. Adsorb the contaminant
    Activated carbon

[a] From Stern, A. C., Boubel, R. W., Turner, B. D., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.

B. Control of Stationary-Source Particulate Matter Emissions

The primary method of control of particulate matter emissions from a process is to change the process to one that does not produce or emit particulate matter. The secondary method is to remove from the process as much as possible of the particulate matter from the effluent gas stream. There are several ways of doing this. The simplest is the use of gravity separators (settling chambers), which allow the heavier (larger) particles to be removed by sedimentation. This method does not remove the lighter (smaller) particles. More separation can be obtained by increasing the number of *g*'s (gravity forces) applied to the particles, by whirling the dusty gas in a centrifugal separator, the most commonly used of which is called a "cyclone" separator. A still higher degree of separation is obtained by passing the dusty gas through an electrostatic precipitator, a filter, or a scrubbler.

An electrostatic precipitator applies an electrical charge to the particles in the gas stream and then passes the stream through a high-voltage electrical field established between a wire and a plate. The charged particles are driven to the plate by the electrical force of the field. They are removed from the plate by rapping or washing it. The efficiency of particulate matter collection of an electrostatic precipitator can be controlled by design. To obtain higher efficiency, a larger, more expensive precipitator is required.

In a filter, the dusty gases are passed through a filter medium, which retains the particles to the extent provided by the medium chosen. Media include sand; gravel beds; natural, synthetic, or glass fiber woven mats; woven fabrics; or felts. The most common are woven fabrics, which are cleaned by shaking, intermittent reverse flow, continuous reverse flow, or air pulses.

Scrubbers wash the particles out of the dusty gas by a variety of mechanisms, including contact with water droplets created by spray nozzles and breakup of liquid streams (venturi scrubbers) or contact with wetted solid surfaces.

C. Control of Stationary-Source Gaseous Emissions

Just as with the control of particulate matter emissions, the primary method of control of gaseous emissions is to change the process to one that does not produce or emit the pollutant gas. The secondary methods are scrubbing, incineration, catalytic burning, and adsorption.

Scrubbing gaseous emissions is similar to scrubbing particulate matter emissions, except that methods may be used that cannot be employed to scrub particulate matter. Particles can clog openings through which gases and scrubbing liquid pass in a scrubber removing only pollutant gases. This allows the use of sieve plate towers and packed towers. Dusty gas must have its particulate matter content removed before being passed to such devices. However, spray towers and venturi scrubbers may be used for removing particulate matter or gaseous pollutants, either separately or simultaneously. Scrubbing liquids other than water can be used.

Combustible gases and vapors can be burned in incinerators (high temperature) or catalytic burners (low temperature) to carbon dioxide and water. Provision has to be made for burning gases and vapors containing chlorine, fluorine, or bromine. (Incinerators can be designated to burn solids or liquids as well as gases and vapors.)

D. Control of Mobile-Source Emissions

Much of air pollution control in the nineteenth century and the first half of the twentieth century was of smoke and cinder emissions from steam locomotives and steamships. This is no longer a major problem because of the displacement of steam locomotives by electric and diesel locomotives and the displacement of steamships by diesel-powered ships. Diesel engines in these applications produce the same kinds of pollution as diesel engines

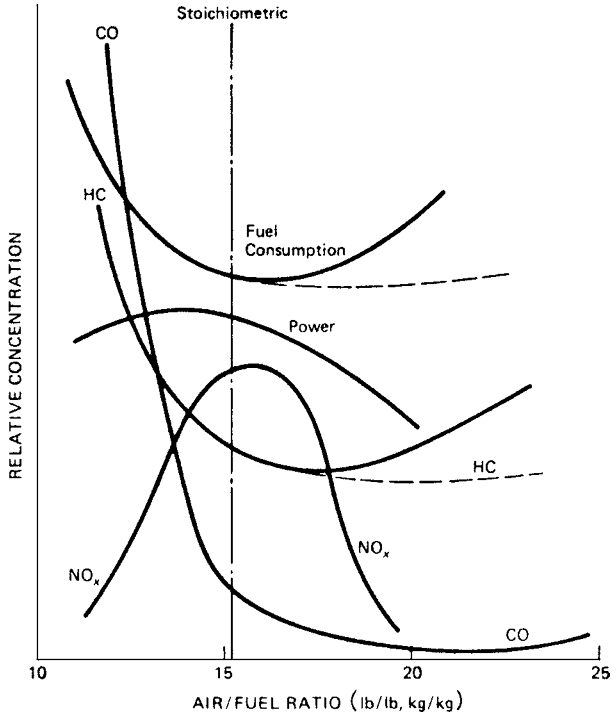


FIGURE 3 Relation of typical engine emission and performance to air/fuel ratio. Solid lines, conventional engine; dashed lines, lean-burn engine. [From Stern, A. C. (ed.) (1977). "Air Pollution," 3rd ed., Vol. 4. Academic Press, New York.]

TABLE X U.S. Federal Primary and Secondary Ambient Air Quality Standards[a,b]

Pollutant	Type of standard	Averaging time	Frequency parameter	Concentration $\mu g/m^3$	Concentration ppm
Sulfur oxides (as sulfur dioxide)	Primary	24 h	Annual maximum[c]	365	0.14
		1 year	Arithmetic mean	80	0.03
	Secondary	3 h	Annual maximum[c]	1,300	0.5
Particulate matter	Primary	24 h	Annual maximum[c]	260	—
		24 h	Annual geometric mean	75	—
	Secondary	24 h	Annual maximum[c]	150	—
		24 h	Annual geometric mean	60[d]	—
Carbon monoxide	Primary and secondary	1 h	Annual maximum[c]	40,000	35
		8 h	Annual maximum[c]	10,000	9
Ozone	Primary and secondary	1 h	Annual maximum[c]	235	0.12
Nitrogen dioxide	Primary and secondary	1 year	Arithmetic mean	100	0.05
Lead	Primary and secondary	3 months	Arithmetic mean	1.5	—

[a] From 40 Code of Federal Regulations, Part 50. U.S. Government Printing Office, Washington, D.C.

[b] National primary ambient air quality standards define levels of air quality that the EPA administrator judges are necessary, with an adequate margin of safety, to protect the public health. National secondary ambient air quality standards define levels of air quality that the administrator judges necessary to protect the public welfare from any known or anticipated adverse effects of a pollutant.

[c] Not to be exceeded more than once per year.

[d] As a guide to be used in assessing implementation plans for achieving the annual maximum 24-h standard.

in trucks and automobiles discussed in the following paragraphs.

Internal combustion (IC) engines for motor vehicles are of two kinds: spark ignition (SI) and compression ignition (diesel). The SI engines use gasoline, alcohol, or compressed natural gas fluids (propane) or compressed natural gas as fuel; introduce the fuel to the cylinder by carburation or injection; and use a spark plug for ignition. Diesel engines use heavier fuels (usually designated diesel fuel), introduce the fuel to the cylinder by injection; and ignite the fuel–air mixture by the heat of compression of the cylinder.

Spark ignition engines can operate either fuel-rich (more fuel than stoichiometrically necessary) or fuel-lean (more air than stoichiometrically necessary). Diesel engines are designed to operate fuel-lean without smoke emission and emit smoke when operated fuel-rich.

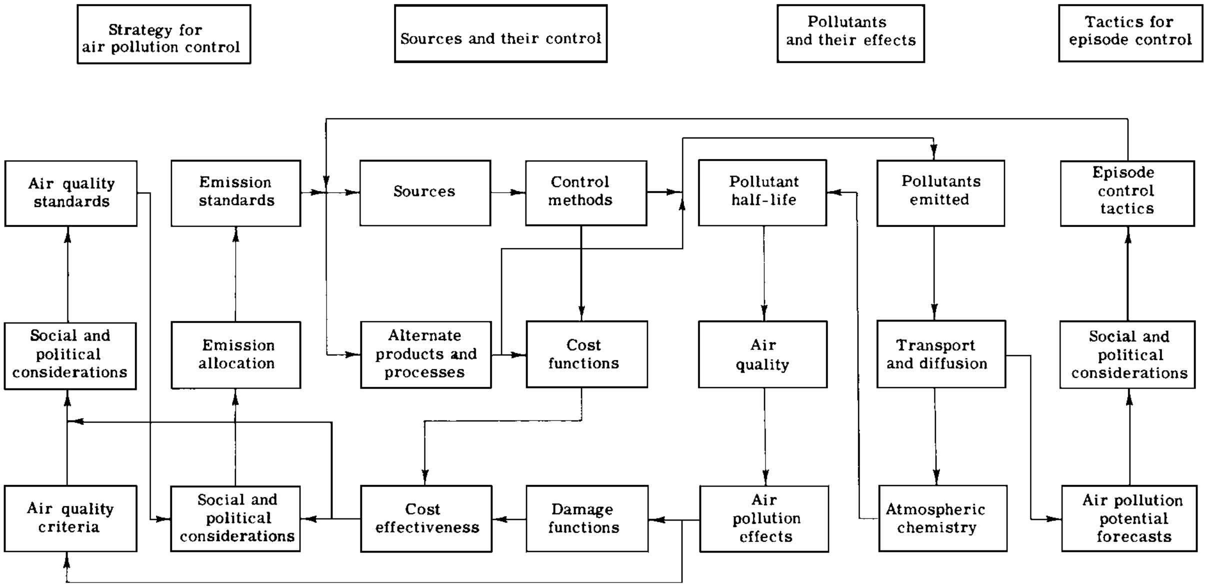


FIGURE 4 A model of the air pollution system. [From Stern, A. C., Boubel, R. W., Turner, D. B., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.]

TABLE XI U.S. Federal Standards of Performance for New Stationary Sources: Fossil-Fuel-Fired Steam Generators[a]

Date and reference[b]	Affected facility	Pollutants regulated	Maximum emission level permitted	Monitoring requirement
Proposed/effective	Coal, coal–wood residue-fired boilers, $>264 \times 10^6$ kJ/h	Particulate matter	43 g/10^6 kJ	No requirement
8/17/71 (35 FR 15704)		Opacity	20%, 27% for 6 min/h	Continuous
		SO_2	516 g/10^6 kJ	Continuous[c]
Promulgated	Coal-fired boilers	NO_x	301 g/10^6 kJ	Continuous[c]
12/23/71 (36 FR 24876)				
Revised	Oil, oil–wood residue-fired boilers, $>264 \times 10^6$ kJ/h	Particulate matter	43 g/10^6 kJ	No requirement
7/26/72 (37 FR 14877)		Opacity	20%, 27%, for 6 min/h	Continuous
10/15/73 (38 FR 28564)		SO_2	344 g/10^6 kJ	Continuous[c]
6/14/74 (39 FR 20790)		NO_x	129 g/10^6 kJ	Continuous[c]
1/16/75 (40 FR 2803)	Gas, gas–wood residue-fired boilers, $>264 \times 10^6$ kJ/h	Particulate matter	43 g/10^6 kJ	No requirement
10/6/75 (40 FR 46250)		Opacity	20%, 27% for 6 min/h	Continuous
12/22/75 (40 FR 59204)		NO_x	869 g/10^6 kJ	Continuous[c]
11/22/76 (41 FR 51397)	Mixed-fossil-fuel-fired boilers, $>264 \times 10^6$ kJ/h	Particulate matter	43 g/10^6 kJ	No requirement
1/31/77 (42 FR 5936)		Opacity	20%, 27% for 6 min/h	Continuous
7/25/77 (42 FR 37936)		SO_2	Prorated	Continuous[c]
8/15/77 (42 FR 41122)		NO_x (except lignite or 25% coal refuse)	Prorated	Continuous[c]
12/5/77 (42 FR 61537)				
3/3/78 (43 FR 8800)				
3/7/78 (43 FR 9276)	Lignite, lignite–wood residue. $>264 \times 10^6$ kJ/h	Particulate matter	43 g/10^6 kJ	No requirement
1/17/79 (44 FR 3491)		Opacity	20%, 27% for 6 min/h	Continuous
6/11/79 (44 FR 33580)		SO_2	516 g/10^6 kJ	Continuous[c]
12/28/79 (44 FR 76786)		NO_x (as of 12/26/76)	344 g/10^6 kJ for North Dakota, South Dakota, and Montana lignite in cyclone-fired units	Continuous[c]
2/6/80 (45 FR 8211)				
5/29/80 (45 FR 36077)				
7/14/80 (45 FR 47146)				
11/13/81 (46 FR 55975)				
11/24/81 (46 FR 57497)				
1/15/82 (47 FR 2314)				

[a] From Stern, A. C., Boubel, R. W., Turner, D. B., and Fox, D. L. (1984). “Fundamentals of Air Pollution,” 2nd ed. Academic Press, Orlando.

[b] FR, *Federal Register.*

[c] There are some exceptions; see appropriate standards.

The principal pollutants emitted by IC engines are carbon monoxide, oxides of nitrogen, unburned hydrocarbons and particulate matter (smoke). The emissions from an SI engine without air pollution controls built into the engine or vehicle are shown in Fig. 3.

Carbon monoxide and unburned hydrocarbon control is achieved by fuel-lean operation, improved carburetor calibration, fuel injection, and catalytic conversion (afterburners). Nitrogen oxide control is achieved by control of combustion chamber temperature via exhaust gas recirculation and by catalytic conversion (afterburners). Additional unburned hydrocarbon control is achieved by control of fuel evaporation from the fuel tank and the carburetor and of cylinder blow-by by closure of the crank case and its venting by a positive crank case ventilation valve back to the carburetor.

Smoke emission from diesel engines is controlled by operation of the engine within its design limits. (Most smoke emission occurs when the engine is overloaded.) Filters for diesel engine exhaust are under development.

V. REGULATORY CONTROL

The two principal regulatory control strategies are (1) those based on establishing and achieving air quality standards—the so-called air quality management strategy—and (2) those based on establishing and achieving emission standards—the so-called best practicable means of control approach. The former is in use in the United States; the latter is being used in Great Britain. Actually, by the 1990s, each strategy had adopted some

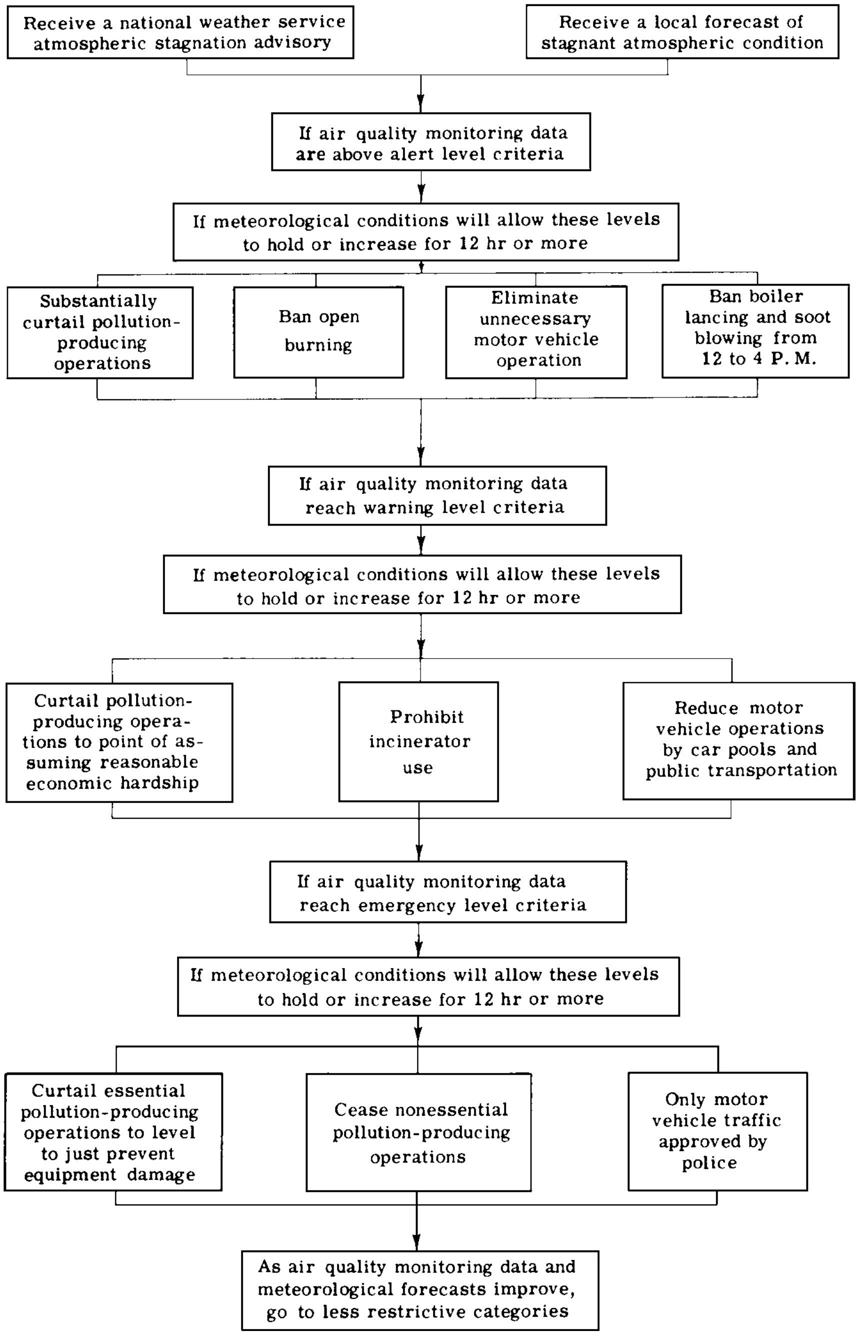


FIGURE 5 Air pollution episode control scenario. [From Stern, A. C., Boubel, R. W., Turner, D. B., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.]

of the best features of the other, so that there was less difference between them than there had been several decades earlier.

A. Air Quality Criteria and Standards

Air quality criteria are statements of the adverse effects of a pollutant at various atmospheric concentrations of that pollutant (Tables VI and VII). They are subject to revision as new data on adverse effects are published and accepted by the scientific and medical communities.

Air quality standards are governmentally promulgated concentrations of a pollutant that are not to be exceeded. A standard usually specifies the method or methods acceptable for measuring the pollutant concentration, the averaging time of the measurement, and the percentage of time that the standard may be exceeded in a specified time period (e.g., a year). It may also specify the time period during a day in which measurements will be acceptable for establishing compliance with the standard.

Some countries have promulgated a large number of air quality standards (e.g., Russia); others a very small number (e.g., the United States; see Table X).

B. Emission Standards and Permits

Emission standards are governmentally promulgated limits on the emission of a pollutant from a process. There are limits on the emission of specific gaseous, vapor, and particulate pollutants from any process and on their emission from specific processes. Some countries (e.g., the United States) officially promulgate and publish their emission standards for new sources for a number of processes and for hazardous pollutants from any process (Table XI). In the United States, emission standards for existing sources are promulgated and published by state and local air pollution control authorities. Some countries (e.g., Russia) do not promulgate or publish emission standards.

Emission standards are expressed in a large number of units, the most common of which is mass of the pollutant emitted per unit volume of effluent gas. The standard usually specifies the method of measurement.

As a general rule, where an emission standard has been promulgated, no new source can be built or an existing source substantially altered without a permit from the air pollution control authority having jurisdiction. The permit is issued only after that authority has satisfied itself that the applicable emission standard will not be exceeded or in some cases, as discussed further later, that the air quality standard for the pollutant will not be exceeded.

Emission standards are derived differently by the air quality management and the best practicable means of control approaches to air pollution control. In the best practicable means of control approach, emission standards are derived by consideration only of the means of engineering control available to control emission from a process. Air quality standards are not involved in the setting of the emission limit, and emission controls of the same stringency are applied to different processes.

In the air quality management approach, emission standards are set in an attempt to achieve a previously set air quality standard (Fig. 4). This requires an allocation step, which applies emission standards of different stringency to different processes to achieve the air quality standard goal. It should be noted that Fig. 4 shows a short-term statistical pathway to control "air pollution episodes" (Fig. 5). There has been less success with this short-term pathway than with the longer term strategic pathway.

One of the requirements of the U.S. Clean Air Act of 1977 is that best available control technology (BACT) be used for new major sources of a specific pollutant that are to be located in an area that has met the national ambient air quality standard (NAAQS) for that pollutant (called an attainment area). BACT is also required for pollutants for which there is no NAAQS [e.g., total reduced sulfur, for which emission limits are specified by a federal new-source performance standard (NSPS)]. BACT must be at least as stringent as NSPS, but is determined on a case-by-case basis.

TABLE XII U.S. Federal PSD Concentration Increments[a]

Pollutant	Increment (μg/m^3)
Class I areas[b]	
Particulate matter	
Annual geometric mean	5
24-h maximum	10
Sulfur dioxide	
Annual arithmetic mean	2
24-h maximum	5
3-h maximum	25
Nitrogen dioxide	
Annual arithmetic mean	2.5
Class II areas[c]	
Particulate matter	
Annual geometric mean	19
24-h maximum	37
Sulfur dioxide	
Annual arithmetic mean	20
24-h maximum	91
3-h maximum	512
Nitrogen dioxide	
Annual arithmetic mean	2.5
Class III areas[d]	
Particulate matter	
Annual geometric mean	37
24-h maximum	75
Sulfur dioxide	
Annual arithmetic mean	40
24-h maximum	182
3-h maximum	700
Nitrogen dioxide	
Annual arithmetic mean	50

[a] From Stern, A. C., Boubel, R. W., Turner, B. D., and Fox, D. L. (1984). "Fundamentals of Air Pollution," 2nd ed. Academic Press, Orlando.

[b] Class I areas are pristine (e.g., national parks, national seashores, natural wilderness areas).

[c] Class II areas are areas where moderate deterioration is allowed. (Unless otherwise designated, all areas are class II.)

[d] Class III areas are specifically designated as heavy industrial.

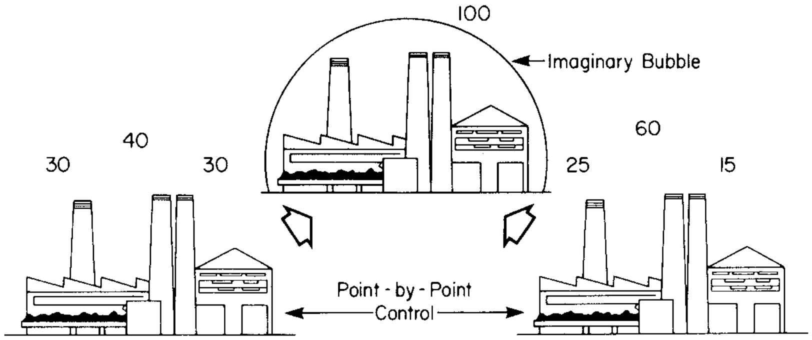


FIGURE 6 Bubble concept. This pollution control concept places an imaginary bubble over an entire industrial plant, evaluating emissions from the facility as a whole instead of requiring control point by point on emission sources. Numbers represent emissions from individual sources, some of which can be fugitive sources, and from the entire industrial plant. [Drawing courtesy of the Chemical Manufacturers Association.]

C. Prevention of Significant Deterioration

The United States has adopted a policy called prevention of significant deteriotation (PSD) (Table XII). Under this policy, no permit for certain specified new sources can be issued if a calculation using a mathematical model indicate that it would cause one of the following:

1. The air quality of a national park, national seashore, or national wilderness area to be exceeded by a specified amount
2. The concentration of a specific pollutant in an "air quality control region," which already exceeds the air quality standard for that pollutant (called a nonattainment area), to be increased

The type of mathematical model to be used and its input parameters are specified in the PSD regulations of the U.S. Environmental Protection Agency (USEPA).

D. Emission Trading

For new sources to be added to a nonattainment area or the capacity of an existing source to be increased, the USEPA has established a policy of allowing emission trading in such areas. Under this policy for a source to be built or enlarged, it must show that it has arranged for a decrease in the emission of that pollutant by other sources greater than the new source would add to the environment. The new source can do this by shutting down or decreasing emission from its own plants or from other plants in the area by purchase of their emission rights or by other means. In addition, new sources or enlarged sources in a nonattainment area must be constructed to meet the lowest emission rate allowed or achieved anywhere without regard for cost or energy usage (LAER). LAER is intended to be more stringent than BACT or NSPS and is determined on a case-by-case basis.

E. Bubble Policy

The USEPA has established a "bubble policy" under which, instead of each source of a specific pollutant in a source complex being required to meet its emission standard, all the sources of that pollutant in the complex can be considered one large source of the pollutant. The complex is then required to meet the emission limit; that is, some components of the complex can individually exceed the limit if compensated for by other components that individually emit less than the individual limit (Fig. 6). The size of the "bubble" is left to administrative discretion.

SEE ALSO THE FOLLOWING ARTICLES

AEROSOLS • COMBUSTION • ENVIRONMENTAL TOXICOLOGY • GREENHOUSE WARMING RESEARCH • HAZARDOUS WASTE INCINERATION • INTERNAL COMBUSTION ENGINES • POLLUTION, AIR • POLLUTION, ENVIRONMENTAL • POLLUTION PREVENTION FROM CHEMICAL PROCESSES • SOIL AND GROUNDWATER POLLUTION • TRANSPORT AND FATE OF CHEMICALS IN THE ENVIRONMENT • WATER POLLUTION

BIBLIOGRAPHY

Calvert, S., and Englund, H. (1984). "Handbook of Air Pollution Technology," Wiley, New York.

"Handbook of Emission Factors (1980–1983), "Vols. 1 and 2. Government Publishing Office, The Hague, The Netherlands.

Reist, P. (1984). "Introduction to Aerosol Science," Macmillan, New York.

Stern, A. (ed.) (1976–1978). "Air Pollution," 3rd ed., 5 vols. Academic Press, New York.

Stern, A. (ed.) (1986). "Air Pollution," Suppl. Vols. 6–8. Academic Press, Orlando.

Stern, A., Boubel, R., Turner, D., and Fox, D. (1984). "Fundamental of Air Pollution," 2nd ed. Academic Press, Orlando.

Pollution, Air

Roy M. Harrison
Rob P. Kinnersley
University of Birmingham

GLOSSARY

Acid rain The deposition of strongly acidic pollutants, especially sulfuric and nitric acids, from the atmosphere to the earth's surface. The term strictly refers only to wet deposition (in rain, snow, etc.) but is often used to include dry deposition processes (i.e., due to settling or turbulent transfer).

Advection The transfer of pollutants by the horizontal motion of a mass of air.

Emissions inventory A compilation, usually allocated into source categories, of pollutant emissions to atmosphere within a specified geographic zone.

Oxides of nitrogen (NO_x) The compounds nitric oxide (NO) and nitrogen dioxide (NO_2); NO_x refers to the combined concentration of the two.

Photochemical smog A form of air pollution caused by chemical reactions of oxides of nitrogen and volatile organic compounds in sunshine, characterized by the formation of elevated concentrations of ozone and of secondary particles, the latter causing a reduction in visibility.

$PM_{2.5}$, PM_{10} The mass concentration of particles in the atmosphere passing a size-selective inlet with a 50% efficiency cutoff at 2.5 and 10 μm, respectively.

Polycyclic aromatic hydrocarbons (PAH) A group of chemical substances comprising two or more fused benzene rings. Some members of the group are carcinogenic.

Primary pollutant A pollutant which is directly emitted from a source rather than formed within the atmosphere.

Secondary pollutant A pollutant formed within the atmosphere from chemical reaction processes rather than emitted directly.

Semivolatile air pollutant A substance such as ammonium nitrate or many of the polycyclic aromatic hydrocarbons which partition within the atmosphere between vapor phase and condensed forms.

Smog, London-type The combination of severe air pollution by smoke and sulfur dioxide, normally from coal burning, with fog, leading to very poor visibility.

Smog, Los Angeles-type See photochemical smog.

Sources and sinks Terms describing the sources of emission and removal pathways, respectively, of pollutants to and from the atmosphere.

Suspended particulate matter Small particles, mainly less than 20 μm in diameter, of either solid or liquid material forming a relatively stable suspension within the atmosphere.

Volatile organic compounds A generic term describing organic compounds of low boiling point present in the atmosphere in the vapor phase.

AIR POLLUTION in cities has been a major problem for many hundreds of years as a result of the combustion of fuels mainly for the purposes of home heating. Records of severe pollution in cities such as London go back to at least 1300 and there was a recognition that air pollution episodes were associated with adverse effects on human health as far back as 1661. Whereas the use of wood and charcoal as fuels was important in early times, it is the large-scale use of coal that was associated over several hundreds of years with periodic episodes of very poor air quality. In developed countries measures have been taken to remedy this problem, including a switch to cleaner fuels and the installation of engineering controls on more major sources of emissions. In less developed countries, however, most notably in China, the problems associated with the combustion of coal in cities persist. The major pollutant emissions are smoke and sulfur dioxide, which can build up to toxic concentrations under conditions of adverse meteorology which limit atmospheric dispersion. There have been a number of notorious episodes of such pollution in the Western world involving coal burning as the major source, such as in London and New York, or a combination of coal burning and industrial emissions, as in the Meuse Valley episode in Belgium in 1930 and Donora, Pennsylvania, in 1948. The December 1952 episode in London was the subject of a government enquiry which found that it had led to 4000 deaths over and above those which would normally have occurred during the period of the episode and immediately afterward. The term *smog* (i.e., smoke and fog) was coined to describe this kind of winter pollution episode. The relationship between the concentrations of smoke, sulfur dioxide, and daily deaths during the 1952 episode appears in Fig. 1.

The instigation of control measures from the 1950s onward led to a steady improvement in the airborne concentrations of smoke (i.e., combustion particles) and sulfur dioxide. Concurrently, however, there were marked increases in the use of road vehicles, and the air pollution climate progressively changed from one dominated by smoke and sulfur dioxide to an atmosphere dominated by the emissions of road vehicles, or, in some areas, their chemical reaction products. In many parts of the world it is the direct emissions from the road vehicles themselves, most notably carbon monoxide, oxides of nitrogen, and hydrocarbons, which represent the problem, mainly through the direct toxicity of these species. However, it was recognized in southern California from the 1950s onward that a new kind of pollution had developed arising from the chemical reaction products of oxides of nitrogen and hydrocarbons in sunshine. This became known as *photochemical smog* because, although its origins were wholly different from that of the London smogs, one of the major manifestations was visibility loss. In the United States such problems have tended to manifest themselves most notably in extended urbanized areas such as Los Angeles and the adjacent region of southern California, whereas in Europe, photochemical smog, which is most recognizable through elevated concentrations of ozone, tends to be a regional problem which is frequently more severe in rural than in urban areas due to long-range transport of pollutants and the time taken to progress chemical reactions.

I. INTRODUCTION

Air pollution problems are recognizable on a wide range of spatial scales. Perhaps the smallest of these is the scale of the street canyon. Within heavily trafficked streets set between continuous rows of high buildings, there exists a possibility to build up high levels of traffic-generated air pollutants due to very restricted atmospheric dispersion. This leads to local pollution hotspots. Moving upward in scale, cities present a pollution climate which is usually quite distinct from that of the surrounding countryside. Often, if background locations are considered (i.e., locations not adjacent to busy highways or major point sources) concentrations of traffic-generated pollutants tend to be rather uniform and significantly higher than those outside the urban area. Concentrations tend to scale with the size of the city, with larger cities able to generate significantly higher concentrations due to the greater source strength. Some pollution problems are most noticeable on the regional scale. These are typified by tropospheric (i.e., ground-level) ozone and acid rain problems. These are connected with secondary pollutants, a term used to describe pollutants which are formed within the atmosphere rather than emitted directly. The regional scale might describe, for example, the eastern United States or western Europe. Finally, long-lived pollutants can affect the atmosphere on

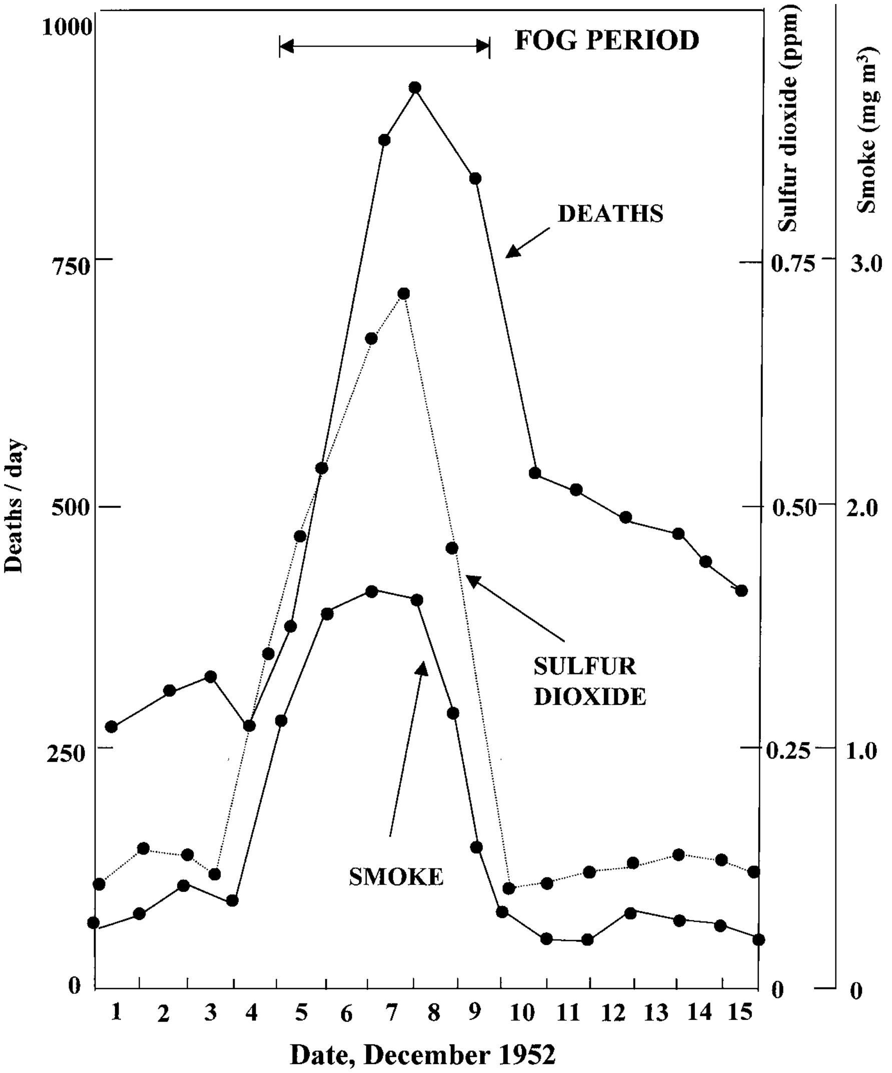


FIGURE 1 Daily concentrations of smoke and sulfur dioxide and daily deaths in the December 1952 London smog.

a continental, hemispheric, or global scale. In order to become globally important, pollutants require a long atmospheric lifetime which will allow them to mix through the atmosphere before being removed by a sink process. Since interhemispheric mixing processes are on a time scale of 1–2 years, it is only substances with very slow atmospheric chemical reaction processes, such as carbon dioxide, which become genuinely globally distributed. Pollutants with very low reactivity in the lower atmosphere, such as the chlorofluorocarbons (also known as CFCs or freons), are able to penetrate into higher layers of the atmosphere, where ultimately they are broken down by photochemical processes involving shorter, more energetic wavelengths of light than are available at ground level. Such long-lived substances can be a particular problem since they contribute not only to global warming processes, but also to the depletion of ozone within the stratosphere.

II. PRIMARY POLLUTANTS

A. Sources and Inventories

The term *primary pollutants* describes those pollutants which are emitted directly to the atmosphere from sources such as road traffic, power plants, and industry. Their emissions are described in emissions inventories which are compiled on spatial scales relating to the nature of the pollutant itself. Thus, for pollutants having a local impact on health, emissions inventories are likely to be compiled on a spatially disaggregated grid basis within a city, whereas for a globally acting pollutant, the location of emission is far less important and inventories more typically describe national totals disaggregated according to source category, which can then be used in the planning of control measures or international negotiations on reductions, such as the Montreal and Kyoto Protocols relating to ozone-depleting substances and global warming, respectively. Inventories are compiled by combining some level of activity, such as the number of kilometers driven by a given kind of vehicle per year together, with an emission factor which describes, for example, the grams of pollutant emitted per kilometer driven.

The main source of *sulfur dioxide* is from the oxidation of sulfur in fuels, most notably coal, which can typically contain between 1% and 3% sulfur by weight, and oil-derived fuels. Crude oils vary greatly in their sulfur content according to their origins and many of the refining processes are now designed to produce fuels such as gasoline and diesel and heating oil which have a very low sulfur content. Residual oils, however, which can have an extremely high sulfur content (several percent sulfur by weight), are commonly used by shipping. The inventory of sulfur dioxide emissions for the United Kingdom shows a predominance of emissions from power plants and other combustion processes. Some industrial processes, such as the smelting of sulfide ores and the production of sulfuric acid, lead to emissions of sulfur dioxide, but in comparison these are very small.

There are several ways of measuring emissions and airborne concentrations of *particulate matter*. Particles are distinguished from gas-phase pollutants by the fact that they comprise solid particles or liquid droplets suspended in the atmosphere and may be separated from the air by filtration. Weighing the mass of collected particles gives a measure known as total suspended particulates (TSP). Most commonly, however, particles are sampled through an inlet with a size cutoff at 10 μm diameter such that only particles of less than 10 μm are collected on the filter and weighed. This size fraction is referred to as PM_{10} and the directly analogous fraction less than 2.5 μm, i.e., $PM_{2.5}$, is also widely measured. If, rather than weighing the filter, the collected particles are assayed according to their capacity to absorb light either through measurement of reflection or transmission of light, then the measurement is referred to as black smoke, and although reported in mass per unit volume units analogous to those for PM_{10}, this measurement depends upon use of a calibration curve based upon the use of coal smoke as a calibrant. Many of the historically collected data for suspended particles are in terms of black smoke, but nowadays this measure is little used.

The sizes of airborne particles range from a few nanometers at the lower end to tens of micrometers at the upper end. Generally speaking (but not exclusively) particles generated from combustion and other high-temperature processes, and thus most anthropogenic emissions, are within the size range less than 2.5 μm in diameter. On the other hand, particles larger than 2.5 μm come largely from mechanical disintegration processes and comprise materials such as particles from sea spray and wind-derived resuspension of soils and road dusts, although some industrial emissions from processes such as quarrying are largely within this size range. Currently, there is a strong focus of interest on particles of less than 100 nm diameter, referred to as the *ultrafine particles*. Toxicological studies suggest that these are appreciably more toxic per unit mass than coarser particles and it is believed that these play a major role in mediating the adverse health effects of airborne particles.

An inventory of sources of PM_{10} particulate matter in the United States appears in Fig. 2 and excludes fugitive sources, such as resuspended dusts from paved and unpaved roads. The emission factors for such processes are still very uncertain, and whereas there is no doubt such processes are important, their exact magnitude is not at all well known. Within Europe, emissions from road traffic feature very highly as a source of primary particles,

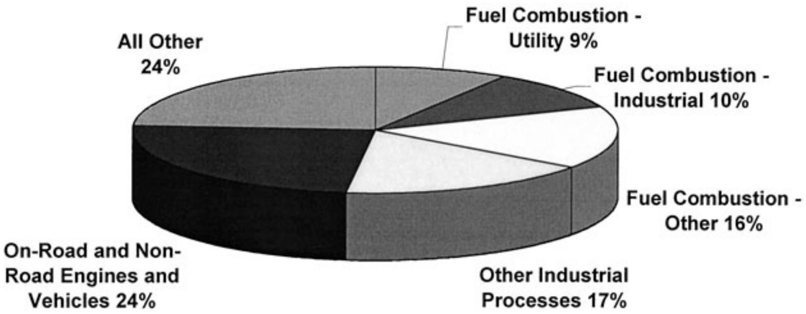


FIGURE 2 Inventory of national particulate matter (PM_{10}) emissions by principal source category in the United States for non-fugitive dust sources in 1997.

especially with cities, and the diesel vehicle, which emits a far greater mass of particles per kilometer driven than an equivalent gasoline-powered vehicle, has become a major focus of concern.

Oxides of nitrogen are pollutants generated predominantly through combustion processes. The major proportion comes from the combination of atmospheric nitrogen and oxygen at the high temperature involved in the combustion process. This reaction, leading to nitric oxide, NO, is strongly temperature sensitive, and emission factors rise sharply accordingly to the temperature of the combustion process. In the case of coal combustion, the fuel has a significant nitrogen content, which is also converted to nitric oxide during combustion. Both emissions and airborne concentrations of oxides of nitrogen are frequently described in terms of the sum of nitric oxide and nitrogen dioxide, referred to as NO_x. The other oxide of nitrogen, N_2O, or nitrous oxide, is not included within NO_x and plays little role in the lower atmosphere other than as a greenhouse gas. Some nitrogen dioxide, NO_2, is formed also in combustion processes, but this typically comprises only about 5% of NO_x emissions. The main source of nitrogen dioxide is the oxidation of NO within the atmosphere, and NO_2 is therefore considered as a secondary pollutant. An inventory of oxides of nitrogen emissions appears in Fig. 3 and shows the very important role of road traffic and power plants as sources of oxides of nitrogen emissions. There is some primary emission of nitrogen dioxide from industrial processes such as nitric acid manufacture, but this is only a very small part of national totals for NO_x.

Carbon monoxide is a product of incomplete combustion of carbonaceous fuels. It is most notable as an emission from gasoline engines, which run very close to a stoichiometric mixture of hydrocarbon fuel and air, compared to the diesel, which runs with substantial excess air and emits far less carbon monoxide per kilometer driven. Carbon monoxide concentrations were a matter of concern in developed countries prior to the adoption of automotive catalytic converters. Once warmed up, these provide a highly efficient oxidation of carbon monoxide and therefore reduce emissions by at least 90%. The emissions inventory for carbon monoxide is dominated by road traffic and by off-road use of internal combustion engines.

The term *volatile organic compounds* (VOC) is used to describe organic substances existing in the atmosphere partitioned almost exclusively within the vapor phase. The largest single component within the atmosphere is normally methane, which is long-lived, with a global background of around 1.8 parts per million. The sources of methane relate mainly to biologically mediated decay processes for organic material, although the anthropogenic contribution to such processes is greater than the natural one; thus, processes such as landfill and some agricultural practices are major sources of methane. There is also a significant contribution from fossil fuel-related sources such as the leakage of methane from natural gas pipelines. An inventory of volatile organic compounds appears in Fig. 4. This attributes considerable importance to road traffic and off-road vehicles, which emit a very wide range of uncombusted and partially combusted hydrocarbon materials, together with solvent use (e.g., in paints) and noncombustion processes, largely industrial processes. Whereas the majority of the mass of volatile organic compounds emitted is in the form of hydrocarbons, other organic compounds such as oxygenates and halocarbons are also significant contributors. Due to the very great diversity of VOCs emitted, policy measures have tended to focus on a smaller number of substances because of their potential to cause adverse effects on health. These include benzene and 1,3-butadiene, which are chemical carcinogens. Reactivity scales have been introduced in order to categorize VOCs in relation to their ability to promote the formation of tropospheric ozone, and relative measures such as the photochemical oxidant creation potential (POCP) are now considered when devising policies for VOC emission control.

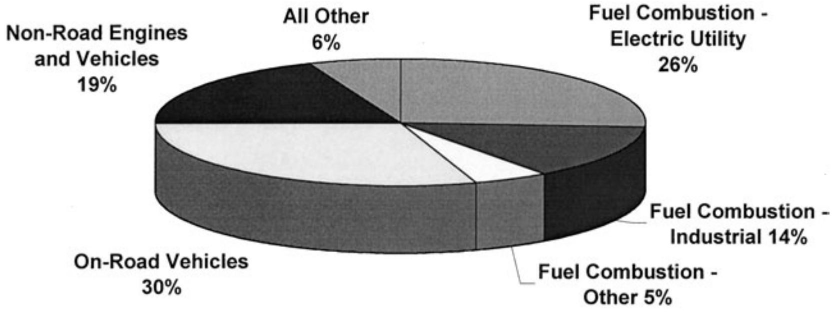


FIGURE 3 Inventory of national oxides of nitrogen (NO_x) emissions by principal source category in the United States for 1997.

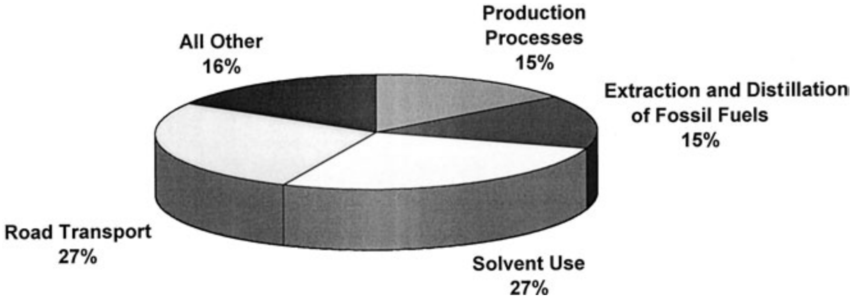


FIGURE 4 Inventory of national emissions of volatile organic compounds by principal source category in the United Kingdom for 1998.

As the molecular weight of organic compounds increases, so their tendency to be present in airborne particles rather than within the vapor phase increases. Very important groups of substances, such as the polycyclic aromatic hydrocarbons and the chlorinated dibenzo-*p*-dioxins/furans, are termed semivolatile since they exist in the atmosphere in partition between particle and vapor phases. Measurements of emissions also need to account for both physical forms.

A pollutant which has generated much public concern is *lead*. This is a multimedia pollutant to which the public is exposed through diet, drinking water, and the atmosphere. The main source of emission to atmosphere was through the use of lead additives, tetramethyl lead and tetraethyl lead, in gasoline. The use of lead in gasoline was curtailed in the United States during the 1970s and in Europe during the 1980s and 1990s. The use of lead in gasoline is not now permitted in most developed countries, although some use still persists in less developed countries, although this is rapidly declining.

National inventories of emissions such as shown in Figs. 2–4 generally attribute importance to a range of source categories. When, however, emissions inventories are compiled for cities, they generally attribute a much increased importance to road traffic emissions due to the concentration of traffic within urban areas and the frequent siting of power plants and major industry outside of conurbations. Given that road traffic emissions occur at ground level, whereas many other sources emit at high level, the air pollution climate of cities is frequently dominated by emissions from road traffic and/or the secondary pollutants formed from their chemical reactions in the atmosphere.

B. Determinants of Airborne Concentrations of Primary Pollutants

A number of factors are influential in determining the concentrations of air pollutants prevailing at ground level. For an individual source, the main factors are the following:

(a) The rate of pollutant emission, termed the source strength. Concentrations are linearly related to the emission rate, the corollary being that a cut in emissions will lead to linear rollback in airborne concentrations deriving from that source.

(b) The height of emissions. One kilogram per second of sulfur dioxide emitted at ground level has a far greater-impact on ground-level concentrations than the same mass of sulfur dioxide emitted from a 100-meter stack. Emissions are highly dispersed by the atmosphere between emission from a high-level source and ground level. The impact of high-level sources on annual average ground-level concentrations is relatively small, although the impact on short-term concentrations during the passage of a plume across a sampling site can be very much higher.

(c) The topography of the site. It has been mentioned above that situations such as street canyons are very unfavorable for the dispersion of pollutants. Severe winter air pollution episodes have typically taken place in valleys, where cold air drainage flows lead to the accumulation of stagnant air and emissions, allowing the buildup of high concentrations.

(d) Meteorological factors. There are a number of meteorological influences upon concentrations of primary air pollutants, but two factors are especially important. The first is horizontal windspeed. If all other factors are constant, the ground-level concentration is inversely proportional to the horizontal windspeed and therefore lower strength wind conditions strongly favor high concentrations. The second factor is mixing in the vertical direction. An atmosphere which is capped by a strong inversion, as in southern California, or where ground-level emissions enter a stable layer of air with little vertical transfer, as in London smogs, causes pollutants to be trapped within a relatively shallow layer of the atmosphere, therefore allowing high concentrations to develop.

(e) Removal processes. These are known as *sinks* and range from removal in rain and other forms of precipitation

TABLE I Pollutant Concentrations from Various Locations

Site	Year	Pollutant	Measurement	Concentration
Los Angeles	2000	Carbon monoxide	Highest 1-hr	6.3 ppm
			Highest 8-hr	4.6 ppm
		Nitrogen dioxide	Highest 1-hr	127 ppb
			Annual mean	37 ppb
		Sulfur dioxide	Highest 1-hr	75 ppb
			Highest 3-hr	36 ppb
			Highest 24-hr	6 ppb
			Annual mean	3 ppb
		Ozone	Highest 1-hr	136 ppb
		PM_{10}	Highest 24-hr	61.0 $\mu g\ m^{-3}$
			Annual mean	36.4 $\mu g\ m^{-3}$
		Lead	Highest 24-hr	0.1 $\mu g\ m^{-3}$
			Quarterly mean	0.05 $\mu g\ m^{-3}$
London	1998	Benzene	Annual mean	1.2 ppb
		1,3-Butadiene	Annual mean	0.23 ppb
		Carbon monoxide	Highest 8-hr	3.2 ppm
		Lead	Annual mean	0.04 $\mu g\ m^{-3}$
		Nitrogen dioxide	Annual mean	34 ppb
			99.8%ile hourly	78 ppb
		Ozone	97%ile 8-hr mean	35 ppb
		PM_{10}	Annual mean	30 $\mu g\ m^{-3}$
			99%ile 24-hr mean	60 $\mu g\ m^{-3}$
		Sulfur dioxide	Annual mean	7 ppb
			99%ile 15-min mean	229 ppb
Cairo	1988–1992	Total suspended particulates	Annual mean	1100 $\mu g\ m^{-3}$
Mexico City	1988–1992	Ozone	Second highest 24-hr mean	400 ppb
Shenyang, China	1988–1992	Sulfur dioxide	Annual mean	84 ppb

(referred to as wet deposition), to uptake by land surfaces, including vegetation and oceans (referred to as dry deposition), and frequently most important, removal by atmospheric chemical reactions. The concentrations of highly reactive substances such as 1,3-butadiene are strongly limited by their rapid reactions with atmospheric oxidants.

C. Airborne Concentrations and Trends

Airborne concentrations typical of a range of localities and averaging times appear in Table I. Developed countries have seen generally very similar trends in the concentrations of primary pollutants, with steep declines in the concentrations of particulate matter (smoke) and sulfur dioxide during the 1960s and 1970s, when a growth in the traffic-generated pollutants, carbon monoxide and oxides of nitrogen, took place. There are very few long-term continuous datasets with which to illustrate these trends. For example, Fig. 5 shows the trend in annual mean sulfur dioxide concentrations measured at County Hall, London, between 1931 and 1985. There are significant gaps in the dataset, with different lines relating to different measurement methods and slightly varying locations. The overall pattern in the data is, however, very clear. Data from the city of Lincoln, U.K., and the surrounding area for sulfur dioxide measured between 1950 and 1990 show that in the early years the urban concentrations greatly exceeded those in the surrounding countryside due to local urban emissions. After about 1980, however, the two datasets become statistically indistinguishable, as the major sources (power plants) were outside of the urban area and affected both the city and countryside in a similar manner. Urban concentrations of sulfur dioxide over this period fell more rapidly than national emissions due to the curtailment of low-level emissions in urban areas while power plant emissions changed rather little.

Measurements of oxides of nitrogen and carbon monoxide started only within the 1970s and then at a limited

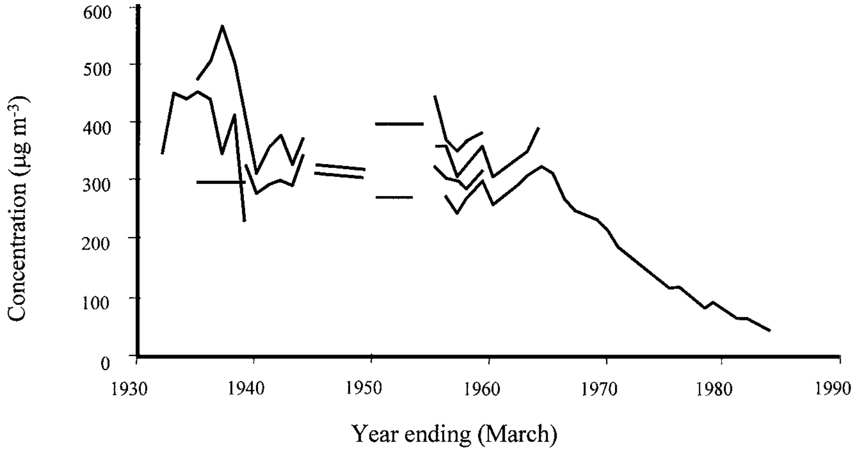


FIGURE 5 Annual mean sulfur dioxide concentrations measured at County Hall, London, between 1931 and 1985. Different lines relate to different measurements methods and slightly varying locations.

number of locations, with measurements of specific hydrocarbons coming substantially later, in the early 1990s typically. Lead monitoring also became a serious activity for monitoring networks in the 1970s, and the data for lead emissions from road traffic and airborne concentrations of lead in central London shown in Fig. 6 are an excellent illustration of linear rollback in action. In this case the vehicular source of lead was dominant and as emissions declined, airborne concentrations reduced correspondingly. The major reduction in the lead content of gasoline at the end of 1985 was not reflected in a proportionate reduction of lead in air by 1996 and this has been attributed to the legacy of lead-rich road surface dusts whose resuspension contributed to airborne lead concentrations in the years following the phase down.

III. SECONDARY POLLUTANTS

A. Air Pollution Chemistry

It is feasible to give only a very limited introduction at this point to atmospheric chemistry essential to air pollution processes, and the interested reader is referred to other sources. Sunshine is the key ingredient driving atmospheric chemical change and two photochemical processes stand out as most influential in air pollution

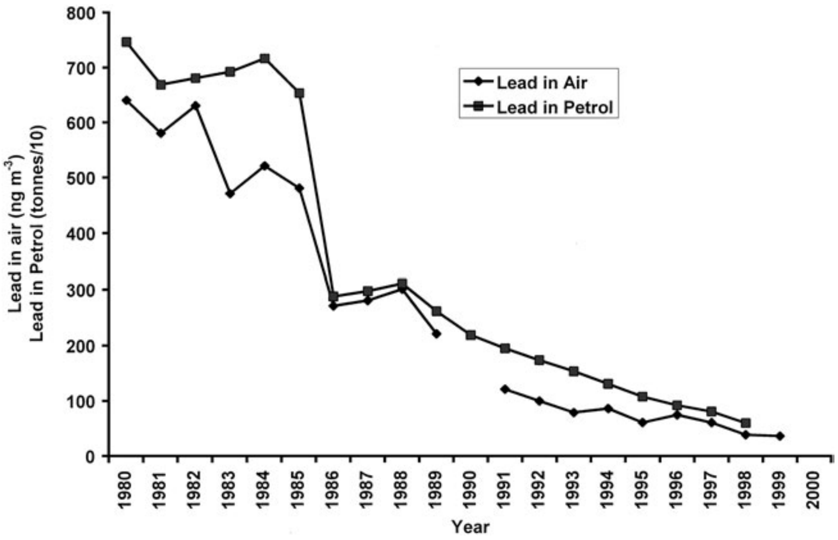


FIGURE 6 Temporal trends in the use of lead in gasoline and airborne lead concentrations in central London from 1980 to 1998.

chemistry. The first is the photolysis of ozone by light at the short-wavelength end of solar radiation penetrating to ground level through the atmosphere. The initial product of ozone photolysis is a singlet-state (excited) oxygen atom, which reacts with water vapor to form hydroxyl radicals:

$$O_3 + h\nu \rightarrow O_2 + O(^1D) \qquad \lambda < 315\,\text{nm}$$
$$O(^1D) + H_2O \rightarrow 2OH.$$

This process is of great importance as a source of the hydroxyl free radical, which is responsible for much of the chemical change occurring within the atmosphere. The other photochemical process of great importance is the photolysis of nitrogen dioxide to give triplet (ground-state) oxygen atoms, which are then able to form ozone:

$$NO_2 + h\nu \rightarrow NO + O(^3P) \qquad \lambda < 420\,\text{nm}$$
$$O(^3P) + O_2 \rightarrow O_3.$$

The ozone can in turn be destroyed by reaction with nitric oxide, thereby regenerating nitrogen dioxide:

$$NO + O_3 \rightarrow NO_2 + O_2.$$

As so far described, photolysis of nitrogen dioxide can give rise to small steady-state concentrations of ozone, which are limited by the reaction with nitric oxide. Concentrations of ozone in city centers tend to be lower than those in adjacent rural areas due to fresh emissions of NO from traffic reacting with ozone, as above. Equating the rate of NO_2 loss by photolysis to the rate of NO_2 formation leads to the following expression for the concentration of ozone:

$$[O_3] = \frac{J_{NO_2}}{k}\frac{[NO_2]}{[NO]},$$

where J_{NO_2} is the sunshine-dependent rate of NO_2 photolysis and k is the rate coefficient for the reaction of NO with O_3.

The hydroxyl radical is key to building up higher concentrations of ozone. It can react with volatile organic compounds and carbon monoxide leading to the formation of peroxy radicals. Consider methane as a simple example of a contributory VOC:

$$CH_4 + OH \rightarrow CH_3 + H_2O$$
$$CH_3 + O_2 \rightarrow CH_3O_2$$
$$CH_3O_2 + NO \rightarrow CH_3O + NO_2.$$

Also, with carbon monoxide,

$$CO + OH \rightarrow CO_2 + H$$
$$H + O_2 \rightarrow HO_2$$
$$NO + HO_2 \rightarrow NO_2 + OH.$$

Thus, in this sequence of processes peroxy radicals, CH_3O_2 and HO_2, formed and converted NO to NO_2 in a catalytic cycle involving regeneration of the hydroxyl radical. In this way, NO-to-NO_2 conversion is achieved without consumption of ozone and therefore much higher concentrations of ozone can be developed. These are typical of photochemical air pollution in all parts of the world.

There are other products which are also typical of photochemical air pollution. One which is highly characteristic of photochemical air pollution processes is peroxyacetyl nitrate (PAN). The formation route is via acetyl radicals (CH_3CO) formed from a number of routes, most notably oxidation of ethanal:

$$CH_3CHO + OH \rightarrow CH_3CO + H_2O$$
$$CH_3CO + O_2 \rightarrow CH_3C(O)OO$$
$$CH_3C(O)OO + NO_2 \rightarrow CH_3C(O)OONO_2\ \text{(PAN)}.$$

Also characteristic are the particulate secondary pollutants sulfate and nitrate, the latter being especially associated with photochemical smog processes. Nitrate formation during daytime occurs through reaction of nitrogen dioxide with the hydroxyl radical, with subsequent neutralization of nitric acid vapor by ammonia to form ammonium nitrate (NH_4NO_3) particles, which equilibrate with their gas-phase precursors and are termed semivolatile:

$$NO_2 + OH \rightarrow HNO_3$$
$$HNO_3(g) + NH_3(g) \leftrightarrow NH_4NO_3\ \text{(s or aq)}.$$

Oxidation of sulfur dioxide is considerably slower (about 1/10 the rate) but also leads progressively to sulfate formation via sulfuric acid:

$$SO_2 + OH \rightarrow HOSO_2$$
$$HOSO_2 + O_2 \rightarrow SO_3 + HO_2$$
$$SO_3 + H_2O \rightarrow H_2SO_4$$
$$H_2SO_4 + 2NH_3 \rightarrow (NH_4)_2SO_4\ \text{(ammonium sulfate)}.$$

There are many other important processes involving the very wide range of VOCs present in polluted atmospheres which can form partially oxidized oxygenate species while leading to breakdown of the VOCs.

B. Airborne Concentrations and Trends

Widespread monitoring of ozone concentrations began in many countries in the 1970s and the interpretation of trends on ozone concentrations is not a trivial matter. The urban to rural differences in ozone alluded to earlier and observable in many parts of the world arise because fresh emissions of nitric oxide from road traffic react with ozone, suppressing its concentration. A decrease in emissions of NO_x from road traffic is therefore likely to cause a small reduction in rural ozone but will cause an increase in urban ozone; in other words the urban and rural concentrations will tend to converge but the trend in the two locations will be in an opposite direction. It is therefore more fruitful to consider trends in the concentration of the sum $[NO_2 + O_3]$. Detailed research in Europe has shown different kinds of behavior in different localities, but broadly speaking in the United Kingdom, episodic peak concentrations have tended to decline over the past 30 years, while long-term average concentrations have tended to increase. The explanation for this is not wholly clear, although it may reflect a changing balance between the Northern Hemisphere background of tropospheric ozone, which is tending to increase, and the locally generated ozone associated with plumes from urban areas, which has tended to decrease.

Recent years have seen substantial cuts in emissions of NO_x in developed countries, which has been reflected in a commensurate reduction in ambient concentrations of NO_x. However, ambient concentrations of nitrogen dioxide have decreased very much less, as this is a secondary pollutant whose concentrations are determined by the oxidizing capacity of the air as well as the emissions of NO_x. Most Western European countries are, however, now seeing a slow decline in nitrogen dioxide in response to substantial cuts in vehicle emissions of NO_x.

There has been considerable debate over the linearity or otherwise of relationships between nitrate and sulfate and their respective precursor gases, NO_x and sulfur dioxide. Currently, the majority of data support the idea of a more or less linear reduction in sulfate with reductions in SO_2 emissions, which have been quite marked in most developed countries. Reductions in NO_x, taking account of all sources, have been very much less, and in Western Europe the inorganic particles are changing from an aerosol dominated by sulfate to one dominated by nitrate. In North America, nitrate dominates the aerosol on the West Coast due to the predominant influence of road traffic emissions within this region, while on the East Coast, sulfate tends to predominate, in part due to the very low ammonia levels, which preclude the formation of substantial ammonium nitrate concentrations, leaving the nitrate in the vapor phase as nitric acid, which is subject to relatively rapid wet and dry deposition processes.

IV. ATMOSPHERIC DISPERSION PROCESSES

A. Point and Line Sources

Meteorology, and hence dispersion processes, are as important in determining the airborne concentrations of pollutants as the emissions themselves. Much work has gone into the development of models for prediction of concentrations of air pollutants downwind of major sources such as power plant chimneys and major highways. Although approaches based on computational fluid dynamics are now coming into use for predicting concentrations within complex urban environments, for most applicants Gaussian plume dispersion models remain state-of-the-art for prediction of downwind concentrations. Such models assume that the spatial distribution of pollutant concentrations within a plume can be described by a Gaussian distribution function, the standard deviation of which increases with downwind travel and is influenced by the stability (i.e., turbulence level) within the atmosphere. Such models are now available as user-friendly packages, but inevitably depend upon the availability of rather sophisticated meteorological data, which are generally available from only a very limited number of sites.

The main factors determining the dispersion of pollutants from a point or line source are the speed of the wind due to the dilution effect and the efficiency of vertical mixing due to turbulence. There is typically a diurnal variation in atmospheric stability, with the atmospheric boundary layer becoming increasingly stable at night, especially under clear skies, leading to very limited dispersion of plumes. During daytime, especially with bright sunshine, there are high levels of turbulence due to heating of the ground, which transfers heat to air parcels in contact with it, which then become very buoyant, rising through the atmosphere and generating vigorous mixing. Concentrations arising from ground-level sources tend to be highest in very stable atmospheric conditions, with little vertical mixing and low windspeeds. In contrast, under these conditions emissions from elevated point sources may well be emitted above the stable layer and thereby penetrate to ground level only very inefficiently.

B. Long-Range Transport

Pollutants particularly suited to transport over long distances within the atmosphere are those that are removed only ineffectively by chemical reactions or wet and dry deposition processes. Examples include fine particulate matter ($PM_{2.5}$) and ozone. In the case of fine particles, the deposition processes are inefficient and the particles may remain airborne for many days. In the case of ozone, dry deposition is relatively efficient but operates mainly during daytime, when photochemistry rapidly replaces depleted ozone. At nighttime, dry deposition occurs from within a shallow, stable layer of air in contact with the ground, while above that layer the ozone remains unaffected by dry deposition and is depleted by only rather slow chemical reaction processes. Once sunshine mixes the atmosphere

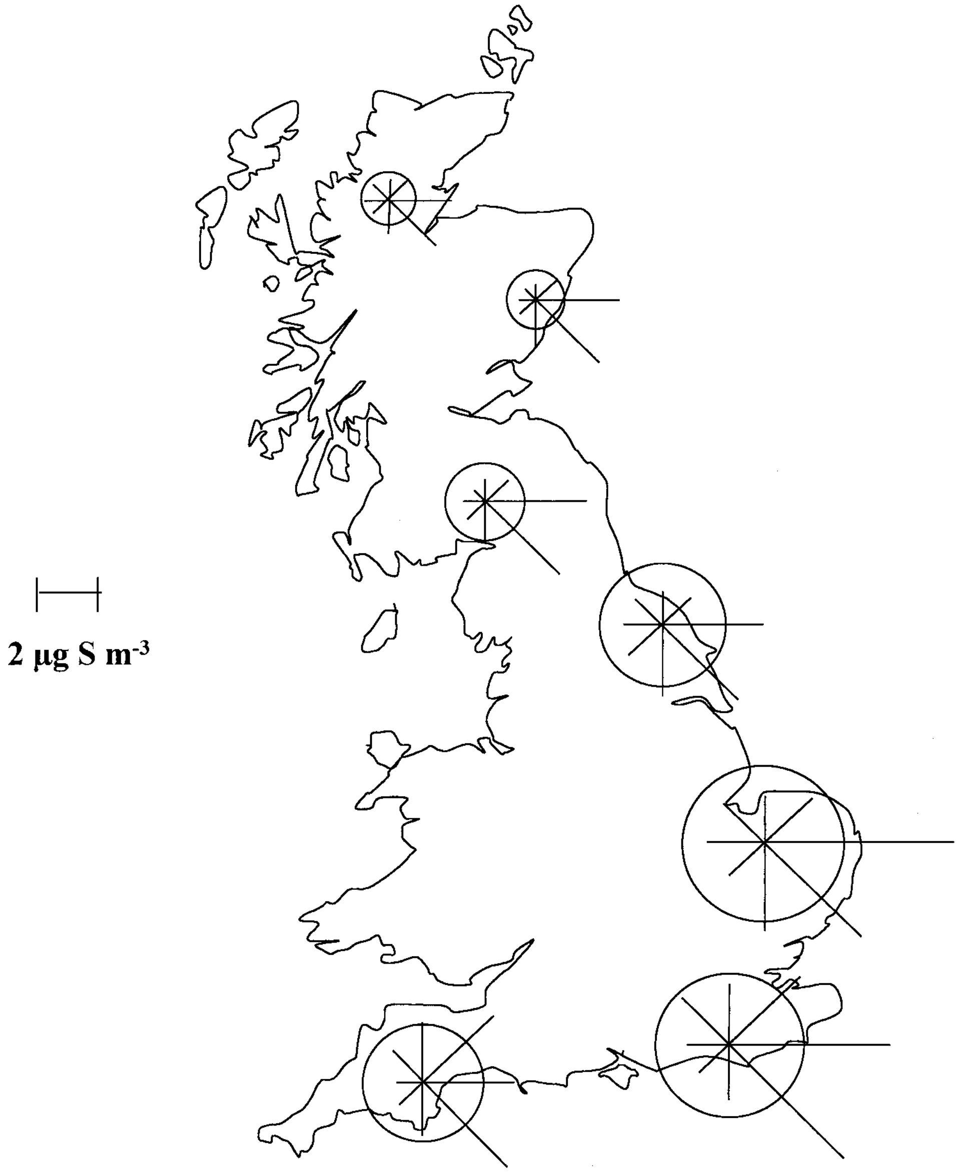


FIGURE 7 Relationship of particulate sulfate measured during 1987–1988 at rural UK sites to airmass trajectory.

during the morning, the influences of dry deposition are very modest.

Another facet of long-range transported pollutants such as sulfate and nitrate particles and ozone is that they are formed only rather slowly within the atmosphere, thus their concentrations are rather low close to the sources of precursors but are higher at large downwind distances. A consequence of the long-range transport phenomenon is that for a country such as the United Kingdom where airmasses may arrive either from maritime or continental sources, by far the highest concentrations of long-lived secondary pollutants arrive in airmasses traveling long distances from continental source regions, typically 1 day or more away. Figure 7 shows the relationship of sulfate concentrations to airmass origins for sites in the United Kingdom. In Scandinavia, which has always had relatively low emission of sulfur and nitrogen oxides, there have been serious acid rain problems due to long-range transport of sulfuric and nitric acids in airmasses arriving from all over Europe. Emissions from the United Kingdom have been a particular problem due to the prevailing southwesterly circulation in the North Atlantic region, tending to transport pollutants from the United Kingdom to Scandinavia, and the very low concentrations of neutralizing ammonia over both the North Sea and over Scandinavia itself.

V. INDOOR AIR QUALITY

At the other extreme of spatial scale, the indoor environment represents a series of distinct microenvironments. Given that the majority of people spend the majority of their time indoors, air quality in these microenvironments plays a major part in determining an individual's exposure to air pollutants. Indoor levels of outdoor air pollutants generally track the outdoor level of the pollutant, with any lag determined chiefly by the ventilation rate of the room. Levels in the indoor environment may be attenuated by losses during entry and by the presence of sinks. Ozone reacts rapidly with a wide range of indoor surfaces, and so is rapidly depleted, levels of solar radiation indoors being insufficient to regenerate it. Particles, too, can sediment out of the atmosphere in the relatively still indoor air, as well as impacting on walls and ceilings. On the other hand, indoor air quality is adversely affected by the presence of polluting activities indoors, the products of which can accumulate in the relatively stagnant indoor air. The presence of people in a room is sufficient to stir up settled dust. Particle monitors attached to people read higher particle levels than stationary monitors in an occupied room. This difference, termed the "personal cloud," can be of significance when trying to quantify the exposure of an individual to particulate matter. Combustion, for cooking and heating, is a significant source of indoor pollutants. Levels of nitrogen oxides and particles rise significantly during cooking. The problem is particularly severe in some developing countries where inefficient stoves burning "dirty" fuels such as wood can give rise to astronomical levels of atmospheric particles. Even electric cookers can generate particles through the direct pyrolysis of food. Carbon monoxide levels are also elevated by combustion indoors. With poorly maintained or inadequately ventilated equipment these levels can become dangerously high, and a number of people die each year from carbon monoxide poisoning. The radioactive gas radon is naturally released from the ground in areas with a predominance of certain types of rock, notably granite. It is normally dispersed and diluted by the wind. However, radon can enter houses built in such an area via seepage into the basement, and without the normal means of dispersion it will accumulate. Radon itself is an unreactive noble gas. Its decay products, however, which are also radioactive, bind readily to atmospheric particles. In this form they can deposit in the lungs of the occupants, accounting for the observed increase in lung cancer incidence in radon-affected areas. Houses in such areas may be protected either actively, by installing extractor fans in the basement area, or passively, by installation of impermeable membranes to the basement floor.

VI. EFFECTS OF AIR POLLUTANTS

A. Effects on Human Health

Exposure to air pollutants is liable to cause a range of adverse effects upon human health. The health outcome associated with a specific pollutant is a function of the mechanism of toxic action of that pollutant. Mechanisms of toxic action include the following:

(a) *Inflammation of lung tissues*. Reactive chemicals such as sulfur dioxide and ozone cause inflammation of the surface tissues of the lung, as evidenced by studies of lung fluids, which show the release of inflammatory mediators and cells. Sulfur dioxide causes a narrowing of the upper conducting airways, most marked in asthmatic subjects, which can lead to breathing difficulties. In the case of ozone, the effects occur also within the deep lung, leading to a reduction in lung capacity. For ozone, this appears to be reversible upon cessation of exposure and people frequently exposed to ozone appear to develop some form of tolerance. A small proportion of people show a markedly higher sensitivity to ozone than the bulk of the population and that proportion is similar among asthmatics and normal subjects. Other pollutants which appear to act through inflammatory mechanisms include nitrogen dioxide and particulate matter, although there are additional mechanisms in the case of the latter.

(b) *Effects via the cardiovascular system*. One of the findings of epidemiological research going back to the December 1952 London smog has been that many of the deaths during or following air pollution episodes are from heart disease rather than respiratory disease. More recent epidemiology in relation to exposure to particulate matter has also attached considerable significance to heart disease as a consequence of particle exposure. The mechanisms are not totally clear, although there is accumulating evidence that exposure to ultrafine particles can increase the viscosity and coagulability of the blood, increasing the risk of heart attacks. Pollutants such as particles and sulfur dioxide can also influence the pattern of heartbeats.

(c) *Oxygen deprivation*. Carbon monoxide binds with hemoglobin in the blood to form carboxyhemoglobin, which limits the ability of the blood to transport oxygen. This in turn puts pressure on the cardiovascular system, exacerbating the symptoms of those with preexisting heart disease.

(d) *Carcinogenesis*. A number of common air pollutants are known to be carcinogenic in humans or in experimental animals. The best known example is that of benzene, which is associated with an increased risk of leukemia. Polycyclic aromatic hydrocarbons (PAH), which are ubiquitous products of combustion processes, are associated with an increase in the incidence of lung cancer. Such effects frequently require long induction periods and have been elucidated only through studies of occupationally exposed workers.

Epidemiological studies linking air pollution exposure with mortality and disease indicate that air pollution exposure has both short-term and long-term effects. The short-term effects are quantified through correlating day-to-day changes in air pollutant concentrations with day-to-day changes in death rates and hospital admissions within specific disease categories, allowing for a lag of up to 5 days between exposure and effect. Such studies have linked particulate matter, sulfur dioxide, nitrogen dioxide, and carbon monoxide with short-term effects, although due to correlations between the concentrations of the various pollutants, there is some residual doubt over whether all of these pollutants are individually responsible for the effects observed in the population. The evidence for long-term effects of air pollution exposure on morbidity and mortality is primarily in the area of exposure to particulate matter. It is expected that long-term exposure to chemical carcinogens would lead to an increase in cancer incidence, but this is very hard to detect in the general population against the background level of cancer due to a wide range of other causes such as smoking.

B. Effects on Vegetation

Exposure of vegetation to air pollutants may produce a range of symptoms in plants, including leaf blemishes, inhibition of growth, and increased susceptibility to other stresses such as disease or insect attack. Where these symptoms occur on commercially grown crops or in recreational areas they may have a significant adverse economic and/or aesthetic effect. In ecologically sensitive areas they may lead to a reduction in biodiversity through the loss of sensitive species. The levels of symptoms in such sensitive species are sometimes used as “biomonitors” for the presence of air pollutants. For example, the particularly susceptible tobacco variety Bel-W3 is used as an indicator for the concentrations of ground-level ozone. Accumulation in food plants via wet or dry deposition offers a significant pathway for air pollutants to enter the food chain, and the rate at which contaminant material is transported from the atmosphere into vegetation is a crucial parameter in food chain contamination models such as those used to predict the effects of chemical or radioactive emissions.

Air pollutants can affect plants both directly and indirectly. The principal phytotoxic air pollutants are ozone, nitric oxide, nitrogen dioxide, sulfur dioxide, fluorides, and peroxyacyl nitrate (PAN). Their toxic effects have been demonstrated convincingly through field measurements and controlled exposure studies. Many of the mechanisms giving rise to the observed effects, however, have not yet been fully elucidated. Research suggests that in many of these cases the link between cause and effect is complex, involving interactions among several processes. The most significant of the indirect effects on plants arise as a consequence of acid deposition.

(a) *Direct effects*. Plants can experience direct toxic effects from contact with various gaseous air pollutants. These generally enter the leaves through the same stomatal pores that allow normal metabolic gas exchange. Pollutants can also be adsorbed onto the waxy cuticle covering the leaf surface, which exhibits cation exchange characteristics and so can actively accumulate contaminants. However, such sorbed contaminants are unlikely to permeate the plant, and so are significant mainly as potential food-chain contaminants rather than direct toxins. Water-soluble gases dissolve in the film of water within the stomatal cavity, where they may be incorporated into existing metabolic pathways within the plant and shift reaction equilibria in such a way as to cause an accumulation of harmful intermediate compounds. For example, direct absorption of nitrogen dioxide can lead to accumulation of toxic nitrite ions. Nitric oxide is also directly toxic to plants. There is some evidence that once dissolved within the plant cell, it can bind to a key site involved in photosynthesis, impairing its function. Sulfur dioxide in high concentrations can also inhibit photosynthesis, while at low concentrations it can interfere with the ability to translocate sugars from the leaves, where they are produced, to the rest of the plant. Fluorine-containing gases, especially hydrogen fluoride, can inhibit photosynthesis even at low concentrations. Ozone produces a series of symptoms associated with oxidative stress. As with the other gases, it enters the leaf via the stomata and dissolves, decomposing to give damaging radical species such as hydroxyl, peroxy, and superoxide. These can cause damage to cell components. One common plant response to such damage is to release the gaseous plant hormone ethylene, and this may itself react with gaseous ozone in the stomatal cavity to yield highly reactive hydroperoxides which can cause further damage to the plant. Symptoms include reduction of growth at lower concentrations, with visible damage at higher concentrations.

(b) *Indirect effects*. Acid deposition to forests results in yellowing (chlorosis) of leaves, decreasing frost tolerence, increasing susceptibility to pest and disease attack, and eventually the death of trees. In the laboratory little direct damage to leaves is observed until simulated rain pH is lowered to around 3. While such acidic rain does occur on occasions, it is unlikely that this is sufficient to explain the incidence of damage. The symptoms described above are also characteristic of nutrient deficiency, and this, together with the observation that the observed effects are worse where soils are already of low nutritional status, has led to the hypothesis that much of the damage to plants from acid deposition arises through nutrient deficiency caused by soil acidification. Acidification occurs when the rate of acid deposition exceeds the buffering capacity of the soil. It results in mobilization of normally soil-bound cations such as magnesium, calcium, zinc, and aluminum, which are then leached rapidly from the soil. Acidic conditions may also inhibit the ability of plants to take up those nutrients which are not leached. Other indirect effects of air pollution on vegetation are to render the plants more susceptible to attack by pests and pathogens. Again, exact mechanisms are unclear, but are thought to involve the premature activation of the plant's stress response.

C. Effects on Materials

The most visible impact of air pollution on our material surroundings has for a long time been the blackening of buildings in industrialized cities and towns. In many cases these buildings have been subjected to decades of exposure to high levels of pollutants, especially smoke and sulfur dioxide, arising from coal burning. The effects of this exposure have been more than cosmetic, leading to serious structural damage. Limestone (calcium carbonate) is particularly susceptible to attack by sulfur dioxide. The resulting calcium sulfate is relatively soluble and can be leached away by rainwater. Metals, too, can be affected by air pollutants. The presence of sulfur dioxide accelerates the corrosion of ferrous metals and of the protective zinc coatings used in galvanizing, while increased corrosion of copper and silver contacts in switch gear can increase their electrical resistance. Since the introduction of measures to control coal burning in urban areas this type of damage has lessened considerably, though smoke from vehicles can still lead to soiling in heavily trafficked areas. Both ozone and nitrogen oxides can lead to the fading of fabric dyes, while rubber is subject to cracking in the presence of ozone, which attacks unsaturated carbon–carbon bonds in the polymer chains of which it is composed. Most rubber goods now contain various countermeasures against ozone cracking, such as added antioxidants or the use of more highly saturated polymer subunits.

D. Effects on Visibility

Most common gaseous air pollutants are invisible, although nitrogen dioxide has a distinct brown color. However, air pollution can have a significant influence on visibility. The principal cause of this is light scattering by suspended particles. Scattering is strongest by particles whose diameters are similar to the wavelengths of visible light, around 200–800 nm. This coincides with the accumulation mode of atmospheric particles, and so "haze" caused by such particles under polluted conditions can build up rapidly and be slow to disperse. In the past, extreme episodes of poor visibility occurred during urban smogs, as epitomized by the London "pea-souper" fogs described to such dramatic effect in Victorian novels. The reduction of urban coal burning in the latter part of the 20th century greatly reduced the occurrence of these, but "photochemical smog" episodes can still lead to a noticeable reduction in visibility. More recently, concern has arisen over the reduction of visibility over longer ranges, in the order of kilometers, in the vast scenic national parks of the United States. This concern has been sufficient for legislation to be introduced which recognizes visibility as an air quality criterion.

VII. AIR QUALITY MANAGEMENT

A. Legislative Strategies for Local Air Quality Management

Most countries have put in place legislative frameworks to control and improve the quality of air. The general pattern of such a strategy appears in Fig. 8. While air quality management strategies have supranational, national, and local components, the final task of managing air quality is usually devolved to the local level, typically that of a city, county, or state.

The key elements of such a strategy include the following:

(a) *Monitoring*. High-quality air pollution monitoring data are central to any air quality management strategy. Such data provide hard evidence of the degree of need for air quality management as well as providing a record of progress in improving the quality of the air. Monitoring networks can also be used to provide public alerts and information when serious episodes are predicted or occurring.

(b) *Air quality standards and objectives*. In most countries a distinction is made between air quality standards and air quality objectives. Air quality standards are generally health-based guidelines which seek to establish the concentrations of air pollutants to which the public can be exposed throughout their lifetime without significant adverse effects at a population level. This does not imply total safety for all members of the population. Such standards

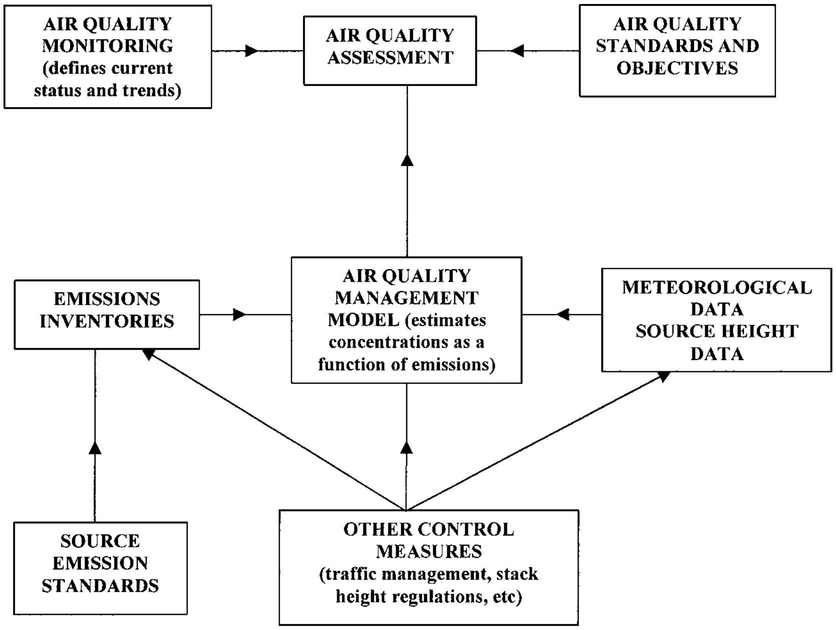


FIGURE 8 Key elements of an air quality management strategy.

TABLE II U.S. National Ambient Air Quality Standards

Pollutant	NAAQS concentration ppm	NAAQS concentration μg/m^3	Standard type[a]
Particulate matter $\leq$10 μm (PM$_{10}$)			
24-hr average		150	Primary and secondary
Annual arithmetic mean		50	Primary and secondary
Particulate matter $\leq$2.5 μm (PM$_{2.5}$)[b]			
24-hr average		65	Primary and secondary
Annual arithmetic mean		15	Primary and secondary
Ozone (O$_3$)			
1-hr average	0.12	235	Primary and secondary
8-hr average[b]	0.08	157	Primary and secondary
Sulfur dioxide (SO$_2$)			
24-hr average	0.14	365	Primary
Annual arithmetic mean	0.03	80	Primary
3-hr average	0.50	1,300	Secondary
Nitrogen dioxide (NO$_2$)			
Annual arithmetic mean	0.053	100	Primary and secondary
Carbon monoxide (CO)			
1-hr average	35	40,000	Primary
8-hr average	9	10,000	Primary
Lead (Pb)			
Quarterly average		1.5	Primary and secondary

[a] Primary standards relate to the protection of human health; secondary to amenity and the environment.

[b] At the time of writing these standards have not been implemented.

TABLE III Air Quality Standards Recommended by the U.K. Expert Panel on Air Quality Standards

Pollutant	Concentration	Averaging time
Particulate matter (PM_{10})	50 μg m^{-3}	24-hr running
Sulfur dioxide	100 ppb (266 μg m^{-3})	15 min
Nitrogen dioxide	150 ppb (286 μg m^{-3})	1 hr
Ozone	50 ppb (100 μg m^{-3})	8-hr running mean
Carbon monoxide	10 ppm (11.6 mg m^{-3})	Annual mean
Benzene	5 ppb (16.25 μg m^{-3})	1-year running mean
1,3-Butadiene	1 ppb (2.25 μg m^{-3})	1-year running mean
Polycyclic aromatic hydrocarbons	0.25 ng m^{-3} benzo(*a*) pyrene	Annual mean

cannot be totally protective, but risks to all but the most susceptible individuals should be negligible at concentrations below the air quality standard. Air quality standards determined by the U.S. Environmental Protection Agency appear in Table II and by the U.K. Expert Panel on Air Quality Standards in Table III. These are set without reference to economic considerations or the practicability of achievement. Many countries therefore adopt air quality objectives which define targets for air quality improvement by a specific target date. Thus, in the United Kingdom the air quality objectives set in the year 2000 provide targets for the years 2003–2008 for various pollutants, including total compliance with the air quality standards for a number of pollutants and percentile compliances in the case of others.

(c) *Source emission standards*. These are set for both mobile sources (i.e., road vehicles) and stationary sources such as industry and power plants. It is necessary to set emissions standards in terms, for example, of grams of pollutant emitted per standard test drive cycle for a vehicle, or kilograms per hour of pollutant from a stationary source. Such standards are usually set by regulatory authorities on the basis of the best available technologies, taking account of affordability. Due to technical innovations, such standards become stricter with the passage of time, which has been notable especially for road vehicles.

(d) *Air quality management models*. Numerical models on the scale of a conurbation or region provide the essential predictive link between emissions and airborne concentrations. Such models include information on the emissions from all known sources, details of pollutant inputs from outside of the model domain, as well as a capability to predict the dispersive properties of the atmosphere. Once the model has been established and validated with monitoring data, it can then be used to predict the impact on air quality of different control strategies.

(e) *Other control measures*. There are many other measures which can be taken by regulatory authorities to impact air quality. These include the regulation of chimney heights, the siting of industrial sources, and the management of road traffic. The informed application of such procedures requires the use of predictive modeling tools, including road traffic models.

B. Supranational and National Air Quality Management Measures

Many key decisions on air quality management are best taken at the national or international level. Air pollutants have no respect for national boundaries and therefore international conventions on long-range transport of air pollutants have been established in a number of parts of the world, which have led to international agreements to limit pollutant emissions. Such agreements subsequently require national implementation effectively to allocate the emissions to different industrial sectors or even to specific companies or operating sites. Emissions standards for road vehicles are generally set on a national (e.g., United States) or international (e.g., European Union) level, although in the United States there are provisions in the legislation to allow individual states to set their own vehicle emission regulations provided they are no less strict than the Federal regulations.

SEE ALSO THE FOLLOWING ARTICLES

Aerosols • Atmospheric Turbulence • Cloud Physics • Combustion • Environmental Geochemistry • Environmental Measurements • Environmental Observation and Forecasting Systems • Pollution Control • Pollution, Environmental • Soil and Groundwater Pollution • Transport and Fate of Chemicals in the Environment • Water Pollution

BIBLIOGRAPHY

Boubel, R. W., Fox, D. L., Turner, D. B., and Stern, A. C. (1994). “Fundamentals of Air Pollution,” 3rd ed., Academic Press.

de Nevers, N. (2000). “Air Pollution Control Engineering,” 2nd ed., McGraw-Hill, New York.

Fenger, J., Hertel, O., and Palmgren, F. (eds.). (1998). “Urban Air Pollution—European Aspects,” Kluwer, New York.

Harrison, R. M. (ed.). (2001). "Pollution: Causes, Effects and Control," 4th ed., Royal Society of Chemistry, London.

Hester, R. E., and Harrison R. M. (ed.). (1995). "Volatile Organic Compounds in the Atmosphere," Royal Society of Chemistry, London.

Holgate, S. T., Samet, J. M., Koren, H. S., and Maynard, R. L. (eds.). (1999). "Air Pollution and Health," Academic Press, New York.

Jacobson, M. Z. (1999). "Fundamentals of Atmospheric Modelling," Cambridge University Press, Cambridge.

Oke, T. R. (1987). "Boundary Layer Climates," 2nd ed., Routledge, London.

Seinfeld, J. H., and Pandis, S. N. (eds.). (1998). "Atmospheric Chemistry and Physics. From Air Pollution to Climate Change," Wiley, New York.

Pollution, Environmental

M. E. Baur
University of California, Los Angeles

GLOSSARY

Biological oxygen demand (BOD) Quantity of oxygen used by microorganisms in a water sample during a standard time interval, usually 5 days. Also called biochemical oxygen demand.

CFCs Abbreviation for chlorofluorocarbons, small aliphatic organic molecules containing the elements carbon, hydrogen, fluorine, and chlorine. These substances have been widely used as refrigerant and propellant gases. They are typically described in terms of their CFC number, which is obtained by subtracting 90 from the three-digit number composed by giving the number of carbon, hydrogen, and fluorine atoms in the molecule in sequence. Thus the CFC number for the compound difluoromethane, CH_2F_2, is $122 - 90 = 32$.

Criteria pollutant Any of the six airborne substances for which National Ambient Air Quality Standards have been established by the Clean Air Act, standards which the Environmental Protection Agency has the responsibility to enforce. At present, the six substances are CO (carbon monoxide), SO_2 (sulfur dioxide), PM-10 (particulate matter with diameter $\leq 10\,\mu$m), lead, NO_2 (nitrogen dioxide), and O_3 (ozone).

Curie (Ci) Unit of intensity of radioactivity, defined as 3.7×10^{10} nuclear disintegrations per second. The SI unit of radioactivity is the becquerel (Bq), defined as 1 nuclear disintegration per second. $1\text{C}_\text{i} = 3.7 \times 10^{10}$ Bq.

Dobson unit (DU) Unit used to describe the amount of ozone present in the atmosphere. The amount of ozone in DU is the thickness in units of 10^{-3} cm of a layer into which the ozone would be compressed if under a pressure of 1 atm. at 0°C.

Free radical Molecular fragment having one or more unpaired electrons. Free radicals are usually highly reactive.

Half-life Time required for half of a set of objects to undergo a specified change. Especially used for the time required for half of a set of unstable nuclei to undergo radioactive decay.

K_{ow} Octanol–water partition coefficient; the ratio of concentration of a substance in octanol solution to that in water solution when the two are in equilibrium. Usually reported as the base 10 logarithm, $\log K_{ow}$.

Quad Unit of world energy generation, a quadrillion or 10^{15} btu (1 btu, or British thermal unit, equals to 1.054 kJ).

Residence time Average time a given chemical species spends in a reservoir. Alternatively the time required for the amount of the species originally in a reservoir to drop to 1/*e* of its initial value if there is no replenishment.

ENVIRONMENTAL POLLUTION deals with the modification of the chemical environment of the earth by human activity; its study encompasses the sources, distribution, and effects of pollutants, that is, chemical substances introduced into the environment in locations and/or at levels not observed in the absence of human intervention. Environmental pollution as a discipline originates in the recognition that the terrestrial soil–hydrosphere–atmosphere–biosphere system is an interlinked set of reservoirs with complex connective pathways, so that modification of chemical conditions in one region can propagate in unintended ways into others, producing unexpected and undesirable effects.

I. INTRODUCTION

Manipulation of the chemical environment by humans may be said to have begun with the domestication of fire. Use of toxic heavy metals, most notably lead and mercury, is believed to have caused deleterious effects on human health in some ancient societies, and deterioration of air quality in medieval urban settings due to use of heavily polluting fuels (coal and wood) is documented. Profound large-scale modification of the chemistry of the earth's soils, waters, and atmosphere began with the industrial revolution of the eighteenth century and accelerated with the development of chemical and nuclear technology such that now no portion of the earth's surface is entirely free of the chemical signs of human activity; indeed, Antarctic ice has proven to be a useful substrate for analysis of historical trends in worldwide use of heavy metals.

The first task of the specialist in environmental pollution studies is identification of the chemical substances that cause deleterious effects. This is followed by an effort to identify the pathways by which these substances have reached their locus of action, to find their sources, and finally to eliminate or minimize those sources. The tools for identification and measurement of the amounts of such substances are those of analytical chemistry, namely, *spectroscopy*, *mass spectrometry*, *gas and liquid chromatography*, *differential thermal analysis*, and *neutron activation analysis*. These subjects are discussed fully elsewhere in this encyclopedia.

Some idea of the scope of human activities that affect the environment can be obtained from Fig. 1, in which data collected by the Environmental Protection Agency (EPA) on the major sources of groundwater pollution are

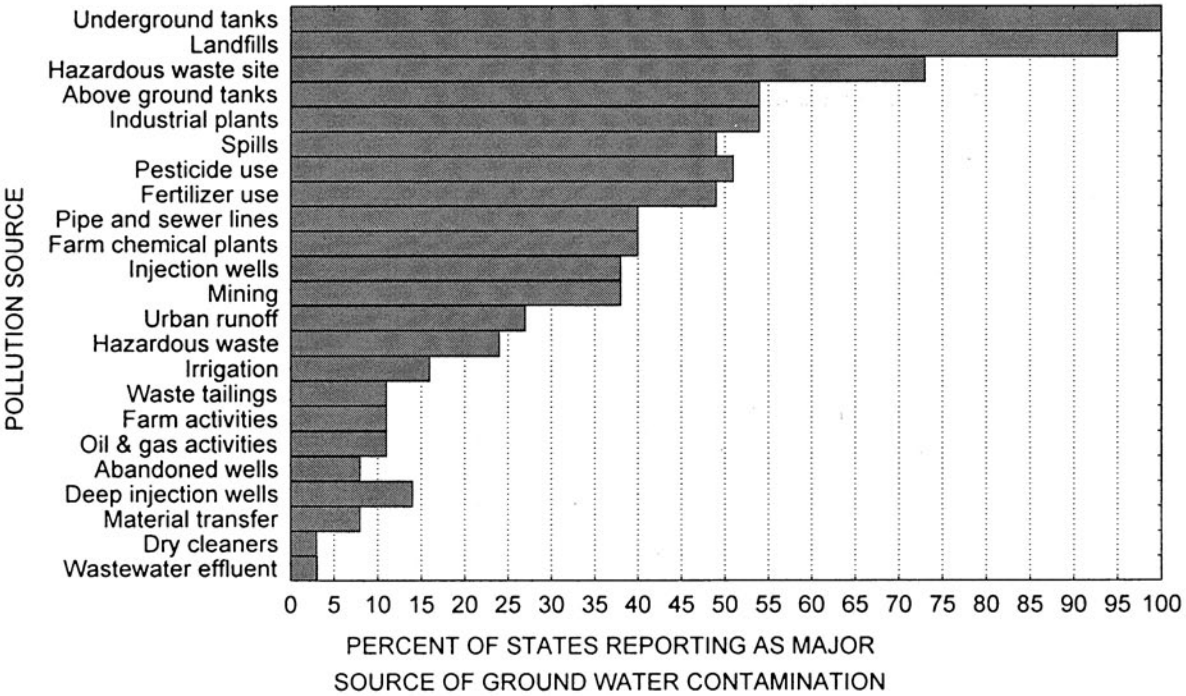


FIGURE 1 Major sources of groundwater contamination as reported by states to the EPA.

replotted as the percentage of states citing each category of activity as a major source. Many of the major known sources of pollutant chemical species are summarized in Section II. This is followed by a survey of the principal reservoirs into which the terrestrial environment is conventionally divided—the atmosphere, the hydrosphere, and the biosphere—with a description of the chemical perturbations due to human activity of known major importance in each. Issues connected with toxic waste disposal in dump sites are touched on in the last section.

It should be understood that a complete and final catalog of pollutants and their effects cannot be given; new substances are constantly entering the terrestrial chemical network, and negative effects often become apparent only after a considerable time. The number of chemical substances for which reasonably complete environmental analysis is available remains a relatively small proportion of the whole in spite of the best efforts of environmental chemists and engineers. Availability of information and data on pollution matters has, however, been much enhanced by the proliferation of environment-related sites on the Internet; there are far too many to attempt to list or categorize here, but the most comprehensive is probably that of the Environmental Protection Agency (http://www.epa.gov).

There are no foolproof procedural algorithms in the field of pollution studies; environmental scientists must employ all their skills to foresee the consequences of even seemingly benign or neutral alterations in the chemical environment, and each pollutant substance has its own particular chemical features. That such foresight is possible was demonstrated by the prediction of deleterious effects on the earth's ozone layer due to the use of CFCs (see Section III.C.1) well in advance of the observation of such effects; for this work Crutzen, Molina, and Rowland were awarded the Nobel Prize in chemistry in 1995.

II. SOURCES OF CHEMICAL POLLUTANTS

A. Processes Related to Energy Generation

1. Combustion

Combustion (the rapid reaction of fuel substances with oxygen) has been employed for warmth, cooking of food, and materials processing since prehistoric times, but it assumed particular importance with the development of the heat engine in the eighteenth century. A heat engine is a cyclic device by means of which a quantity of heat generated in a high-temperature source is extracted and partially converted to useful work (a portion of this heat, usually 60–70% of it, is necessarily lost to a cold reservoir). The efficiency of a heat engine, that is, the ratio of useful work generated to total heat extracted, increases with the temperature difference between the hot and cold reservoir, other things being equal; it is thus advantageous to operate an engine cycle with as high a temperature as possible in the hot reservoir. This is the principal reason why the classical steam engine, the key to the industrial revolution, is now rarely employed; it has a maximum operating temperature not much in excess of 100°C and a typical efficiency of about 8%, whereas much higher peak temperatures and efficiencies (25–45%) are achieved in internal combustion engines (automobile or Otto engine, diesel engine, turbine).

The chemical process of combustion can be represented as

$$\mathrm{CH}_x\mathrm{O}_y + (x/4 - y/2 + 1)\mathrm{O}_2 \rightarrow \mathrm{CO}_2 + (x/2)\mathrm{H}_2\mathrm{O},$$

where $\mathrm{CH}_x\mathrm{O}_y$ is the generic stoichiometric formula for a fuel molecule. For high-grade coal, $x, y \approx 0$; for petroleum hydrocarbons, $x \approx 2$ and $y \approx 0$; for cellulose (wood, biomass), $x = 2, y = 1$; for ethyl alcohol, $x = 3, y = {}^1\!/_2$; for natural gas (methane), $x = 4, y = 0$. The amount of heat from combustion per mole of carbon varies substantially between these fuels: about 300 kJ for coal, wood, and biomass; 600 kJ for petroleum and ethyl alcohol; 900 kJ for methane. In the simplified scheme above, combustion appears innocuous, because both CO_2 and H_2O are normal components of the atmosphere. However, the amount of CO_2 thus generated and deposited in the atmosphere, principally since 1900, is believed to be sufficient to substantially modify earth's heat balance (Section III.C.1). In 1990, world production of energy by combustion of fossil fuels (coal, petroleum, natural gas) was 295 quads; this was expected to rise to about 365 quads by 2000; estimated emission of CO_2 to the atmosphere from this energy production was 5×10^{14} moles in 1990 and was expected to be about 6×10^{14} moles in 2000 (these emission factors correspond to a mean emission rate of 1 mole carbon as CO_2 per 590 kJ of heat produced). Also, engine combustion is not complete, and varying amounts of CO (carbon monoxide), HC (residual hydrocarbons), and elemental C (particulate carbon, or soot) are released in the exhaust gases. The CO is directly toxic, because it bonds to hemoglobin, the oxygen-transporting pigment in blood, displacing oxygen and resulting in asphyxiation. The volatile organic compounds (VOCs) in the HC mix enter the gas phase directly, whereas less volatile compounds may attach to particles (soot, dust, ash, aerosols) and remain in the atmosphere for a considerable time. The elemental C in the exhaust gas adds to the background burden of particulates. Inhalation of the HC species, whether as VOCs in the gas phase or adsorbed species on particulates, can be directly harmful; further many of them contribute to the complex chemistry of smog (Section III.C.1).

Two other characteristics of combustion are of importance for environmental pollution:

1. At the elevated temperature encountered in some internal combustion engines, most notably the Otto cycle engine employed in conventional automobiles, the major constituents of air, N_2 (nitrogen) and O_2 (oxygen), partially react to form NO, nitric oxide. NO together with other oxides of nitrogen are collectively termed NO_x; they are important in the development of photochemical smog and acid rain, as well as being undesirable components of the atmosphere in their own right.
2. Fossil fuels contain impurities that enter the atmosphere on combustion and cause deleterious effects. The most important of these is sulfur, which occurs in coals and petroleum hydrocarbons at levels up to several percent by weight; combustion yields the gaseous species SO_2 and other oxides of sulfur, collectively denoted SO_x. Atmospheric chemical reactions convert SO_2 to SO_3 and then, in the presence of water, to H_2SO_4, sulfuric acid, a principal component of acid rain, snow, and fog (Section IV.B). In addition to natural impurities, fuel additives whose function is to increase the smoothness and efficiency of engine operation may be present. The most notable such additive from the standpoint of pollution is lead, in the form of the compound $Pb(C_2H_5)_4$, lead tetraethyl, added to gasoline to increase octane (antiknock) rating. Lead from "ethyl" gasoline is believed to have been a major contributor to the lead burden in the bodies of urban residents in developed countries. During the 1980s its use was gradually discontinued in the United States, where it is now prohibited.

Various types of internal combustion engines (gasoline or Otto cycle, diesel, turbine or Joule-Brayton cycle) have assumed an important role in modern technology because of their relative simplicity and high efficiency. Efforts to reduce the levels of atmospheric pollution resulting from their operation encounter the obstacle that increased efficiency and reduction of all pollutant species in exhaust tend to be mutually inconsistent. For example, in the standard Otto cycle engine of gasoline-powered vehicles, a high operating temperature favors efficiency but also increases NO_x production; manipulation of the air–fuel mixture ratio to reduce NO_x tends to augment HC and CO levels, as shown in Fig. 2. The diesel engine has a longer period of fuel burn than does the standard Otto cycle engine and hence is able to have higher efficiency at a lower maximum operating temperature and produce relatively little NO_x; for this reason diesel-powered vehicles have been able to escape the severe emission control regulations applied to gasoline-powered vehicles in the United States. However the slow fuel burning in diesel engines, which extends through much of the expansion phase of the engine cycle, results in the formation of oxygen-poor zones in which fuel pyrolysis occurs, generating large quantities of particulate carbon (soot) in the exhaust. Because this soot has the capacity to adsorb carcinogenic molecules from the exhaust and transport them into the human lung, it appears likely that stringent control of diesel emissions will ultimately be undertaken.

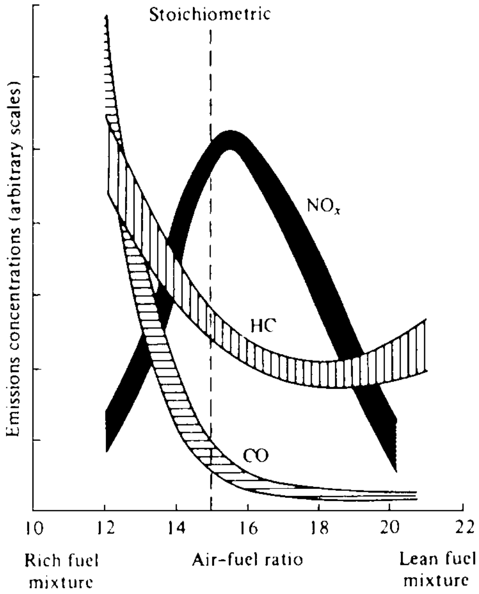


FIGURE 2 Effect of air–fuel ratio on exhaust emissions. [From Hodges, L. (1977). "Environment Pollution," 2nd ed. Holt, Rinehart & Winston, New York. Reproduced by permission of source, Battelle Memorial Laboratories, Columbus, Ohio.]

Considerable improvement in the amounts of NO_x, CO, and HC released to the environment by gasoline-powered vehicles has been achieved by improved engine design, modification of fuels (maintaining octane levels by increasing the amount of branched aliphatic hydrocarbons, while eliminating lead tetraethyl and reducing the amounts of olefin and aromatic compounds present), and the introduction of catalyst devices in the exhaust pathway, which increase the oxidation of CO and HC and promote the decomposition of NO back to N_2 and O_2. Figure 3 illustrates trends in the release of five of the criteria pollutants—the original criteria pollutant list included VOCs, since replaced by ozone, O_3—between 1970, when the imposition of stringent emission control measures was in its initial phases, and 1981 when these measures were well established. Important improvement in the emissions of VOCs and CO (largely from vehicles) and of SO_x and particulates (largely from stationary sources such as power plants and factories) is evident. However, NO_x levels continued to rise during this interval. Subsequently, introduction of additional control measures succeeded

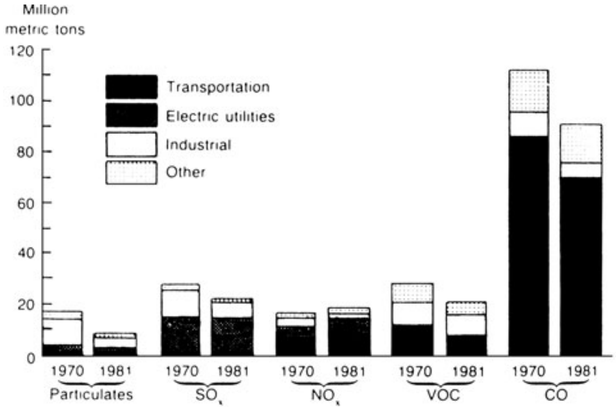


FIGURE 3 Air emissions by source and type of pollutant, 1970 and 1981. [From Conservation Foundation. (1984) "State of the Environment, an Assessment at Mid-Decade," Washington, D.C. Reproduced by permission of the Conservation Foundation.]

in first holding NO_x production constant and then reducing it somewhat, but this remains a serious problem area.

2. Nuclear Reactions

The development of techniques for the controlled release of nuclear energy in reactors after World War II was regarded as an important advance in the direction of nonpolluting energy production. Unfortunately, new environmental problems have arisen from the operation of such reactors; these include the release of radioactive materials in leak episodes or occasional catastrophic accidents, and the disposal of waste materials.

The only nuclear process that has been found practicable for energy production so far is the fission of a radioactive atomic nucleus, either naturally occurring (^{235}U) or artificially generated (^{239}Pu), to release two or three neutrons whose capture by other nucleii induces further energy-releasing fission events in a chain reaction:

$$^{235}\mathrm{U} + {}^{1}\mathrm{n}\ (\text{neutron}) \rightarrow \text{two daughter nucleii} + 3\,{}^{1}\mathrm{n} + \text{energy}.$$

The *fuel* in a fission reactor, typically mainly the oxide of nonreacting ^{238}U somewhat enriched to contain several percent of ^{235}U, is contained in rods jacketed with a *cladding* (steel or zirconium). For capture of neutron to occur efficiently, they must be slowed (thermalized) by a *moderator* (water, heavy water, graphite) that surrounds the rods. *Control rods* of a neutron absorber (boron, cadmium) remove enough of the thermal neutrons to hold the reactor in a steady state. The heat released is removed from the *reactor core* by a *coolant*, usually pressurized water in U.S. reactors. Heat is transferred from this primary coolant to water in a secondary loop outside the core, where it then operates as the heat source in a standard turbine engine, generating electricity.

Fission of a ^{235}U nucleus releases about 200 MeV (million electron volts) or 3.2×10^{-11} J of energy, of which about 30% can be captured as useful work. A reactor therefore must consume about 5 moles (1.2 kg) of ^{235}U fuel and produce about 10 moles of radioactive daughter species, per megawatt (MW) of electric power generating capacity per year. The daughter species are mainly short lived and undergo further decay with emission of neutrons, α particles (^{4}He nucleii), β and β^{+} particles (electrons and positrons), and γ radiation (high-energy electromagnetic radiation) to produce longer lived further product nucleii. Irradiation of the moderator, control rods, coolant, and structural material in the reactor core also generates radioactive nuclides; reactor operation thus results in a substantial multiplication of the amount of radioactive material existing on the earth. About 35 tons of spent fuel (^{238}U, remnant ^{235}U, daughter nucleii produced in the fission process) are generated each year by a 1000-MW reactor. This spent fuel is reprocessed to retrieve all usable material, but about 7 tons per year must be disposed of as high-level solid waste, together, eventually, with the irradiated structural and control material from the reactor core.

Short-lived daughter nuclides represent an environmental hazard only where accidental discharge of material from a reactor occurs. Of particular concern in such cases are ^{131}I (half-life 8.07 days) and ^{136}Cs (half-life 13 days), both β emitters that efficiently enter the food chain and are thereby transmitted to animals and humans. Medium- and long-lived species represent a hazard in leakage episodes as well, but are mainly of concern because they must be kept in secure storage for long periods of time. Reactors typically generate about 1700 MCi (megacuries) of such species per 1000 MW of power generated per year, mostly in the form of ^{90}Sr (670 MCi/1000 MW-year; half-life 40.4 years), ^{137}Cs (930 MCi/1000 MW-year; half-life 43.2 years), ^{85}Kr (95 MCi/1000 MW-year; half-life 15.2 years), and ^{3}H or tritium (6 MCi/1000 MW-year; half-life 17.7 years). The total long-term steady-state radiation load from these species, representing a balance between generation and decay, will be approximately 10^5 MCi per 1000 MW of reactor power generated. As of 1998, there were 437 nuclear reactors in operation worldwide, generating a total of 3.51×10^5 MW; 107 of these, generating 1.0×10^5 MW, were in the United States. The long-term steady-state radiation load from these reactors is then approximately 3.5×10^7 MCi. For comparison, the natural background radiation loading of the earth's surface, primarily due to ^{40}K and ^{87}Rb, is approximately 5×10^5 MCi. A roughly similar amount of radiation results from impingement of cosmic rays on the earth. Thus disposal over the earth of the radioactive wastes generated by the present reactor population would increase this radiation burden by a factor of about 35. Evidently the proper storage and containment of these radioactive wastes is an extremely important pollution problem (Section VI). In addition to the nuclear waste resulting from electric power generation, large quantities of radioactive material generated by the nuclear weapons programs of the United States and the former Soviet Union have been released into the environment. In the United States, a total of about 2.6 MCi of radioactivity has been discharged into soils, streams, or surface ponds, or has been injected underground, mainly at Oak Ridge, Tennessee; Hanford, Washington; and Savanna River, Georgia. The situation is far worse in the former Soviet Union where a total discharge of more than 1700 MCi into the environment is estimated, mainly at Krasnoyarsk, Tomsk, and near Chelyabinsk, in Siberia.

In addition to the problem of disposal of the waste they generate, nuclear reactors are subject to unintended releases of radioactivity. If coolant flow through the core fails (a loss-of-coolant accident or LOCA), core temperatures may rise sufficiently to cause fuel and control rods to melt, thus destroying the means of controlling heat generation. Water in contact with the core under such circumstances may react with molten metal, forming an explosive mix of hydrogen and oxygen. Essentially this occurred at the Three Mile Island reactor near Harrisburg, Pennsylvania, on March 29, 1979. While an explosion was averted, the reactor core was partially exposed and emitted radioactive gases and water to the environment. A total of about 0.1 MCi of radioactive gas was emitted, and some of the population in the vicinity probably received radiation doses several times the average yearly dose from natural background radiation. A subsequent excess infant mortality of about 600 individuals has been inferred from public health statistics for Pennsylvania and New York.

A much more serious incident occurred on April 26, 1986, at one of four graphite-moderated reactors at Chernobyl near Kiev in the former Soviet Union (now Chornobyl in Ukraine). The management of the reactor complex had decided to carry out a test of backup power generators as part of the shutdown process of the reactor for refueling; this required that the shutdown procedure be such as to keep the reactor power from dropping too quickly. A crew that expected to carry out this procedure was in place, but unfortunately the test was delayed until late at night when that crew had been replaced by a new crew lacking adequate preparation for the delicate operations involved in the test. The new crew allowed power to drop too rapidly, then attempted to raise the power level back up—in essence, to restart the reactor. To accomplish this they overrode the reactor's safety controls and withdrew the control rods from the reactor core to an excessive extent. This resulted in an uncontrollable runaway production of heat in the core, meltdown of the fuel and control rods, and a massive explosion that blew out the roof of the reactor (which had no containment vessel) and allowed the contents of the core to vent into the environment. The graphite moderator of the core ignited and continued to burn for several days, while release of radioactivity continued. About 50 persons, workers at the reactor and emergency personnel, died of radiation exposure incurred during the incident; eventually the entire region downwind of the reactor was evacuated, but not until several days had passed. Thousands of persons who lived near the plant received large doses resulting in significant increases in radiation-related illness and early death, especially from childhood leukemia. Total release of radioactivity was probably about 50 MCi, which was spread widely through the western portions of the former Soviet Union and northern and central Europe. The runaway nuclear reaction in the melted reactor core was finally brought under control by air-dropping enough borax on the mass to absorb neutrons effectively.

Releases of radiation such as those of Three Mile Island and Chernobyl will no doubt continue to occur from time to time and cause damage to local human populations and ecological networks. A nuclear war would produce a far more serious perturbation of the world environment. Explosion of a 1-megaton nuclear bomb generates 1000–100,000 MCi of radiation, depending on conditions; the employment of, say, 100 such weapons would add 10^5–10^7 MCi to the radiation burden of the terrestrial atmosphere and surface, increasing the average dose of the human population by as much as 20 times, with many individuals receiving much greater amounts than that. The result would be radiation sickness in some individuals and significant gene damage, increase in incidence of cancer, and susceptibility to debilitating infection in all. Similar effects on plants and animals would result in widespread disruption of ecological networks.

B. Processes Related to Mineral Recovery and Processing

Fossil fuels are generated by the slow chemical modification or *maturation* of deposits of organic matter at moderate temperatures (100–200°C) in suitable mineral substrates (usually shales or limestones). The nature of the fuel formed depends on the ratio of carbon, hydrogen, and oxygen in the starting material; organic debris with a high H/C and low O/C ratio, as provided by bacteria and algae, tends to produce petroleum, while that with a low H/C and high O/C ratio from woody plants leads to peat or coals of various rank. Methane (CH_4) gas is formed to some extent in conjunction with both petroleum and coals, but especially as an end product of prolonged maturation. Because all living cells contain nitrogen and sulfur, these elements occur in combination in varying proportions in fossil fuels. A typical approximate stoichiometric formula for a bituminous coal, for instance, is $C_{100}H_{85}S_2N_{1.5}O_{9.5}$.

Coal recovery involves the use of large amounts of water for cooling, washing, and lubrication, followed by discharge as effluent. During recovery and processing a portion of the sulfur content of coal is oxidized to SO_x and dissolves in the effluent, producing sulfurous acid (H_2SO_3) and sulfuric acid (H_2SO_4). Release of such contaminated water in the environment causes large-scale acidification of lakes, streams, and groundwater. The problem is particularly severe with open-pit mining or strip mining unless steps are taken to sequester that part of the sulfur (usually about half) in the form of iron pyrite, FeS_2, which is easily oxidized to SO_x in air.

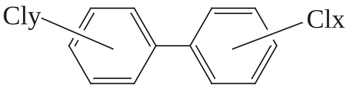
Production and refining of petroleum hydrocarbons release large quantities of pollutant chemicals into the environment. These include saltwater brines, which accompany crude oil in deposits and are pumped out with the latter in wells, various sulfur compounds, and nitrogenated and halogenated hydrocarbons. Refining operations and evaporation of petroleum in transfer and storage release about 7×10^{10} kg of volatile hydrocarbons into the atmosphere per year globally, about 2×10^9 kg in the United States. Transport of petroleum by ship is accompanied by leakage of hydrocarbons into the marine environment, both as a result of ships rinsing their tanks with seawater after emptying, and through accidents resulting in massive spills such as that involving the *Exxon Valdez*, which struck a reef in Prince William Sound, Alaska, on March 24, 1989, and released about 3×10^7 kg (10 million gallons) of crude petroleum into the local waters. Total oil spill amounts are estimated at 5×10^9 kg per year worldwide. The effects of hydrocarbon contamination of water on the biota are not well understood, although prompt kills of fish and birds from massive discharges are frequent and evident. Bacterial degradation of petroleum in natural waters is most efficient with straight-chain aliphatic hydrocarbons and aromatics; thus crude oil spills are probably a smaller long-term perturbation of ecosystems than are spills of refined petroleum enriched in branched compounds.

Ore mining presents many of the same dangers of harmful chemical modification of the environment as does fossil fuel recovery. Sulfur-containing minerals frequently accompany ores and, unless properly recovered and treated, promote the formation of acid waters and contamination of the atmosphere with sulfur dioxide (SO_2) and hydrogen sulfide (H_2S). Acidic sludges and slurries are produced in conjunction with the recovery and processing of iron, copper, zinc, and lead; cyanide salts are used in the recovery of gold and silver and are frequently discharged into the environment.

C. Industrial Processes

Industrial operations are an important source of pollutants in both the atmosphere and the hydrosphere. The principal contributor to atmospheric pollution is combustion, stationary plant combustion being a major source of atmospheric SO_x (Fig. 3). Virtually all industrial operations have an impact on the composition of natural waters; it is not possible here to do more than give an overview. The totality of industry in the United States is estimated to generate 5×10^{10} m^3 (5×10^{13} liters or 1.4×10^{13} gallons) of wastewater annually, containing 8×10^9 kg of solids. The chemical industry is an important source of organic chemicals; these enter the environment mainly by discharge of spent solvents and solvent evaporation. The latter contributes about 10^{10} kg of material worldwide per year.

A particularly troublesome group of such organic chemicals is the polychlorinated biphenyls (PCBs):



which are used as solvents, plasticizers, and insulating fluids because of their exceptional chemical and thermal stability. However, they thus have long persistence in natural waters. They are believed to cause an increased incidence of birth defects and possibly cancer and to adversely affect resistance to disease in animals and humans. Their manufacture in the United States was discontinued in 1977. An especially severe example of PCB pollution is that in the Hudson River in New York State, into which large quantities of transformer oil containing a high percentage of PCBs were dunped in the 1950–1976 period. The PCBs remain adsorbed to bottom sediments and are slowly leaching out, maintaining a concentration of about 10 ppm in river water downstream of the dump zone. Other organic solvents of importance in industry that are known health hazards include benzene (C_6H_6), a skin irritant, carcinogen, and general systemic poison, and tetrachloroethylene or TCE (C_2Cl_4), a mutagen and carcinogen. Mineral processing and plating industries generate wastes containing cadmium (high human toxicity due to its tendency to replace zinc in enzymes), copper (slight toxicity to humans and animals, toxic to plants), cyanides (acute toxins), mercury (acute and chronic toxicity to humans), silver (slight toxicity to humans), and zinc (plant toxin at high levels).

Paper and pulp processing operations have severe impacts on the aqueous environment. The manufacture of paper requires that the cellulose fraction in wood be separated from branched aromatic compounds (collectively termed *lignin*) that stabilize the wood structure. Treatment with solutions of bisulfite ion (HSO_3^-) or a strongly alkaline solution of NaOH and Na_2S is used for this purpose. The resulting effluent is a concentrated dark mix of organic waste, sulfur compounds, and alkali that changes the pH, oxidation–reduction balance, and oxygen content of the waters into which it is released. Textile processing plants have similar, if less pronounced, effects.

Especially important in its impact on human health has been the widespread use of asbestos in construction, insulation, and transport (in brake linings). Asbestos is a generic term for several fibrous minerals. The most common is chrysotile, a serpentine silicate with the approximate composition of $Mg_3Si_2O_5(OH)_4$. A second form is amphibole asbestos; like chrysotile, it is a hydrated chain silicate structure, but it has more varied composition, its principal chemical difference from the latter being that it contains a substantial amount of aluminum ion. The utility of asbestos lies in its resistance to heat and fire and the ease with which it can be fabricated into thin flexible sheets. Mining and processing of the mineral generate airborne microfibers, which are easily inhaled by those working with or using asbestos products. A typical such fiber (length 11 μm) is shown in Fig. 4. These small fibers evade the filtering apparatus of the upper respiratory tract and pass into the alveoli of the lung, where they lodge permanently. Either by chemical action or direct mechanical abrasion (the precise causative mechanism is not established) they cause breakdown of lung tissue, leading to impairment of respiratory function (asbestosis), lung cancer, or in some cases mesothelioma (a cancer of the lung lining that is only known to occur in individuals exposed to asbestos). Although these effects were known or strongly suspected by the 1920s, production and use of asbestos continued to be actively promoted until about 1975. Its use is now greatly reduced in the United States.

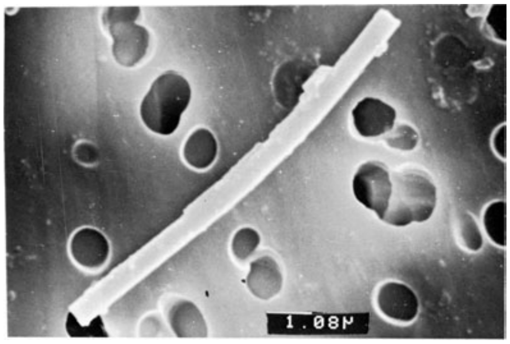


FIGURE 4 Asbestos microfiber, length about 11 μm. Electron micrograph, × 9300.

D. Agriculture

Effects on the chemical environment associated with food production can be broadly divided into two areas: those due to the use of biological promoters (fertilizers) and those due to biological inhibitors (biocides for the removal of unwanted plants, animals, or microorganisms).

The limiting factor in growth of crop plants is usually a low level of one or more soil ingredients—nitrogen, phosphorus, or an essential trace mineral—or of availability of water. Increased yield is achieved by augmenting the limiting ingredient. Common synthetic fertilizers used to increase availability of nitrogen and/or phosphorus in soils include ammonium nitrate (NH_4NO_3), urea ($(NH_2)_2CO$), sodium nitrate ($NaNO_3$), and ammonium phosphate ($NH_4H_2PO_4$) and polyphosphates. Because all of these materials are highly water soluble, a portion is leached from the soils to which they are applied and enters the hydrosphere, where it tends to promote changes in the chemical composition and ecology of lakes, streams, and rivers. Irrigation water frequently contains high quantities of dissolved salts, which remain in the soil and cause loss of fertility; in some cases trace minerals in irrigation water build up to toxic levels. This has occurred with selenium in some areas of the Central Valley of California.

A wide array of biocides is used in agriculture. General systemic poisons such as arsenicals have been largely supplanted by compounds in the chlorinated hydrocarbon, carbamate, or organophosphorus groups for use against insects, soil parasites, and weeds. Structural diagrams of representative members of each group are shown in Fig. 5. The insecticides pass through the insect cuticle (epidermis) and exert their toxic function primarily through disruption of nerve function; many organophosphorus insecticides are related to nerve gases developed for military use in World War II. Figure 5 includes the diagram for the most famous, or infamous, of the chlorinated hydrocarbons, DDT, although this substance can no longer be legally manufactured or marketed in the United States. DDT was first introduced on a mass scale as a delousing agent employed on soldiers and civilians in Naples, Italy, in 1943 to stop a serious typhus epidemic. Its success in this and other wartime applications led to its virtually universal adoption as the insecticide of choice in the United States through the 1940s and 1950s. However the massive introduction of DDT into the environment had disastrous effects on many populations of wild birds; DDT in avian bodies interferes with calcium metabolism and makes it impossible for birds to produce eggs with sufficiently thick shells to withstand incubation. For this reason, use of this

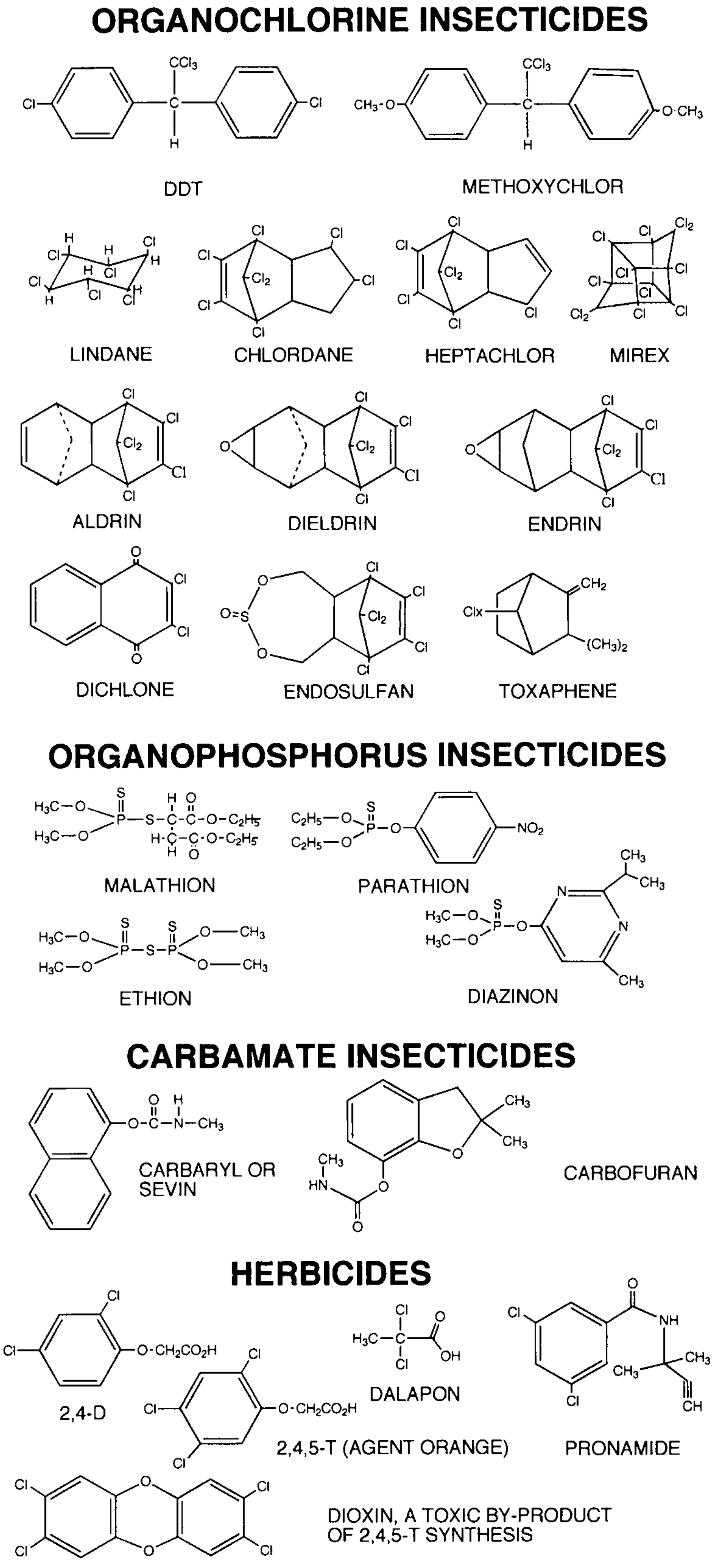


FIGURE 5 Chemical structures of representative biocides.

TABLE I Properties of Some Common Biocides

Category of biocide	Log K_{ow}	Persistence in river water
Chlorinated hydrocarbons		
Aldrin	6.5	80% (2 weeks), 40% (4 weeks), 20% (8 weeks)
Chlordane	5.5	85% (2 weeks), 85% (8 weeks)
Dieldrin	5.3	100% (1 week), 100% (8 weeks)
Endosulfan	3.8	30% (1 week), 5% (2 weeks), 0% (4 weeks)
Endrin	4.6	100% (8 weeks), 80% (16 weeks)
Heptachlor	5.3	25% (1 week), 0% (2 weeks)
Methoxychlor	4.7-5.1	50% (>3 weeks)
Carbamates		
Carbaryl	2.4	90% (1 week), 0% (8 weeks)
Carbofuran	2.3	25% (8 weeks)
Organophosphorus		
Diazinon	2.8	0% (12 weeks)
Malathion	2.4	10% (2 weeks)
Parathion	3.8	50% (1 week), 0% (4 weeks)
Herbicide		
2,4-D	2.6-3.3	50% (1–7 weeks)

chemical has been terminated in the United States. Large deposits of DDT remain in some environmental reservoirs, however, most notably the waters and sediments of San Pedro Bay, south of Los Angeles, due to accumulated waste runoff from one of the former major DDT users. Another example of unintended consequences from the use of a chemical is indicated in Fig. 5; the herbicide 2,4,5-T, or Agent Orange, was employed broadly by the U.S. military during the Vietnam War to destroy mangrove wetlands thought to be employed as sanctuaries by enemy forces. However a by-product of 2,4,5-T synthesis is the extremely toxic compound *dioxin*, and this compound was probably present as a contaminant in much Agent Orange supplied for use in the war. It is believed that many military and civilian personnel involved in the application of Agent Orange or otherwise in proximity to the areas in which it was used have suffered long-term disabling illnesses as a result.

Movement of biocides from the point of application through water runoff into the broader environment is evidently a potential source of ecological disruption and harm to the human population. Most of the organophosphorus compounds undergo rapid biological-assisted breakdown in natural waters and soils, but the halogenated hydrocarbons tend to be much more resistent to degradation. Also, the latter are mostly strongly *lipophilic*, with low solubility in water but high solubility in soils or the lipid component of plants and animals, as indicated by their large log K_{ow} values. Hence the halogenated hydrocarbon pesticides tend to be adsorbed to soils and in the biosphere and to be eliminated into water from these reservoirs only very slowly. Table I shows some of the more common biocides together with their log K_{ow} values and their persistence in river water in terms of percent remaining in the water after the stated time interval since release.

A new form of control of agricultural pests that has become possible with the development of genetic engineering techniques in the 1990s is the incorporation of foreign genes into crop plants, endowing them with the capacity to manufacture chemicals with insecticidal function. The implications of introducing such artificial mutants are only beginning to be explored; if previous experience with chemical manipulation of the environment is any guide, it may be expected that many undesirable, but so far unpredictable, consequences will occur. One result already noted is that pollen produced by corn genetically engineered to produce such insecticidal agents has proven to be lethal to Monarch butterflies that encounter it. Because these butterflies, generally considered a valuable and beautiful part of the arthropod biota, are already greatly endangered by loss of habitat, this is a matter of grave concern. More such effects, affecting other desirable plant and animal species, may be expected.

E. Domestic Operations

Certain very stable species that penetrate the environment and produce significant chemical modifications there are widely applied for domestic purposes in developed countries. Synthetic detergents have largely replaced soaps for cleaning and washing. A detergent consists of a surfactant, a chemical species that lowers the surface tension of water, facilitating the emulsification (dispersal in very small droplets) of dirt and stains, plus builders, compound that complex divalent ions (which make water "hard"). Early surfactants were chemically very stable and tended to persist and enter natural water supplies, causing foaming and other deleterious effects. "Biodegradable" (less stable) surfactants are now employed. Many builders, notably polyphosphates, act as nutrients for plants and microorganisms in natural waters that they eventually reach, thus perturbing the ecological balance.

CFCs (chlorofluorocarbons or *Freons*) such as Freon-11 (CCl_3F) and Freon-12 (CCl_2F_2) have been widely used as the heat transfer fluids in refrigerators and as a propellant medium in spray cans because of their low cost, absence of toxicity, and exceptional chemical stability. These compounds permeate the lower atmosphere and gradually migrate to the stratosphere, where they cause a reduction in the ozone layer, which shields the surface of the earth from dangerous ultraviolet radiation (Section III.C.2).

III. THE ATMOSPHERE

A. Natural Structure and Chemical Composition

Table II gives the standard sea-level composition of dry air in terms of mole fraction (i.e., mixing ratio) and actual total number of moles in the atmosphere as a whole for each species, assuming vertical chemical uniformity; those gases that contribute to the greenhouse effect are noted. In addition to the components shown in this table, natural air contains water vapor in variable amounts, ranging in mole fraction from near 0 to 0.03.

This composition represents the culmination of an evolutionary process unlike that of any other known planet. The oxygen in the atmosphere is believed to derive almost entirely from the photosynthetic action of green plants and was absent before this form of photosynthesis arose, probably 2.5–3 × 10^9 years ago. The concentration of CO_2, on the other hand, is low compared with that in the atmospheres of Mars and, in particular, Venus; this is because most of the earth's CO_2 inventory is in solution in seawater (3×10^{18} moles) or incorporated in carbonate minerals (6×10^{21} moles).

The composition of the atmosphere has probably varied over even recent geologic time, and it is not clear precisely what mechanisms act to control concentrations of the various components. It is therefore difficult to say how much human activity will perturb the atmospheric environment. A major such perturbation has been the addition of about 2.5×10^{16} moles of CO_2 from combustion of fossil fuels in the last 150 years. This has caused the atmospheric concentration to rise from 290 to 360 ppm (est. 2000), which corresponds to a net increase of only 1.3×10^{16} moles in the atmosphere, that is, about half of the amount of CO_2 generated by fossil fuel burning remains in the atmosphere. Most of the remainder has probably been taken up into the seawater reservoir, consistent with the observation that atmosphere–hydrosphere CO_2 transfer is efficient, with the residence time of CO_2 in the atmosphere being only 7 years. Whether this degree of partitioning of added CO_2 between the atmosphere and oceans will continue in the long run is unknown.

Concern has sometimes been expressed about the possibility of depletion of the Earth's oxygen supply if deforestation on land and kills of marine algae from oil spills were to substantially reduce the photosynthetic production of O_2. In the short term, this is not a significant danger. Fossil fuel combustion requires about 4×10^{14} moles of O_2 per year, and the amount of O_2 consumed by the respiration of animals and by inorganic reactions with minerals is not

TABLE II Composition of Dry Air

Chemical species	Molecular weight	Mole fraction	Total atmospheric amount (moles)
Nitrogen (N_2)	28.01	78.084	1.42×10^{20}
Oxygen (O_2)	32.00	20.946	3.81×10^{19}
Argon (Ar)	39.94	0.934	1.70×10^{18}
Carbon dioxide (CO_2)[a]	44.01	3.6×10^{-4} (360 ppm) (est 2000)	6.5×10^{16}
Neon (Ne)	20.12	1.8×10^{-5} (18 ppm)	3.3×10^{15}
Helium (He)	4.01	5.2×10^{-6} (5 ppm)	9.5×10^{14}
Methane (CH_4)[a]	16.04	1.6×10^{-6} (1.6 ppm) (est 2000)	2.9×10^{14}
Krypton (Kr)	83.80	1.1×10^{-6} (1.1 ppm)	2.0×10^{14}
Nitrous oxide (N_2O)[a]	44.01	5×10^{-7} (0.5 ppm)	9×10^{13}
Hydrogen (H_2)	2.02	5×10^{-7} (0.5 ppm)	9×10^{13}
Xenon (Xe)	131.30	8×10^{-8} (0.08 ppm)	1.5×10^{13}

[a] Greenhouse gas.

more than a few percent of this; so even if all production of O_2 were to cease, the inventory in the atmosphere would suffice to sustain human activity for some 10,000 years before a 10% reduction in oxygen concentration occurred.

The pressure of the atmosphere decreases uniformly with altitude, but the temperature exhibits complex vertical structure. It decreases through the lowest portion, or *troposphere*, to an inversion point, the *tropopause*, whose altitude varies from about 10 km at the poles to 15 km at the equator. Above the tropopause, temperature increases with altitude through the *stratosphere*; this heating is mainly due to absorption of solar ultraviolet energy by the dissociation of ozone:

$$\text{solar UV radiation}$$

$$O_3 \rightarrow O_2 + O.$$

A second inversion occurs at a height of about 90 km, at the *mesopause*, between the *mesosphere* and *thermosphere*; the heating in the latter is mainly due to absorption of far-UV solar radiation by dissociation of N_2 and O_2. These temperature inversions separate the atmosphere into distinct reservoirs, since they act as barriers to convective mixing; material passes between troposphere and stratosphere mainly by the relatively slow process of diffusion. Local temperature inversions also occur in the lower troposphere to form regional reservoirs in which pollutant chemicals can build up to high concentrations; a well-known local inversion phenomenon of this sort is that which occurs at elevations ranging from 300 m to 2 km over the Los Angeles basin in Southern California.

B. Lower Atmosphere Effects

The introduction of species from fossil fuel combustion (Section II.A.1) into the lower atmosphere leads to an extensive series of chemical reactions, most of which are of free-radical type. They can be divided into three categories: daylight (D) reactions, requiring solar near-UV radiation; night (N) reactions, which are important in the absence of sunlight; and 24-hr (DN) reactions.

The fundamental light-requiring process is

$$NO_2 + h\nu \text{ (below 398 nm)} \rightarrow NO + O. \qquad \text{(D)}$$

The O atom then forms ozone in the presence of a third body M:

$$O + O_2 + M \rightarrow O_3 + M, \qquad \text{(DN)}$$

and ozone reacts with NO to form NO_2:

$$NO + O_3 \rightarrow NO_2 + O_2. \qquad \text{(DN)}$$

Most emission of NO_x by combustion is in the form of NO, and direct air oxidation of NO to NO_2 is too slow to be of much significance; at typical concentrations the formation of O_3 as an intermediate permits this oxidation to occur on a timescale of 1 hr or so in typical urban air. A further D process is

$$O_3 + h\nu \text{ (below 308 nm)} \rightarrow O^* + O_2, \qquad \text{(D)}$$

where O^* denotes an excited state (the ^1D state) of the oxygen atom, followed by

$$O^* + H_2O \rightarrow 2OH\cdot, \qquad \text{(DN)}$$

whereas at night ozone reacts further with NO_2:

$$O_3 + NO_2 \rightarrow O_2 + NO_3. \qquad \text{(N)}$$

A further source of the hydroxyl radical OH· is

$$2NO_2 + H_2O \rightarrow HNO_2 + HNO_3, \qquad \text{(DN)}$$
$$HNO_2 + h\nu \text{ (295–410 nm)} \rightarrow OH\cdot + NO. \qquad \text{(D)}$$

The nitrate radical formed at night leads to production of nitric acid, HNO_3:

$$NO_3 + NO_2 \rightarrow N_2O_5, \qquad \text{(N)}$$
$$N_2O_5 + H_2O \rightarrow 2HNO_3. \qquad \text{(N)}$$

Nitric acid is also produced from OH· :

$$OH\cdot + NO_2 \rightarrow HNO_3. \qquad \text{(DN)}$$

Thus solar irradiation of a mix of air and NO leads to a buildup of the oxidant O_3, the reactive free radical OH· and the strongly acidic HNO_3; after sunset, OH· production ceases but that of NO_3 begins, leading through N_2O_5 to HNO_3. The latter acidifies water droplets, causing the formation of rain, snow, mist, or fogs with low (i.e., acid) pH. Ozone and OH· undergo further reactions in the presence of hydrocarbons RH (i.e., volatile organic compounds or VOCs):

$$OH\cdot + RH \rightarrow R\cdot + H_2O,$$

$$R\cdot + RCHO \rightarrow RH + RCO\cdot,$$

$$R\cdot + O_2 \rightarrow RO_2\cdot,$$

$$RO_2\cdot + NO \rightarrow RO\cdot + NO_2,$$

$$RCO\cdot + O_2 \rightarrow RC(O)OO\cdot \text{ (peroxyacyl radical)},$$

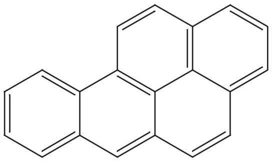
$$RC(O)OO\cdot + NO_2 \rightarrow RC(O)OONO_2$$

$$\text{(peroxyacyl nitrate or PAN)}.$$

Thus the presence of organic molecules acts to speed the production of NO_2. The PAN compounds formed in the last step are potent eye irritants; it is the mix of such substances with O_3 and the brown NO_2 that constitutes "smog." Carcinogenic nitrosamines can be formed from secondary amines with HNO_2:

$$R_2NH + HNO_2 \rightarrow R_2NNO\text{(nitrosamine)} + H_2O.$$

Carcinogenic condensed-ring hydrocarbons such as benzo(*a*)pyrene



are emitted in the exhaust mix of unburned HC; they undergo further reactions with O_3 or N_2O_5 to yield mutagenic epoxides or nitro compounds.

Particulates, especially the soot emitted in incomplete combustion or by diesel engines, play a role in the pollution chemistry of the lower atmosphere. They adsorb carcinogens of low volatility such as the condensed aromatics, thereby retaining them in the atmospheric reservoir available for inhalation, and also catalyze the further reaction of these species with oxidants.

Conversion of SO_2 to SO_3, leading to sulfuric acid formation, is slow in the gas phase but is probably accelerated when the SO_2 is adsorbed on soot grains. In this way particulates contribute to acid precipitation. Extremely fine particulates also seem to play a role in the generation of smog, perhaps by adsorbing and stabilizing organic free radicals R· and RO·, which serve as intermediates in the production of ozone and other oxidant species. Hence there is increased interest in the possibility of designating very fine particulates (with diameters of a few microns or less) as a separate category of pollutant to be subjected to EPA monitoring and control.

C. Upper Atmosphere Effects

1. Ozone Layer

Ozone, objectionable when present in the lower atmosphere, is a vital ingredient of the upper atmosphere (stratosphere); by absorbing UV light at wavelengths below about 340 nm it protects the biota from exposure to this damaging radiation and maintains the temperature inversion at the tropopause, which is important in shaping terrestrial meteorological patterns. Ozone is produced in the upper atmosphere by

$$O_2 + h\nu \text{ (below 242 nm)} \rightarrow O + O,$$

$$O_2 + O + M \text{ (third body)} \rightarrow O_3 + M.$$

Maximum O_3 concentrations of as much as 10 ppm are attained around 25 km of altitude. The total amount in the air column is conventionally measured in *Dobson units* (DU), with about 300–350 DU being the usual amount of O_3 present in the stratosphere in the absence of chemical perturbations. Such perturbations are important, since the amount of ozone present is determined by a steady-state balance between the rate of production by the above mechanism and the rate of removal processes, either by absorption of ultraviolet light or reaction with such species as O and OH· :

$$O_3 + O \rightarrow O_2 + O_2,$$

$$O_3 + HO\cdot \rightarrow O_2 + HOO\cdot.$$

The introduction of free-radical species X such as atomic chlorine or bromine or nitric oxide, NO, can speed the decomposition of O_3 by the reactions

$$O_3 + X \rightarrow O_2 + XO \quad \text{and} \quad O + XO \rightarrow O_2 + X,$$

thereby reducing its steady-state concentration in the stratosphere. One possible source for such ozone-destroying species would be NO_x introduced into the stratosphere by supersonic jet aircraft, missiles, or (in the event of nuclear war) by large explosions. However, of more present concern is the effect of CFCs (Section II.E). These compounds are so stable in the troposphere that a large fraction of them when released in the lower atmosphere will eventually migrate to the stratosphere by diffusion. Once in the stratosphere, their carbon–chlorine bonds are readily dissociated by UV light to form free radicals, for example:

$$CCl_2F_2\text{(Freon-12)} + h\nu \rightarrow CClF_2\cdot + Cl\cdot.$$

The original analysis of this situation in 1974 suggested that on the basis of the amounts of CFCs already present in the terrestrial environment, average ozone levels in the stratosphere might decrease by as much as 15%. The ensuing controversy led to a number of other estimates, both lower and higher. However, the reality of the effect became evident when it was noted in the mid-1980s that the ozone layer over the Antarctic continent was developing an "ozone hole" in the spring, that is, a region in the stratosphere in which the ozone was essentially absent. Observations made at the Halley Bay Research Station for the month of October (Antarctic spring) through the 1958–1992 period are plotted in Fig. 6; they show that the 1988–1992 average October ozone amounts were approximately half those found at the start of the measurement period.

The explanation for this unexpectedly sharp decrease in ozone lies in a particular phenomenon of the Antarctic winter; it is cold enough in winter in the stratosphere over Antarctica to form *polar stratospheric clouds* (PSCs) of nitric acid hydrate, $HNO_3 \cdot xH_2O$, where $x \approx 3$. The normal end fate of chlorine atoms in the stratosphere is to be incorporated in hydrochloric acid, HCl, which does

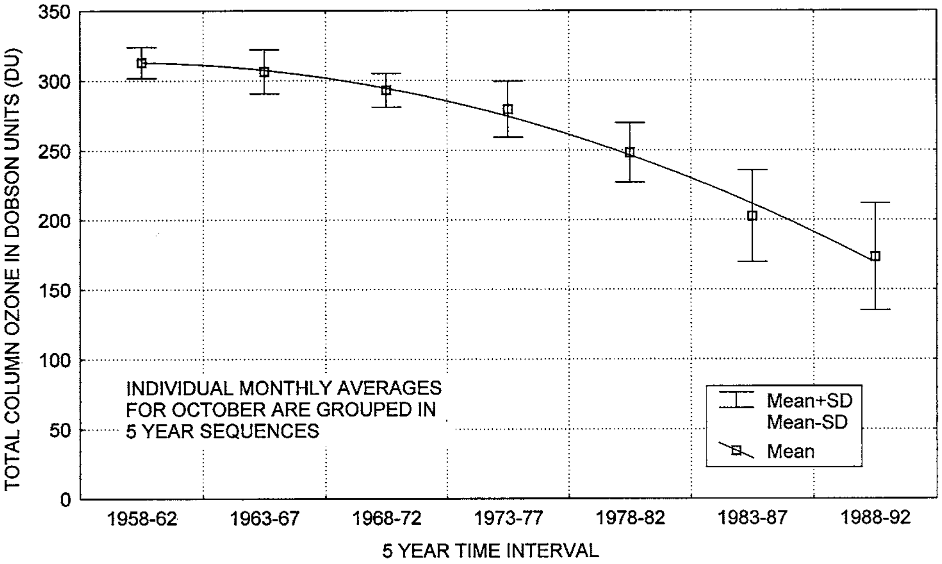


FIGURE 6 Total column ozone observed over Antarctica in spring from Halley Bay.

not photodissociate and eventually is removed in precipitation. However, HCl in the Antarctic stratosphere will adsorb onto PSCs and can thereby react with chlorine nitrate:

$$\mathrm{HCl} + \mathrm{ClONO_2} \rightarrow \mathrm{Cl_2} + \mathrm{HNO_3},$$

producing a molecule, Cl_2, which is released from the PSC in the spring and is efficiently photodissociated to reestablish a high concentration of Cl· atoms and resume the process of ozone destruction. The dramatic nature of the "ozone hole" phenomenon removed all doubt as to the seriousness of the danger to the ozone layer and led to the convening of an international meeting at Montreal in September 1987, at which it was agreed to place restrictions on ozone-destroying CFCs; the residence time of CFCs in the stratosphere is sufficiently long (50–100 years) that depletion of the ozone layer will continue for many decades to come. The ozone hole over Antarctica may be expected to continue for many decades to come, permitting a greatly increased amount of solar UV radiation to reach the Antarctic continent and the seas around it. This appears to be damaging the ability of marine phytoplankton to grow and reproduce; the resulting reduction in primary production and availability of nutrients in the food chain could well prove catastrophic for the maintenance of the world's marine ecosystem. Development of a similar, if less pronounced, ozone hole in northern latitudes is also now occurring, and similar effects may be expected, with the additional factor of enhanced risk of skin cancer, cataract formation, and other UV-radiation caused pathologies in the dense Northern Hemisphere human population.

2. Greenhouse Effect

The temperature of the earth's surface represents a balance of incoming solar energy, largely in the visible and ultraviolet regions of the spectrum, against outgoing radiation from the atmosphere, hydrosphere, and solid surface of the earth, largely in the infrared. This balance is sketched in Fig. 7.

The temperature that a blackbody must have to radiate away exactly the amount of energy that the earth receives from the sun is only 255 K (−18°C); the actual average temperature of the earth's surface (15°C) is higher because infrared radiation is partially trapped in the atmosphere and then reradiated back to the ground. That is, the atmosphere is relatively transparent to short-wavelength light and relatively opaque to infrared light; this is termed the *greenhouse effect*. The bulk of this effect is due to water vapor; but absorption by CO_2 between 12 and 16.3 μm contributes about a quarter of the 33 K excess greenhouse temperature. The increase in atmospheric CO_2 level (Section II.A) of about 24% in the last 150 years should enhance the greenhouse effect and lead to a warming of the earth's surface by a few degrees Fahrenheit. However, because of the natural variability in terrestrial temperatures and the presence of other factors that can influence the radiation balance, the reality, or at least the magnitude, of this increase remains a subject of debate. It is also possible that other trace atmospheric species that are increasing due to human activity, notably methane (CH_4) (the amount of which in the atmosphere has doubled in the last 150 years, from about 0.75 ppm to the present value of 1.6 ppm)

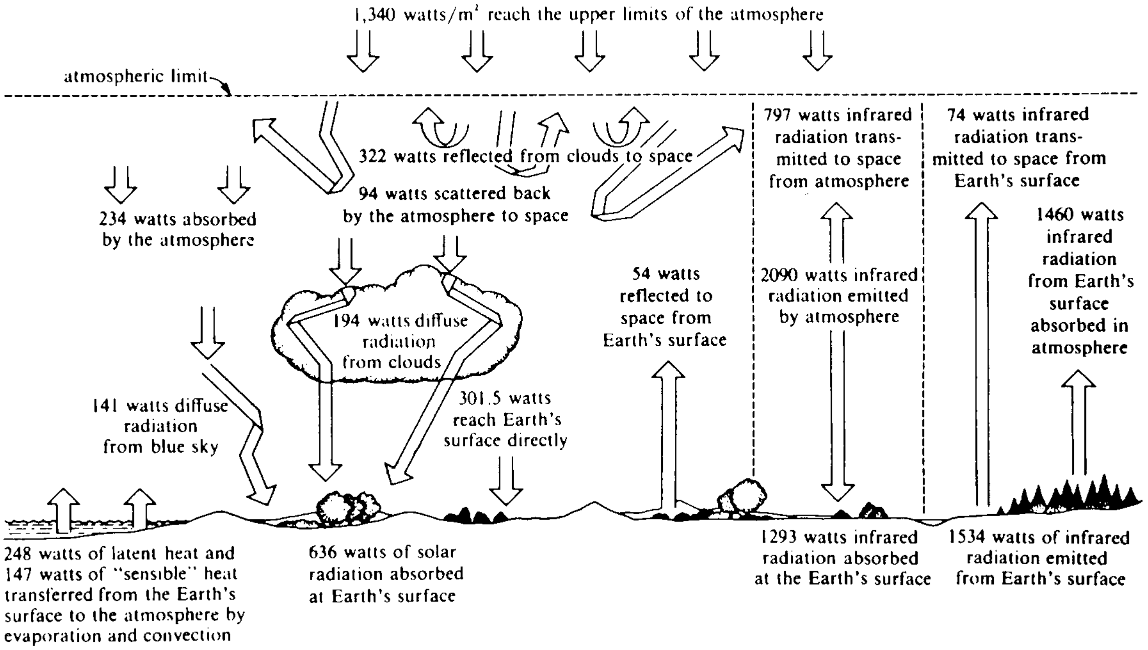


FIGURE 7 Terrestrial radiation energy balance. Overall solar energy flux is 340 W/m^2. [From Manahan, S. E. (1984). "Environmental Chemistry," 4th ed., Willard Grant Press, Boston. Reproduced by permission.]

and the same CFCs implicated in stratospheric ozone depletion (Section III.C.1), will add to the greenhouse effect. Whether or not it is due to the increase in greenhouse gases resulting from human activities, however, there can be no serious doubt as to the reality of a global warming during the last century. An advance of the tree line to the north in the Arctic has been observed; studies of subsurface soil temperatures have demonstrated a significant worldwide rise in temperatures during the last several decades; glaciers and the polar ice caps have diminished markedly during the last century; and analysis of climatological records indicates that since 1976 this warming trend has accelerated to a rate of 4 degrees (Fahrenheit) per century. Continuation of this trend would have profound consequences for the distribution of climatic zones on the earth, result in a massive rise in sea level with inundation of large areas along coastlines and would probably be catastrophic for the global ecosystem, for most plant and animal species will not be able to adapt or migrate rapidly enough to accommodate to such extreme temperature changes.

It should also be understood that a profound perturbation of the earth's heat balance would be a probable consequence of nuclear war, were such an event ever to occur. The explosions in such a war might inject enough soot into the upper atmosphere to greatly reduce the amount of solar radiation reaching the surface for many weeks or months. The resulting temperature drop has been termed "nuclear winter." A similar cooling may have occurred at the end of the Cretaceous period, 65 million years ago, as a result of dust produced by the collision of a massive meteorite with the earth and may have contributed to the wave of extinctions in the biota at that time.

IV. THE HYDROSPHERE

A. Natural Structure and Chemistry

Whereas the atmosphere is uniform laterally and varies in structure only with altitude, the hydrosphere consists of many distinct reservoirs of differing size. These are the seas and oceans, with a limited range of variation in composition and a definite vertical structure; lakes, with a wide range of chemical variation and also vertical stratification; rivers, with highly variable composition but little structure; precipitation (rain, snow, fog, mist), whose only important compositional variable is pH (degree of acidity); and groundwater, whose composition depends strongly the character of the soils or minerals with which it is in contact.

The compositions of seawater and average river water with respect to major solution species are given in Table III with concentration in molarity (M). The composition of seawater appears to have changed little since mid-Precambrian time, about 2×10^9 years ago, although the pH may have varied slightly. The principal factor regulating the concentrations of the major

TABLE III Compositions of Seawater and Average River Water, in Molarity (*M*)

Species	Seawater concentration (surface)	Average river water concentration
Na^+	0.47	2.7×10^{-4}
K^+	0.010	5.9×10^{-5}
Ca^{2+}	0.010	3.8×10^{-4}
Mg^{2+}	0.054	3.4×10^{-4}
Cl^-	0.56	2.2×10^{-4}
SO_4^{2-}	0.028	1.2×10^{-4}
HCO_3^-	2.4×10^{-3}	1×10^{-3}
CO_3^{2-}	2.7×10^{-5}	1×10^{-4}
Total nitrogen	2.0×10^{-5}	
Total phosphorus	1.5×10^{-6}	
pH	8.15	7.5–8.5

ionic species is probably chemical exchange with mantle magma at midocean crustal spreading centers, and there is no evidence for any human-related effect on bulk oceanic composition.

Oceans are divided vertically into the

1. *Photic zone*, the top few meters into which light penetrates and photosynthesis is possible;
2. Remainder of the surface or *mixed layer*, extending to about 100 m depth, in which temperature, composition, and nutrient levels are somewhat variable and convective mixing occurs;
3. *Thermocline* layer, from 100 to about 1500 m, in which the temperature decreases sharply; and
4. *Deep ocean*, below 1500 m, with essentially uniform temperature near 3°C.

The thermocline acts as a barrier separating reservoirs of the surface and deep ocean zones. Chemical perturbations associated with human activity mainly impact only on the surface zone.

Lakes also exhibit thermal stratification. In the summer months a warm upper layer called the epilimnion is separated by a thermocline from a cool bottom zone or hypolimnion. Autumn cooling of the epilimnion may make the surface water denser than the bottom water, provoking a turnover and bringing chemical species from bottom sediments to the top. Such recirculation can have severe biological consequences; rapid turnover of Lake Nyos in Cameroon brought so much water supersaturated with CO_2 to the surface that a CO_2 cloud formed and suffocated many area residents.

B. Perturbation of pH

Human activity has not so far affected global oceanic pH, although discharges of acidic wastes into shallow near-shore waters can cause destructive local pH fluctuations. However discharge of SO_x and NO_x from combustion into the atmosphere (Sections II.A.1 and III.B) leads to the generation of H_2SO_4 (sulfuric acid) and HNO_3 (nitric acid), which dissolve in precipitation. Such acidic precipitation (“acid rain”) has acted to markedly lower the pH of lake and river water in large areas of North America and Europe. Natural rainwater has a pH of about 5.6 because some atmospheric CO_2 dissolves in it to form the weak acid H_2CO_3 (carbonic acid); but both sulfuric and nitric acids are strong acids and their impact on the pH of rainwater is far greater for a given molar quantity dissolved than is that of carbonic acid. Because of their presence, the average pH of rainfall in the northeastern United States and eastern Canada has decreased to about 4.5, and values as low as 2 have been observed. Lakes, rivers, and groundwater in the affected areas are becoming more acidic, resulting in decrease or disappearance of fish and other aquatic organisms and damage to forests because lowered pH in the soil results in loss of nutrients by leaching and the mobilization of aluminum (Al^{3+}) ion, which is toxic to plants. Alleviation is difficult, since the excess acidity may be carried long distances by air mass movement and it is not generally possible to pinpoint the source of the acidity with any precision.

C. Perturbation of Redox Balance

The chemical state of a water body is sensitive to its redox (oxidation–reduction) level, which in turn depends on the amount of dissolved oxygen (DO) the water contains—9 mg O_2/liter for water saturated with air at 25°C. If nutrient levels and oxygen-consuming biological activity are low in a body of water undergoing efficient exchange with the atmosphere (as in a rapidly flowing stream), DO levels remain near the maximum and the water body remains aerobic (Fig. 8, top); but if excessive nutrient material requiring oxygen for its utilization is introduced, the DO may be exhausted and the body converted to an anaerobic cycle (Fig. 8, bottom), characterized by microbial fermentation and sulfur metabolism. Addition of nutrients by the runoff of fertilizer from agricultural land (Section II.D) or of industrial and domestic waste in sewage from urban areas thus works to enhance the biological oxygen demand (BOD) in adjacent water bodies and causes them to become partially or entirely anaerobic. Addition of such nutrients to lakes accelerates their *eutrophication*, that is, their conversion to anaerobic bogs. Anaerobic conditions are frequently

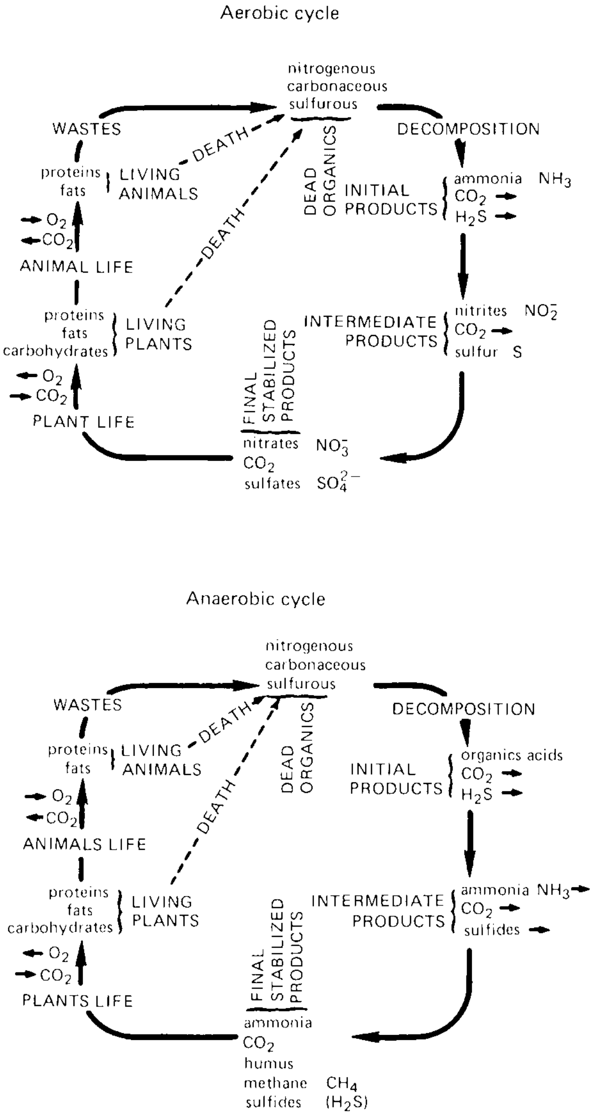


FIGURE 8 Aerobic and anaerobic nutrient cycles. [From Vesilind, P. A. (1975). "Environmental Pollution and Control," Ann Arbor Science, Ann Arbor, MI. Reproduced by permission.]

encountered under natural conditions as well, in bodies of water with limited circulation such as the Black Sea.

D. Perturbation by Toxic Materials

The overtly toxic materials present in the hydrosphere include biocides, industrial solvents, by-products of the petroleum industry, and heavy metals. Materials in the first three classes, most notably chlorinated hydrocarbons such as DDT and the PCBs, undergo only slow microbial degradation and are present in sufficient concentration in many aquatic environments to severely affect the biota. In favorable cases metals may be removed by deposition as insoluble precipitates in the sediments. However, they may remain in the hydrosphere because of the formation of complexes with other pollutant species or with naturally produced chemical substances. An example is the mobilization of mercury (Hg) in ocean waters. Mercury is used widely in industry, especially as a fungicide; it had been assumed that the dumping of mercury-containing wastes in ocean water would be harmless because the mercury would be sequestered as the insoluble salt mercuric sulfide, HgS, in bottom sediments. Instead some of this mercury reacted with methylcobalamin, a compound produced by anaerobic microorganisms abundant in such sediments, to form soluble methylmercury compounds such as $Hg(CH_3)Cl$, which can then enter the food chain. Because methylmercury is lipophilic, it tends to accumulate in the fatty tissues of fish and other aquatic organisms, and consumption of these organisms, by humans has led to serious episodes of mercury poisoning.

V. THE BIOSPHERE

The biosphere is both a chemical reservoir in its own right and the fundamental entity of concern in judging the importance of environmental chemical perturbations. Many chemical substances entering the global environment have accumulated in high concentrations in living tissue and produced severe effects on animal or plant populations. Examples are the buildup of mercury in certain marine fish and of various heavy metals in shellfish (Section IV.D). Perhaps the most dramatic is the tendency of lipophilic chlorinated hydrocarbons to accumulate in the body fats of animals that feed on insects or plants to which these substances have been applied. This appears to be strictly a matter of relative solubilities in water and in lipids of these substances, and their K_{ow} values are generally a good guide to their expected degree of bioaccumulation. One especially notable example of the deleterious effects of such accumulation of pesticides in animal tissue was the inhibitory effect of DDT on calcium metabolism in birds and the consequent disastrous thinning of their egg shells (Section II.D). This particular problem has been remediated by halting production and use of DDT, at least in most of the developed world, and some of the avian populations affected have rebounded; but this must not blind us to the fact that worldwide a mass extinction of biota rivaling and perhaps exceeding that at the end of the Cretaceous period 65 million years ago is occurring and indeed intensifying. The reasons for this catastrophe are not always well understood, but both loss of habitat and chemical pollution are surely playing a role. It has been suggested that the massive worldwide die-off of amphibian species (frogs, salamanders, newts) that is now occurring may be due to increased

UV radiation reaching the earth's surface because of thinning of the ozone layer; the almost total elimination of many species of fish in the North Atlantic Ocean probably is mainly due to overfishing, with frank pollution effects also playing a role; the *Waldsterbe*, or forest death, which has destroyed much of the woodlands of Central Europe is probably due to hydrosphere acidification and widespread heavy metal contamination; the death of conifers along the mountain ridges of Central and Southern California is almost certainly a result mainly of high oxone concentrations in smog intruding further and further into these areas as development of land for homes and industries occurs. As noted above (Section III.C.2), global warming is likely to greatly exacerbate the destruction of biota and the breakdown of regional ecological networks and disastrous loss of biodiversity. Clearly it is now a matter of the highest importance both to clearly delineate the reasons for the depletion of the world ecosystem and the roles played by pollution in this process, and to take far more ambitious and wide-ranging steps both to control and reduce future pollution and to clean up that already generated, so far as may be possible.

VI. TOXIC WASTES

It will be appreciated from earlier parts of this article that a large portion of the chemical pollution of the environment is due to the discharge of waste materials: combustion exhaust, effluents from mining, spent solvents, and by-products of chemical manufacture. In some cases these wastes have been found to have unexpectedly high toxicity, an example being that of dioxin (Section II.D), a by-product in the manufacture of herbicides; in addition to its probable role in causing illness among Vietnam War veterans, it has been involved in several mass poisoning episodes of humans and animals.

Uncontrolled discharge of chemical wastes to the environment is now recognized as unacceptable, and accordingly methods for their elimination or sequestering are receiving much attention. Elimination by selective chemical transformation is possible in the case of chemical compounds, but this is usually very expensive; combustion is generally undesirable because some incompletely burned material will be released to the atmosphere and hydrosphere. The alternative is storage in a dump site until biological degradation can occur. Exposure of wastes in the dump site to the atmosphere to permit fast aerobic microbial activity may entail transfer of toxic volatiles to the atmosphere and cannot usually be permitted. Therefore mainly slow anaerobic degradation must be employed, sometimes generating further toxic (or explosive) substances in unpredictable fashion. Thus it is necessary that the toxic dump be well sealed, that is, be a reservoir with a very long residence time for all contents (very slow transfer to the environment) to minimize migration of such wastes away from their site of deposition into the atmosphere and hydrosphere.

Of particular concern is the contamination of groundwater (underground aquifers) by wastes leaching through soils from surface dump sites (Fig. 9). The removal of pollutants from such aquifers is difficult and expensive;

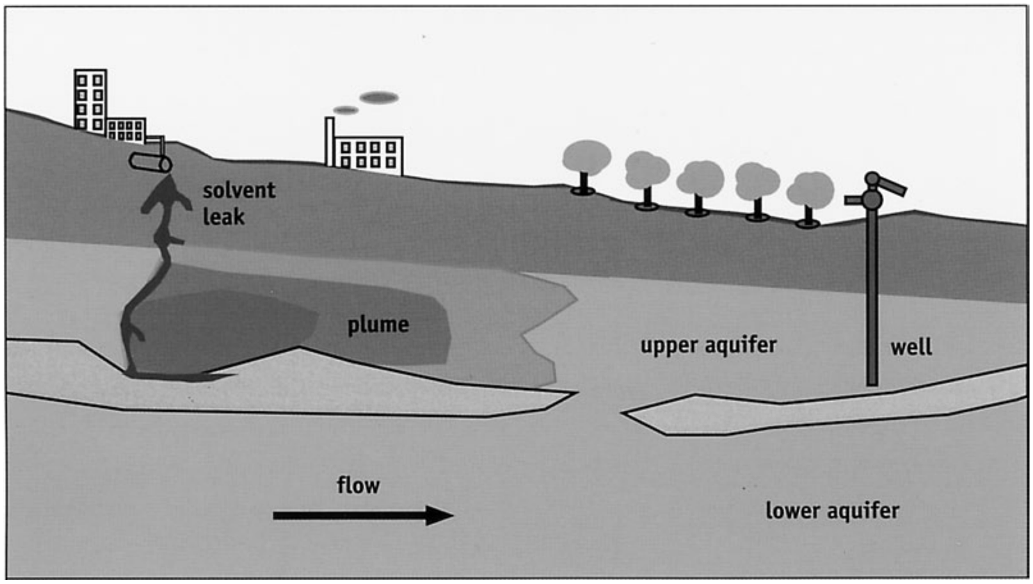


FIGURE 9 Migration of pollutant substances into groundwater. [From Harmon, Thomas C. (1999). "Southern California Environmental Report Card," UCLA Institute of the Environment, Los Angeles. Reproduced by permission.]

many of the pollutant substances have large values of K_{ow} and hence tend to adhere tenaciously to organic material in soil particles, transferring only slowly to groundwater and thereby constituting a reservoir to maintain long-term contamination of the latter. It is thus far better to prevent migration of pollutants away from an improperly prepare disposal site than to attempt a cleanup after aquifer contamination has occurred. However, prior to the recognition of the importance of proper chemical waste disposal, large quantities of such materials were simply dumped into soils, shallow unlined impoundments, or leaking subsurface storage tanks. This resulted in many areas in which industrial or agricultural pollutants were routinely discharged into the environment becoming grossly polluted, to the point where other uses of them became impossible and human health and safety (not to speak of that of other members of the biota) became severely imperiled. To address this problem, the U.S. Congress in 1980 enacted the Comprehensive Environmental Response, Compensation, and Liabilities Act (CERCLA), which among other things created a fund (“Superfund”) to clean up such grossly polluted areas, now designated as Superfund sites. The process of cleanup, however, has generally turned out to be slower, more difficult, and more expensive than originally expected, and many Superfund sites are still unremediated.

Nuclear waste (Section II.A.2) presents even more serious problems than does chemical waste. No method of elimination is possible; radioactive material must be sequestered from the environment until it decays, which for some common by-products of nuclear reactor operation will take thousands of years. Continual exposure to radiation promotes deterioration of materials, so nuclear wastes held in aboveground storage or in subsurface tanks must frequently be transferred to new containers. Burial of such wastes has frequently been discussed as a long-term solution to the problem of their disposal, but finding geological formations sufficiently stable and remote from aquifers into which the wastes might migrate has proven to be very difficult. At present, attention is focused on Yucca Mountain, Nevada, as a possibly suitable long-term burial site for high-level nuclear wastes, but the proposal to establish such a site there is contentious. A site for burial of midlevel (transuranic) wastes, the Waste Isolation Pilot Plant, or WIPP, has been under development (and the subject of much litigation) near Carlsbad, New Mexico, since the mid-1980s. The wastes to be deposited at WIPP would be entombed in a thick salt stratum, which seems to be reasonably stable and far from the nearest groundwater. The first wastes were to be delivered to WIPP in late 1999, but attempts to block the opening of the facility continue. It must be concluded that the problem of safely disposing of nuclear waste remains unsolved.

SEE ALSO THE FOLLOWING ARTICLES

Aerosols • Combustion • Environmental Radioactivity • environmental Toxicology • Pollution, Air • Pollution Control • Pollution Prevention from Chemical Processes • Radiation Sources • Soil and Groundwater Pollution • Transport and Fate of Chemicals in the Environment • Wastewater Treatment and Water Reclamation

BIBLIOGRAPHY

Baird, C. (1998). “Environmental Chemistry,” 2nd ed., W. H. Freeman and Company, New York.

Howard, P. H., ed. (1989). “Handbook of Environmental Fate and Exposure Data for Organic Chemicals,” Vols. 1–4, Lewis Publishers, Chelsea, MI.

Manahan, S. E. (1999). “Environmental Chemistry,” 7th ed., CRC Press, Boca Raton, FL.

Masters, G. M. (1991). “Introduction to Environmental Engineering and Science,” Prentice-Hall, Englewood Cliffs, NJ.

Pitts, P. J. F., and Pitts, J. N., Jr. (1986). “Atmospheric Chemistry,” Wiley/Interscience, New York.

Seinfeld, J. H. (1986). “Atmospheric Chemistry and Physics of Air Pollution,” John Wiley & Sons, New York.

Spiro, T. G., and Stigliani, W. M. (1996). “Chemistry of the Environment,” Prentice-Hall, Englewood Cliffs, NJ.

Radioactive Waste Disposal (Geology)

Saeid Saeb
RockSol Consulting Group, Inc.

Stanley J. Patchet
Westinghouse TRU Solutions, LLC

GLOSSARY

Disposal Placement of waste in a form and condition such that it cannot be easily retrieved.

Repository An isolation system for nuclear wastes, including the surrounding enclosing geological medium, the introduced fill materials, and the waste packages.

Salt Rock made up of the mineral halite (NaCl) formed at various periods in earth history by the evaporation of sea water.

Seal A component in an isolation system that prevents or substantially delays the release of radionuclides to the surrounding environment.

Storage Placement of waste in a form and condition that it can be easily retrieved.

NUCLEAR waste is generated through civil, military, and research use of nuclear materials. There is a strong, worldwide consensus that the best and safest long-term option for dealing with nuclear waste is geological disposal. While the technical and scientific communities believe that deep geological disposal is safe and ethical and the necessary technology is available and can be deployed now, the public seems much more skeptical. The main hurdle now is gaining public and political confidence in the safety of a deep geological disposal program and of the sites selected by that program. There must be recognition within the technical and scientific community that implementing any form of nuclear waste disposal is determined not only by technical or regulatory processes but also by broad public acceptance.

I. RADIOACTIVE WASTE

Disposal of nuclear waste, generated through civil, military, and research use of nuclear materials, is the final step in the nuclear cycle. The cycle starts at the uranium mines and eventually results in spent nuclear fuel, surplus plutonium and materials that are contaminated during processing. Wastes are generated at processing plants, fuel fabrication facilities, fuel reprocessing facilities, commercial nuclear power plants, reactors that produce materials for nuclear weapons, reactors on military nuclear-powered ships, and medical and research activities. High-level radioactive wastes in a variety of forms are created when spent nuclear fuel is treated chemically to separate uranium and plutonium. The production of nuclear weapons generates a variety of wastes.

Radioactive wastes, like other wastes, present a variety of behaviors that must be considered in safe disposal. Some, for example, generate considerable heat while others remain radioactive for very long periods so disposal regulations require a design that isolates the wastes from public access and the accessible environment for tens of thousand years. This performance requirement is unprecedented and presents a wide variety of technical and professional challenges and dilemmas. At minimum the design must consider the governing mechanisms affecting repository performance over such long time scales and deal with them in a manner acceptable to both the regulator and the general public. While nuclear wastes present a significant disposal problem, the problem must be kept in context. Many other common industrial wastes are toxic and are not biodegradable thus presenting equally difficult disposal specifications.

Currently, worldwide, there are 437 nuclear power reactors operating, 25 reactors under construction, and 79 reactors ordered or planned around the world. These reactors generate spent fuel waste which is initially stored in cooling ponds or dry containers at the reactor sites. The capacity of this initial storage is limited as it is usually intended to serve only the short-term needs of the reactor. Some countries have decided to retire their reactors gradually due to uncertainties about the future disposal of their waste and the limited capacity of reactor site temporary storage. While this approach stops the generation of additional future wastes it does not address the disposal of existing and decommissioning wastes. Leaving these wastes in the reactor site short-term storage facilities can only be considered temporary and is not an option protective of human health and the environment.

Contaminated wastes derive from the fabrication of weapon components and related activities also. Further, the end of Cold War resulted in closing weapons production facilities and the creation of decontamination products, surplus plutonium, and other radioactive materials. These materials have to be safely disposed so that they do not cause any harm to public health or the environment. Moreover, they have to be secured so that they cannot be easily recovered and reused.

II. DISPOSAL ALTERNATIVES

Alternative concepts for nuclear waste disposal have been studied extensively. These alternatives include continued storage, disposal in the sub-seabed, sending the waste to outer space, and partitioning and transmutation. Each of these alternatives has its benefits and risks. However, “there is worldwide consensus that the best, safest long-term option for dealing with HLW is geological isolation. Although the scientific community has high confidence that the general strategy of geological isolation is the best one to pursue, the challenges are formidable.” (NRC, 1990) In this article storage is defined as the placement of waste in a form and condition such that it can be easily and quickly retrieved. Disposal is the placement of waste in a form and condition such that it cannot be retrieved, but it can be recovered albeit at great difficulty and cost. By definition and agreement, any form of geological isolation meets this definition of disposal. Some concepts do not permit either retrieval or recovery.

There are arguments that nuclear waste should be stored for several generations in order to allow time to learn more about the geologic disposal and to take advantage of new technologies that may be developed in the future. Ideally this keeps all options open for future generations but it also imposes all the cost and responsibility for monitoring stored waste (as well as much of the eventual total disposal costs) on future generations too. Furthermore, there is high risk of accidents or security breaches when nuclear waste is scattered around in a large number of temporary sites for a long period of time. Safety then depends on social stability, a fragile container, and material accountability is much more complex and difficult to ensure.

Sub-seabed disposal is an alternative that has been extensively studied. It can be considered a type of geologic disposal. This alternative has certain merits, since many areas of the ocean floor have been stable for millions of years. At present the major obstacle for this disposal option is obtaining the necessary international agreements, relationships, and responsibilities. It seems most unlikely that these could be obtained considering that many individual countries are internally divided about the acceptability of disposal alternatives. At present this alternative does not offer simple or easy retrieval, obtaining site characterization information of the currently demanded standard and quality will be very difficult and costly, and performance assessment modeling will be contentious.

Transferring the waste into space has also been proposed as an alternative to waste disposal in or near the biosphere. The technological capability to transfer waste into the space is not currently available so the associated risks and costs are not well established. For example, it would be necessary to ensure that mission failure does not disperse waste into the atmosphere and that the waste packages can be retrieved no matter where they fall. Like the sub-seabed option a variety of international agreements, etc, will probably be required before implementation. At present this does not appear to be a viable alternative.

When the use of nuclear reactors for commercial power was beginning, recycling of spent fuel was considered to be a source of uranium and plutonium. During the 1970s the United States decided, for a variety of reasons, to dispose of spent fuel and high-level wastes without reprocessing. This fuel utilization process, called the once-through fuel cycle was chosen because the recovered plutonium could effect the proliferation of nuclear weapons. Furthermore, advances in technology for locating high-grade uranium ores and the end of Cold War led to abundant supplies of low-cost and surplus-enriched uranium. This makes the once-through cycle significantly more economical than recycling of spent fuel. Other countries such as France and United Kingdom continue to reprocess spent fuel and to stockpile recovered plutonium for future use.

Several countries are conducting research on partitioning and transmutation of radionuclides in order to create wastes with shorter half-lives which therefore pose less of a challenge for long-term isolation. Reprocessing of spent fuel is an important application of this option. Reprocessing, in turn, generates it's own wastes and contamination products. A report from National Research Council (NRC, 1996) ". . . found no evidence that application of advanced separation and transmutation technologies have sufficient benefit to delay the development of the first permanent repository for commercial spent fuel." Partitioning and transmutation should be viewed as a technology within the waste management process rather than as its end point.

There is a strong, worldwide consensus that the best and safest long-term option is geological disposal. The technical and scientific communities also believe that although the technology is changing and concepts for deep geological repositories have made significant progress, the necessary technology is available and can be deployed now, and that deep geological disposal is safe and ethical. Site characterization field techniques are advanced and allow accurate measurement of parameters likely to be required in any reasonable performance assessment. Computer hardware and software have sufficient power to perform and present a mind-numbing amount of analyses.

III. PUBLIC PERCEPTION AND ACCEPTANCE

While the technical and scientific communities may agree that deep geologic disposal is safe and ethical, the public seems much more skeptical. The main hurdle now is gaining public and political confidence in the safety of a deep geological disposal program and of the sites selected by that program.

The waste management community needs to communicate to the general public in simple yet precise terms how it obtains scientific and technical consensus that safe disposal can be achieved. If this communication is to be effective, it requires the involvement of the public in all aspects of the program. There must be recognition within the technical and scientific community that implementing any form of nuclear waste disposal is determined not only by technical or regulatory processes but also by broad public acceptance. There are no technical solutions for sociopolitical problems. Establishing and building public confidence is a slow but vital exercise. It starts with local acceptance of the concept and continues with public and regulatory review and involvement as the concept develops. A stepwise approach allows public review and feedback, thus strengthening public and political confidence in the process, the program, and the eventual site and facility.

At present public opinion seems to desire some means of reversibility or retrievability, in case repository performance proves to be faulty at some point in the future. Retrieveability essentially is in direct conflict with permanent isolation and also adds costs, complexity, and additional exposure for disposal workers. Nevertheless, in view of the apparent public desire to keep the future options open, some measure of retrievability has to be maintained. Therefore most regulations worldwide now recommend consideration of or require retrievability for the whole operational period and often for periods of hundreds of years after closure. To preserve these options and still provide a permanent solution, the United States Department of Energy (DOE) is designing the Yucca Mountain Project as a monitored geologic repository that permits, but does not require, the retrieval of waste. U.S. Nuclear Regulatory Commission regulations require that a geologic repository be designed such that the waste is retrievable at any time up to 50 years after emplacement. For a long-life facility this means that the earliest disposed waste must be retrievable for periods perhaps several times this value. This gives maximum flexibility to future generations, since they make the choice of closing and sealing the repository as early as the regulations permit or of keeping it open and continuing to monitor it (perhaps for hundreds of years). The price to be paid is additional exposure potential, cost, and complexity.

IV. SITE SELECTION AND PERFORMANCE PREDICTION

In nuclear waste repositories, mechanical, hydrological, chemical, thermal, and biological systems are coupled, the degree ranging from slight to tight. The behavior of the repository as a whole is dependent on the interaction of these systems. However, the long-term performance of a deep geological nuclear waste repository is largely governed by the hydrologic characteristics of the host rock mass. Fluid flow, either through pore space or fractures, is the only relevant physical means by which the radioactive material can be released in to the biosphere once the repository is sealed.

A. Site Characterization

Site characterization is a part of the greater process of determining the acceptability of long-term repository performance. In this sense it is simply an application of the classic engineering design process. It is unique in many regards, however. It is unprecedented in the demands it places on detail and accuracy in field work, quality assurance, regulatory compliance, and public oversight. In every case suitable and acceptable techniques and methodologies are available but rarely are they used in this combination. The end effect is a great increase in the time required to perform and complete field and test programs and, following from that, a great increase in costs. Careful consideration must be given to these needs. It is easy to pass beyond the point where the increase in confidence and value is not worth the additional resource allocation.

More subtly, it is often easy to mistake greater and greater concern for accuracy and completeness, the desire to be “conservative,” the perceived need to provide defendable parameters with real progress in characterizing and assessing the eventual performance of a proposed repository. Site characterization and related activities have been and are being performed at numerous sites and in every case there is clear evidence that currently available techniques and methods provide adequate and relevant field data. These data, combined with sophisticated computer models, allow the theoretical evaluation of current conditions and the estimation of likely future behavior. There is some evidence that human construction activities with natural materials endure for the time periods considered but these are not normally considered adequate as proof in this application.

B. Hydrology

There are several key elements in any repository characterization and assessment exercise. In every case the most likely release path involves fluid transport of the waste to the accessible environment. This may or may not involve human intrusion. Any characterization program therefore must completely evaluate the hydrology of the site and how this is likely to change over long time periods. The program should also ensure that all alternative interpretations are considered and be able to explain either why the alternative does not impact performance or that it is not a reasonable alternative. The program must evaluate not only the detailed flow in and around the repository but also the local and regional hydrological regimes.

C. Waste Inventory

In addition to the site characteristics, the repository design must take into account the waste inventory to be disposed and from that the hazardous and radioactive source terms. From these it is possible to establish other relevant design parameters such as heat loading, decay/longevity periods, and occupational exposure limits. These can vary greatly depending on the waste materials involved. At WIPP, for example, the main waste form is plutonium contamination and hence the heat loads are negligible but the repository must have a long performance life. A repository for spent fuel such as Yucca Mountain may have a shorter performance life but will have to deal with high heat loads and complex material handling requirements. The waste form will also affect secondary performance elements such as gas generation and solubility in ground waters.

D. Seals and Closures

Deep geological repositories also require seals in the access ways. The mineral industry has considerable experience in the design and construction of plugs to resist high pressures. What is lacking is experience in the longevity of such plugs. The approach usually adopted therefore is provision of multiple, redundant plugs using natural and engineered materials in simple, robust configurations. It may also be necessary or desirable to place internal plugs or seals. These may vary from immediate borehole or emplacement closures to those isolating sections or panels within the repository. For physically smaller situations it may be possible to perform *in situ* testing but for larger instances (e.g., panel closures) this is not really practical since the amount and type of instrumentation required to validate performance may actually impact that performance. In any case it is impossible to perform such tests for the time spans over which the component must perform acceptably.

E. Computer Analyses

Very large-scale computing, now possible through the use of faster computer chips, parallel processing or network clusters, enables complex models of the mechanical, hydrological, thermal, biological, and chemical systems governing the behavior of a nuclear waste repository. These capabilities also enable better and more effective presentation of the results of computer analyses and if appropriately done enables clear presentation to the public.

F. Design Process

The design process for a deep geological nuclear waste repository will follow the same essential steps as for any complex facility presenting a significant risk to public health and the environment. Regulators, apparently responding to public demand, have set extremely demanding performance requirements that are often based on risk and probability. It does not seem logical to demand a higher performance standard for a facility that in extreme circumstances might cause a possible health risk than for a facility, such as the Channel Tunnel, where any of a variety of problems could cause the immediate death of many hundreds of people. The public must accept some responsibility for this situation. It is willing to accept considerable voluntary risk when it believes it is in control of its fate but seems unwilling to accept any involuntary risk when its fate is apparently in the hands of others, that fate is in the distant future, and, especially, is difficult to comprehend.

It is unfortunate that in promulgating regulations setting acceptable performance levels, sight is often lost of the real objective of the exercise. The purpose of performance assessment is to answer three questions. (1) What can go wrong? (2) How likely is it to go wrong? (3) What are the consequences? Answers to these questions, if based on reasonable measurements and values and compared to the risks of current conditions and activities, should present a means for the public to consider and decide on the acceptability of that risk. Instead, the results are presented in a complex form which while statistically satisfying is completely unintelligible by the public. It seems reasonable to conclude that the current process does not deal with the real risks to human health and the environment since it delays effective disposal and subjects the public in areas where the waste is currently stored to unnecessary potential exposure.

G. Acceptable Performance

A proposed repository should be evaluated on the best available data and a set of assumptions about the future consistent with the best scientific and performance assessments possible. These should be based on actual site, local, and regional conditions. Regulatory agencies clearly have the responsibility to require a comprehensive and thorough assessment but they should not enter into the characterization and assessment arena by issuing regulations that arbitrarily and prescriptively set site characteristics and parameters. Having said that, the designer is faced with the fact that such arbitrary and prescriptive regulations do exist. The site characterization work-plan and performance assessments therefore must be carefully designed to ensure that the best available data are available to support design and permitting activities. It may be possible to eliminate the concern (e.g., extract all known resources) or design around it (e.g., compartment the disposal facility).

At present elements of the public are skeptical about the ability of users of nuclear materials to safely dispose of their waste materials. This situation arises from a number of factors. The first is that most people associate the word nuclear only with weapons of awesome power and destructiveness. Even those who accept the need for such weapons are fearful of the processes which created them. The second is the approach to protection of human health and the environment that focuses on obtaining the best final solution at any cost rather than achieving any type of significant, albeit stepwise, improvement. Finally, the presentation of data and results of assessments are rarely if ever presented in a format intelligible to the layperson and which puts the risks associated with a course of action in a readily understood context. Good presentation and contextual associations may be the means by which a fearful, involuntary public is educated to become an understanding, questioning, yet eventually accepting, one.

H. Step-Wise Approach to Characterization and Performance Prediction

The process of site characterization and performance assessment for nuclear waste disposal facilities should include the following steps.

- Establish the need and demand for a repository in quantitative terms. This essentially describes the inventory of wastes to be disposed of. This should be as specific as possible since uncertainty will be interpreted as deviousness. It may be preferable to start small with a well-defined waste inventory and do a good job with that than to try to be all things to all people and never get started at all.
- Establish a performance envelope based on current regulations and best management practices. In many

cases the regulatory regime is like a game with a constantly moving goal line, in which case this may be virtually impossible to achieve. The only recourse then is to establish an arbitrary performance envelope, work for its acceptance, and modify as changes occur.

- Describe current capacities, capabilities, and facilities where wastes are stored and disposed in detail sufficient to allow a detailed, comparable assessment of their ability to protect human health and the environment. This will establish a base of current conditions against which a proposed repository can be evaluated in comparative terms by the public. The probability that WIPP would induce one cancer death is one chance in a trillion is meaningless and irrelevant to the public since they have no basis on which to evaluate it. Numbers such as these always need a context or, better, should be restated in terms of a comparison to a commonly-accepted voluntary risk (e.g., driving or flying)
- Describe alternative industries appropriate to the proposed location and evaluate the risks and the benefits to the community presented by this alternative use. Transportation of nuclear wastes is often stated to be unacceptably risky yet industries which require the movement of large quantities of hazardous materials are considered quite acceptable. It is often stated that nuclear waste disposal will create a situation which repels other economic development. In reality it is probably like a magnet, which repels but also attracts. A careful comparison of economic, cultural, and community effects will be a significant element in developing an understanding by the public of the total impact of a repository.
- Develop several alternative configurations and sites for the proposed repository. Evaluate each sufficiently to permit a valid comparison to the current situation and to assess the complete impact on the region. The objective is to allow a fair comparison of the risks involved with the benefits achieved. The intent is not to find the best solution, but the one that offers the greatest risk reduction with the greatest overall benefit at the least cost. It is important for the public to appreciate that there are risks and costs associated with delaying the decision to pursue alternative disposal options.

Once the basic data are available it should be possible to cycle through this sequence quickly and cheaply. This technique would be useful in reviewing alternatives with localities which may be interested in a nuclear waste repository as a vehicle for economic development. At the same time it would also clearly show to those communities where waste is currently stored the immediate benefits they derive from a better total waste disposal program.

V. PROJECTS

Nuclear waste disposal facilities have been planned and are being planned around the world. Followings are some of the projects that have planned a nuclear waste repository program and made considerable progress toward implementing it.

A. United States—Waste Isolation Pilot Plant

The Waste Isolation Pilot Plant (WIPP) is the first nuclear waste repository certified by the United States Environmental Protection Agency (EPA). Figure 1 presents a three-dimensional view of the current and expected configuration of the WIPP facility. The purpose of the WIPP is to isolate wastes contaminated with long-lived, mostly alpha-emitting radionuclides resulting from the development and production of nuclear weapons. The WIPP regulatory regime is historically complex, U.S. Congress through the Land Withdrawal Act determined that the Environmental Protection Agency (EPA) would establish, and determine compliance with, radiation release limits for WIPP. Because Congress did not exempt WIPP from compliance with the Resource Recovery and Conservation Act (RCRA), WIPP must also comply with the regulations for the disposal of hazardous wastes, that is whenever radioactive wastes are mixed with wastes regulated under RCRA. In New Mexico RCRA is implemented by the New Mexico Environment Department (NMED).

Regulation by two separate and distinct regulators has led to a morass of paper, much of which is duplicative. It is argued that the two agencies cover different periods, the EPA the long-term and the NMED the short-term, but in reality there is extensive overlap and redundancy which is aggravated by day to day operational regulation by yet another set of agencies. On top of the regulatory agencies, there is a vast number of oversight groups each with differing qualifications and agendas. This type of regulatory complexity must be specifically recognized and addressed by any repository owner or operator since it will claim an inordinate amount of time and effort, especially during the characterization and certification period.

Success in moving through the characterization, design, construction, certification, and operation stages resulted from:

- The continuous, vocal support of the local community
- The dogged persistence to comply with a cumbersome, complex, and unique regulatory certification process

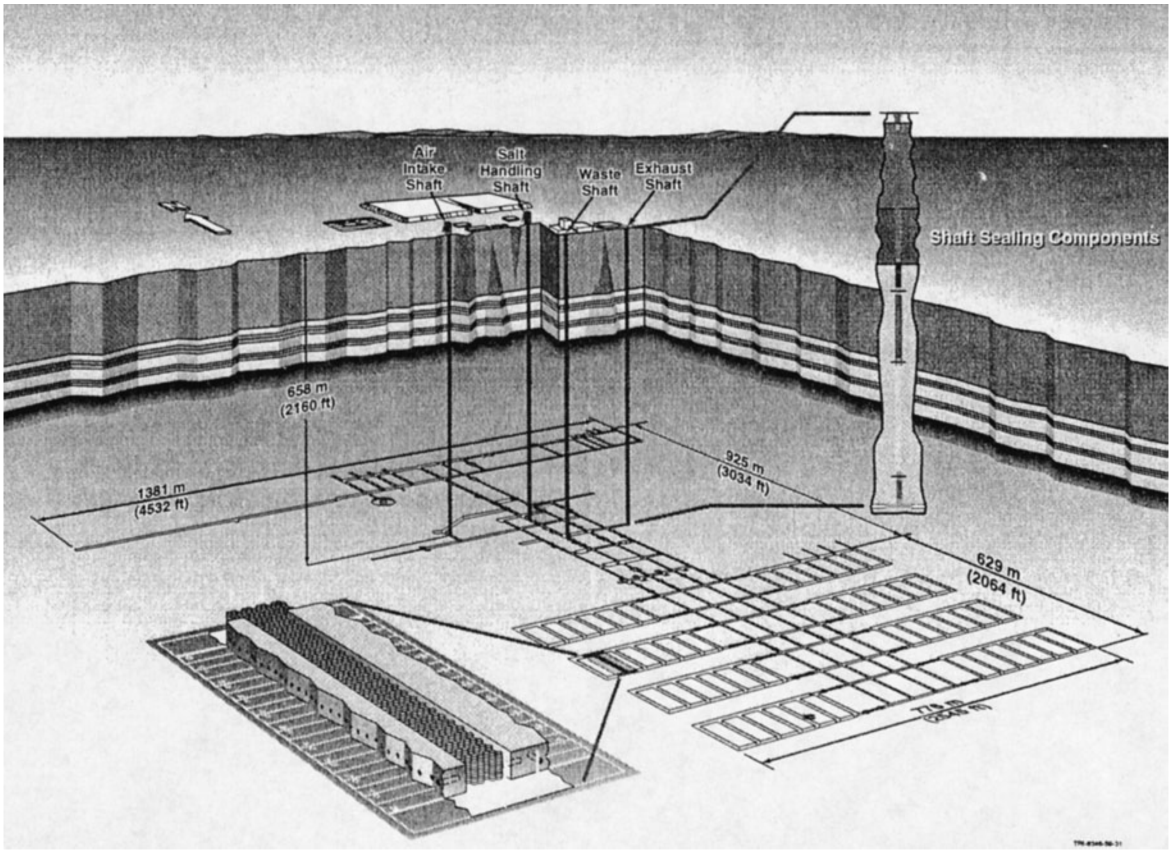


FIGURE 1 Three-dimensional view of the Waste Isolation Pilot Plant.

- An excellent natural setting for such a repository
- A rigorous, transparent scientific and technical evaluation
- A facility with multiple, redundant safety systems.

The Department of Energy was required to demonstrate that the probability of transport of contaminants from the waste panels to the accessible environment remains below regulatory limits for a period of 10,000 years. Demonstration of compliance with the regulations was achieved through quantitative calculations of fluid flow and contaminant transport through a vertical distance from just below the disposal area horizon to the surface, and horizontally through an area extending approximately 40 km^2 around the repository footprint. Releases are calculated within the framework of conceptual models developed for processes affecting releases by a suite of numerical codes which implement these models. All codes and data implemented within the WIPP performance assessment system adhere to a strict quality assurance program that meets both EPA requirements and independent peer review. This program provides assurance to the regulator and to the public that the calculations demonstrating safe operation of the WIPP are based on sound scientific practices. The quality assurance program also requires that parameters which comprise input to the numerical codes be traceable to a known source.

B. Yucca Mountain Project—United States

The U.S. Department of Energy has been studying a site at Yucca Mountain, Nevada, to determine if it is a suitable site for a geologic repository for spent nuclear fuel and high-level radioactive waste. The DOE believes that Yucca Mountain is a promising site and is preparing to recommend the site to the President as a repository in 2001. The preliminary repository design includes a long-lived waste package and takes advantage of the desert environment and geologic features of Yucca Mountain. The natural and engineered barriers can keep water away from the waste for thousands of years.

C. Finland

A community in Finland has agreed to host a repository and an application has been made for a siting permit. There are two underground storage sites currently operating at two nuclear power plant sites. Repositories for low- and intermediate-level radioactive waste started operation in 1992.

D. Sweden

Sweden has a transport system, a repository for operational waste, an interim storage facility for spent fuel, a canister laboratory and an underground hard rock laboratory. Sweden began siting studies for a deep repository in 1992 and test drilling at two sites is expected to start in 2002. The initial operation of a deep repository is expected around 2012. The Aspo Hard Rock Laboratory provides an important scientific and technical basis for the implementation and operation of a future repository in Sweden.

E. Canada

Canada has 22 operating nuclear reactors that generate about 14% of the country's electricity. To meet the needs of the waste disposal research and development program, Canada constructed the Underground Research Laboratory (URL) to provide a representative geological environment in which to conduct large-scale multidisciplinary experiments. A comprehensive program of geotechnical and geomechanical studies at the URL has contributed to defining a robust conceptual design for an underground repository. Results from research at the URL were used in the assessment of the feasibility and safety of deep geological disposal as documented in an Environmental Impact Statement.

After a public review by an Environmental Assessment Panel, the Panel acknowledged that from a technical perspective, the safety of the Canadian concept for nuclear fuel waste disposal was, on balance, adequately demonstrated. However, the Panel recommended that broad public support be demonstrated before proceeding to repository site selection. Work continues at the URL, and the intent of this work is to increase confidence in proposed methodologies for designing repository excavations and engineered barrier systems.

F. Japan

Japan plans to reprocess its commercial waste and immobilize the final waste products in a glass matrix. After a period of 30 to 50 years of temporary storage, the vitrified waste will then be disposed of in a deep geological formation. The organization responsible for managing high-level waste disposal will be established around the year 2000. It will then start site selection and characterization, demonstrate disposal technology, establish the necessary legal infrastructure, prepare relevant licensing applications, and construct the repository with the objective of starting repository operation by about 2035.

The disposal program in Japan is currently in the R&D phase. Japan is using two mines as underground research laboratories for conducting research on the geomechanical aspects of repository behavior. Two new underground research laboratories, in hard and soft rock, are being considered.

G. Germany

Germany is re-examining its policies toward nuclear waste disposal. There is an intense public debate as how to proceed with the disposal of nuclear waste. There is public opposition to moving waste to an interim storage facility near Gorlben, the site of the German candidate repository and the decision on whether to proceed with development of a repository at Gorlben has been delayed. Operation at the Morslen repository has been interrupted and Konrad repository awaits its final license.

H. United Kingdom

The United Kingdom conducted extensive studies at a site near Sellafield in west Cumbria. This included a major site investigation, repository shaft and cavern analysis, coupled hydromechanical system models, and continuum and discontinuum predictions of shaft stability. The future of UK program has been uncertain since March 1977, when the appeal against the adverse underground laboratory planning decision was rejected. However, in 1999 a House of Lords Select Committee conducted a full-scale inquiry into the management and disposal of radioactive waste. It recommended that the government should act without delay in the formation of a new commission to develop and oversee strategy, and of a new company to design, build, and operate a repository.

I. Others

Spain has stopped its siting program. Holland has suspended its work on disposal projects for 100 years.

SEE ALSO THE FOLLOWING ARTICLES

Environmental Radioactivity • Environmental Toxicology • Nuclear Fuel Cycles • Nuclear Reactor Materials and Fuels • Nuclear Safeguards • Radiation Shielding and Protection • Radioactive Wastes • Radioactivity

BIBLIOGRAPHY

Nuclear Energy Agency (1999). “Geological Disposal of Radioactive Waste, Review of the Developments in the Last Decade,” AEN/NEA, France.

Saeb, S., and Francke, C. (eds.) (1999). “Proceedings of the International Workshop on the Rock Mechanics of Nuclear Waste Repositories,” American Rock Mechanics Association, Alexandria, VA.

Board on Radioactive Waste Management, National Research Council (1999). “Discussion Papers prepared for the Workshop on Disposition of High-Level Radioactive Waste Through Geological Isolation,” National Academy Press, Washington, DC.

U.S. Department of Energy, Office of Civilian Radioactive Waste Management (1998). “Viability Assessment of a Repository at Yucca Mountain, Overview,” U.S. Department of Energy, Yucca Mountain Site Characterization Office, Las Vegas, Nevada.

NRC (1957). “The Disposal of Radioactive Waste on Land: Publication 519,” National Academy Press, Washington, DC.

NRC (1990). “Rethinking High-Level Radioactive Waste Disposal: A Position Statement of the Board on Radioactive Waste Management,” National Academy Press, Washington, DC.

NRC (1995). “Technical Bases for Yucca Mountain Standards,” National Academy Press, Washington, DC.

Witherspoon, P. A. (ed.) (1996). “Geologic Problems in Radioactive Waste Isolation, Second Worldwide Review,” Lawrence Berkeley National Laboratory, University of California, Berkeley, CA.

Soil and Groundwater Pollution

Michael J. Barcelona
University of Michigan

Gary A. Robbins
University of Connecticut

GLOSSARY

Advection The movement of ground water in the saturated zone in response to differences in groundwater elevation, hydraulic head or pressure differences.

Dispersion A hydrologic process which affects mixing of solutes at the grain scale owing to differences in flow path, and rates, which tend to elongate and broaden contaminant plume dimensions.

Redox Reactions Oxidation and reduction processes which affects the stability and persistence of electron acceptors (e.g., oxygen, ferric iron, nitrate, etc.) and donors (e.g., organic matter, ferrous iron, sulfide, etc.) often mediated by microbial activity.

Saturated Zone The subsurface environment where the void space in rock or soil is filled with ground water.

Sorption A physicochemical phenomenon which retards the movement of solutes due to adsorptive and absorptive partitioning on aquifer sediments.

Unsaturated (Vadose) Zone The subsurface environment from the surface soil zone to the capillary fringe above the zone of saturation.

THE CONTAMINATION of soil and groundwater with organic, inorganic, and microbial constituents has occurred for many centuries due to improper handling of waste materials ranging from septic wastes to those arising from products of commerce. In the last century, the rise of industrial manufacturing complexes, synthetic chemical/petrochemical, metal working and mining processes have led to the release of complex mixtures of contaminants. Chemical contamination of shallow soil and groundwater by petrochemical, pesticides, radionuclides, metals, and agricultural chemicals is recognized as a major potential human and environmental health risk due primarily to ground water use for drinking water. The reduction of groundwater supply from over use and degradation of groundwater quality by contamination in many countries have become major issues bearing on public health and socioeconomic development. Water resource development and protection programs have responded to these challenges over the past 2 decades with the application of a diverse array of options to ameliorate long-term impacts.

The subsurface is a dynamic environment extending downward from the earth's surface. It is composed of soils or sediments overlying bedrock. It forms the framework for physical, chemical, and biological processes that influence the fate and transport of contaminants. A multitude of surface and near-surface sources of contamination exist which vary in extent, frequency, and magnitude of potential human or environmental risk that may persist for years to many decades. Contaminant mixtures ranging from soluble compounds and biochemical transformation products are transported in the subsurface by ground water. During their transport, contaminants are attenuated or transformed into other constituents by hydrogeochemical processes. Scientific and engineering approaches to subsurface contamination have matured from initial detection and assessment efforts to current programs, which incorporate integrated hydrogeochemical and microbial remediation approaches in the context of the three-dimensional nature of geologic conditions and contamination.

I. NATURE OF THE SUBSURFACE ENVIRONMENT

A. Soils and Sediments

Soils are the *in-situ* products of physical and chemical weathering of rocks or preexisting soils. The type of soil and its thickness are largely functions of the parent material and the climate (whether humid or arid). As shown in Fig. 1, soils are aggregates composed of solid materials (rock fragments, minerals, or organic matter) and void space that may be filled partially or wholly with water. Sediments are soils that have been transported away from their source and deposited. Sediments may be transported by gravity movement down slope, surface water, such as streams or rivers, wind, or glacial ice. In pollution investigations, soils and sediments are typically classified using the Unified Soil Classification System. In this system soils are classified on the basis of grain size and grain size distribution. From a pollution standpoint the most critical properties of soils are listed in Table I. These properties influence the migration of water and contaminants in the soil.

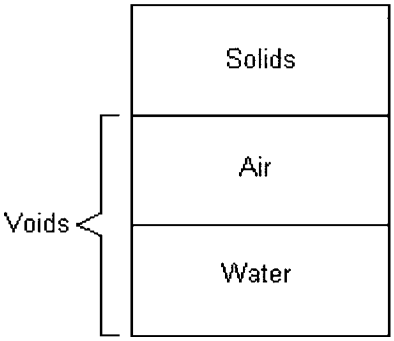


FIGURE 1 Soil phases.

TABLE I Properties of Soils and Sediments that Influence Subsurface Pollution Migration

Properties
Mineralogy
Grain shape
Grain size and distribution
Density
Porosity (amount of void space)
Soil moisture
Organic content
Degree of cementation or compaction
Anisotropy
Heterogeneity

B. Rocks

When soils or sediments are compacted or cemented they form a type of rock called a sedimentary rock. Sedimentary rocks also include rocks formed by chemical precipitation (e.g., limestone) and organic deposition (e.g., coal). If these rocks are exposed to high temperatures and pressures resulting from large-scale earth processes, they can be recrystallized into metamorphic rocks (e.g., slate, marble). If rocks are subject to melting and then cooling, they form igneous rocks (e.g., granite).

C. Geologic Setting/Conceptual Model

The geologic setting includes the types of soils, sediments and rocks, their properties, and their three-dimensional distribution in space. In evaluating the migration of pollutants in groundwater, the geologic setting is analogous to the plumbing in a house. It controls the direction of groundwater and contaminant movement. A key objective in conducting site investigations of pollution entails obtaining information on subsurface geologic conditions. This information aids in the development of a conceptual model used in evaluating contaminant conditions and developing a remedial strategy. Historically, the scope of geologic investigations conducted at contamination sites was an extension of the type of investigation conducted for construction purposes. We have since learned to apply new techniques and methods to help expedite site investigations and reduce their costs. Significant advances have also been made in investigating fracture conditions in rock.

D. Hydrogeologic Settings

Figure 2 shows an idealized profile of the subsurface. From the surface down to the water table, the soil void space is partially filled with water and is termed the unsaturated zone. Below the water table, the soil void space is completely filled with water. This zone is termed the

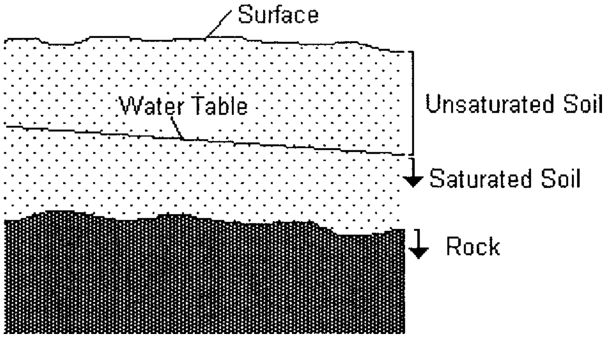


FIGURE 2 Subsurface profile.

saturated zone and the water in the saturated zone is called ground water. Just above the water table, the pore space may be completely filled with water but the water is held by capillarity. This zone is called the capillary fringe. Water infiltrates down into the soil to the water table under gravity and capillarity and then continues to migrate as groundwater. The groundwater environment is dynamic and ground water is always moving. In general, ground water moves from areas of high elevation to areas of low elevation. The exact path that the groundwater and pollution will take is dependent on how groundwater elevation changes and the geologic setting. For example, in igneous and metamorphic rocks, ground water flow is restricted to fractures and faults in the rock (see Fig. 3). The movement of the groundwater is then dependent on the orientation and aperture (opening) width of fractures, fracture continuity, and the spatial variation in ground water elevation (hydraulic gradient). In soils, sediments, and sedimentary rocks groundwater and pollution move predominately through the pore spaces between grains. In some cases these materials may also be fractured and flow through both fractures and pore space occurs simultaneously. In recent years, we have learned that if soils or rocks are exposed for long periods of time to pollution, the pollutants can become trapped in dead-end pore (or fracture) spaces by diffusion. Remedial efforts may be required for a long period to allow pollutants to diffuse out of the dead-end pore spaces.

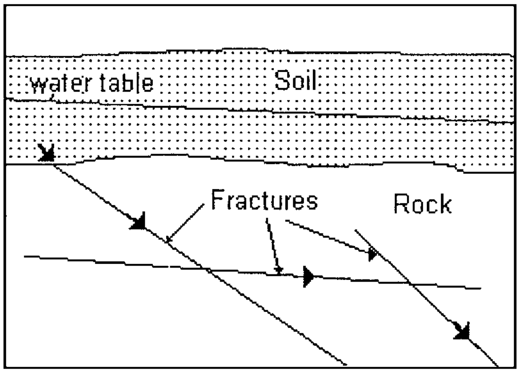


FIGURE 3 Ground water flow restricted to fractures in crystalline rock.

Typically, groundwater pollution is characterized through the installation and sampling of wells. We have since learned that wells may provide misleading information owing to vertical concentration averaging. In evaluating the hydrogeologic setting we have learned that a three-dimensional point of view is required to understand the flow of groundwater and migration of pollution in the subsurface. This requires installing multilevel samplers in soil or sediments and conducting packer testing in wells in fractured rock.

II. SOURCES OF SUBSURFACE CONTAMINATION

A. Source Type and Magnitude

Subsurface contamination can occur due to releases of chemical and microbial constituents from both natural and anthropogenic sources. Contaminant sources therefore vary in extent and magnitude by virtue of their origin. Natural sources may be exemplified by point releases from economic mineral deposits (e.g., petroleum seeps) or trace elements contained in regional geologic materials (e.g., release of arsenic due to natural reduction processes). Similarly, anthropogenic sources include point releases from underground fuel or solvent storage tanks as well as local to regional sources, such as waste impoundments, landfills, mining activities, or excessive ground water withdrawals resulting in the intrusion of saline waters or brines. It should be recognized that due to the interplay between hydrologic processes and the nature of contaminant releases, the demarcation between point and nonpoint sources may be indistinct, particularly in historical areas of commercial, industrial, and human activity.

A number of authors have provided detailed lists of contaminant sources based on the origin as well as the scale of consequences. The focus of the present treatment is on anthropogenic sources, which range over multiple scales of source extent and contaminant impacts (Johnson *et al.*, 2000; Westrick, 1990).

Table II contains a categorization of sources of contamination developed by the USEPA (1976) and Office of Technology Assessment (1984) as referenced by Canter *et al.* (1988). The categories are based on a functional classification of various activities. It is evident from the previously cited categorization that many engineered systems designed to contain potentially toxic or hazardous products and wastes need to be considered in the inventory of sources. This point is borne out by the detailed listing

TABLE II Sources of Groundwater Contamination

Category	Type of source (example)
I.	Sources designed to discharge substances (Injection wells, land applications)
II.	Sources designed to store, treat, and/or dispose of substances, discharge through planned release (landfills, impoundments, tanks)
III.	Sources designed to retain substances during transport or transmission (pipelines, motor transportation)
IV.	Sources discharging substances as consequences of other planned activities
V.	Sources providing a conduit or inducing discharge through altered flow patterns (oil, gas, or water wells)
VI.	Naturally occurring sources whose discharge is created and/or exacerbated by human activity (natural leaching, saltwater intrusion, ground water–surface water interactions)

of sources with the estimated amount of material involved contained in Table III.

It should be noted that the number of facilities and the approximate amounts of material involved are subject to considerable uncertainty, which is common to both data sources. In any one category, the specific design, age, release characteristics, and the resident hydrogeologic setting all exert influence on the duration and extent of potential releases to soil or groundwater. It is entirely possible that some of the sources represent little or no continuing threat to the environment or human health since regulations have been passed requiring upgrading or replacement of critical components. An example of this view is the status of underground storage tanks for fuels which were to be removed or upgraded by the fall of 1998 by regulation (USEPA, 1976).

Certain sources were identified as "important" in the Office of Technology Assessment report due to their number, diversity, areal extent and toxicity. These sources have been designated by a star (*) preceding their place in the table. In general, these sources represent those which are numerous, widespread, with multiple types of contaminants which may pose a threat to human or environmental health.

Sources can only be generally linked to specific chemical or microbial contaminants and their potential effects on human and environmental health. The picture at local, regional, and national scales may be blurred by chemical contaminants, which appear in several categories (e.g., benzene or trichloroethylene which have been associated both with landfill leachates and leaking storage tanks). Also, natural biodegradation processes act on primary contaminants to yield transformation products (e.g., vinyl chloride and dichloroethylene from trichloroethylene or perchloroethylene).

The correspondence of sources and the potential threat from specific contaminants can be better appreciated by considering sources as a function of scale, i.e., either point or nonpoint sources. Point sources are those which are confined or generally release contaminants in a single, geographic area. Example of point sources are septic or underground storage tanks, landfills, or surface impoundments. Nonpoint sources may affect large geographic areas by the nature of their release characteristics. The principal examples of nonpoint sources are agricultural applications of fertilizers and pesticides in rural areas. In this case, millions of acres dedicated to selected crops (e.g., cotton, corn, soybeans, wheat, etc.) receive annual applications of pesticides or nutrients (i.e., nitrate, urea, phosphates). These substances may infiltrate with groundwater recharge, resulting in widespread contamination.

Table IV contains a selective review of contaminants from point and nonpoint sources with a general estimate of persistence and risk to human environmental health. The persistence and risk estimates are generalized on the basis of subsurface transport and fate and risks to surface water environments or drinking water.

Identification of contaminant source types and major contaminants are really the initial steps in the evaluation of the potential for soil and groundwater contamination. From this point, a number of factors having to do with the hydrogeologic settings, pollutant fate, and transport and potential exposure points (i.e., sensitive aquatic environments and drinking water sources) should be considered. A number of source and site prioritization methods have been proposed for use in assessment of pollution potential (Aller *et al.*, 1985; Canter, 1985; Canter *et al.*, 1988) for product and waste management activities. The end result of these applications of these methods is a composite score, usually based on a conceptual model of pollution potential with which to select sites or prioritize further assessment efforts.

B. Extent of Subsurface Contamination

The extent and magnitude of subsurface contamination are largely dependent on source characteristics, contaminant properties, and the influence of hydrologic and biogeochemical processes on their transport and fate. Contaminants of concern to human or environmental health may originate from diverse sources and be transported differentially. For example, benzene and trichloroethylene detected in groundwater due to storage tank releases or landfills are soluble contaminants with major differences in contaminant behavior. Benzene, irrespective of source characteristics, may migrate in ground water but is known to be microbially degraded, particularly under aerobic conditions, to organic acids and other less harmful metabolites. Trichloroethylene on the other hand, is transformed by microbial reductive dechlorination processes to

TABLE III Ground Water/Contamination Sources Categories and Extent

Source	Approximate number of facilities	Approximate amount of material[a]
Category I		
∗ Subsurface percolation	~16 to 20 M	>800 bgy
Injection wells	>0.5 M	~450 to 600 bgy
∗ Land applications sites	>3,000	~4 Mtons/yr
Category II		
∗ Landfills	>95,000	>280 Mtons/yr
Open dumps	~2,400	~10 bgy
∗ Surface impoundments	>180,000	>800 bgy
∗ Waste tailings	?	~580 Mtons/yr
∗ Underground storage tanks	>2.5 M	>35 bgy
∗ Radioactive disposal sites	~31	~3.7 Myd^3
Category III		
Pipelines	>900,000 miles	>280 bgy
∗ Materials transport/transfer	>16,000 spills	~14 Mtons/yr
Category IV		
∗ Irrigation practices	~50–60 M acres^2	~169 M acre-ft
∗ Pesticide applications	~280 M acres^2	~0.3 Mtons/yr
∗ Fertilizer applications	~229 M acres^2	~42 Mtons/yr
∗ Urban runoff	~21–33 M acres	?
∗ Mining/mine drainage	>80,000 active on >4 M acres	~108 bgy
Animal feeding	~1935 sites	~8 Mtons/yr
Category V		
Oil production wells	~548,000 active >2 M abandoned	?
∗ Other wells	?	?
Category VI	?	?
Ground water–surface water interactions	?	?
∗ Natural leaching	?	?
Saltwater intrusion	?	?

[a] Units: bgy, billions of gallons/yr; bg, billion gallons; Mtons/yr, millions of tons/yr; Myd^3, millions of yd^3.

dichloroethylene and vinyl chloride, which are far more mobile and potentially toxic contaminants.

The dimensions of subsurface transport and transformation coupled with the properties of intermediate by-products further complicate risk assessments as well as the application of technologies to manage potential health or environmental impacts. For example, contaminant characteristics as well as associated risks differ significantly between source areas and the dissolved plume. Fortunately, predictive models incorporating fate and transport considerations have been developed to cope with these complications given adequate source and site characterization data. These models are widely distributed and are discussed in depth in Section III.

Over the past 2 decades, local discoveries of impaired soil or groundwater quality conditions have led to national, regional, and state contaminant assessment programs (Council on Environmental Quality, 1981). Assessment programs initially focused on drinking water quality and organic contaminant detection frequency. They were later directed at the extent of contamination of the subsurface using both public water supply wells and those installed for monitoring purposes at known sites of contamination. Early drinking water quality assessment programs should not be interpreted solely as indicative of the quality of the resource, since finished rather than raw drinking water quality was measured. In this respect, they also included the effects of drinking water treatment and disinfection. Disinfection of drinking water by either chlorine compounds or ozone has been shown to result in the production of disinfection by-products from reactions with natural organic matter including haloform compounds (e.g., chloroform, bromoform, dichloromethane etc.).

Assessment results directed toward the contamination of the resource have been conducted in increasing numbers at the national, regional, and state levels. Selected examples of the results of resource contamination assessment

TABLE IV Major Contaminants from Selected Point and Nonpoint Sources with General Estimates of Persistence and Risk

		General[a]	
Source type	**Selected major contaminants**	**Persistence**	**Risk**
Point			
Septic tanks	Ammonia (NH_3)-N	L	M
	[Nitrate (NO_3^-); Nitrite (NO_2^-)]	[H;L]	[M;L]
Underground storage tanks	Fuels; petroleum hydrocarbons	M	M
	[benzene, methyl-t-butyl ether]	[M;H]	[H;M]
	Chlorinated solvents	H	H
	[percholoroethylene; trichloroethylene]	[H;H]	[H;H]
	Breakdown products	[H;M]	[M;H]
	[dichloroethylene, vinyl chloride]		
Landfills/hazardous waste sites	Fuels, solvents	(M;H)	(M;H)
	Trace metals [Chromium, Cadmium, Lead]	[M;H;H]	[M;H;H]
	[Arsenic, Copper, Mercury]	[M;H;H]	[M-H;M;H]
	Organic constituents		
	[polychlorinated biphenyls, phenols, polycyclic aromatic hydrocarbons]	[H;M;H]	[H;M;M]
	Inorganic constituents		
	[HCN]	[H]	[H]
Surface wastewater impoundments	Inorganic and organic constituents (as above)	(as above)	(as above)
Nonpoint			
Agricultural chemicals	Insecticides		
	[Chlordane; Diazinon]	[H;L]	[M;L]
	Herbicides		
	[Atrazine; 2-4D]	[M;L]	[M;L]
	Fertilizers		
	[NO_3^-, NH_3]	[H;L]	[M;L]
Natural soil/sediment constituents	Arsenic, Selenium	[M;M]	[M-H;M-H]
	Radon	[H]	[M-H]

[a] Risk Designations: [H]—high, [M]—medium, [L]—low.

methods from nonpoint sources are detailed in the following with an emphasis on contaminants of concern. It is important to recognize that land use (both past and present) and hydrogeology exert major effects on the occurrence of contaminants as well as their spatial distributions. For example, nitrate and boron concentrations in groundwater were evaluated by Eckhardt *et al.* (1989) for undeveloped, developed (sewered and unsewered), and agricultural land uses. They found that nitrate concentrations were not significantly different among the various uses. Boron concentrations on the other hand were significantly higher in recently sewered and agricultural areas than in unsewered areas but not from long-term sewered areas. The results suggest little difference between nitrate and boron concentrations in shallow groundwater in areas with agricultural or septic tank uses. (Other studies have shown significantly higher nitrate concentrations in both areas of high-density septic tank uses compared to undeveloped areas (Bicki *et al.*, 1984), as well as in shallow groundwater aquifers in agricultural areas (Hallberg, 1986).)

In general, shallow groundwater resources are considered more vulnerable to surface sources of contamination, particularly water table aquifers which are overlain by transmissive soils and thin permeable unsaturated zones. Soil properties or the presence of macropores have been linked to higher pesticide leaching potential and groundwater contamination by DBCP (1,2 dibromochloropropane) in agricultural regions of California (Teso *et al.*, 1988). The conditions in the Netherlands' shallow groundwater resources underscore the vulnerability of these formations particularly in urban and agriculturally intensive areas. In this summary, nitrate, pesticides, and selected trace metals (e.g. Zn, Pb, Cu, and Cd) were found to be significantly higher in shallow groundwater in areas of intensive development or agricultural use. An example of this correspondence for nitrate is shown in Figs. 4a and b, where shallow aquifers with permeable overburden, consistently, showed the highest levels of nitrate. This type of pattern of agricultural land use, groundwater vulnerability, and contamination by septic, fertilizer, and pesticide contamination has been noted in a number of studies (Goetsch *et al.*, 1992; Mason *et al.*, 1987; USEPA, 1982).

Volatile organic compounds, most often associated with fuel and solvent sources (e.g., storage tanks, spills, waste landfills, and abandoned waste sites) have been the subject of a number of groundwater and drinking water investigations (Johnson *et al.*, 2000; Squillace *et al.*, 1996; Westrick, 1990). Here again, the coincidence of high levels of contaminants in shallow ground water and intensive human activities is a clear result.

In the Groundwater Supply Survey reviewed by Westrick (1990), approximately 340 large (i.e., >10,000 people) and 600 small (i.e., <10,000 people) systems were sampled. Sampling was done in both a random and nonrandom fashion. The most frequently observed compounds were chlorinated solvents (e.g., tetrachloroethylene, trichloroethylene, trichloroethane, dichloroethylene, carbon tetrachloride, etc.) and aromatic hydrocarbons (e.g., benzene, toluene, ethyl benzene, and xylenes). Positive results for contamination detection were found in 28% (random) and 37% of (nonrandom) large system samples. Small systems' positives were 17 and 22%, respectively. About half of the water supplies were positive for more than one volatile compound and multiple chlorinated compounds often were often detected in individual samples. This at least in part reflects the breakdown of higher chlorinated compounds to intermediate products like dichloroethylene. The presence of aromatic

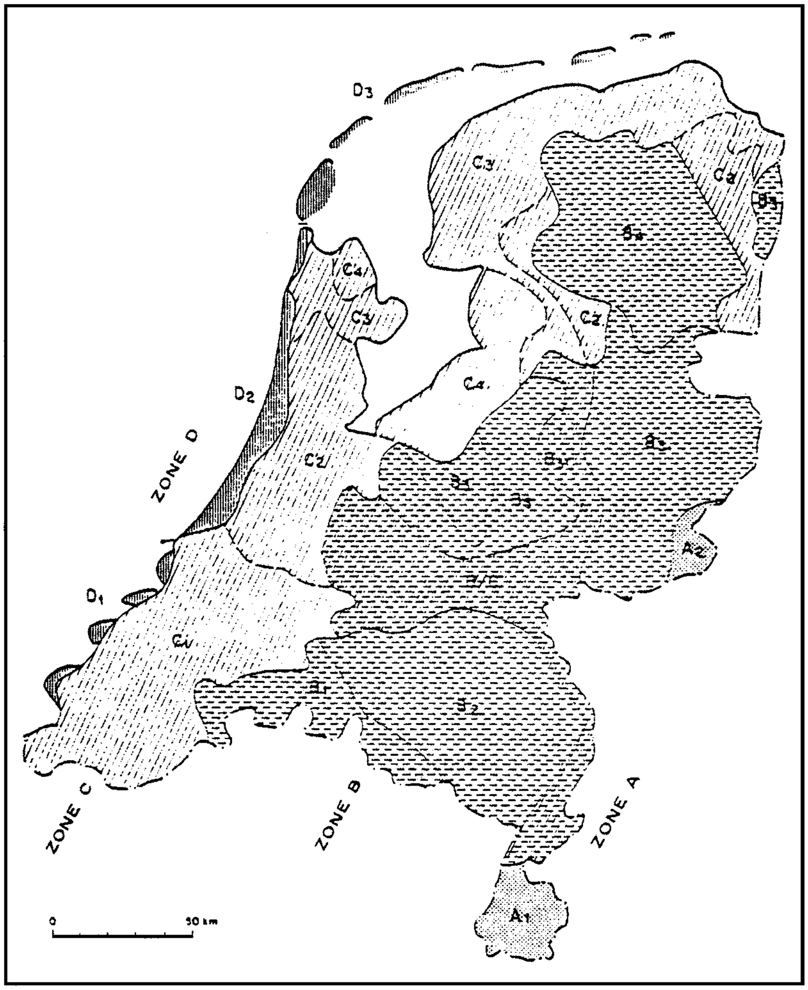


FIGURE 4a Major hydrologic and soil conditions in The Netherlands: (A) higher zone with bedrock close to the land surface; (B) higher freely draining areas with predominantly sandy soils at the land surface; (C) low artificially drained polder areas with mainly clayey and peaty deposits; and (D) higher freely draining sandy coastal dune belt. (From "Regional Ground Water Quality," Copyright 1993, Van Nostrand Reinhold, New York. Reprinted by permission of John Wiley & Sons, Inc.)

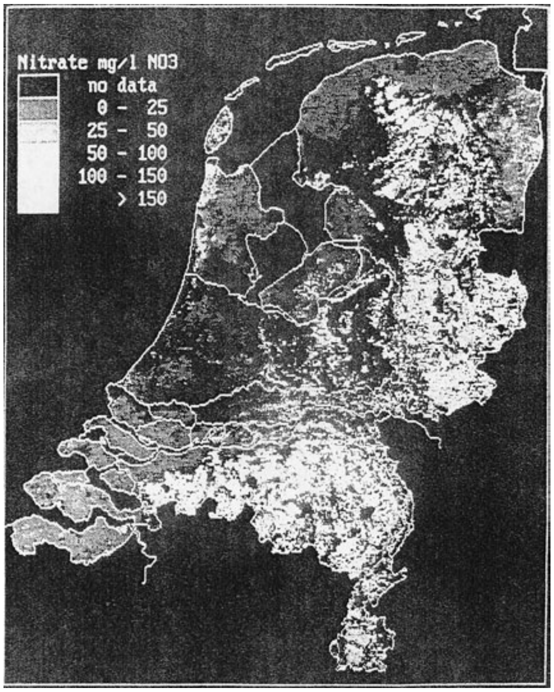


FIGURE 4b Nitrate concentration in shallow ground water based in 1989 data (mg/L nitrate-N). (From "Regional Ground Water Quality," Copyright 1993, Van Nostrand Reinhold, New York. Reprinted with permission of John Wiley & Sons, Inc.)

hydrocarbons seemed to point to contributions from underground storage tanks at service stations in developed areas.

A more recent U.S. Geological Survey-Nation Water Quality Assessment program study of the occurrence of the fuel additive MTBE in shallow groundwater supports this view. Urban areas had positive MTBE results in 27% of the samples while the frequency in agricultural areas was only 1.3%. MTBE was not necessarily found with the aromatic fuel hydrocarbons. This in part can be explained by the greater stability and mobility of MTBE relative to the aromatics. Johnson *et al.* (2000) brought up the question of just how many water supply wells have yet to be contaminated by MTBE given that it has only been in widespread use for the last 2 decades. The fact that the results of these surveys are similar supports the contention that ground water quality, especially where groundwater use and intensive development coincide, needs to be protected if the resource is to continue to yield potable water for a large fraction of our population.

Two major types of contamination of concern are associated with microbial pathogens (e.g., bacteria, virus, cryptospiridium) and chemical or radiogenic elements native to the subsurface environments. Microbial contamination of water supply is a worldwide problem which has far more damaging immediate consequences in causing water borne disease than those arising from chemical sources. Groundwater supplies have traditionally been considered safe from pathogens owing to their limited persistence and mobility in the subsurface. More recent work has shown that viable bacterial and viral mobility can be quite significant and water supply operators should exercise care in well-head protection from septic influences as well as in disinfection practices.

Natural distributions of elements in subsurface geologic formations can give rise to ground water or soil zone contamination. Two examples of note are the generation of radioactive decay products (e.g., radon gas, radium) from natural thorium and uranium, and the release of naturally occurring arsenic or selenium from earth materials.

Wanty and Nordstrom (1993) have reviewed the origins and occurrence of natural radionuclides in ground water and soil gas. They recommend monitoring vigilance particularly for radium and radon gas, which undergo further decay via emission of (beta) and (alpha) radiation. A number of field studies have identified areas in the United States with potentially significant problems with these radionuclides (Gundersen and Wanty, 1993). In many cases, more active ventilation of the subfloors or basements of buildings can minimize radon exposures, while radium problems may require water supply treatment at the plant or the use of alternative supply.

Chemical contamination resulting from the natural occurrence of selected elements in subsurface materials has been recognized as a potentially broad problem in certain areas. For example, under reducing conditions, arsenic III is soluble in the H_3AsO_3 and H_2AsO_3 forms in natural waters. Results from leaching of arsenic minerals within areas of the United States and the world which have naturally high aquifer sediment levels of arsenic levels may be high enough to cause neurological damage, liver damage, and cancers via drinking water. Most of the U.S. reports of elevated arsenic levels in groundwater have been limited to regional studies with relatively small sample sizes. It is notable that, owing to the differential mobility of oxidized versus reduced arsenic and selenium (the opposite trend holds in that reduced Se forms are far less soluble or mobile than oxidized forms), evaluations of contamination must pay careful attention to groundwater age, depth, and geochemistry.

The main questions that need to be considered for many intensive agricultural and urban areas are: How do we confront current contaminated conditions and protect that portion of the resource which has yet to be compromised? The answer to both questions lies in public policy to support the reasoned stewardship of the resource and bringing to bear the best available technologies to protect and remediate known contamination sources and plumes.

III. SUBSURFACE PROCESSES WHICH GOVERN CONTAMINANT BEHAVIOR

A. Transport

In the unsaturated zone, dissolved contaminants migrate in conjunction with infiltrating water. As such, they migrate under the influence of gravity and capillarity. The migration of contaminants in the unsaturated zone is largely controlled by the amounts of precipitation and the unsaturated permeability of the soils and sediments. Another mechanism for the transport of volatile contaminants in the unsaturated zone is vapor diffusion. Chemicals with low boiling points (fuels and organic solvents) can readily volatilize into soil air. Through gaseous diffusion these chemicals can spread in the unsaturated zone and partition into soil pore water or the groundwater. The ultimate fate of volatile organic chemicals will depend on their origin and extent as free-phase liquids (e.g., LNAPL—lighter than aqueous phase, or DNAPL—more dense than aqueous phase) and to which they are involved in chemical or microbial processes. Examples of various release scenarios are shown in Figs. 5a through e.

Once in the saturated zone, dissolved contaminants migrate from their source principally by advection and dispersion. Advection refers to the movement of groundwater and contaminants owing to spatial or temporal variations in hydraulic head (water level elevation). Advection is described by Darcy's law. Variations in hydraulic head may be caused by surface elevation differences, spatial variations in recharge or discharge locations, or be induced by well pumpage.

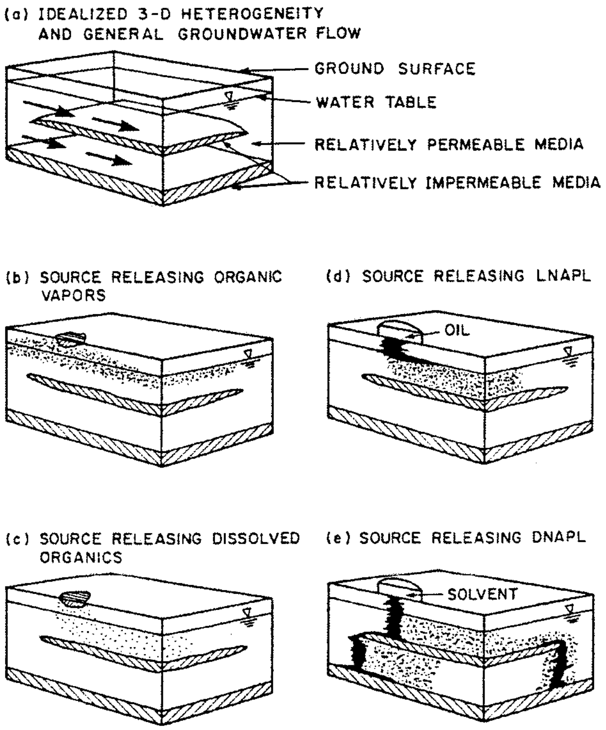


FIGURE 5 Organic contaminant distribution resulting from interactions between geologic heterogeneities and types of sources. Stippling represents dissolved or vapor phase contaminants. Solid black represents nonaqueous phase liquids (NAPLs).

Dispersion refers to the spreading of contaminants owing to velocity variations. Although contaminant spreading can occur by diffusion this is generally not important, except under low permeable conditions where the ground water velocity is very low. Dispersion acts to spread contamination in three dimensions along the ground water flow path. As such contaminant concentrations are reduced with distance from the source. Despite much research on dispersion since the 1950s, capability to quantify dispersion remains illusive. Difficulties stem from apparent scaling of dispersion with contaminant travel distance due to heterogeneities, concentration averaging in wells, or not accounting for the three-dimensional nature of contaminant spreading.

Advection and dispersion of contaminants can be modeled using analytical and numerical methods. Advancement of computer technology now permits the use of complex numerical computational methods on portable computers. Modeling is helpful in evaluating factors that control contaminant migration, assessing the extent of contamination, and evaluating the effectiveness of remedial measures.

B. Reactivity

1. Chemical

Chemicals that migrate in groundwater which are not attenuated by chemical and biological processes are said to be conservative. Conservative chemicals move at the rate of the ground water. Reactive chemicals are those that are attenuated in velocity or are mass dissipated. Table V lists pertinent reactions that can influence contaminant migration. Whether a given reaction is important depends on the chemistry of the contaminant, redox conditions, and the nature of the solids that make up the soil or rock. For example, organic contaminants such as benzene may be velocity attenuated by adsorption on to organic matter in the soil. The extent of adsorption is, among other factors, dependent on the amount of organic matter present. The migration of chromium depends on oxidation-reduction conditions and pH.

These factors control the speciation of chromium. Under oxidizing conditions and low pH, the stable chromium species is hexavalent chromium(+6). Hexavalent chromium binds with oxygen to form a negatively

TABLE V Soil and Vadose Zone Treatment Methods

Method category	Target contaminants	Residual
Excavation-incineration	Organic compounds metals, inorganic compounds, miscellaneous contaminants	Ash to appropriate landfill
Excavation-fixation or Stabilization	All of the above	Landfilling of residue
Excavation-vitrification	Long–lived radionuclides	Appropriate repository
Slurry walls or grout curtains (constructed in place)	Many types of contaminant sources	Isolation of contaminants from vadose zone and upper saturated zone. Requires monitoring
Soil thermal treating via injection of steam, inductive heating, etc. . .	Volatile organic compounds	Vacuum capture and treatment of all gases
Soil vapor extraction and/or venting (introduction of air)	Volatile organic compounds and aerobically degraded nonvolatile organic compounds	Treatment of off gases

charged chromate ion that is highly mobile in groundwater. Under reduced conditions trivalent chromium (+3) is stable and can precipitate out of solution as an oxyhydroxide or adsorb readily on to minerals or organic matter.

The velocity of contaminants may also be influenced by the diffusion of contaminants into dead space pores or fractures (where the ground water velocity is essentially negligible) or within grains. These processes may become significant in attenuating the movement of contaminants at sites where soil and rocks have been exposed to contamination for long periods.

Radioactive substances migrating in ground water undergo radioactive decay. If decay rates are high (half-lives are short) relative to the groundwater velocity, radioactive decay is an effective mass attenuation mechanism.

2. Biological

Bacteria can have a significant influence on the mass attenuation of organic and inorganic chemicals. Bacteria in the subsurface can utilize organic contaminants as a food source, particularly hydrocarbons in fuels. Bacteria are known to play a major role in attenuating the aromatic hydrocarbons from gasoline releases. The oxidation of organic chemicals by bacteria requires electron acceptors. Under aerobic conditions, oxygen is the electron acceptor. Under anaerobic conditions, iron (Fe III), manganese (Mn IV), nitrogen (Nitrate), and sulfur (Sulfate) act as electron acceptors. Under highly reduced conditions, methanogens can utilize organic contaminants in the production of methane. The utilization of organic contaminants by bacteria generally results in changes to the overall water quality (e.g., increase in pH, decrease in Eh, increases in total dissolved solids, reduced metal species, alkalinity). Although biological degradation of contaminants is complex in detail, it is often simulated as a first-order process (Domenico, 1987).

IV. REMEDIATION OF SOIL AND WATER CONTAMINATION

Groundwater resources provide approximately one half of the nation's water supply as well as a major fraction of water used for self-supplied industry. Since both the consequences of human activities and the natural leaching processes may compromise the quality of water, treatment is frequently necessary prior to use. Preferably, the treatment would be accomplished by engineered treatment plants prior to distribution to the consumer. However, this is not always possible, particularly in cases where consumers are geographically distant from the resource or where distribution system treatment other than at the point of use is too expensive.

In this sense, impaired water quality can exert a major influence on quantity. In addition further resource damage must be minimized to ensure the quality of future water supply. Treatment of contaminated public resources has been required by statute when the responsible party(ies) are known or the USEPA assumes responsibility for cleanup when the sites of sources have been abandoned. Over the past 2 decades, numerous approaches have been developed to contain contaminated groundwater plumes, treat or destroy contaminants in the subsurface (i.e., *in situ*) or above ground, or manage risk to human or environmental health by active monitoring of ongoing natural processes which diminish contaminant concentrations. These approaches range from those which are technology based such as soil venting, vapor extraction, groundwater pumping with treatment above ground, or *in-situ* methods; to those based on observed plume behavior relative to known exposure points. In the latter case, called natural attenuation, model predictions are made for the outcome of natural processes such as advection/dispersion induced dilution, sorption on aquifer sediments, volatilization, and biodegradation. Natural attenuation remedies require active monitoring as well as periodic reviews of progress toward cleanup or risk management goods. It is implied in regulatory guidance for natural attenuation remedies that steps will be taken to isolate or remove contaminant sources, particularly in cases where dissolved plumes are expanding.

Remediation of near-surface contamination sources frequently requires consideration of vadose zone residuals as well as nonaqueous phase liquids in the capillary fringe or fully saturated environment of groundwater. Though cleanup technologies directed at the vadose and saturated zones are discussed separately here, it should be emphasized that thorough cleanup of the subsurface requires a combined approach. Also, with the exception of pump and treat technologies, the track records of many existing or developing treatment technologies are limited to 10 to 20 years at most. Many of these methods have been found to have limited effectiveness, though research and development efforts are progressing.

A. Soil (Vadose) Zone Methods

Many of the sources of soil and groundwater contamination occur at or near the land surface. With this in mind, source removal implies the removal, destruction, or isolation of contaminated soil which would otherwise represent a long-term source of contaminants to groundwater. Where these approaches are feasible and cost effective, they are clearly preferable to attempt to destroy contaminants in-place. Table VI provides a selective summary of soil and vadose zone treatment methods ranging from excavation and removal to contaminant treatment in place. Above-ground treatments favor total contaminant destruction rather than isolation of contaminants except in the cases of radionuclides for which only time will cause to dissipate the radiation.

In-situ methods are generally less effective than above-ground remedies. However, they minimize surface disruption of utilities, infrastructure, etc., and may extend well below the depth of excavation capability. The construction of grout (clay) curtains or slurry walls has been shown to be effective in source control efforts particularly in formations which are sand sized or greater. The *in-situ* thermal methods have been used successfully in pilot-scale to mobilize organic compounds with boiling points less than 150°C even when the contaminants are in a nonaqueous phase liquid.

TABLE VI Chemical and Microbial Processes Influencing Contaminant Attenuation

Velocity attenuation	Adsorption, ion exchange
	Dead space diffusion
	Intragranular diffusion
Mass attenuation	Radioactive decay
	Biodegradation (coupled with reduction of oxygen, iron, manganese, sulfur, nitrogen or methanogenous)
	Precipitation (resulting from pH or redox changes)

One category of the vadose zone treatment technologies which has been applied in relatively numerous instances is soil vapor extraction (SVE) (with or without the introduction of air) (Hutzler *et al.*, 1989; Johnson *et al.*, 1990, 1994). The most effective applications of SVE employ careful control of applied vacuum through designed extraction points about which are arrayed air introduction points to ensure airflow through the most contaminated zones. Figure 6a shows a typical SVE process diagram (NRC, 1994). Success with SVE methods has been more difficult to achieve where toxic nonvolatile compounds are present in the source, or where permeability of the subsurface limits the zone influence of gas flow.

Vadose zone treatment methods focused on source areas can effectively limit the long-term release of contaminants

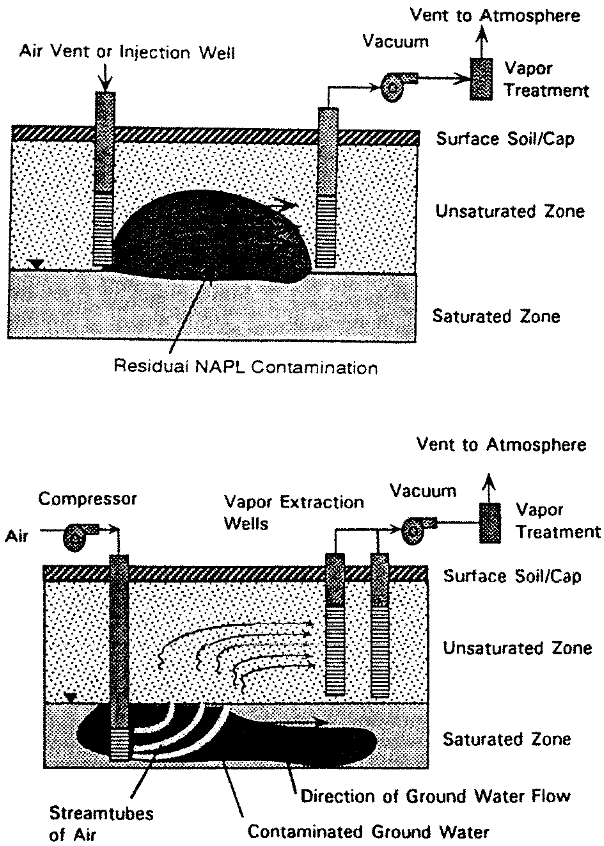


FIGURE 6 Process control diagram for (a) SVE, (b) air sparging, (c) direct chemical injection, (d) geochemical enhancement for bioremediation. [Reprinted with permission from "Alternatvies for Groundwater Cleanup," 1994, by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, D.C.]

to soil gas or ground water. As such, their operation in concert with plume control technologies has shown promise for many mobile, persistent contaminants. Above-ground treatments of off-gases from SVE systems include thermal or catalytic oxidation, combustion in internal combustion engines, and activated carbon adsorption prior to release to the atmosphere.

B. Groundwater (Saturated) Zone Methods

The transport of contaminants in soil gas and groundwater represent major pathways of contaminant exposure to humans or sensitive ecosystems. The vadose zone treatment methods discussed previously exert control over further contaminant release to soil gas when applied and monitored carefully.

Effective treatment or containment of dissolved contaminant plumes and NAPL phases in the capillary fringe or saturated zones represent a continuing challenge to remediation technologies. Effective source control measures depend on careful delineation (in three dimensions) of contaminants, and characterization efforts may limit the success of saturated zone contaminant treatment or control measures. Selected treatment or containment methods for the saturated zone are provided in Table VII. Flow barriers constructed by slurry wall, grout injection, or sheet pile methods have been used in some instances to channel or contain contaminated groundwater usually in combination with pumping control methods.

Among the first saturated zone treatment and containment methods applied for LNAPL contaminants (i.e., lighter than aqueous phase liquids) were interceptor trenches that penetrated the upper portion of the near source saturated zone. In these cases, the trench provides a collection point for hydrocarbon contaminants by skimming, discrete liquid pumping, or sorption for subsequent treatment or re-use.

Pump and treat systems were employed early in the development of groundwater remediation approaches. Pump and treat methods frequently involve pumping groundwater from contaminated portions of the plume to the surface for treatment by air-stripping volatile organic compounds. Off-gases can be treated as in SVE treatment streams. Though they frequently failed to remove significant contaminant mass, these systems often succeeded in controlling plume migration to off-site boundaries or potential exposure points. Plume control and a more rapid approach to groundwater cleanup goals may be achievable when pumping and treatment are followed by reinjection.

Air sparging technologies seek to introduce air (or O_2) bubbles into the upper saturated zone to strip volatile organics from the plume to be captured in the vadose zone, add O_2 to stimulate degradation processes, and influence water flow patterns for enhanced extraction or reinjection. A process diagram for air sparging is shown in Fig. 6b (NRC, 1994). Research has shown that the formation of gas stream tubes for escaping sparged air limited the zone of influence of these systems (Pankow and Cherry, 1996). It is clear that the technology is best suited for degradable organic compounds in plume rather than source areas.

The realization that contaminant mass removal was limited in many instances where the preceding technologies were employed led to the development of more aggressive treatment approaches. In many cases, the newer technologies incorporated the more successful flow control and plume containment advantages of previous work. It should be recognized that continuing improvements in site characterization monitoring approaches enabled better

TABLE VII Saturated Zone Treatment Methods

Method category	Target contaminants	Residual
Slurry wall/grout curtain/sheet pile	Flow control	None
Interceptor trenches	Lighter than aqueous phase Liquids–LNAPLs	Oil/H_2O separation sorption
Pump and treat (with or without reinjection)	Dissolved contaminants	Air stripping and/or carbon sorption for organic compounds off-gas treatment
Air sparging with or without soil vapor extraction or control pumping	Volatile organic compounds	Off-gas treatment and carbon sorption
Reactive permeable barriers	Organic compounds and selected trace metals	Various polishing treatments as necessary
Injection of surfactants, direct chemical reductants, or oxidants with or without control pumping	Organic compounds and selected trace metals	Treatment of pumped effluent
In-situ enhanced bioremediation injection of electron acceptors, donors, and nutrients	Solubilized mobilized or degradable organic compounds and selected trace metals	Treatment of pumped effluent
In-situ bioaugmentation introduction of microorganisms, nutrients, cosubstrates, electron donor/acceptors	Degradable organic compounds and selected trace metals	Treatment of pumped effluent

resolution of contaminant source and plume dimensions by the late 1980s.

Reactive barrier plume treatment methods represent a standard of comparison for *in-situ* treatment of contaminants that are recalcitrant to other mitigation methods. The initial work in North America issued from observations that zero-valent iron (Fe^0 filings) could cause the dechlorination of dissolved tetrachloroethylene and trichloroethylene.

The placement of these materials in the path of contaminant plumes as permeable reactive barriers followed logically. The technology has been employed at numerous locations in the United States and Canada for chlorinated solvents as well as for toxic oxidized trace metals (e.g., Cr VI). Many of these systems have achieved sustained dechlorination and reductive capacity after 10 years of service.

Further *in-situ* treatments have been developed to solubilize or mobilize contaminants, particularly NAPLs by the use of surfactants, or to oxidize or reduce contaminants directly. All of these technologies require reagent injection and extraction flow control and may demand aboveground polishing treatments to assure success. These techniques are emerging from pilot scale to field applications with rather encouraging results (Bedient *et al.*, 1999; Bjerg *et al.*, 2000; Sabatini *et al.*, 1995).

The injection of chemicals into the subsurface to directly treat subsurface contamination often occurs with significant local changes in subsurface geochemistry. For example, oxidant injection (e.g., magnesium peroxide or hydrogen peroxide) can result in increased oxygen concentrations, which may stimulate natural aerobic microbial activity. The enhancement of conditions for microbial degradation of contaminants supports microbial restoration of the subsurface resource. Process diagrams for direct chemical introduction and geochemical enhancement approaches for bioremediation are shown in Figs. 6c and 6d, respectively.

More recent refinements of biostimulation technologies have led to the introduction of microorganisms capable of contaminant degradation along with engineered geochemical adjustments. The organisms introduced in these schemes have been isolated from previously contaminated sites or waste-water treatment plants and are not genetically modified bacteria. Although currently at the research stage, bioaugmentation techniques have shown promise for contaminants, which either fail to degrade or undergo transformation to more mobile and toxic byproducts under natural conditions. Working in concert with natural flow conditions, the success of these technologies rests on the use of low-maintenance injection methods to sustain favorable degradation conditions and microbial viability.

V. CONCLUDING COMMENTS

The field of contaminant hydrogeology in the last 2 decades has incorporated new concepts in mathematical geology, geostatistics, subsurface microbiology, toxicology, biogeochemistry, and environmental engineering. Exciting developments have been made in the detection, assessment, mitigation, and elimination of subsurface contamination, largely in response to the need to meet challenges posed by expedient chemical waste and product handling methods. While significant challenges remain to be met, the tools exist to avoid further contamination of soil and groundwater and minimize risk to human and environmental health.

In the near future, scientists and engineers will be drawn further into advanced resource protection efforts particularly where water quantity and quality concerns clash with the demands of population and economic growth across the globe.

SEE ALSO THE FOLLOWING ARTICLES

DRINKING WATER QUALITY AND TREATMENT • ENVIRONMENTAL GEOCHEMISTRY • ENVIRONMENTAL MEASUREMENTS • ENVIRONMENTAL OBSERVATION AND FORECASTING SYSTEMS • POLLUTION, AIR • POLLUTION CONTROL • POLLUTION, ENVIRONMENTAL • SOIL MECHANICS • SOIL PHYSICS • TRANSPORT AND FATE OF CHEMICALS IN THE ENVIRONMENT • WATER POLLUTION • WATER RESOURCES

BIBLIOGRAPHY

Aller, L., *et al.* (1985). "Drastic," A Standard System for Evaluating Ground Water Pollution Potential Using Hydrogeologic Setting, USEPA, R. S. Kerr Environmental Research Laboratory, EPA/60012-851018. Ada, OK.

Alley, W. M., Reilly, E., and Franke, O. L. (2000). "Sustainability of Ground Water Resources," U.S. Geological Survey Circular 11 86, Denver, CO.

Bedient, P. B., Rifai, H. S., and Newell, C. J. (1999). "Ground Water Contamination, Transport and Remediation," Prentice Hall PTR, Upper Saddle River, NJ.

Bhattacharya, P. O., Chatterjee, and Jacks, G. (1997). "Occurrence of As-contaminated ground water in Allurid Aquifers from Delta Plains, Eastern India: Options for safe drinking water supply," *Res. Devel.* **13**(1), 79–92.

Bjerg, P. C., Engesgaard, P., and Krom, T. D. (eds.) (2000). "Ground Water 2000," Proceedings of the International Conference on Ground Water Research, Copenhagen, DK, A. A. Balkema, Rotterdam, Netherlands.

Bicki, T. J., Brown, R. B., Collins, M. E., Mansell, R. S., and Rothwell, D. F. (1984). "Input of On-Site Sewage Disposal Systems on Surface and Ground Water Quality," Florida Dept. of Health and Rehabilitation Services, Tallahassee, FL.

Canter, L. W. (1985). Methods for assessment of ground water pollution potential, *In* "Ground Water Quality" (C. H. Ward, W. Giger, and P. C. McCarty, eds.), pp. 270–306, J. Wiley, New York.

Canter, L. W., Knox, R. C., and Fairchild, D. M. (1988). "Ground Water Quality Protection," Lewis Publishers Inc., Chelsea, MI.

Council on Environmental Quality (1981). Contamination of Toxic Organic Chemicals, Washington, D.C.

Driscoll, F. W. (1986). "Ground Water and Wells," Johnson Division, St. Paul, MN.

Dubrovsky, N. M., Neil, J. M., Welker, M. C., and Evenson, K. D. (1991). Geochemical Relations and Distribution of Selected Trace Elements in Ground Water of Northern Part of the Western San Joaquin Balley. U.S.G.S. Water Supply Paper #2380.

Eckhardt, D. A., Siwiec, S. F., and Cauller, S. J. (1989). Regional appraisal of ground water quality in five different land-use areas Long Island, N.Y. U.S.G.S. Water Resources Investigations Report 88-4220, Reston, VA.

Gundersen, L. C. S., and Wanty, R. B. (eds.) (1993). "Field Studies of Radon in Rocks, Soils and Water," C. K. Smoley, CRC Press, Boca Raton, FL.

Hallberg, G. R. (1986). "Overview of Agricultural Chemicals in Ground Water. Agricultural Impactson Ground Water," pp. 1–63, National Ground Water Assoc., Westerville, OH.

Hutzler, N. J., Murphy, B. E., and Gierke, J. S. (1989). "State of Technology, Review Soil Vapor Extraction Systems," USEPA Risk Reduction Engineering Lab, Cincinnati, OH.

Johnson, P. C., Baehr, A., Brown, R. A., Hinchee, R., and Hoag, G. (1994). Innovative Site Remediation Technology, Vacuum Vapor Extraction, American Academy of Environmental Engineers.

Johnson, P. C., Stanley, C. C., Kemblowsi, M. W., Byers, D. L., and Culthart, T. D. (1990). "A practical approach to the design, operation and monitoring of *in-situ* soil-venting systems," *Ground Water Monit. Rev.* **16**(2), 159–178.

Johnson, R., Pankow, J., Bender, D., Price, C., and Zogorski, J. (2000). MTBE, To what extent will past releases contaminate community water supply wells? *Environ. Sci. Technol.* 210A–217A.

Mason, R. E. *et al.* (1987). National Pesticide Survey, Pilot Evaluation Technical Report. USEPA, Research Triangle Institute, RTI/7801/06-02F, Research Triangle Park, NC.

Office of Technology Assessment (1984). Protecting the Nation's Ground Water from Contamination, Report to Congress, OTA-0-33. Washington, D.C.

Pankow, J. F., and Cherry, J. A. (1996). "Dense Chlorinated Solvents and Other DNAPLS in Ground Water," Waterloo Press, Waterloo, ONT, Canada.

Robbins, G. A., and Butler, M. A. (1997). Recommended Guidelines for Multilevel Sampling of Soil and Ground Water in Conducting Expedited Site Investigations at Underground Storage Tank Sites in Connecticut, Connecticut Department of Environmental Protection, Hartford, CT.

Teso, R. R., Younglove, T., Peterson, M. R., Sheeks, D. C., and Gallavan, R. E. (1988). "Soil taxonomy and surveys: Classification areal sensitivity to pesticide contamination of ground water," *Journey Soil Water Conserv.* **43**(4), 348–352.

USEPA (1976). RCRA Subtitle I and Sections 7003, 9003(h), 9005 and 9006.

van Deijvenbooden, W. (1993). "Ground Water Quality Monitoring in the Netherlands," Ch. 21 (W. M. Alley, ed.), Ground Water Quality, Van Nostrand Reinhold, New York.

Wanty, R. B., and Nordstrom, D. K. (1993). Natural radionuclides, Ch. 17. *In* "Ground Water Quality" (W. M. Alley, ed.), Van Nostrand Reinhold, New York.

Westrick, J. J. (1990). National Surveys of Volatile Organic Compounds in Ground and Surface Water. *In* "Significance and Treatment of Volatile Organic Compounds in Water Supplies" (N. M. Ram, R. F. Christman, and K. P. Cantor, eds.), pp. 103–138, Lewis Publishers, Chelsea, MI.

Transport and Fate of Chemicals in the Environment

Frank Wania
University of Toronto

GLOSSARY

Advection A process in which the molecules of a chemical move due to their presence in a phase that is moving.

Chemical transformation A process in which the chemical identity of a molecule is changed by the breaking and forming of covalent bonds. In the environment most chemical transformations are a result of the action of light (photochemical transformations) and microorganisms (biochemical transformations).

Diffusion A process in which a chemical moves from regions of higher concentration to regions of lower concentration. The process is called *molecular diffusion* if the chemical movement is due to the random walk of the molecules within the phase, and *eddy diffusion* if the chemical movement is due to the turbulent movement of the phase.

Environmental chemical Either a synthetic chemical that is entirely of human origin and has been released into the environment or a chemical that occurs naturally but whose occurrence in the environment has been enhanced by human activity beyond natural levels.

Equilibrium phase distribution The distribution of the molecules of a chemical between any two phases determined by the relative compatibility of the chemical with each phase.

Sorption A process in which a dissolved or vapor phase molecule becomes associated with a solid phase. We speak of adsorption if the sorption occurs on the surface of the solid, and absorption if the sorption occurs into the three dimensional matrix of the solid.

THE FATE OF A CHEMICAL comprises all the processes that it can undergo when released to the environment. The two major types of fate processes are transport and distribution processes that leave the chemical's identity unchanged and transformation processes that convert the chemical into a different chemical entity. The use of the word “fate” in this context reflects our interest in a chemical’s behavior over its entire “life-time” in the environment from its initial release until it has reached its final destination (i.e., is degraded or has reached a final resting place). A quantitative and qualitative understanding of a chemical’s fate is important because it determines the nature and extent of any potential detrimental effect that a deliberately or accidentally released chemical or its reaction products may have on humans and their environment. A major goal of studying the fate of environmental chemicals is to relate it quantitatively to molecular chemical characteristics. How a chemical is going to behave in the environment, and what potential effects it thus may have, can then be inferred from the chemical’s structure.

Detrimental effects of environmental chemicals may manifest themselves in a variety of ways. Accordingly, all sorts of chemicals can become environmental contaminants (Fig. 1). Sulfur dioxide emitted from metal smelters is subject to atmospheric processes leading to the formation of acid precipitation, which can have series consequences for both aquatic life and plant growth. An excess input of phosphate from detergents and sewage into a lake can lead to eutrophication and oxygen depletion killing much of the aquatic life in that lake. Nitric oxide and volatile hydrocarbons such as benzene and ethene from automobile exhaust can undergo a series of atmospheric reactions to form photochemical smog with negative implications for human health, plant growth, and visibility. The role of the chlorofluorocarbons in the destruction of the stratospheric ozone layer and the resultant increase in ultraviolet (UV) exposure is well known. Pollutants such as methyl mercury or the pesticides DDT and toxaphene have been shown to accumulate in aquatic food chains and cause effects on the health of top predators. Fuel additives such as methyl tertiary-butyl ether (MTBE) have contaminated groundwater resources by leaking from underground storage tanks. A quantitative knowledge of the environmental fate and transport of these and other chemical substances in the environment is a major element in preventing, assessing, mitigating, and remediating their detrimental effects.

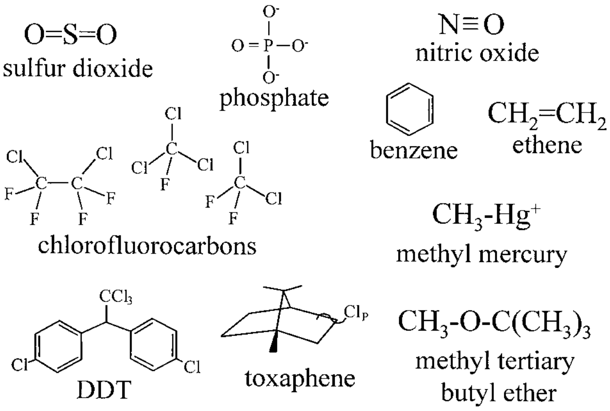


FIGURE 1 Molecular structures of selected environmental chemicals.

I. BASIC CONCEPTS OF CHEMICAL FATE AND TRANSPORT IN THE ENVIRONMENT

A. Environmental Spheres and Phases

The environment is often seen as being composed of environmental spheres, specifically the atmosphere, the hydrosphere, the terrestrial environment, and the biosphere (Fig. 2). Environmental chemicals can occur in all these spheres. If a chemical occurs exclusively or predominantly in one of these spheres, its environmental fate can be reduced to the fate in that particular phase. For example, some very volatile chemicals occur exclusively in the atmosphere, whereas some very water-soluble chemicals are restricted to the aqueous phase. They are sometimes referred to as air pollutants or water contaminants and a distinct sphere-specific approach to such chemicals’ fate and transport may be appropriate. Most chemicals, however, do occur in more than one of the environmental spheres in notable amounts, and a single-compartment perspective is insufficient to capture their environmental behavior.

Despite the obvious differences between the spheres, many of the processes and mechanisms that control chemical fate are common to all of them. In all spheres, transport can occur by advection and diffusion, and many of the chemical reactions are similar in the various spheres. One reason for the similarities is that the environmental spheres are composites of the same principal phases. The most important of these environmental phases are

- Gas phase
- Aqueous phase
- Solid mineral phases
- Various organic phases

The environmental spheres vary primarily in terms of the relative size of these phases (Fig. 3) and often also their chemical composition. Obviously, the gas phase is the dominant component of the atmosphere, but the aqueous phase is present in the form of hydrometeors (cloud and rain droplets, ice crystals), as are organic and mineral phases in the form of suspended particulate matter, called aerosols. And even though the volume taken up by these hydrometeors and aerosols may be miniscule compared to that of the gas phase, atmospheric chemical fate is often

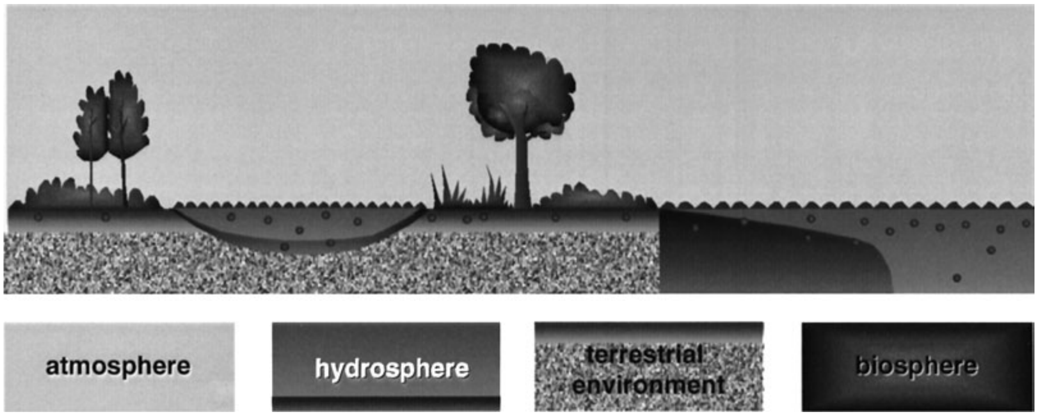


FIGURE 2 Environmental chemicals can occur in all the major spheres making up the natural environment.

controlled by the presence of these nongaseous phases. Similarly, the fate of environmental chemicals in the marine and freshwater environment is strongly influenced by the mineral and organic matter suspended in or bordering onto the aqueous phase, even though the latter dominates by volume. Sometimes the gas phase is present in water bodies in the form of gas bubbles. Soils (i.e., the unsaturated subsurface environment) constitute the most obvious multiphase system, the pore space left by mineral and organic matter being filled to a variable degree by water and air. The saturated subsurface environment, finally, is characterized by the absence of a gas phase.

The organic phase in all spheres is composed of living and dead organic matter, the latter derived from plants, microorganisms, and animals at various stages of decomposition and geological transformation. The living organic matter is sometimes referred to as the biosphere.

Each of the phases provide a distinct "setting" for an environmental chemical; that is, it is either surrounded by water molecules or by a variety of organic functional groups, is in contact with a liquid or solid surface, or is only intermittently coming in contact with other molecules. This "setting" determines the chemical potential in a phase and thus its tendency to remain or leave that phase and its tendency to react with other chemical species.

The perception of the environment as being a composite of phases is useful, because chemists and chemical engineers have amassed significant knowledge on:

- Molecular interactions of various chemicals with aqueous, organic, and mineral phases
- Equilibrium distribution of chemical between such phases
- Transport behavior of chemicals within and between such phases
- Reactions that chemicals undergo in such phases

The challenge in applying this knowledge to the study of the environmental fate of chemicals lies in the immense complexity and variability of the natural environment. The phases that make up the natural environment are quite unlike many systems studied in a laboratory or industrial setting. The composition of natural phases is often heterogeneous, even on an extremely small spatial scale.

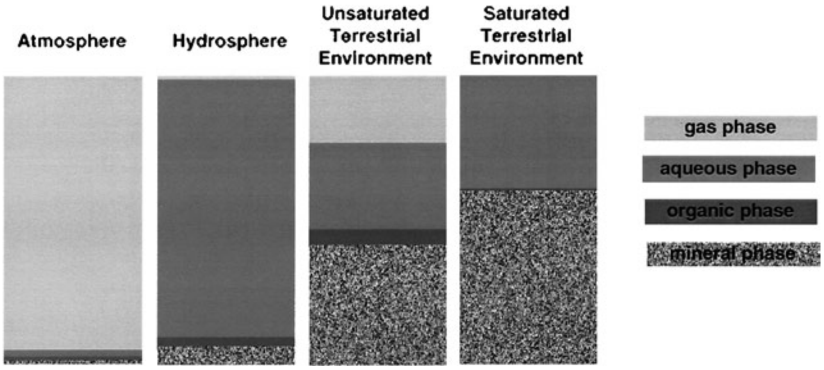


FIGURE 3 The relative phase composition of the major environmental spheres.

The exact chemical composition of virtually any natural environmental phase thus remains elusive. Many natural phases periodically undergo further changes in time (e.g., as a function of season or over longer time periods). Some phases, such as algae blooms or a seasonal snow cover, even have only a temporary existence.

There are, however, simplifying aspects. With the exceptions of the higher atmosphere and the deep sea, the pressure conditions encountered in the environment vary only over a fairly small range. The same is true for the temperature conditions, which range only between approximately −50°C and +50°C. Also, the concentration of the chemical of interest is often small, allowing simplifying assumptions, such as infinite dilution and ideal gas behavior.

The extent of spatial concentration variability in a phase is dependent on the relative time scale of transport and transformation within that phase. The faster the internal transport processes are relative to transformation processes, the more homogeneous the concentration distribution is likely to be. Whereas the time scale of transport is determined by the transport characteristics of the phase (e.g., the dynamics of the atmosphere or the hydrodynamics of a water body), the time scale of transformation tends to be different for each chemical. This implies that one phase may be homogeneous with respect to one chemical but may show significant concentration gradients for another. For example, some very long-lived environmental chemicals such as the chlorofluorocarbons show virtually uniform concentrations in the entire global atmosphere, whereas highly reactive chemicals such as unsaturated hydrocarbons experience immense concentration differences in both space and time. In fact, the degree of spatial variability of concentrations in a phase can be used to estimate removal rates from that phase. For example, the residence time of a chemical in the water column of a lake can be deduced from the observed degree of mixing (i.e., the measured gradients in water concentration).

B. Interphase Transfer and Chemical Cycling

An important aspect of environmental fate are transfer processes, which convey chemical from one phase to another (Fig. 4). Of particular importance are transfer processes between the atmosphere and the Earth's surface and between the water column and underlying sediments. For example, a pesticide molecule may evaporate from an agricultural field into the atmosphere and be deposited to a lake or the ocean further downwind. A heavy metal may be transferred between the water column of a lake and its surface sediment as particles settle and resuspend in response to varying biological activity and water turbulence.

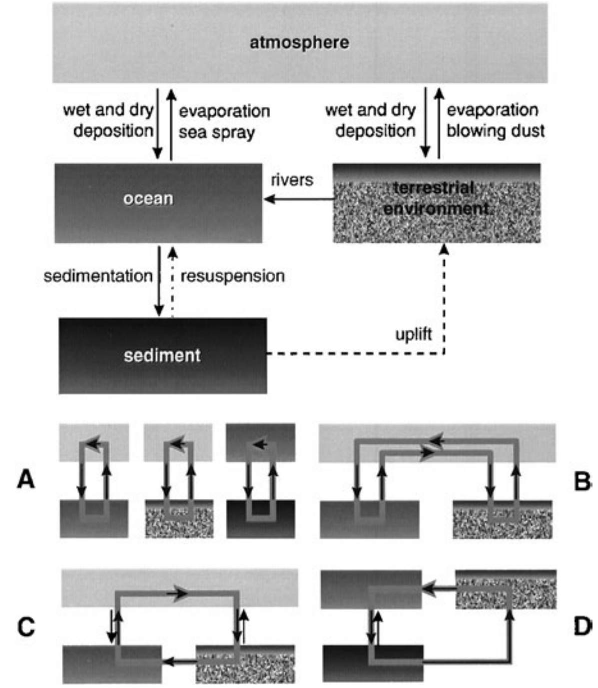


FIGURE 4 Chemical transfer processes between environmental phases can lead to the cycling of chemicals on different time scales.

Cycles of transfer between two and more phases occur if these transfer processes occur repeatedly and reversibly (Figs. 4A and B). Some long-lived environmental chemicals of intermediate volatility, such as mercury and persistent organic pollutants, are believed to undergo numerous cycles between the atmosphere and the Earth's surface. Such cyclical behavior is often linked to cycles in environmental conditions on a seasonal or diurnal time scale. For example, cycles of evaporation and deposition of chemicals across the atmosphere–soil or atmosphere–water interface are often explained by diurnal or seasonal temperature fluctuations. Similarly, seasonal cycles of water–sediment transfer can be caused by the seasonal turnover and stratification characteristics of lakes, the seasonal growth of algae and the resultant settling of organic particle, and the occurrence of seasonal storms strong enough to resuspend surface sediments.

Other cycles involve more than two environmental spheres and span longer time scales. For example, in addition to undergoing atmosphere–surface exchange with the marine and terrestrial system, a long-lived chemical may be transferred from land to ocean via riverine flow, creating a cycle of potential transport from land to ocean to atmosphere and back to the land (Fig. 4C). On a geological time scale, marine sedimentation, continental drift and uplift, and erosion can create yet another type of cycle (Fig. 4D). Such cycles are particularly common and well studied for elements (e.g., sulfur, nitrogen, or lead) and are then referred to biogeochemical cycles. Synthetic chemicals can undergo similar cycles, if they have the appropriate characteristics such as a sufficiently long persistence.

C. Mass Balancing

A key concept in describing environmental chemical fate is the principle of mass preservation. A chemical in a particular location at a specific time can remain at that location, can be transported elsewhere, or can be transformed into another chemical. A mass balance can be formulated for a specific subsection of the environment, called the *system* or *control volume*. Defining a system boundary involves a decision of what is considered a part of the system and what is part of its surroundings. The mass balance then accounts for how much chemical crosses the system boundary and how much chemical is generated and lost within the system during a particular time interval (Fig. 5):

$$\Delta M = \Delta M_{\text{transport}} + \Delta M_{\text{reaction}} = (M_{\text{in}} - M_{\text{out}}) + (M_{\text{production}} - M_{\text{degradation}})$$

M_{in} and M_{out} are the amounts of chemical transported into and out of the control volume during that time interval, and $M_{\text{production}}$ and $M_{\text{degradation}}$ are the amounts produced by sources and eliminated by sinks within the control volume.

More commonly, the mass balance is formulated for an infinitesimally small time interval using rate expressions N in units of amount per time, thus:

$$\frac{\mathrm{d}M}{\mathrm{d}t} = (N_{\text{in}} - N_{\text{out}}) + (N_{\text{production}} - N_{\text{degradation}})$$

The total amount of a chemical in a control volume divided by its total loss rate (by degradation and transport beyond system boundaries) is the average residence time of the chemical in the system. It is a useful measure for the time scale within which a system can respond to changes in chemical input.

The system definition is arbitrary and thus normally based on matters of convenience; it may be selected such that the boundary conditions (the chemical fluxes across the system boundaries) are easily established. A control volume can take any size. If the control volume becomes infinitesimally small, the mass balance equation takes the form:

$$\frac{\partial C}{\partial t} = \left(\frac{\partial C}{\partial t}\right)_{\text{transport}} + \left(\frac{\partial C}{\partial t}\right)_{\text{reaction}}$$

The two above formulations of mass preservation differ only with respect to the size of the control volume. Because in the latter case the system is of an unspecified, infinitesimal size, each term is expressed as amount per unit time per unit volume. The term $\partial C/\partial t$ expresses the change of a chemical's concentration at one specific point in space in an environmental phase, which can be brought about by transport of chemical to or from that point, or by formation or degradation of that chemical at that point. Expressions for describing these transport and transformation processes will be presented below in Sections III and IV.

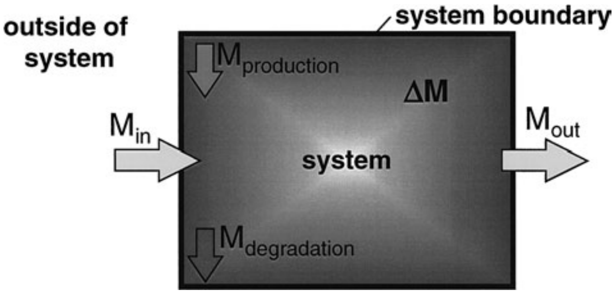


FIGURE 5 A mass balance can be applied to any subsection of the environment. That subsection is called a *control volume*.

D. Thermodynamic vs. Kinetic Factors

As in any laboratory system, chemicals in the natural environment seek to reach a thermodynamically favorable state of equilibrium. Equilibrium distribution coefficients between environmental phases and equilibrium reaction constants between various forms and species of environmental chemicals and elements are employed to quantitatively express this thermodynamically favorable state. Dynamic limitations, however, often prevent a behavior based on chemical equilibrium from being observed in the environment. Chemicals may seek to establish an equilibrium distribution in the environment, yet the time required for establishing that distribution may be much longer than the overall environmental lifetime of the chemical or the phase. For example, plant foliage has shown a very high capacity for sorbing highly nonpolar organic substances, yet a description based on equilibrium partitioning between gas phase and plant leave is often inappropriate, because the delivery of such contaminants to foliage occurs too slowly to establish equilibrium during a single growing season. Similarly, a chemical transformation may be thermodynamically favorable yet proceed so slowly that it is negligible from a quantitative point of view. That is particularly relevant for transformations involving reductions and oxidations. It is thus often necessary to take into account the kinetics of distribution and transformation processes. Yet, even for processes that are kinetically controlled, equilibrium considerations are useful as they indicate the direction of reversible chemical processes as well as the final state the system strives to reach. Also, equilibrium partition characteristics often influence the kinetics of transport. As will be shown in Section III.E, the kinetics of diffusive gas exchange for organic chemicals are strongly influenced by, and thus cannot be expressed quantitatively without, the air–water equilibrium partitioning constant.

II. ENVIRONMENTAL PHASE DISTRIBUTION

The distribution of chemical between environmental phases is at the core of understanding environmental fate. Examples of environmental phase distributions are those between air and water, between atmospheric particles and the gas phase, between plants and air, between soil and air, between suspended matter and the dissolved phase in water, and between groundwater and subsurface solids. On a fundamental level, the distribution behavior of a chemical determines where a chemical is residing in the environment. It further influences the nature and extent of the transport and transformation processes it will experience. The distribution of a chemical between gas phase and particle phase in the atmosphere not only determines by which mechanism and how fast the chemical is being deposited to the Earth's surface, but also further determines the type and rate of reactions that it will experience in the atmosphere. A chemical in a water body experiences very different behavior depending on whether it is dissolved in water or whether it sorbs to colloidal or solid matter suspended in the water column. Similarly, the mobility and reactivity of a contaminant in the subsurface environment depend strongly on its distribution between water and solids.

A. Quantitative Description of Phase Distribution

Many environmental distribution equilibria can be understood in terms of phase partitioning; that is, the environmental chemical is distributed between two bulk phases according to its dissolution properties. Partitioning can occur not only into liquid phases but also into polymeric phases including natural organic matter, if that polymer is sufficiently flexible and porous to allow the environmental chemical to penetrate the inside of the polymeric phase. Partitioning between two different phases, X and Y, is typically expressed using concentration ratios K_{XY}:

$$K_{XY} = \frac{C_X}{C_Y}$$

If the concentrations in the two phases are reflecting equilibrium conditions, the K_{XY} is called an *equilibrium partitioning coefficient*. For many neutral organic chemicals and phases, such coefficients are constant over a wide range of concentrations (i.e., show linear sorption behavior).

If one of the phases is a solid, the distribution is more appropriately described in terms of adsorption. Such adsorption can be a combination of unspecific adsorption of the environmental chemical to the solid surface, electrostatic interactions between the charged chemical and oppositely charged surface sites, and the covalent reaction of the chemical with reactive sites at the surface. Quantitative expressions rely again on concentration ratios, but equilibrium adsorption coefficients are less likely to be constant over wide concentration ranges. These relationships thus need to be characterized by nonlinear adsorption isotherms, such as those formulated by Langmuir and Freundlich.

Environmental phase distributions of elements or inorganic chemicals usually involve different chemical species and therefore speciation reactions. For example, different species of an element such as mercury have different vapor pressure and solubility. Elemental mercury, Hg(0), is fairly volatile and only sparingly soluble in water, whereas oxidized Hg(II) complexes are much less volatile but more water soluble. The distribution of mercury among the phases of air, water, and solid will thus depend on its speciation, which in turn is influenced by variable conditions of the environment, including pH, redox conditions, and the presence of other chemical species. This is approached quantitatively using equilibrium reaction constants for the various speciation reactions and illustrated using distribution diagrams that delineate the major prevalent species as a function of pH or pE, or both.

B. Distributions Involving Real Environmental Phase

By measuring concentration ratios, environmental phase distributions for chemicals can be empirically determined and characterized, both in the field and the laboratory. Examples are sorption coefficients K_D between water and soil or sediment solids or coefficients K_P describing the distribution between the gas phase and atmospheric particulate matter. Going beyond such empirical descriptions requires the derivation of relationships of general validity that allow the quantitative description of phase distributions involving environmental chemicals or environmental phases for which no such empirical data exist.

Some environmental phase distributions closely resemble distributions involving pure or chemically well-defined phases. For example, the equilibrium distribution between the gas and aqueous phase is a well-defined characteristic of a chemical, described by Henry's law constant or the air–water partition coefficient K_{AW}. Further, the influence of properties of the water phase such as the electrolyte concentration, pH, and temperature on the air–water equilibrium distribution is well understood for many chemicals. This allows the quantitative prediction of how chemicals distribute between the gas phase and many environmentally relevant aqueous phases such as rain droplets and lake and ocean water. Similarly, the adsorption of many chemicals on chemically well-defined mineral matter, such as quartz sand and various clay minerals, has been determined, as has the influence of selected environmental conditions such as temperature and relative humidity on these sorption equilibria.

Most environmentally relevant phase equilibria, however, refer to less well-defined or heterogeneous phases, such as plant foliage, aerosols, or soil particles. It may often be possible to distinguish between the various components contributing to the distribution of chemicals into such phases, such as the mineral and organic matter. Nevertheless, the immense heterogeneity of environmental phases in space and time makes an exact treatment of chemical distribution processes in the environment very difficult. A particularly complex, yet important issue is the description of chemical distribution into natural organic matter, which often has no clearly defined chemical identity. Two complementary approaches exist for tackling the issue of chemical distribution into heterogeneous environmental phases. The first seeks to simplify in order to obtain uncertain, yet practical, descriptions of phase distribution, whereas the second seeks to investigate the heterogeneous composition of the natural phases in order to gain a fundamental mechanistic understanding of its influence on the phase distribution of environmental chemicals.

Many environmental phase equilibria—in particular, those involving a phase transition to the gas phase—are strongly influenced by temperature. Higher temperatures favor partitioning into the gas phase, whereas sorption to solid surfaces and dissolution in aqueous and organic phases are enhanced at lower temperatures. Environmental temperature differences on both a spatial and temporal scale can thus result in variations in environmental partitioning. Gas–particle equilibria are shifted toward the particle phase under cold atmospheric conditions. Diurnal and seasonal cycles in the atmospheric deposition and evaporation of environmental chemicals of intermediate volatility have been observed and explained by the temperature influence on partitioning.

C. Correlation with Partitioning Equilibria Involving Pure Phases

Many simplifying approaches rely on the correlation of environmentally relevant phase equilibria with well-defined partitioning equilibria between pure phases. The pure phases most important to environmental fate considerations are the pure liquid substance, the ideal gas phase, the pure aqueous phase, and a pure organic solvent phase. Figure 6 depicts the relevant phase equilibria. The organic solvent that has been most widely, though not exclusively, used in the context of environmental fate is 1-octanol. It has been applied as a surrogate for all types of organic phases, including humic substances in soils and sediments, organic matter in aerosols, lipid tissues in animals, and cuticular material in plant foliage.

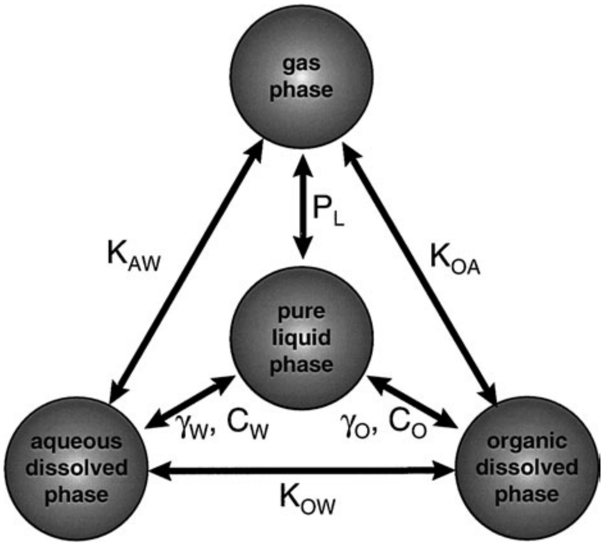


FIGURE 6 Phase distribution equilibria involving pure phases with importance for describing environmental phase partitioning. K_{AW}, air–water partition coefficient; K_{OW}, octanol–water partition coefficient; K_{OA}, octanol–air partition coefficient; P_L, (supercooled) liquid vapor pressure; C_W and C_O, saturation solubility in water and octanol, respectively; γ_W and γ_O, activity coefficient in water and octanol, respectively.

The correlations usually take the form of linear free-energy relationships (LFERs) of the following type:

$$\log K_{XY} = m \cdot \log K_{AB} + b$$

where K_{XY} is the partition coefficient for an environmentally relevant phase pair, X and Y; K_{AB} is a partition coefficient involving two pure phases, A and B; and m and b are empirical regression coefficients. Examples of such regressions are given in Table I. It should be noted that K_{AB} includes physical–chemical properties such as vapor pressure and aqueous solubility, as one of the two phases A and B may be the pure liquid (or solid) phase. The relevant physical–chemical properties can be determined in the laboratory and are known and tabulated for many environmentally relevant chemicals. Even if they are not known, a variety of empirical and theoretical methods, called quantitative structure-property relationships (QSPRs), exist that predict such parameters from chemical structure.

TABLE I One-Parameter LFERs Used To Describe Important Environmental Phase Distribution Equilibria[a]

Gas/particle partitioning:
$\log K_P = m \cdot \log P_L + b$
$\log K_P = m \cdot \log K_{OA} + b$
Organic carbon/water or organic matter/water partitioning:
$\log K_{OC} = m \cdot \log K_{OW} + b$
$\log K_{OC} = m \cdot \log C_W + b$
Lipid/water partitioning:
$\log K_{LW} = m \cdot \log K_{OW} + b$
$\log K_{LW} = m \cdot \log C_W + b$
Foliage/air partitioning:
$\log K_{FA} = m \cdot \log K_{OA} + b$
Soil/air partitioning:
$\log K_{SA} = m \cdot \log K_{OA} + b$

[a] For definition of symbols, see Fig. 6.

K_{XY} is often normalized to the fraction of the environmental phase that is believed to contribute predominantly to chemical uptake. Sorption coefficients of nonpolar organic environmental chemicals between water and soil and sediment solids K_D are thus often normalized by the organic matter or organic carbon content, yielding normalized organic matter–water K_{OM} or organic carbon–water partition coefficients K_{OC}. The distribution coefficients between the gas phase and atmospheric particles are sometimes normalized by the organic matter content. Similarly, partition coefficients between water and aquatic organisms are often normalized by the lipid content, as are those between the gas phase and foliage.

The primary limitations of these approaches are that:

- They ignore much of the variability of environmental phases. For example, not all variability in K_D among various soils and sediments is explained by the organic matter content, nor is all the variability in plant–air partitioning explained by the lipid content of the plant tissue.
- Pure phases such as 1-octanol only approximate the partitioning properties of natural phases. The regression coefficients m and b are thus generally valid only within a group of similar chemicals. For example, the regressions between the aerosol–air partition coefficient K_P and vapor pressure are different for polycyclic aromatic hydrocarbons and chlorinated hydrocarbons, even though both are relatively nonpolar substance groups. Empirical regressions for polar chemicals are largely missing.
- They assume that the phase distribution is a reversible partitioning process that is linear (i.e., does not show a dependence on chemical concentration).

The focus of these approaches is on practicality—unexplained variability in partitioning properties within a certain range is accepted. Activity coefficients γ_O of many organic chemicals in organic matter and organic solvents are usually not very large (<10). The mistake made by approximating a natural organic phase by an organic solvent such as 1-octanol is therefore often less than an order of magnitude. It turns out that the environmental fate of chemicals can sometimes be described reasonably well, if much of the complexity of partitioning into heterogeneous natural material is ignored.

D. Mechanistic Understanding of Environmental Phase Equilibria

Other approaches aim for a more mechanistic understanding of environmental phase partitioning. This often involves identification of the detailed characteristics, structure, and chemical composition of environmental phases. Much effort is, for example, devoted to determining the exact chemical composition of the organic matter present in aerosols in order to understand the gas–particle partitioning of organic compounds. Similar efforts are directed towards the characterization of the natural organic matter present in soils and sediments. Largely composed of heterogeneous macromolecules and polymeric humic substances, such phases defy exact chemical identification. Nevertheless, the relative abundance of various functional entities within such organic phases or the glass/rubber transition behavior of environmental polymers can be determined and correlated with their distribution properties. Finally, the heterogeneity of the elemental composition and other chemical characteristics of surfaces and phase aggregates are being mapped and investigated at ever smaller scales.

Other routes of investigation seek to gain mechanistic understanding of phase distribution from carefully designed laboratory experiments. Examples are the detailed mechanistic interpretation of adsorption isotherms and sorption/desorption kinetics. Yet other approaches seek to correlate environmentally relevant phase equilibria, experimentally determined for a large and varied set of solutes, directly with a series of descriptors indicative of intermolecular interactions. There are a variety of such approaches, but they have in common the use of multiple linear regression analysis on a number of parameters accounting for van der Waals (dispersive and dipole interactions) and Lewis acid-base interactions (hydrogen bonds) between solute and solvent. The linear solvation-energy relationship (LSER) approach, for example, uses five experimental molecular descriptors to describe the various solutes. Other approaches seek to assign similar descriptors for the phases.

E. Uptake of Environmental Chemicals in Organisms

The distribution process of an environmental chemical between an organism and its surrounding phase is of particular interest, because it often controls the detrimental effect that the chemical may have on that organism or those feeding on it. The two most important distribution processes of this type are those between aquatic organisms and the surrounding water phase and between plants and the surrounding atmosphere. This distribution process is referred to as *bioconcentration* if the organism takes up chemicals directly from its surrounding phase, whereas it is called *bioaccumulation* if the uptake additionally occurs through food. These processes are of most concern for nonpolar organic substances, as these have a particular affinity for biological lipid phases. In fact, bioconcentration of such substances in aquatic organisms is often regarded as an equilibrium partitioning process between the water and the lipid phase of the organisms (Fig. 7A). The corresponding lipid–water partition coefficient K_{LW}, which is also referred to as the bioconcentration factor (BCF), can then be estimated from empirical linear free-energy relationships with the K_{OW} or the water solubility of a substance (Table I). Such relationships are not valid for compounds that are easily metabolized by the organism.

Bioaccumulation (chemical uptake with the food) becomes important in terrestrial organisms and higher aquatic organisms such as fish and mammals. It too can be interpreted as a distribution process between the food in the gastrointestinal tract and the organism's body tissues. Compounds that are both highly nonpolar and resistant to metabolic transformation may be subject to biomagnification; that is, an organism can accumulate concentrations that are higher than those in its food. This can lead to an accumulative effect within a food chain, with higher concentrations being achieved in organisms feeding at higher trophic levels. The quantification of these processes requires kinetic approaches that describe the various uptake and depuration processes of a chemical in an organism. This is sometimes done in models that formulate a mass balance around an organism, which thus constitutes a control volume (Fig. 7B). More sophisticated, physiologically based models seek a mechanistic understanding of the chemical transport processes across the gill membrane and in the gastrointestinal tract, as well as between the various body tissues (Fig. 7C).

III. ENVIRONMENTAL TRANSPORT PROCESSES

The two primary modes of chemical transport are advection and diffusion. Advective transport occurs when the phase in which the chemical resides is moving and the chemical thus undertakes the same movement. Diffusive transport of chemical occurs as a result of random or seemingly random mixing processes.

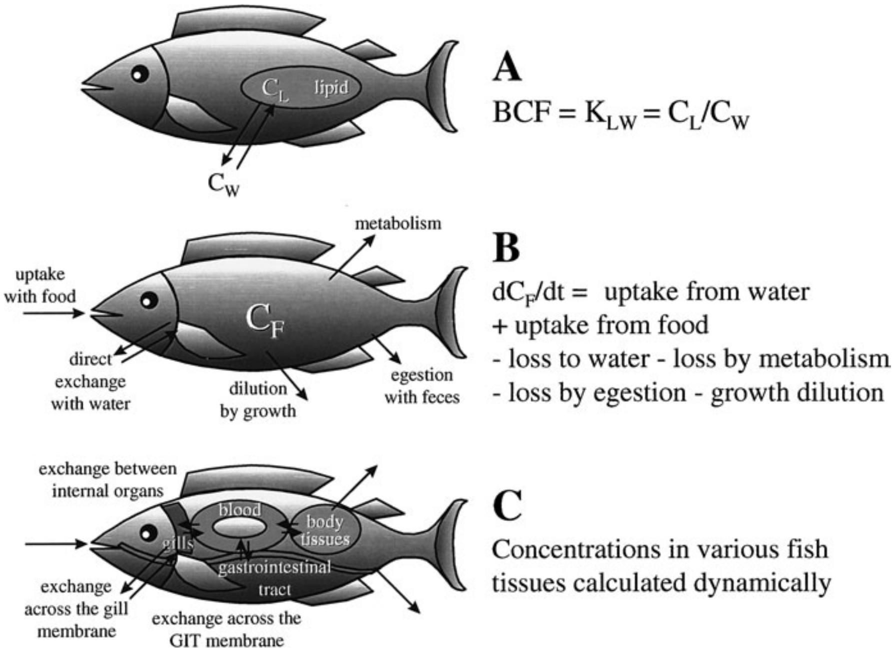


FIGURE 7 Various approaches to quantifying the uptake of a chemical in an aquatic organism.

A. Advective Transport

The movement of air, water, and solid material drives advective transport of environmental chemicals. The uneven distribution of solar energy across the globe is the driving force for the large-scale advection of air and ocean water masses. By supplying the energy for the evaporation of water, the sun is also driving the hydrological cycle. Advective transport of environmental chemicals with moving air and water masses occurs on a wide range of scales. A plume drifting from a smokestack is an obvious illustration of local atmospheric advection, whereas haze episodes in the Arctic are testament to the potential for long-range atmospheric transport of environmental chemicals. Similarly, oceanic currents have advected radionuclides released from nuclear reprocessing plants into the Irish Sea all the way to the Arctic Ocean basin. The hydrological cycle is another major agent in advecting environmental chemicals. Falling precipitation, surface and subsurface runoff, and rivers and streams can all carry environmental chemicals, either in dissolved form or attached to solids and colloids.

Less obvious advective transport processes are those that involve the movement of phases within a phase. Particles in the atmosphere are advected together with moving air masses, but gravitational forces may also move them relative to the surrounding gas-phase molecules. Environmental chemicals associated with these particles are thus being advected with the particles. The same applies to water percolating downward or rising by capillary forces through soil or the gravitational settling of particulate matter from the surface ocean to the deep sea across the halocline. The latter process is an important vector for particulate organic carbon to the deep sea and thus also for environmental chemicals associated with that organic matter.

The mathematical description of advective transport processes involves the multiplication of a transport flux of the advected medium, G (e.g., in mass or volume per time unit) with the concentration of the environmental chemical in that medium, C (in units of amount per mass or volume) to yield an advective flux (in units of amount per time unit):

$$N_{\text{advection}} = G \cdot C$$

For an infinitesimally small control volume, advective transport is the product of the phase velocity with the concentration gradient in space:

$$\left(\frac{\partial C}{\partial t}\right)_{\text{advection}} = -\left(v_x \frac{\partial C}{\partial x} + v_y \frac{\partial C}{\partial y} + v_z \frac{\partial C}{\partial z}\right)$$

where x, y, and z are the Cartesian coordinates in space, and v_x, v_y, and v_z are the velocity components in each of these three directions.

B. Molecular Diffusion

The random movement of the molecules in a liquid or gas leads to the movement of chemicals from regions of high concentration to regions of low concentration, even if the fluid itself is stagnant. This process is called *molecular diffusion*, and the mathematical description of a diffusive flux (in units of amount per time unit) is based on Fick's first law:

$$N = -D \cdot A \cdot \frac{\mathrm{d}C}{\mathrm{d}x}$$

where D is the molecular diffusion coefficient (in units of length squared per time), A is the area across which diffusion occurs, and $\mathrm{d}C/\mathrm{d}x$ is the concentration gradient in space. D is a function of both the diffusing molecule and the phase. It decreases with molecular size and is much larger in the gas phase than in liquid phases. The negative sign is required to convert a negative concentration gradient into a positive flux. In an infinitesimally small control volume, we obtain Fick's second law:

$$\left(\frac{\partial C}{\partial t}\right)_{\text{diffusion}} = \frac{\partial}{\partial x}\left(D_x \cdot \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y \cdot \frac{\partial C}{\partial y}\right) + \frac{\partial}{\partial z}\left(D_z \cdot \frac{\partial C}{\partial z}\right)$$

where D_x, D_y, and D_z are the molecular diffusion coefficients in the three spatial dimensions. This simplifies to:

$$\left(\frac{\partial C}{\partial t}\right)_{\text{diffusion}} = D \cdot \left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2}\right)$$

if the molecular diffusivity D does not vary with direction (i.e., is isotropic) or in space, which is usually the case. Except over very short distances, molecular diffusion is a very slow process. It takes a molecule half a day to diffuse a distance of one meter in air, and more than 10 years to cover the same distance in the aqueous phase. Despite this slowness, molecular diffusion becomes important in microenvironments, where phase movement is very limited. This is the case immediately adjacent to interfaces or in very fine pores in soil and sediment particles.

C. Turbulent or Eddy Diffusion

Due to the slowness of molecular diffusion, it is usually the movement of the phase itself that contributes to the mixing of chemicals in environmental fluids. Random or seemingly random phase movement is referred to as turbulence and leads to concentration homogenization analogous to the result of molecular diffusion. This mixing process is called *eddy* or *turbulent diffusion*, an eddy being the element of the fluid that is moving randomly. Eddies have widely variable sizes, and large eddies usually have a fine structure of smaller eddies. Fluid movement in the atmosphere and in rivers, lakes, and oceans is always turbulent. The source of turbulence in environmental fluids is the shear forces at the phase boundaries. For example, turbulence in a lake is caused by wind shear at the air–water interface and by the shear resulting from the flow of water along the water–sediment interface.

It turns out that turbulent diffusion can be described with Fick's laws of diffusion that were introduced in the previous section, except that the molecular diffusion coefficient is to be replaced by an eddy or turbulent diffusivity E. In contrast to molecular diffusivities, eddy diffusivities are dependent only on the phase motion and are thus identical for the transport of different chemicals and even for the transport of heat. What part of the movement of a turbulent fluid is considered to contribute to mean advective motion and what is random fluctuation (and therefore interpreted as turbulent diffusion) depends on the spatial and temporal scale of the system under investigation. This implies that eddy diffusion coefficients are scale dependent, increasing with system size. Eddy diffusivities in the ocean and atmosphere are typically anisotropic, having much large values in the horizontal than in the vertical dimension. One reason is that the horizontal extension of these spheres is much larger than their vertical extension, which is limited to approximately 10 km. The density stratification of large water bodies further limits turbulence in the vertical dimension, as does a temperature inversion in the atmosphere. Eddy diffusivities in water bodies and the atmosphere can be empirically determined with the help of tracer compounds. These are naturally occurring or deliberately released compounds with well-established sources and sinks. Their concentrations are easily measured so that their dispersion can be observed readily.

D. The Continuity Equation

When we insert the expressions for advective and diffusive transport into the mass balance equation from Section I.C, we yield the following equation:

$$\frac{\partial C}{\partial t} = \underbrace{-\left(v_x \cdot \frac{\partial C}{\partial x} + v_y \cdot \frac{\partial C}{\partial y} + v_z \cdot \frac{\partial C}{\partial z}\right)}_{\text{advection}}$$
$$+ \underbrace{\frac{\partial}{\partial x}\left(E_x \cdot \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(E_y \cdot \frac{\partial C}{\partial y}\right) + \frac{\partial}{\partial z}\left(E_z \cdot \frac{\partial C}{\partial z}\right)}_{\text{diffusion}}$$
$$+ \underbrace{J - R}_{\text{reaction}}$$

which is referred to as the advection-diffusion-reaction equation or continuity equation. J and R are production and degradation terms, respectively. $\partial C/\partial t$ describes the change of concentration at one point in space, which can be brought about (1) if a part of the medium with a different concentration is carried to that point by advective motion, (2) if there is a spatial concentration gradient in the medium leading to turbulent diffusive transport of chemical toward or away from that point, or (3) if chemical is formed or degraded at that point. If we retain only one dimension and assume that the eddy diffusion coefficient does not vary with space and that degradation follows a first order kinetics this simplifies to:

$$\frac{\partial C}{\partial t} = -v \cdot \frac{\partial C}{\partial x} + E \cdot \frac{\partial^2 C}{\partial x^2} + J - C \cdot k$$

This equation is used to describe chemical fate in the atmosphere, in water bodies, and in the subsurface environment, whenever spatial variability is large and the use of box models that assume homogeneous conditions is inappropriate. It is the relative magnitude of the parameters v, E, and k that determines the character of a chemical's fate in a phase. One way to illustrate this is the calculation of the term $E \cdot \mathrm{k/v}^2$ (sometimes called the Damköhler number). If that term is much larger than 1, diffusive processes are fast relative to advective processes within the life-time of the chemical indicated by 1/k. A value much smaller than 1, on the other hand, indicates that advective transport is faster than diffusive transport within that time scale.

E. Transport Processes Across Phase Boundaries

Whereas Section I.B highlighted the importance of understanding chemical transfer between environmental phases, the discussion in this section has so far been limited to transport processes within a phase. Similar to transfer within phases, transfer of chemicals between phases can take place by diffusive and advective processes. Advective transfer between phases occurs if a subphase itself is transferred from one phase to another. Wet atmospheric deposition processes are of this type, as the falling hydrometeors are transferring both dissolved and particle-bound chemical species to the Earth's surface when reaching a terrestrial or aqueous surface. The gravitational settling of particles in water to the underlying sediment is another example.

Diffusive transport of chemicals also occurs across phase boundaries. However, it is no longer a concentration gradient that serves to describe this process, but a difference in chemical potential. Diffusive exchange across the air–water interface may serve as an example to illustrate how rates of diffusive interfacial chemical transfer are being derived. A common conceptual approach

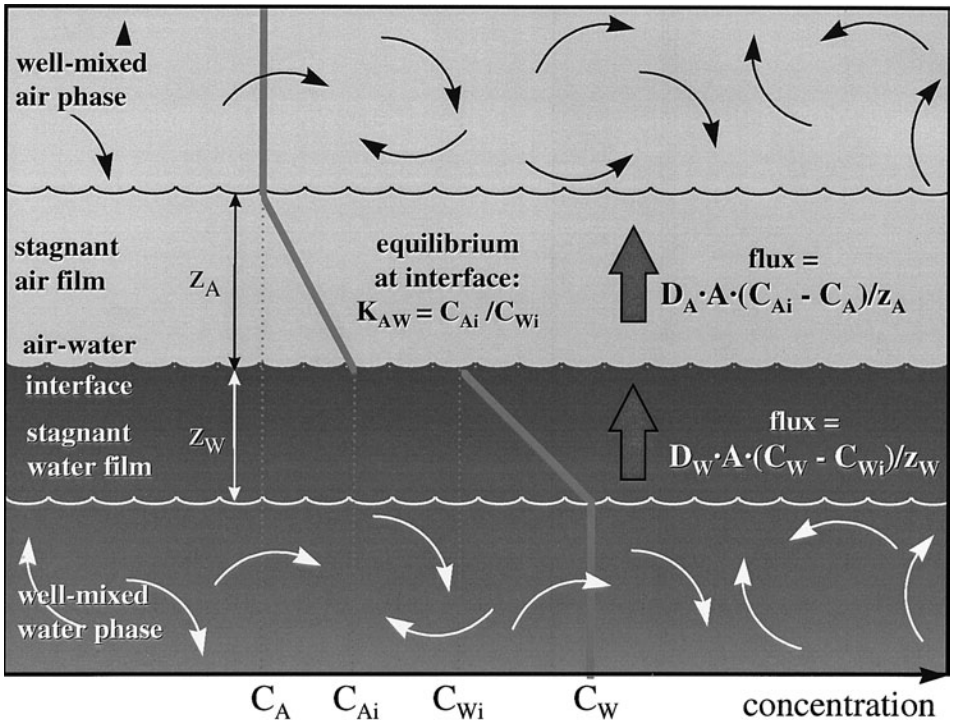


FIGURE 8 Two film model of the diffusive exchange of gases across the air-water interface.

to interfacial diffusion is the stagnant two-film model (Fig. 8). It assumes that turbulent diffusion results in fairly homogeneous concentrations C_A and C_W in the two bulk phases, whereas concentration gradients are restricted to the two regions adjacent to the interface, where two stagnant layers or films of thickness z_A and z_W develop. Molecular diffusion across these two films in series is assumed to constitute the rate-limiting step in air–water gas exchange. The layer of air molecules directly at the interface (concentration C_{Ai}) is assumed to be in equilibrium with the adjacent layer of water molecules (concentration C_{Wi}), implying that the ratio of the interfacial concentrations is equivalent to the Henry's law constant K_{AW} of the substance. At steady state, the rate of diffusion across both films is identical and we can apply Fick's law to derive the interfacial flux N (in units of amount per time) across area A of:

$$N = D_W \cdot A \cdot \frac{(C_W - C_{Wi})}{z_W} = D_A \cdot A \cdot \frac{(C_{Ai} - C_A)}{z_A}$$

Using $K_{AW} = C_{Ai}/C_{Wi}$, the interfacial concentrations can be eliminated:

$$N = \frac{1}{\dfrac{z_W}{D_W} + \dfrac{z_A}{D_A \cdot K_{AW}}} \cdot \left(C_W - \frac{C_A}{K_{AW}}\right)$$

$$= \frac{1}{\dfrac{z_W K_{AW}}{D_W} + \dfrac{z_A}{D_A}} \cdot (C_W K_{AW} - C_A)$$

The first term in this equation is a mass transfer coefficient, which can be interpreted as a composite of partial transfer velocities in the two stagnant films, D_W/z_W and D_A/z_A. As the thickness of the stagnant films decreases with increasing wind speed, these mass transfer coefficients increase in high winds, speeding up diffusive gas exchange. The equation reveals that the air–water partition coefficient K_{AW} of an environmental chemical controls whether diffusion through the stagnant air film or through the stagnant water film is the rate-controlling step. For chemicals with large K_{AW}, $D_A \cdot K_{AW}/z_A$ is much larger than D_W/z_W, and transfer through the water boundary layer is the rate-controlling step. The resistance in the air phase, on the other hand, limits air–water diffusion of chemicals with a small K_{AW}.

On a large scale, the bulk phases air and water can no longer be assumed homogeneous and additional resistances to air–water transfer need to be considered, such as the boundary layer in the atmosphere or temperature and salinity stratification in lakes and the surface ocean. Similar two-layer approaches have been developed to describe the diffusive transfer across other phase boundaries, such as the air–foliage, air–soil, and water–sediment interface. A complicating factor in multiphase systems such as soils and sediments is that the transport to the phase boundary can take place in more than one subphase. The exchange of an environmental chemical between two phases is often

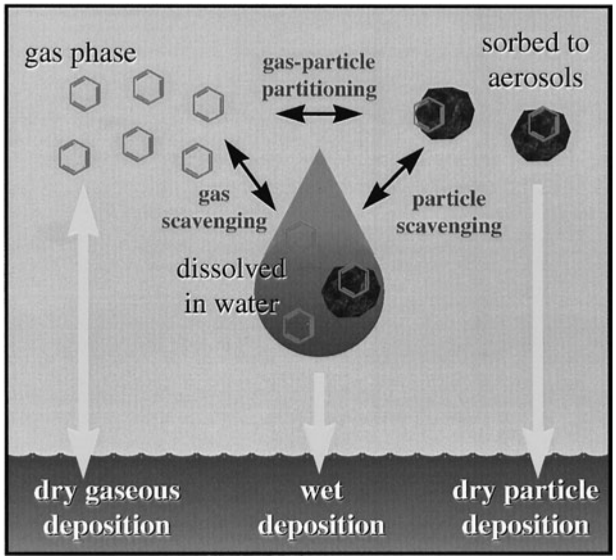


FIGURE 9 The various atmospheric deposition pathways of environmental chemicals.

the sum of various diffusive and advective transport processes. For example, the transfer from the atmosphere to the water surface can occur by diffusive gas exchange, precipitation scavenging of dissolved and particle-bound species, and direct dry deposition of the chemical bound to particles (Fig. 9).

IV. ENVIRONMENTAL TRANSFORMATION PROCESSES

Transformation processes or reactions change the structure and thus the identity of a chemical. Chemical bonds are being broken and new ones are being made, resulting in the formation of one or more products. Such processes are important for the environmental fate of chemicals for primarily two reasons:

1. The rate of chemical transformation determines the rate of disappearance of a chemical from the environment and, in turn, its spatial distribution. Only very persistent chemicals can be distributed widely. The gasoline additive MTBE is widely dispersed in contaminated aquifers because it is fairly resistant to degradation in groundwater. The pesticide DDT can accumulate in remote regions, in organisms, and in the food chain because it is extremely persistent. And chlorofluorocarbons could only reach the stratosphere because of an extraordinary resistance to undergo reactions in the troposphere.
2. The products of transformation processes may have environmental consequences; that is, the transformation may initiate events leading to negative effects. For example, upon oxidation, reactive atmospheric hydrocarbons have the potential to contribute to the formation of tropospheric ozone and atmospheric particles, both major urban air pollution issues. Similarly, stratospheric ozone depletion is a result of reactions that the fluorochlorocarbons are undergoing once they reach the polar stratosphere.

Whereas the former requires mostly a kinetic or quantitative understanding of transformation processes, the latter asks for a mechanistic process understanding.

It transpires that we are sometimes encountering conflicting requirements: On the one hand, humans strive to produce less persistent chemicals to prevent their widespread dispersal and accumulation. For example, pesticides are deliberately designed not to persist in the environment to assure that hazardous residues remain neither on the agricultural produce nor in the environment. On the other hand, they strive to emit less reactive compounds to prevent the negative effects resulting from the transformations. For example, fuels for combustion engines are being designed to lead to emissions of chemicals with low ozone-forming potential (i.e., less reactive and thus more persistent chemicals).

The two primary forces leading to chemical transformations in the environment are sunlight and mircoorganisms, although some chemicals react also in the absence of light and organisms. Photochemical transformations can occur in the gas phase, in the aqueous phase and when chemicals are sorbed to sunlit surfaces. Biotic transformations are primarily mediated by microorganisms in natural waters, soils, and sediments but occur also in higher plants and animals.

A. Chemical Transformation Processes

Transformation processes that occur in the absence of light and organisms are often simply referred to as chemical reactions. In the environment, such reactions are normally occurring in aqueous solution. The occurrence and rate of chemical reactions are often influenced by other properties of these solutions. Potentially important variables include temperature, pH, redox conditions, ionic strength, and the concentration of other solutes and solids.

Some organic environmental chemicals can undergo nonreductive chemical reactions. Chemical species in the environment that may chemically react with organic compounds are mostly inorganic nucleophiles such as water, hydroxy ions, halogen ions, sulfate, and acetate. In contrast, reactive electrophiles tend to be short-lived in the environment outside organisms and in the absence of light. Because of its ubiquity and abundance, water is the most important environmental nucleophile. Nucleophilic substitutions with water (i.e., hydrolyses) are therefore among the most common chemical transformation processes. Hydrolyses tend to yield more polar chemicals than the educts. Examples of such reactions are the substitution of halogen at saturated carbon atoms or hydrolytic reactions of acid derivatives. The latter includes carboxylic acid esters and amides, carbamates, and phosphoric esters.

Only few substances and functional groups are oxidized or reduced abiotically, so that chemical oxidations in the environment are only important for easily oxidizable compounds. As it is often not known which species act as electron donor and electron acceptor in an environmental redox reaction, reaction pathway and kinetics can often not be generalized. Light, microorganisms, or wildfires promote most oxidations in the environment.

B. Photochemical Transformation Processes

The light from the sun is a very powerful agent for transforming chemicals in the environment. The energy of the ultraviolet radiation emitted by the sun is in the same order of magnitude as the binding energies of many covalent bonds. However, much of that high-energy radiation has been absorbed in the upper atmosphere before reaching the earth's surface. In addition to acting upon gas phase chemicals in the atmosphere, chemicals dissolved in hydrometeors and in the surface layer of water bodies and those absorbed on the surfaces of aerosols, plant leaves, and soil particles are exposed to sunlight. Photochemical reactions in the environment thus can be gas phase, aqueous phase, or heterogeneous reactions. For an environmental chemical to be transformed as a result of exposure to sunlight, a transfer of the light energy to the chemical is required. This can occur by three distinct pathways (Fig. 10).

During *direct photolysis*, the environmental chemical itself absorbs the radiation and is promoted to an excited electronic state, which may cause it to undergo a reaction. An organic chemical needs to have chromophores that absorb light within the wavelength region of the sun, in order to experience direct photolysis. Organic substances that absorb at wavelengths greater than 290 nm include aldehydes, nitroaromatic compounds, and chemicals with highly conjugated π-electron systems, in particular, polyaromatic substances. Examples of important atmospheric photolysis reactions are the dissociations of ozone, nitrogen oxide, and aldehydes:

$$O_3 \xrightarrow{h\nu} O_2 + {}^1O \xrightarrow{H_2O} 2OH^{\bullet} + O_2$$

$$NO^{\bullet}_2 \xrightarrow{h\nu} NO^{\bullet} + {}^3O \xrightarrow{O_2} O_3 + NO^{\bullet}$$

$$RCHO \xrightarrow{h\nu} R^{\bullet} + CHO^{\bullet} \xrightarrow{2O_2} ROO^{\bullet} + HOO^{\bullet} + CO$$

As can be seen from these examples, the photodissociation products react with molecules abundant in the environment such as molecular oxygen and water to form reactive species such as the hydroxyl radical $HO^{\bullet}$, the hydroperoxy radical $HOO^{\bullet}$, and the alkylperoxy radicals $ROO^{\bullet}$. The likelihood that after light absorption a photolytic transformation actually occurs is expressed in the quantum yield. Direct photolysis quantum yields for various chemicals can be measured and are a function of the state of the chemical (gas phase, aqueous phase, sorbed) and to a certain degree also the wavelength. Given the quantum yield, the actual direct photolysis rate of a compound in a particular sunlit phase can be estimated from the light intensity, which is both a factor of meteorological conditions and the attenuation of light during passage of the phase.

During *sensitized photolysis*, another chemical, called a *photosensitizer*, absorbs the radiation and transfers the

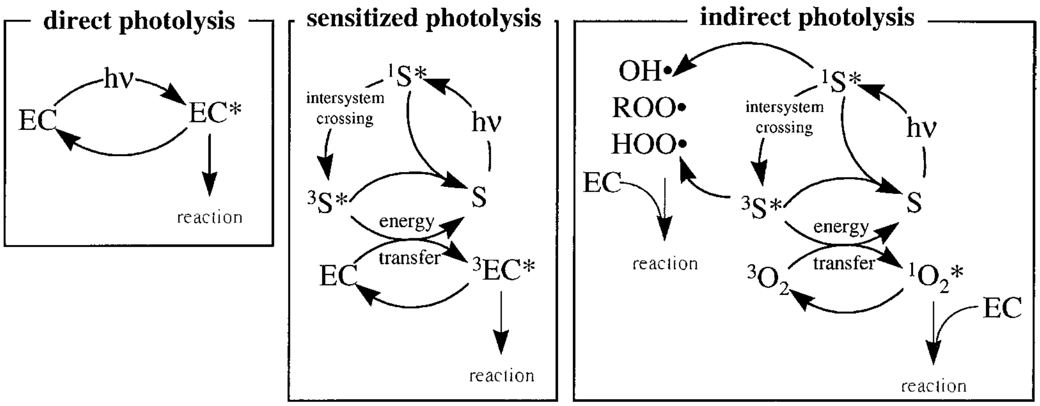


FIGURE 10 The three photochemical reaction pathways of an environmental chemical, *EC*. *S* is a sensitizer.

energy to the environmental chemical, which then reacts. This is of particular relevance in aqueous systems with high concentrations of light-absorbing dissolved organic matter (i.e., in yellowish and brownish waters). It is the aromatic and quinoid structures in these organic materials that can act as photosensitizers. The importance of sensitized photolysis in the environment is limited because of the abundance of molecular oxygen, which competes with the environmental chemicals in the reaction with the excited chromophor and is much easier excited to a higher state than most environmental organic chemicals.

The third and by far the most important pathway in both atmosphere and sunlit water bodies is the formation of reactive species (*photoreactants*) by sunlight, which then react with the environmental chemical. The examples given above show the type of reactive species and the mechanisms of formation. The reactive species may be a radical formed as a product of a reaction of the light-absorbing species or chromophor, before or after intersystem crossing to the triplet state, or it may be reactive singlet oxygen formed in a photosensitization. The most important reactive species in the atmosphere is the hydroxyl radical, which is mostly originating from the photodissociation of ozone, shown above. The hydroxyl radical is also an important reactive species in water, where it is formed from a variety of mechanisms, including the photolysis of nitrate (NO_3^-), nitrite (NO_2^-), and hydrogen peroxide (H_2O_2).

The hydroxyl radical is responsible for the oxidation of most reduced or partially oxidized compounds emitted into the atmosphere from both natural and anthropogenic sources. This includes nitrogen dioxide (forming nitric acid), the hydrides (CH_4, H_2S, NH_3), hydrocarbons other than methane, and even semivolatile species such as the polychlorinated biphenyls. The reaction can take the form of $HO^\bullet$ addition to double bonds or the abstraction of an H by the $HO^\bullet$, which in both cases leads to the formation of radicals that undergo further transformations. Only organic substances without double bonds or an H to abstract (e.g., the chlorofluorocarbons) do not react with $HO^\bullet$ and persist in the troposphere. Such substances eventually will rise to the stratosphere. Also, in the aqueous phase $HO^\bullet$ is an important photoreactant, acting as an electrophilic oxidant for many organic compounds. Heterogeneous photolytic processes are still poorly understood. Light-absorption characteristics, quantum yields, and reactivity with photoreactants of sorbed species may differ from those in the gas and dissolved phase.

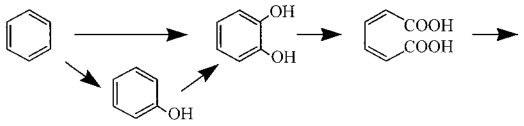
C. Biological Transformation Processes

Together with sunlight, microorganisms are the most potent agents in the transformation of environmental organic chemicals. Plants and animals also have the capability to transform many environmental chemicals, but the rates of transformation tend to be negligible for the overall environmental fate of a substance, unless the focus is on the particular organism. Although microorganisms, which includes bacteria, protozoans, fungi, and microalgae, are ubiquitous in the environment, the occurrence and kinetics of particular biologically mediated transformation processes are dependent on the specific composition of a microbial community at a particular point in space and time. That composition in turn is strongly affected by environmental conditions such as temperature, pH, ionic strength, and oxygen concentration.

Most microbially mediated reactions are redox reactions. The redox conditions in environmental phases are themselves influenced by microbially mediated processes. This becomes particularly obvious in waterlogged environmental phases containing degradable organic material, such as deeper sediment layers. As oxygen cannot be easily replenished in such systems, an ecological redox sequence develops, either in space or in time. Initially, in the presence of oxygen, oxidation of organic compounds occurs by aerobic respiration. As the oxygen becomes depleted, anaerobic microorganisms are using nitrate ions as the oxidant in a process called *denitrification*. When all the nitrate has been depleted, bacterial reduction of manganese and iron hydroxides takes place. Once these oxidants have been consumed and the redox potential drops even further, sulfate respiration and eventually fermentation occurs in the form of methanogenesis. The energy released in the respiration of organic matter decreases as oxygen is depleted and the microbial community shifts to the use of other, less potent oxidants. The oxidative biodegradation of a compound therefore tends to decrease along the ecological redox sequence. Some organic compounds, however, are more easily transformed under reducing conditions. Examples are the reductive dehalogenation of chlorinated alkenes and aromatic compounds.

Microbial transformations of environmental chemicals are enzymatically mediated. The enzymatic systems of organisms have evolved to catalyze transformations of naturally occurring substances. Environmental chemicals that occur naturally are therefore readily transformed. Chemicals that have a structure that closely resembles that of naturally organic substances may also be transformed and channeled into regular metabolic pathways because many enzymes exhibit imperfect substrate specificity. Chemical properties that tend to reduce biotransformation are high molecular weight, low water solubility, aromatic rings, a large amount of branching, and halogen substitutions.

Microorganisms strive to gain energy from the transformation of organic substances. The biodegradation potential of a chemical is thus often related to the energy derived from its degradation. The strategy that evolved as a result is usually an initial oxidative or hydrolytic step, which seeks to convert the environmental organic chemical into something more polar, with the resulting product possibly fitting into a common metabolic pathway. An example is the hydroxylation of benzene to catechol, sometimes via phenol, followed by aromatic ring cleavage:



Most organisms have some relatively unspecific enzymes available for the sole purpose of transforming foreign compounds into more polar substrates. If the product cannot be transformed further (dead-end metabolism), the higher polarity of the product facilitates its return to an aqueous environment in water-soluble, sometimes conjugated form. Co-metabolism is said to occur if an organism transforms a chemical without deriving energy from it. Sometimes several organisms cooperate in the degradation of a compound. Consortia of microorganisms act jointly, each species being responsible for specific steps in the transformation leading towards mineralization (i.e., complete conversion into simple inorganic species).

Several steps can be limiting the rate of microbial transformation of an environmental chemical: The rate-limiting step may be the delivery of the chemical or that of a co-substrate to the transforming enzyme. This in turn involves delivery to the microbial cell, uptake across the cell membrane, and intracellular transport to the site of the degrading enzymes. If transport to the metabolic site is rate limiting, first-order kinetics often apply and the rate of transformation is linearly related to the concentration. If the enzyme is constitutive (i.e., always present), the rate may be limited by the specific interactions of the environmental chemical and the enzyme, which has traditionally been described with a Michaelis–Menten equation. If the concentration of the environmental chemical is small compared to the half saturation constant, and if the microbial cell density can be assumed to be constant, the Michaelis-Menten kinetic also approaches a first-order kinetic. The mere observation of first-order kinetics thus provides little mechanistic information about the rate-limiting steps in the biotransformation of an environmental chemical. Finally, if an environmental chemical is quite abundant and also easily metabolized, it may become the limiting factor in the growth of a microbial population. Then the rate of transformations needs to take into account the growth of the cell population (Monod kinetics). There are other factors that may delay the onset of degradation or slow down the initial rate of degradation, such as the need to induce the production of certain enzymes.

V. THE STUDY OF CHEMICAL FATE AND TRANSPORT IN THE ENVIRONMENT

The fate and transport of chemicals in the environment is investigated both in the field and the laboratory. Numerical models are often employed to synthesize what is known about the fate of a chemical in a particular environmental system.

A. Field Measurements of Environmental Chemical Fate and Transport

Field studies of an environmental chemical's behavior in the environment usually involve in one way or the other the determination of its concentrations in an environmental phase, making use of the various classical and instrumental techniques of analytical chemistry. The difficulty in this task lies in the often very small concentrations of the environmental chemical and the presence of large number of potential interferences. Often the concentration measurement is complicated by the desire to distinguish between different species of the chemical (e.g., various forms of a heavy metals) or different physical state (e.g., the dissolved and sorbed phase in water). We distinguish between techniques that determine the chemical concentration directly in an environmental phase and techniques that rely on the sampling and transfer of subsection of a phase to the laboratory for quantification.

Whereas a concentration value itself reveals little about a chemical's environmental fate, mechanistic and kinetic process understanding can be gained from the interpretation of several concentration measurements. The variability of concentrations and speciation in space and time on a variety of scales can provide considerable insight into fate processes. For example, the extent of variability indicates the average lifetime of chemicals in a phase and therefore information on loss processes. Concentration fluctuations on a diurnal or seasonal scale can help to identify the influence of various environmental variables on chemical fate. Concentrations determined in several phases allow the estimation of distribution coefficients, thermodynamic equilibrium states, and the direction of interfacial fluxes. A whole array of techniques exists to quantify chemical fluxes in the environment, such as those of gaseous, particle-bound, and dissolved chemicals between the atmosphere and the Earth's surface. The determination of the concentrations of various chemicals species or transformation products reveals the nature and often also the kinetics of transformation processes. Changes of the chemical fate in the past can be studied through the measurement of concentrations in archived samples, which includes "natural" archives such as lake sediments, peat bogs, and glacier ice.

B. Laboratory Measurements of Environmental Chemical Fate and Transport

Field results are often difficult to interpret unequivocally and quantitatively, because of the immense complexity and variability of the natural environment. Laboratory experiments, therefore, have an important place in the study of environmental chemical fate and transport by facilitating the focus on specific fate processes under controlled and simplified conditions. Phase distribution, transport, and transformation processes are studied within controlled laboratory settings. The understanding of chemical phase distribution relies on the measurements of basic physical–chemical properties, partitioning coefficients, and equilibrium sorption isotherms. Examples of investigations in the kinetics of transport are studies of the sorption and desorption rates between water and natural organic matter, the rate of uptake in plants in controlled growth chambers, or the air–water mass transfer coefficients in wind–wave tanks. Laboratory studies are indispensable in the understanding of the mechanisms and kinetics of environmental transformations. Photochemical and microbially mediated reactions can be studied in controlled laboratory systems, allowing much easier identification of transformation rates and products than in the environment. The major limitation of the laboratory studies is a potential lack of relevance of the findings for real environmental settings. Some studies try to address this by recreating fairly complex artificial environments in the laboratory or in the open or by conducting controlled field experiments within the natural environment. Examples of the former are so-called mesocosms simulating chemical fate in artificial multiphase systems and atmospheric chamber experiments simulating the reactions of several chemical species. Examples of the latter include experiments with controlled, deliberate release of chemicals into the environment.

C. Synthesis of Chemical Fate and Transport in the Environment in Models

It is often difficult to comprehend how a particular environmental chemical behaves in the environment given the multitude of information on its distribution, transport, and transformation characteristics. The various fate processes interact and compete with each other in complex and sometimes not very intuitive ways. Especially when a quantitative understanding of environmental fate is required, the available information is therefore often synthesized in conceptual and mathematical models.

Most models of chemical fate are based on the principle of the preservation of mass and thus build upon the mass balance equations introduced in Sections I.C and III.D. The two primary approaches are box models and continuous models. Box or compartmental models are based on simple mass balance equations and account for chemical fluxes in and out of a control volume, which is considered homogeneous in terms of environmental characteristics and concentrations (Fig. 5). A box model often contains several compartments, representing either parts of the same phase or different phases. Partitioning and sorption coefficients describe the distribution of chemicals between the phases of one box or between different boxes, whereas kinetic expressions are used to describe transformations in each of the boxes and chemical transport from one box to the other. Continuous models on the other hand are based on analytical and numerical solution of the continuity equation for specific starting and boundary conditions. They are mostly used when describing chemical fate in one sphere. There are thus atmospheric, hydrodynamic, and groundwater dispersion models describing chemical transport and transformation in the atmosphere; rivers, lakes, and oceans; and the subsurface environment, respectively.

SEE ALSO THE FOLLOWING ARTICLES

AEROSOLS • BIOENERGETICS • CARBON CYCLE • ENVIRONMENTAL MEASUREMENTS • ENVIRONMENTAL OBSERVATION AND FORECASTING SYSTEMS • NITROGEN CYCLE, ATMOSPHERIC • OCEAN-ATMOSPHERIC EXCHANGE

BIBLIOGRAPHY

Finlayson-Pitts, B. J., and Pitts, J. N., Jr. (2000). "Chemistry of the Upper and Lower Atmosphere: Theory, Experiments and Applications," Academic Press, San Diego, CA.

Goss, K. U., and Schwarzenbach, R. P. (2001). "Linear free energy relationships used to evaluate equilibrium partitioning of organic compounds." *Environ. Sci. Technol.* **35,** 1–9.

Hemond, H. F., and Fechner-Levy, E. J. (2000). "Chemical Fate and Transport in the Environment," Academic Press, San Diego, CA.

Mackay, D. (2001). "Multimedia Environmental Models—The Fugacity Approach," Lewis Publishers, Boca Raton, FL.

Manning, W. J., ed. (1999). "Special anniversary issue: issues in environmental pollution," *Environ. Pollut.* **100,** 1–245.

Morel, F. M. M., and Hering, J. G. (1993). "Principles and Applications of Aquatic Chemistry," Wiley-Interscience, New York.

Schwarzenbach, R. P., Gschwend, P. M., and Imboden, D. M. (1993). "Environmental Organic Chemistry," Wiley-Interscience, New York.

Seinfeld, J. H., and Pandis, S. N. (1998). "Atmospheric Chemistry and Physics: From Air Pollution to Climate Change," Wiley-Interscience, New York.

Stumm, W., and Morgan, J. J. (1996). Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters," Wiley-Interscience, New York.

Thibodeaux, L. J. (1996). "Environmental Chemodynamics," Wiley-Interscience, New York.

Valsaraj, K. T. (1995). "Elements of Environmental Engineering: Thermodynamics and Kinetics," CRC Press, Boca Raton, FL.

Wania, F. (1999). "On the origin of elevated levels of persistent chemicals in the environment." *Environ. Sci. Pollut. Res.* **6,** 11–19.

Water Resources

Edward D. Schroeder
University of California, Davis

GLOSSARY

Biochemical oxygen demand The quantity of oxygen required by bacteria to metabolize organic matter in water.

Groundwater Water that accumulates in the soil. Where accumulation is great enough to saturate the soil pores in quantities great enough to be extracted economically, the region is termed a groundwater aquifer.

Hydrograph A graph of flow in a stream versus time.

Hydrologic cycle The pathway through which global water cycles from the oceans through the atmosphere, precipitation, runoff, infiltration, and back to the oceans.

Hydrosphere The total quantity and distribution of global water.

Suspended solids Particulate matter carried in water and measured as the dry residue on a filter.

Transpiration Loss of water vapor from plants. Evapotranspiration is the combined evaporation from surface puddles and transpiration over an area.

Turbidity Cloudiness in water resulting from the presence of colloidal particles.

Watershed An area that drains to a specific point.

WATER is fundamental to life, and human societies cannot survive without adequate supplies of water. Living organisms are over 70% water by mass; the large majority of organisms live in water and obtain their sustenance from water; and terrestrial organisms can survive for only brief periods without water. Our natural environment has been shaped to a great extent by the occurrence, the availability, and the physical, chemical, and biological characteristics of water. Water has shaped the physical characteristics of our environment through glacial action, floods, and erosion. Reservoirs and canals have been used since ancient times to control floods, improve transportation, and change the availability of water. Today, approximately 19% of the global supply of electricity is generated with water, and approximately 12% of global food supply is grown using stored water.

I. THE ROLE OF WATER

The importance of water availability is illustrated by comparing the ecology of the Florida Everglades and Death Valley in California, and the importance of the chemical and physical characteristics of water can be illustrated by considering Lake Superior, the Dead Sea, and Antarctica. We tend to set relative values on particular types of environments; for example, most of us would say that Lake Superior is "superior" to the Dead Sea. However, each environment provides competitive advantages for particular types of life. The diversity in life forms resulting from the range in availability and characteristics of water on our planet appears to be a significant factor in stability of the worldwide ecosystem.

Usefulness is strongly related to the chemical characteristics of water. For example, rainwater usually contains minimal concentrations of dissolved ions, is poorly buffered, and has a pH of about 5.6. As a result, rainwater tends to be slightly corrosive, generally good for irrigating plants, and is generally thought to be good tasting. Subsurface water (more commonly called groundwater) is often relatively high in dissolved matter and may have concentrations of ions that make it unsuitable for drinking or irrigation of some crops. The high salt concentrations of seawater are required by some species of organisms but are unsuitable for consumption by terrestrial animals and for irrigation of most crops.

Most organisms that live in water are microscopic bacteria, algae, protozoans, and invertebrates that serve as bottom rungs in the food chain and as primary degraders of organic material. However, water also serves as a conduit for the transmission of many diseases caused by microorganisms. In most cases, the pathogenic organisms are added to water and are not natural members of the aquatic community. This is particularly true of human viruses and pathogenic bacteria. However, some diseases such as schistosomiasis, filariasis, and guinea worm are caused by organisms whose life cycle includes periods in water.

A. Distribution and Availability of Water

Worldwide distribution of water by type is shown in Table I. Of the total global volume of water, about 1.5 trillion km^3, only 2% (28,000,000 km^3) is freshwater, that is, water usable for consumption and for agriculture. The definition of freshwater is imprecise but can be nominally defined as water having total dissolved matter concentrations of less than 1500 mg/L. As indicated in Table II, 85% of global freshwater is frozen. Groundwater, that is, water completely filling pores in soil, gravel, and sand, accounts for 4000 km^3, or 14% of the global freshwater. Much of the total groundwater resources are either physically inaccessible or economically unavailable. Surface waters, rivers, lakes, and reservoirs comprise approximately 155,000 km^3, less than 1% of the total freshwater volume. Extraction of freshwater for municipal, agricultural, and industrial use places considerable stress on the available water resources. About 1300 km^3/day of freshwater is used in the United States alone. Clearly, the worldwide demand for freshwater would rapidly exhaust the available volumes if the system were not dynamic. Although the total amount of global water is constant, there is a continuous movement from the oceans to atmospheric water, through evaporation, to surface water, through precipitation, and to groundwater, through infiltration. Thus extracted water is replaced in a process known as the *hydrologic cycle*. Moreover, extracted water is returned to the system through wastewater discharges, agricultural drainage, and evaporation.

TABLE I Worldwide Distribution of Water by Type

Type	Volume, 1000 km^3	%
World oceans	1,370,323	94.2
Groundwater (fresh and saline)	60,000	4.1
Glaciers	24.000	1.65
Lakes and reservoirs	280	.019
Soil moisture	85	0.006
Atmospheric water	14	0.001
River water	1.2	0.001
Total	1,454,700	

[From L'vovich, M. I. (1979). "World Water Resources and Their Future" (R. L. Nase, translator), American Geophysical Union, Washington, DC.]

B. The Role of Water Quality

Water quality is as important as water availability in supporting ecosystems and human society. The two most

TABLE II Worldwide Distribution of Freshwater by Type

Type	Volume, 1000 km^3	% of total freshwater	% of total type of water
Glaciers	24,000	85	100
Groundwater	4,000	14	6.7
Lakes and reservoirs	155	0.6	55.4
Soil moisture	83	0.3	97.7
Atmospheric water	14	0.05	100
River water	1.2	0.004	100
Totals	28,253.2	100	

[From L'vovich, M. I. (1979). "World Water Resources and Their Future" (R. L. Nase, translator), American Geophysical Union, Washington, DC; Heath, R. C. (1982). "Basic Groundwater Hydrology," USGS, Water Supply Paper No. 2220.]

important water quality parameters are the total dissolved solids (TDS) concentration and temperature. Dissolved matter in water is principally composed of four cations (calcium, magnesium, potassium, and sodium) and for anions (bicarbonate, chloride, nitrate, and sulfate), although smaller quantities of virtually every element and compound can be found in water at some location on the earth. The relative concentrations of the major ions can be important. For example, water having high ratios of sodium to calcium and magnesium cause clay soils to swell and become impermeable, and low permeability soils are unsuitable for agriculture. However, the total salt concentration is the principal determinant of the type of life that can be supported. Organisms have adapted to specific ranges of TDS, either physiologically or with respect to food and nutrient sources. Thus we find different plants, animals, and microorganisms living in fresh, brackish, and saline waters. Some organisms such as the brine shrimp (*Artemia*) and halophilic algae and bacteria grow in the Dead Sea and Great Salt Lake where salt concentrations are up to 25%.

The ability of organisms to grow, compete, and resist disease is strongly impacted by water temperature. Species of plants, animals, and microorganisms tend to do best in rather specific temperature bands or seasonal temperature patterns. Fish are perhaps the most familiar examples of organisms that are distributed according to temperature patterns. Trout and pike are typical coldwater gamefish while bass do better in warmer waters of shallow lakes and slow moving streams. Anadromous fish, such as salmon, wait for temperatures to reach specific values before proceeding upstream to spawn.

Physical, chemical, and biological water quality characteristics are important in determining the suitability of water for municipal, agricultural, and industrial use. Municipal water supplies must be essentially free of suspended matter, have low concentrations of specific ions and compounds, and be free of pathogenic organisms. As noted above, TDS concentrations and the ratio of sodium to calcium and magnesium are limiting factors in agricultural water use. Additionally, certain contaminants, such as boron, place constraints on the type of crop that can be grown. The principal use of water in industry is for cooling, where scaling and corrosion are the principal water quality issues.

C. Spatial and Temporal Distribution of Water Resources

Spatial and temporal distribution of water resources has been a definitive factor in the development of human societies. Although the land was extremely arid, ancient Egypt was able to develop because the source of the Nile River, the Ethiopian plateau, receives heavy rains between April and October. Despite the fact that precipitation is badly located both spatially and temporally, large-scale agricultural development in California has been possible, through the development of storage and distribution facilities. British agriculture depends very little on seasonal water storage and irrigation because rainfall is generally plentiful throughout the growing season.

Historically, floods have been one of the principal factors controlling the development of communities and societies. The flooding Nile eroded nutrient-rich upland soils and deposited them in the delta, allowing Egyptian agriculture to flourish. However, more common effects of floods have been to limit use of land on flood plains and to wreak calamitous damage during unusually large flood events. During the 20th century major efforts were made to develop flood control systems in major river basins throughout the world. These efforts combined dams, reservoirs and levees to temporarily store flood waters, protect land along river banks, and move flood waters rapidly to the ocean. One result has been increased flood plain development, faster storm runoff, and higher peak flows during floods. A second result has been loss of habitat for many aquatic species. For example, floods clean out spawning gravels and deposit new gravels eroded from upstream hills. When dams are constructed, flood peaks are decreased, which decreases cleaning, and newly eroded gravels are deposited in the reservoir formed by the dam. Such unintended results of flood control measures have increased understanding of system complexity and raised questions about the sustainability of modern society.

II. MOVEMENT OF WATER IN THE HYDROSPHERE

The distribution of water shown in Table II can be considered a snapshot in time. Small changes in the distribution occur on a seasonal basis, and major changes occur over periods of centuries and millennia. During the Pleistocene Epoch (the most recent ice age ending approximately 10,000 years ago), the volume of water stored as ice was several times greater than at present, the earth was more arid, with regions such as Florida being sandy deserts, and the mean sea level elevation was about 130 m lower than at present. The global climate, and hence the distribution of water in space and time, will continue to change on several time scales and the distribution of water will vary correspondingly. Currently, there is a general retreat of global ice that is thought to be at least partially due to anthropogenic causes, such as the production of greenhouse gases. While the effects of the warming trend are not completely clear, most predictions include increased variability in weather and changes in annual precipitation patterns as well as a rise in mean sea level of 0.35 to 0.55 m by 2100. Efforts to predict impacts of increased mean sea levels is the subject of considerable research at the present time.

A. Short-Term Water Resource Dynamics

Human society must necessarily focus on much shorter time scales than those that result in global climate change. Two principal areas are of importance in water resources: provision of water for human activities or environmental management and management of runoff or flood control. The time scale associated with water supply is defined by annual rainfall patterns, particularly the length of dry periods. Flood management requires consideration of the intensity and volumes of precipitation, and frequency of particular storms. Thus flood management facilities must be designed for the maximum storm that can reasonably be expected.

Consideration of the movement of water on a shorter time scale is built around the hydrologic cycle, which provides a conceptual picture of the short-term dynamics of the system. Time scale is an important factor. Surface water dynamics have time scales of hours and days, while subsurface (groundwater) dynamics have time scales of months and years. For example, flood waves move down rivers at velocities of several km/hr while groundwater surface elevations may change by a few meters during the course of an irrigation season and tens of meters during periods of continued overdraft.

B. The Hydrologic Cycle

The hydrologic cycle, depicted in Fig. 1, involves five fundamental steps: evapotranspiration, transport of water vapor, condensation/precipitation, overland flow/percolation, and channel runoff. Evapotransporation is the vaporization of water through the combined processes of evaporation from surface water and transporation from plants. Sublimation of frozen water in glaciers is relatively small but is effectively included in the general category. The amount of water vapor that can be held in the atmosphere is a function of the ambient temperature and can be approximated using Eqs. (1) and (2)

$$\rho_S = 0.622\frac{e_S}{RT} \tag{1}$$

$$e_S = 2.7489 \times 10^7 \exp\left(-\frac{4278.6}{T - 30.37}\right) \tag{2}$$

where ρ_S = saturation density of water vapor, kg/m^3,
e_S = saturation vapor pressure, kN/m^2,
R = dry air gas constant, 0.287 kN · m/kg · K,
T = temperature, K.

Thus, at 20°C (293.16 K) the saturation vapor pressure is approximately 2.33 kN/m^2 and the amount of water vapor that can be held in the air is 0.017 kg/m^3. At 30°C the saturation vapor density is 0.030 kg/m^3 and at 10°C the saturation vapor density is 0.009 kg/m^3. Thus warm air (e.g., 30°C air) containing 0.017 kg/m^3 of water vapor will be unsaturated. Cooling the air to 20°C will result in the air being saturated. Further cooling to 10°C will result

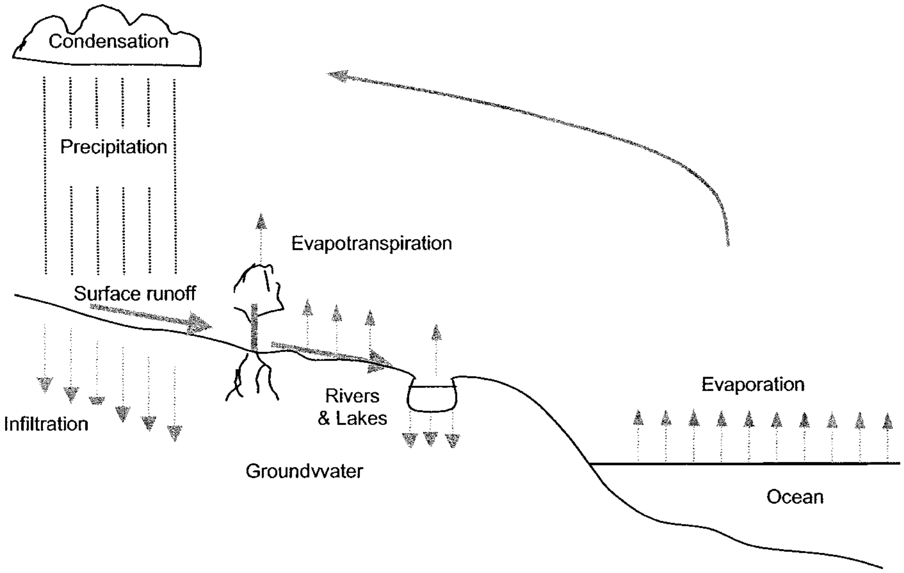


FIGURE 1 The hydrologic cycle.

in super saturation by approximately 0.008 kg/m^3. Water vapor in supersaturated air condenses around microscopic particles that precipitate as rainfall. Thus cooling of moist air will result in precipitation if the temperature falls below the saturation temperature for a particular water density (20°C in the above example). The saturation temperature is called the *dew point*.

Pressure gradients in the atmosphere resulting from uneven heating and cooling result in winds that move water vapor with other gases. Air expands and cools as it rises (e.g., due to the *orographic* lift when winds move air over mountains). Air is also cooled by passing over cold surfaces (e.g., snow fields) and due to thermal convection. The density of wet air is less than that of dry air. Wet air may be warmed by radiation from warm surfaces, causing it to decrease in density and rise through cooler overlying air. Expansion occurs as the pressure decreases and the temperature drops to the dew point. Condensation of the water vapor releases the latent heat of vaporization, resulting in decreased air density, further rising, and further condensation. Thermal convection is responsible for the formation of *cumulus* clouds which have a clearly defined base and fluffy, cottony tops.

Depending on the temperature and other meteorologic conditions, the form of the precipitation is rain, snow, sleet, or hail. Precipitation reaching the surface infiltrates into the soil, becomes surface runoff, is temporarily stored as snow, or falls directly onto surface waters such as the oceans, lakes, or rivers. Most river systems discharge to the world oceans, but a few discharge to lakes such as the Great Salt Lake and the Dead, Caspian, and Aral Seas that have no surface outlets. Water that infiltrates into the soil may be assimilated by plants, percolate into groundwater aquifers, and/or resurface as springs and become part of the surface runoff. Estimates of residence times in each phase of the hydrologic cycle are given in Table III.

TABLE III Residence Times of Water Molecules in the Hydrologic Cycle

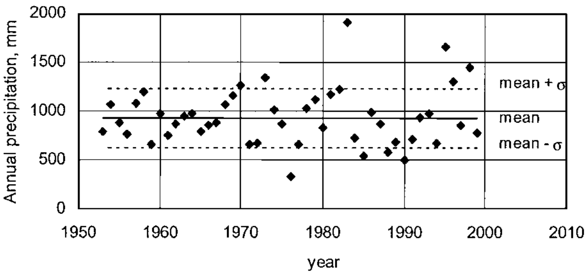
Location	Residence time
Atmosphere	10 days
Rivers (speed 1 m/sec)	2 weeks
Soil moisture	2 weeks to 1 year
Largest lakes	10 to 1,000 years
Shallow groundwater (speed 1–10 m/sec)	10 to 100 years
Mixed layer of oceans (150 m depth)	120 years
World ocean	3,000 years
Deep groundwater	up to 10,000 years
Antarctic ice cap	10,000 years

[Source: Heath, R. C. (1982). "Basic Groundwater Hydrology," USGS, Water Supply Paper No. 2220.]

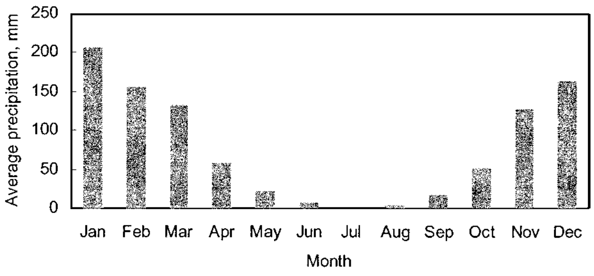
Although evapotranspiration is continuous, both evaporation and transpiration are functions of temperature, and hence are seasonal. The largest source of atmospheric moisture is evaporation from the oceans. Evaporation from surface waters is a very small fraction of the total, with transpiration from plants being somewhat more significant. Of course, transpiration requires growing plants and stops altogether when annual plants die, or when deciduous trees lose their leaves. However, some transpiration occurs all year from perennials and nondeciduous trees.

Movement of water vapor in the atmosphere is a function of weather patterns that result in generally repeatable annual precipitation cycles. However, precipitation is highly stochastic on a short-term basis, and annual total precipitation can be highly variable, as indicated in Fig. 2.

The form of precipitation, rain, ice, or snow, is important because rain begins moving through the ground or over the surface on impact while frozen precipitation is stored on the surface for a period of time. Regions receiving large amounts of snow have spring floods as melting occurs. Many arid regions, such as the southwestern United States, depend on snow melt from mountains for water during the long dry summers.



a. Annual total precipitation in mm at the Hopland Experiment Station



b. Monthly mean precipitation at the Hopland Experiment Station

FIGURE 2 Precipitation patterns at the headquarters weather station of University of California Agricultural Experiment Station, Hopland, CA. Elevation is 244 m. Annual total precipitation values are shown in (a) with mean and standard deviation lines. Monthly averages shown in (b) are for the period 1953 through 1999.

Rate of infiltration into soil is dependent on the recent storm history and soil characteristics. When storms are frequent, sorption sites on soil particles remain saturated and infiltration is solely due to flow of water downward through the soil. Additionally, smaller pores remain full due to capillary forces, and the available drainage cross section is reduced.

C. Return Period

Considerable effort is made to evaluate the return period of storms, that is, the probability that a storm of a particular magnitude will occur within a given period of time. The magnitude of storms varies considerably, and even the maximum storm that might occur annually is highly variable. An example of the variation in storm magnitudes is given by the annual maximum storm flows in the Don River, shown in Fig. 3. The simplest and most common method of characterizing storms is by estimation of the volume of water precipitated and the time over which the precipitation took place. An additional factor is the area over which the precipitation occurred. The importance of the quantity of water precipitated is quite obvious. Importance of storm duration results from the fact that the rate of runoff is directly related to the storm intensity. For example, flood damage resulting from a 25-mm storm occurring during a 1-hr period may be considerably greater than the damage resulting from a 100-mm storm occurring during a 24-hr period. The area over which precipitation occurs is a complicating factor in defining the return period. Storms are rarely uniform in intensity and duration over an entire drainage basin. Hence, judgments must be made in defining areas of impact and methods of averaging. In summary, storm return period has at least two dimensions, volume and intensity. A third factor, area of impact, increases the uncertainty of the calculations.

Return period is estimated by classifying all recorded storms in a given period of time and ranking them according to frequency. In the simplest approach to defining return period, the largest storm has occurred once and hence the return period is n years. The second largest storm has been equaled or exceeded once and the return period is $n/2$ years. Return periods of smaller storms are defined in a similar manner. When precipitation records are adequate, more sophisticated estimates of storm frequency can be applied using appropriate statistical distributions. Usually normal or log-normal distributions provide adequate descriptions of storm frequency. However, distributions that include skewness, such as the Pearson Type III and Gumbel, may be used as storm magnitudes have a lower bound. Application of the distributions allows estimating the probability of a storm of a given magnitude occurring within a selected time increment. It is important to remember that in frequency analysis storms are assumed to be random events. Thus the occurrence of a storm of a particular magnitude in a given time period does not change the probability that a storm of equal magnitude will occur in the next time period. Thus a 100-year storm has the same probability (0.01) of occurring in any given year. Correspondingly, it is not surprising to have several successive years in which the total rainfall is less than or greater than average. Additionally, return periods and other predictive analyses are based on the available records. Each year the record increases and the estimates of average rainfall, 100-year storm magnitude, and other values change.

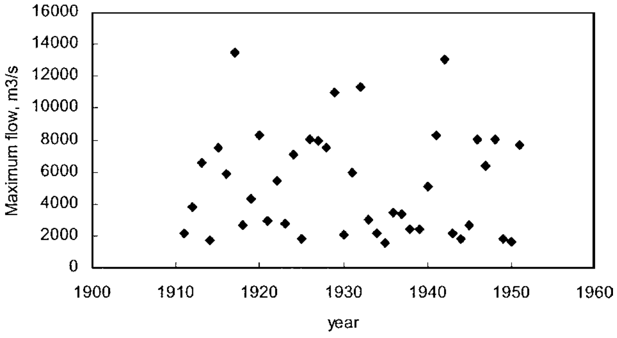


FIGURE 3 Maximum annual flows in the Don River at the Razderskaya gauging station for the years 1911 to 1951.

Relationships between storm intensity and duration are developed using return period as the reference. By plotting (or sorting numerically) storms according to intensity (in mm/hr) and duration (in minutes), the relationship between storms of a given magnitude or return period and the intensity or duration can be determined. This information can be used as the basis for design capacity of flood control facilities.

D. Overland Flow

Precipitation reaching the ground surface is sorbed onto soil surfaces and infiltrates into the soil. When the precipitation rate is greater than the rates of sorption and infiltration, water accumulates and begins to move as free-surface flow. Conceptually, the flow is no different than free-surface flow in engineered facilities. Gravity is the principal driving force, and inertial and pressure forces can be neglected. Gravitational force is nearly balanced by bed friction. Flow rates can be estimated using the kinematic wave equation

$$Q = \alpha y^m \tag{3}$$

where Q = volumetric flow rate, m^3/sec,
y = depth of flow, m, and
α, m = kinematic wave parameters.

The Manning equation for open channel flow is of this form:

$$q = \frac{1}{n}\bar{s}^{1/2}\bar{y}^{5/3} \tag{4}$$

where q = volumetric flow rate per unit width, $\mathrm{m}^3/\mathrm{m} \cdot \mathrm{sec}$,
n = Manning roughness coefficient,
$\bar{s}$ = mean slope, m/m, and
$\bar{y}$ = mean flow depth, m.

Values of the roughness coefficient, n, range from about 0.01 for smooth pavement to 0.4 for dense shrubbery and grass.

Boundary conditions for overland flow are typically difficult to describe, and the flow is inherently nonuniform and non-steady-state. Ground surfaces are uneven, rivulets and channels develop, plants and debris retard the flow and make friction factor estimates highly variable, lateral boundaries are not constant, and slopes change. Each of these factors can be incorporated in overland flow models, but the coefficients are averaged over discrete areas of the drainage area or watershed. The finer the grid, the more accurate a predictive overland flow model can be. However, obtaining coefficient values for drainage basins is difficult and costly. The values are functions of past precipitation history and are affected by changes in land use, accumulation of debris, and many other factors. Consequently, there is a great deal of empiricism involved in applying overland flow relationships.

E. Infiltration and Groundwater Flow

Water below the ground surface may be sorbed onto the surface of soil particles, flow downward as a thin liquid film through unsaturated pores, flow in locally saturated pores, be held in place by capillary forces, or be stored in a saturated zone called an aquifer, as indicated in Fig. 4. Water in aquifers is often moving as the result of pressure gradients generated by springs or by pumping.

Soil above the water table is generally unsaturated; the pores between soil particles are at most locally full, and a significant fraction of the pore volume is filled with air. Some moisture is bound to soil by osmotic forces and a certain amount is held above the water table in small pores by capillary forces. However, most of the water entering the subsurface migrates toward the saturated zone under the dominant action of gravitational forces.

Subsurface flow is nearly always laminar, and thus there are similarities between flow in saturated and unsaturated zones. Because of the heterogeneous nature of the subsurface, flows are extremely difficult to describe and nearly all models are based on Darcy's law

$$v_S = -K_C \frac{\Delta h}{\Delta L} \tag{5}$$

where v_S = superficial velocity, m/sec,
K_C = hydraulic conductivity, m/sec,
Δh = head loss, m,
ΔL = length of flow path, m.

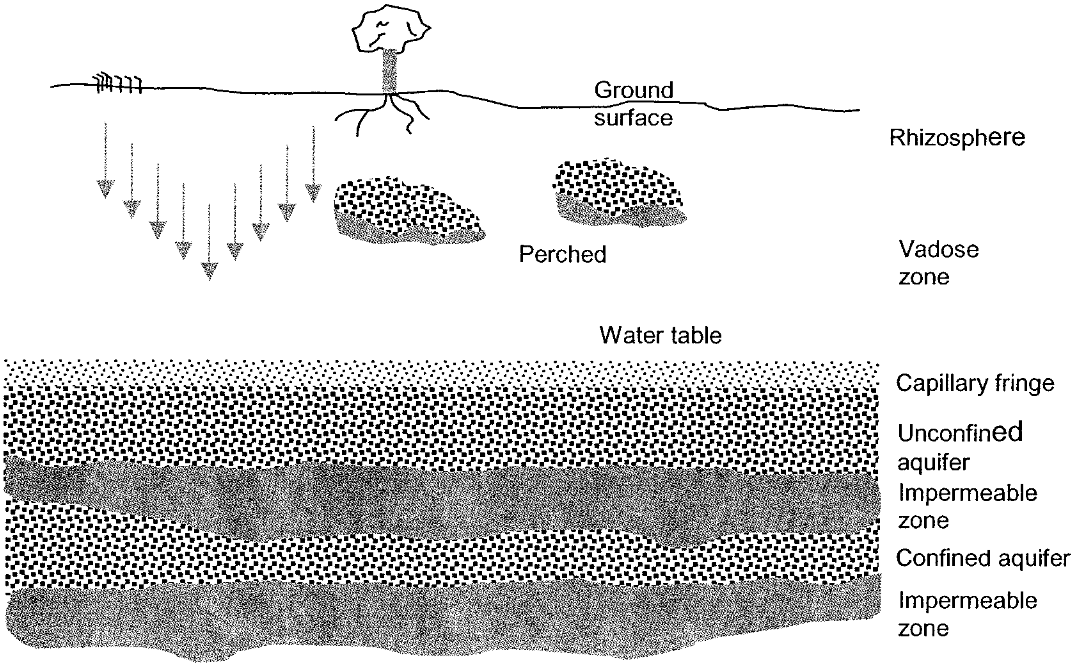


FIGURE 4 Movement of water below the ground surface.

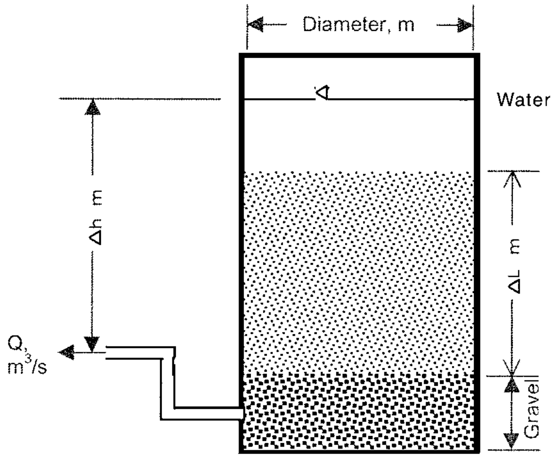


FIGURE 5 Apparatus for measuring head loss in soil sample. The gravel layer has much less resistance to flow than the soil, and head loss through the gravel can be considered insignificant.

Superficial velocity is determined by dividing the volumetric flow rate by the total cross-sectional area. Pore area is approximated by the porosity, and average pore velocity is usually estimated by dividing the superficial velocity by the porosity. Pore velocity is important in estimating transport time of chemical contaminants.

Application of Darcy's law is illustrated for saturated flow in the test system sketch shown in Fig. 5. Note that head loss in the gravel support layer is assumed to be negligible.

Flow in the subsurface is always three-dimensional and the hydraulic conductivity, K_C, is a function of position and local soil characteristics. Because flow is laminar, the hydraulic conductivity should be a function of pore size and viscosity, and a typical expression is given by

$$K_C = Cd^2\frac{\rho g}{\mu} \tag{6}$$

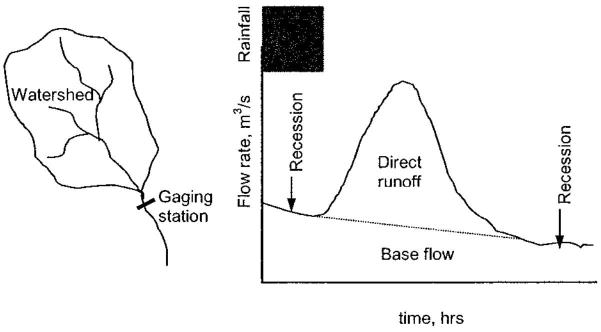
where C = proportionality coefficient,
d = soil grain size, m,
ρ = density of water, kg/m^3, and
μ = viscosity of water, N · s/m^2.

Values of K_C range from 10^{-8} m/sec for fine clays to 10^{-3} m/sec for coarse sands.

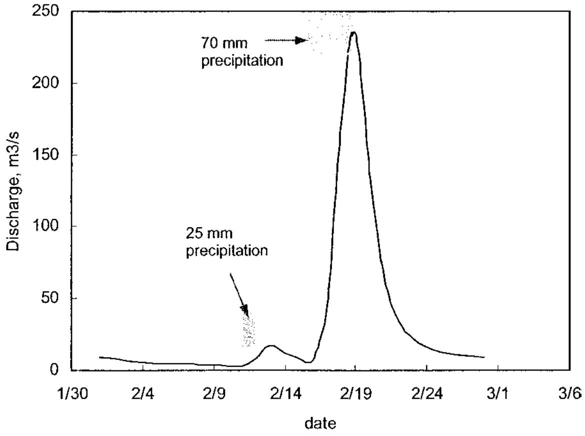
Flow in the unsaturated vadose zone is best described using values of hydraulic conductivity obtained through experimental observation. Not surprisingly, prediction of unsaturated flow rates is considerably less satisfactory than prediction of rates in the saturated zone.

F. Storm Hydrographs

Storm hydrographs are used to describe the response of streams and rivers to overland flow discharges. Water flowing across the land surface eventually comes together in channels to form creeks and rivers. At any point selected, the flow versus time response to a storm will show a relatively rapid rise, a peak, and a gradually declining recession back to the *base* flow fed by springs, as indicated in Fig. 6. The shape of the hydrograph is a function of the size, slope, and physical characteristics of the drainage area. For example, flow from parking lots characteristically peaks rapidly because of the relatively short flow distances, smooth surfaces, and high slopes (required to minimize accumulation of water on the surface). Flow rates in rivers change much more slowly because overland flow velocities are generally slower and distances traversed are larger than in urban settings. An example is provided in Fig. 6b for the Portage River in northern Ohio. Note that the response time to the two storms is fairly rapid but the hydrographs are of the order of days in length.



a. Characteristic flow pattern at a gaging station following a storm



b. Flow of the Portage River at Woodville, Ohio during April, 1997.

FIGURE 6 Characteristic storm hydrograph for a drainage basin and actual flow records for the Portage River at Woodville, Sandusky County, OH for the month of April, 1997.

Storm hydrograph characteristics are reasonably consistent at particular locations, and this fact allows the development of *unit* hydrographs that are characteristic of one unit of rainfall on the upstream drainage area in a unit time period (usually 1 hr). Unit hydrographs can be combined to construct hydrographs for storms of any volume and any duration. Unit hydrograph theory is based on the assumption that discharge is a linear function of precipitation. Although the assumption is not valid, the unit hydrograph concept is useful and can often provide satisfactory estimates of response.

Hydrographs for locations downstream of a gauging station can be constructed by a process known as flood routing which incorporates a stream flow model and the equation of continuity to predict flow rates. The stream flow model must incorporate physical characteristics of the channel. A number of routing models are available as computer software packages.

G. Watershed Models

Watershed models integrate overland flow, infiltration, streamflow, and routing to provide estimates of stream flow as a function of storm characteristics. The most widely used watershed models are distributed by the Hydrologic Engineering Center of the U.S. Army Corps of Engineers. Models require calibration for each application. Such models have become essential tools for prediction of flows and river depths resulting from storms.

General-use watershed models include features such as reservoir storage, variable land characteristics, and soil moisture as a function of time. Inclusion of such features allows application of watershed models for reservoir operation and evaluation of alternative strategies for responding to a range of storm scenarios. For example, watershed models are used to predict the impact of large storms, the potential for local or regional flooding, and areas that will be most severely impacted. At present, construction is discouraged in areas that will be inundated by floods having a 100-year return period. Thus flood control facilities, such as levees and reservoirs, are constructed to contain such a flood, and significant damage can be expected if a flood exceeding the 100-year flood occurs.

H. Impacts of Land Development

Urban development results in significant changes in watershed characteristics. The most significant features are covering of soil by pavement and buildings and installation of stormwater collection systems. Pavement and buildings prevent infiltration and for this reason the fraction of precipitation that becomes surface runoff is much greater in urban than in rural areas. Covering soil with pavement and buildings decreases erosion considerably. Paved surfaces, roofs, concrete pipes, and channels are much less resistant to flow than grass, forests, and fields. Consequently, both the time to peak flow and the overall watershed drainage time decrease with urbanization.

Each of these changes can be considered a benefit or a detriment, depending on one's viewpoint. Increasing the fraction of precipitation that becomes surface runoff can be useful if storage facilities are available and the water can be used in some manner. However, in arid areas that receive the most precipitation during the winter, the result is a loss of groundwater recharge. Rapid drainage of urban areas is highly desirable. For example, accumulation of water in streets is a traffic hazard. However, downstream flood flow peaks increase with urbanization and in some cases the damage to downstream facilities and ecosystems may be high. Erosion control is generally considered to be a benefit. However, in southern California where beach sand is supplied by erosion of sedimentary material in the surrounding mountains, urbanization has resulted in a need to buy sand to replace that washed away by tidal action.

III. WATER QUALITY

Water quality is defined by physical, chemical, and biological characteristics and the intended use. Pure liquid water, that is, pure H_2O without dissolved ions, compounds, particles, and gases, does not exist in nature. From the moment that liquid droplets condense in the atmosphere, materials begin to accumulate (see Table IV). Raindrops quickly reach equilibrium with atmospheric gases, most notably oxygen, nitrogen, and carbon dioxide. Minerals present in the atmosphere as a result of evaporation of spray over the ocean and entrainment of dust by winds dissolve in the drops as they fall. Combustion processes emit sulfur dioxide and oxides of nitrogen which dissolve in raindrops and are converted to mineral acids. Local application of pesticide and herbicide sprays may result in temporary presence of these compounds. Surprisingly, the urban application of pesticides and herbicides is often a larger contributor than that from agriculture. Concentrations of selected contaminants found in rainfall at various locations in California are given in Table IV. Note that the pesticides chlorpyrifos and diazinon are much higher in the Lodi samples because of their use as dormant sprays in orchards.

After reaching the ground surface, water comes in contact with an array of organic and inorganic materials, particles are carried by overland flow, and materials dissolve into the water. The characteristics of water change continually during transport through the hydrologic cycle.

TABLE IV Concentration of Selected Constituents in Rainfall at Five Locations in California during February 1999

Constituent		Bakersfield	Los Angeles	Riverside	Lodi	Orange County
Copper, total	(μg/L)	3.5	5.8	2.8	4.0	2.4
Lead, total	(μg/L)	3	6.6	0.9	3.3	2.7
Zinc, total	(μg/L)	34.0	51.9	15.5	34.0	100.7
Total Kejldahl nitrogen	(mg/L)	1.03	0.76	1.00	1.08	0.85
Nitrate	(mg/L)	0.56	0.02	0.33	0.87	1.98
Phosphorus	(mg/L)	0.03	0.04	0.03	0.11	0.05
Chlorpyrifos	(μg/L)	0.07	0.02	0.01	0.87	0.04
Diazinon	(μg/L)	0.09	0.09	0.03	0.65	0.15

[Source, California Department of Transportation.]

A. Water Quality as Defined by Water Use

Because of the huge range of materials that may be present in *natural* water, any definition of quality must be related to intended use. Thus a water that is quite acceptable for drinking may not be acceptable for irrigation of citrus trees because of the presence of boron. Water quality for shipping is defined by viscosity and perhaps corrosiveness. When water temperature drops below freezing, the physical state of water becomes a problem for ships. Survival of bacteria living in water requires that carbon and energy sources as well as inorganic nutrients be present. The materials that serve as nutrients for the bacteria are generally very unwelcome in public drinking water supplies. Boron is not regulated in drinking water supplies but is a problem at concentrations of 1 mg/L in water used to irrigate citrus trees. We need to think of maintenance of environmental quality, particularly the aquatic habitat, as a water quality issue and therefore environmental management is a type of use. Unfortunately, the aquatic organisms do not have uniform water quality requirements. Many live in ecological niches provided by local water quality characteristics, and changes in water quality can impact the aquatic community is surprising ways. Construction of the Aswan High Dam on the Nile River in Egypt, completed in 1964, has provided one of the best examples of the complexity of aquatic ecosystems. The dam and its reservoir, Lake Nassar, were conceived of as a solution for water resources management, flood control, and electric power requirements for Egypt. Unexpected results have included increases in the incidence of schistosomiasis, a northward movement of malaria, and sharp decreases in the sardine, anchovy, and shrimp population in the Mediterranean Sea. Increases in schistosomiasis and malaria resulted from the decreases in flooding and consequent stability of water levels. Loss of the fisheries resulted from trapping of nutrient-laden sediments behind the dam. The nutrients were necessary to maintain benthic organism and algal populations that were food sources for the sardines, anchovies, and shrimp.

When we speak of water quality, our first thought is the water we drink. In the United States, Congress has enacted regulations on water quality under the Safe Drinking Water Act of 1986 and its later amendments. The requirements are summarized as the National Primary Drinking Water Standards. These standards are based on safety for human consumption and do not include factors related to corrosive or scaling properties of water. Thus it is possible for water that is “acceptable” to cause major problems in the distribution system by corroding pipes or plugging pipes with scale. The water quality requirements of the distribution system are different from those of the consumer.

B. Physical Characteristics of Water

For most people, physical characteristics such as temperature, turbidity, solids, color, and odor form the first impression of the water quality. Streams that are warm, turbid, dark-colored, and odorous would often be assumed to be polluted and perhaps unhealthful. Correspondingly, streams that are cool, clear, and odorless are generally assumed to be of good quality and pollution free. In fact, most natural waters are somewhat cloudy at best and may be colored by tannins and humic matter from decaying plants. Sloughs, particularly in tropical regions, have characteristic odors associated with decaying vegetation. However, the ecological health of such systems may be very good. Mountain streams receiving acid mine drainage may be virtually sterile and contain toxic concentrations of metals but may also be extremely clear, cold, and odorless. Thus the physical characteristics of water must be interpreted in terms of specific situations, and other characteristics may be more important.

1. Temperature

Temperature is a major water quality characteristic because of the relationship of temperature and the type of organisms that can compete, because the solubility of solids and gases in water is a function of temperature, and because the properties of water are functions of temperature. Organisms at all levels of complexity are adapted to certain ranges of temperature. Growth and respiration rates of aquatic organisms increase with temperature. Fish provide a familiar example in that cold-water species, such as trout, are found in mountain streams and lakes of the temperate zone, and warm-water species, such as bass, are found in warmer regions. Cold- and warm-water fish may also be found in the same streams or lakes, but occupying different zones. For example, cold-water fish will be found in the deeper portions of a lake while warm-water fish will be found along the edges in shallow water were temperatures are higher.

2. Turbidity

Turbidity in water results from the presence of colloidal particles that scatter light. As a result, objects in water become indistinct. Lake Tahoe on the California–Nevada border, Crater Lake in Oregon, and the reefs in the Caribbean Sea are famous for the clarity, or lack of turbidity, of their water. Shoreline development and wastewater discharges increase particle density and support growth of phytoplankton that increases turbidity. Decreased light penetration and energy absorption in the near surface layers affect both the aquatic ecology and the heating and cooling of the water.

Turbidity is an important characteristic in drinking water supplies because microorganisms attached to particles are more likely to survive disinfection. Thus drinking water treatment is focused on removal of colloidal particles and the production of extremely clear water—which is also aesthetically pleasing.

In natural waters, a Secchi disk is used to assess turbidity. The black and white Secchi disk is lowered into the water until the markings become indistinguishable and the depth is recorded as the Secchi depth. Extremely clear lakes may have Secchi depths of 30 to 40 m while extremely turbid lakes will have Secchi depths of less than 1 m. Turbidity in drinking water is measured by determining light transmission using standard light sources and reported in nephelometric turbidity units (NTUs). Drinking water standards require turbidities of less than 1 NTU. While only a qualitative relationship exists between turbidity units and Secchi depth, ultraclear natural waters are nearly always below 5 NTU. The United States National Primary Drinking Water Standard for turbidity is 1.0 NTU.

3. Solids

Solids in water include all contaminants other than gases and are classified according to state, size, size distribution, and chemical characteristics. The most widely used measurements of solids in water and their significance are summarized in Table V, and the size distribution of solids is summarized in Fig. 7.

Three classifications, suspended, dissolved, and volatile solids, are of particular importance in defining water quality. Suspended solids and dissolved solids are measures of size, while volatile solids is based on a procedure—burning. Filtration through a standard 1.2-μm pore size filter is used to define suspended and dissolved solids. The practical definition of suspended solids is particles larger

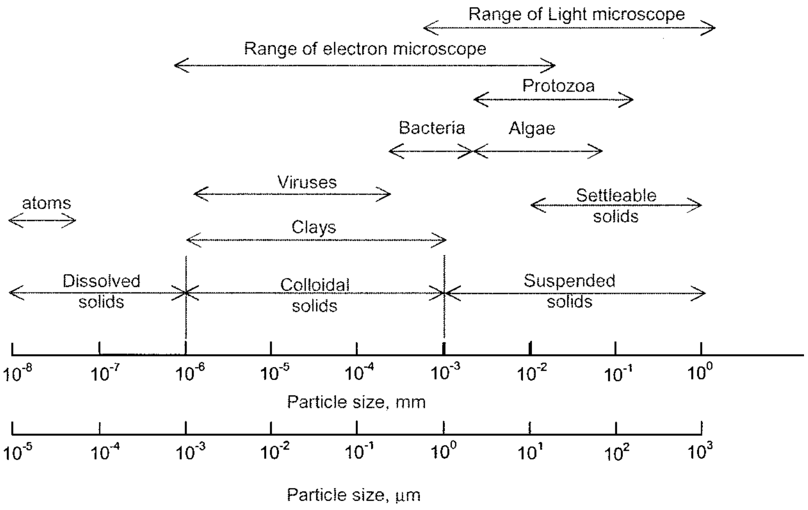


FIGURE 7 Size distribution of solids in water.

TABLE V Standard Measures of Solids Content of Water

Test	Description and use
Settleable solids (mg/L)	A measure of the volume of sludge that can be produced by gravity settling. The test is used to predict sludge accumulation in treatment facilities.
Total solids (TS, mg/L)	The residue remaining after drying at 103°C. Total solids measurements are used in conjunction with other measurements.
Total volatile solids (TVS, mg/L)	The difference between TS and the residue remaining after burning at 550°C. TVS is a measure of the organic solids present in a water.
Suspended solids (SS, mg/L)	The residue retained on a 1.2-μm pore size filter dried at 103°C. Suspended solids are regulated in wastewater discharges, urban stormwater discharges, and runoff from construction sites.
Volatile suspended solids (VSS, mg/L)	The difference between SS and the residue remaining after burning at 550°C. VSS are used to estimate the organic suspended matter concentration and as a measure of microbial concentration in wastewater treatment.
Total dissolved solids (TDS, mg/L)	The solids passing through the standard 1.2-μm pore size filter. TDS includes colloidal particles and thus might be better defined as *filterable* solids. TDS can be measured by drying filtrate at 103°C or as the difference between total solids and suspended solids (TS–SS).

Note. Detailed descriptions can be found in Water Environment Federation (1998). “Standard Methods for the Examination of Water and Wastewater,” 20th edition, Alexandria, VA.

than 1 μm because the standard filter retains most of the particles down to that size. Included are most, but not all, bacteria, algae, and protozoa. Organic particles larger than 100 μm and dense particles (e.g., clay, silt, and sand) larger than 10 μm settle relatively rapidly and are removed in moderately quiescent conditions. Thus suspended solids are closely related to, but not synonymous with, settleable solids. Settleable material forms shoals and sludge banks in natural waters, fills reservoirs, and clogs wastewater collection systems. Removal of settleable material is a major task in both water and wastewater treatment. In addition to being less aesthetically pleasing, water containing suspended matter is less suitable for drinking because pathogenic organisms are shielded from disinfectants.

Dissolved solids, as measured by filtration, include colloidal matter, as noted above. In most natural waters nearly all of the material making up the TDS is inorganic. The principal organic component is complex, high molecular weight humic material that is present at concentrations up to a few mg/L. Concentrations of dissolved solids in surface waters range from about 10 mg/L in lakes and streams at the upper end of watersheds to several hundred mg/L at the lower end of watersheds. As streams progress through watersheds, agricultural runoff, municipal and industrial waste discharges, and erosion add dissolved matter. For example, municipal use of water adds 200 to 300 mg/L of TDS to extracted waters. Total dissolved solids concentrations in the Mississippi River increases from about 170 mg/L at Grand Rapids, Minnesota, to about 250 mg/L at St. Francisville, Louisiana. Dissolved solid concentrations may vary seasonally due to factors such as snowmelt, seasonal agricultural practices, and weather. Groundwaters tend to have higher dissolved solids concentrations than surface waters because of the time of contact with solutes sources (see Table III). Concentrations of dissolved matter found in groundwaters varies from below 100 mg/L to values more concentrated than seawater. Waters having dissolved solids concentrations greater than 1000 mg/L are considered unacceptable as municipal water supplies.

Volatile solids, the material that combusts at a temperature of 550°C, is used as a measure of the organic content. Like suspended and dissolved solids, volatile solids is a lumped parameter and no information is provided about the chemical characteristics of material. However, knowing the volatile solids content of a water provides information about the source and history of the water. High volatile solids concentrations are associated with decaying organic matter such as might be found in swamps and in streams receiving poorly treated wastewaters.

4. Color

Color in water is imparted by dissolved material such as tannins from decaying plants. As a result, colored water is most often brownish in tint. Colloidal organic matter found in wastewater will give receiving waters a gray color if concentrations are high, and swamp waters are often black due to the presence of ferrous sulfide precipitates. True colors are due to dissolved materials but in practice, colors associated with colloidal material are included as a characteristic of water.

Color is usually associated with poor water quality by the public. In general this assessment is correct, although many natural waters that are suitable as municipal sources are slightly colored.

C. Chemical Characteristics of Water

The chemical characteristics of water are most commonly described by the concentrations of a limited number of dissolved inorganic ions, lumped parameters such as alkalinity, acidity, hardness, conductivity, the aqueous carbon dioxide concentration, the radioactivity, and lumped measures of the organic content such as biochemical oxygen demand, as indicated in Table VI. Concentrations of other inorganic ions and specific organic compounds are important in relation to particular uses. The most common issues with respect to concentrations of particular ions and compounds are related to toxicity. For example, the maximum concentration limit (MCL) for arsenic in drinking water of 50 μg/L is based on the average ability to excrete approximately 900 μg of arsenic per day, the probable intake of arsenic from other sources (principally food), and the assumption that we will follow recommendations to drink two liters of water per day. MCLs in surface waters are often set on the basis of toxicity to the most sensitive species present. The water flea *Cerodaphnia* is commonly used as a test organism for toxicity in water because of its place in the food chain and ease of cultivation in the laboratory. Sensitivity of *Cerodaphnia* to the organophosphate pesticide diazinon is based on the concentration at which 50% of *Cerodaphnia* die in 48 hr (LC_{50}) which is approximately 360 to 600 ng/L. Note that precipitation concentrations of diazinon shown in Table IV are in this range. Stream concentrations in areas where diazinon is used are often several times this value.

1. Chemical Characteristics of Water—Inorganic Matter

The most common parameters used to define the inorganic chemical characteristics of water are given in Table VI. As noted above, four cations (calcium, magnesium, sodium, and potassium) and four anions (bicarbonate, chloride, sulfate, and nitrate) comprise most of the inorganic matter

TABLE VI General Measures of Chemical Characteristics of Water Quality

Measure	Abbreviation/ definition	Description
Cations		
Calcium	Ca^{+2}	The four major cations and anions comprise over 99% of the inorganic chemical mass of most freshwaters. The principal chemical characteristics associated with water are most agricultural and domestic uses of defined by these ions.
Magnesium	Mg^{+2}	
Sodium	Na^{+}	
Potassium	K^{+}	
Anions		
Bicarbonate	$HCO_3{}^{-}$	
Chloride	Cl^{-}	
Nitrate	$NO_3{}^{-}$	
Sulfate	$SO_4{}^{-2}$	
pH	$-\log_{10}[H^{+}]$	Living organisms survive within fairly narrow pH ranges. The pH value is important in determining the tendency of a water to corrode or scale in distribution systems.
Alkalinity	$[HCO_3]+2[CO_3{}^{-2}]+[OH^{-}]-[H^{+}]$	Alkalinity measures the ability of a water to neutralize acid.
Hardness	$2([Ca^{+2}]+[Mg^{+2}])$	A measure of the tendency of a water to form scale and consume soap.
Gas		
Carbon dioxide	CO_2	A measure of the corrosiveness of a water and used in determination of chemical requirements in water treatment.
Conductivity	μS/cm	Approximately proportional to the TDS of a water.
Organics		
Biochemical oxygen demand	BOD	A measure of the concentration of biodegradable organic material present using units of oxygen consumed.
Chemical oxygen demand	COD	A measure of the concentration of total organic material present using units of oxygen consumed.
Total organic carbon	TOC	A measure of the total organic carbon present.

found in natural waters. Typical water quality analyses are focused on these ions and three other parameters, pH, alkalinity, and hardness. Approximately 20 minor ionic species are significant in defining water quality. Several species are important because of toxicity. Examples include arsenic (As^-) and copper (Cu^{+2}). Other minor ionic species, such as ammonium (NH_4^+), nitrate (NO_3^-), and phosphate (PO_4^{-3}), are nutrients for plants and algae. Growth of aquatic plants and blooms of algae result from the presence of nutrients. As bodies of water *age*, nutrients accumulate and the water becomes more eutrophic, that is, capable of supporting larger quantities of organisms.

Some inorganic ionic species found in water are primarily from anthropogenic sources. The majority of these ions are heavy metals such as cadmium (Ca^{+2}), chromium (Cr^{+3}, Cr^{+6}), lead (Pb^{+2}), and mercury (Hg^{+2}) that tend to accumulate in living organisms until toxic levels are attained.

a. pH. Liquid water is a weak acid and disassociates into the hydrogen (H^+) and hydroxide (OH) ions. The equilibrium is strongly toward water within the environmental temperature spectrum. However, the presence of other chemicals can alter the equilibrium significantly. In most cases the equilibrium state is characterized by the negative log of the molar hydrogen ion concentration, which is defined as the pH of the water. Pure water at 25°C has a pH of 7, which means that the hydrogen ion concentration is 10^{-7} mol/L, a very small number. The molar concentration of undisassociated water will be approximately 55.5 mol/L.

The activity of the hydrogen ion makes the pH value very important. Waters below pH 7 are increasingly corrosive to pipes. The majority of aquatic organisms require pH values between 6.5 and 9. Precipitation reactions are characteristically strong functions of pH.

b. Alkalinity. Alkalinity has been defined in Table VI and mentioned in connection with other water quality parameters. The quantitative definition of alkalinity is

$$A = \left[HCO_3^-\right] + 2\left[CO_3^{-2}\right] + \left[OH^-\right] - \left[H^+\right] \qquad (7)$$

where $A =$ alkalinity, eq/L, and $[i] =$ molar concentration of chemical species i.

Note that the equivalent weights of the bicarbonate, hydroxide, and hydrogen ions are the same as the gram molecular weights and that the equivalent weight of carbonate is one-half the gram molecular weight. Hence the units of alkalinity are in equivalents per liter. Because constituent concentrations are generally in the millimolar range, alkalinity is usually reported in milliequivelents per liter (meq/L).

The pH of most natural waters falls between 6 and 8.5, a region in which the equilibrium between bicarbonate and carbonate is strongly in favor of bicarbonate. In most cases the bicarbonate concentration is far greater than the hydroxyl or hydrogen ion concentrations. Thus for most situations the alkalinity is approximately equal to the bicarbonate concentration.

c. Hardness. As noted in Table VI, hardness is a measure of the tendency of a water to form scale and to consume soap. Hardness is defined as the sum of the Ca^{+2} and Mg^{+2} concentrations measured in equivalents per liter. Because both ions have a charge of +2, the relationship can be expressed algebraically as

$$H = 2([Ca^{+2}] + [Mg^{+2}]) \qquad (8)$$

where $H =$ hardness, eq/L.

Scale formation results from the precipitation of $CaCO_3$ and $Mg(OH)_2$. Precipitation of $CaCO_3$ is a particular problem in heated water systems such as steam power generation systems, industrial cooling systems, and home hot water systems because the equilibrium between HCO_3^- and CO_3^{-2} increasingly favors CO_3^{-2} as the temperature increases. Although the solubility of $CaCO_3$ increases with increasing temperature, the increase in CO_3^{-2} availability often results in precipitation and scaling. Both Ca^{+2} and Mg^{+2} react with soap to form a precipitate. The result is that hard waters require more soap for cleaning purposes. Substitution of synthetic detergents has greatly reduced the problem.

d. Major cations. The major cations, Ca^{+2}, Mg^{+2}, Na^+, and K^+ are virtually always the dominant chemical species in natural waters. The balance between the divalent cations and sodium is particularly important in agriculture because sodium has a very large radius of hydration. Clays in soils have a net negative surface charge which attracts cations. When the hydrated sodium ion sorbs onto the clay surface the clay swells, soil pores become smaller, and soil permeability decreases. Waters with a high ratio of divalent cations (i.e., Ca^{+2} and Mg^{+2}) to Na^+ ions are highly desirable for irrigation. All four major cations are essential nutrients for life. The United States Food and Drug Administration has reported that dietary intake of sodium should be between 500 and 2400 mg per day. However, sodium has been implicated in hypertension and thus people with high blood pressure, heart disease, or a family history of stroke are often advised to drink low sodium water as one method of minimizing their sodium intake.

e. Major anions. As described above, HCO_3^- is a key constituent in regulation of the pH of natural waters, and because of the equilibrium relationship with CO_3^{-2}, HCO_3^-

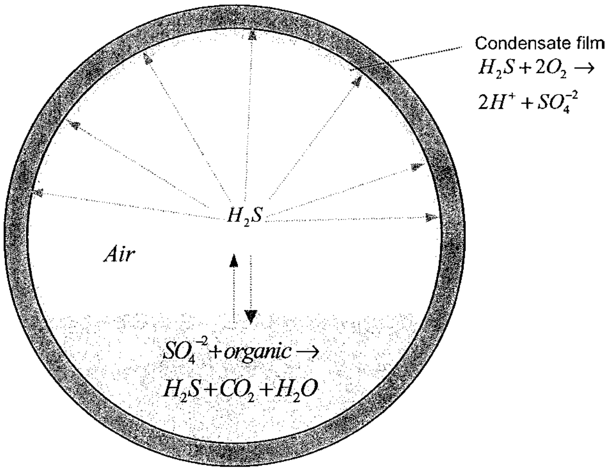


FIGURE 8 Reactions and transport processes involved in crown rot in wastewater collection systems. Note that the wastewater must be anaerobic for the sulfate reduction reactions to occur.

is a key factor determining the scaling properties of a water. Sulfate in water is derived chiefly from mineral sources but undergoes microbial reduction to sulfide (S^{-2}) in the absence of oxygen. A number of bacterial species use sulfide as an energy source, oxidizing it to sulfate. The oxidation process also releases two hydrogen ions per mole of sulfide oxidized, resulting in the formation of sulfuric acid. The sulfur cycle is a particular problem in wastewater collection systems where it is responsible for *crown rot*. In the crown rot process (see Fig. 8) anaerobic wastewater partially fills a pipe. Sulfate present in the wastewater serves as an electron acceptor for the oxidation of organic material and is reduced to hydrogen sulfide gas (H_2S). The H_2S partitions into the air and then into the aerobic condensate film on the pipe crown where it serves as an energy source for the sulfide oxidizing bacteria. The condensate is poorly buffered, and the pH will often drop to values between 1 and 2 as a result of the sulfide oxidation. Corrosion of the pipe crown results from the acidic conditions. Crown rot is a serious problem in collection systems, causing tens of millions of dollars per year in damage.

Nitrate is an important plant nutrient, as noted above, and is also the cause of methemoglobinemia. Normally the pH of the human stomach is low enough to prevent growth of nitrate reducing bacteria. However, in infants under six months old and adults with depressed gastric acid production the stomach pH is higher. Nitrite produced through denitrification reacts with the iron in hemoglobin to form methemoglobin, a form that does not carry oxygen. Accumulation of methemoglobin results in the disease methemoglobinemia, which can be fatal. Nitrate concentrations as nitrogen above 10 mg/L have been implicated in methemoglobinemia. Nitrate has been shown to be involved in the microbial production of carcinogenic nitrosamines and an apparent relationship between high nitrate levels and gastric cancer has been shown in epidemiological studies.

f. Minor ionic species in water. Virtually every mineral will be found in water at some location in some concentration. Ionic species found at lower concentration and with less frequency than the major ionic species are termed minor ionic species. In some cases, where solubility is very low, the term trace species is used. However, many of the minor ionic species are extremely important in defining water quality. Arsenic was used above as an example in the setting of drinking water standards. A number of minor ions found in water have human activities as the principal sources. Examples include phosphorus and the heavy metals cadmium, chromium, copper, lead, mercury, and zinc.

g. Water quality of rivers and groundwaters. Comparison of the winter and summer chemical characteristics of the Grand River, a small stream in Northeastern Ohio, provides some insights into the importance of inorganic ionic species and the relationship between chemical and biological characteristics of water. Overall, the Grand River is low in dissolved constituents and would be considered of high chemical quality. The data presented in Table VII are not completely consistent. For example, no information is given on the calcium concentrations, which are certainly significant, and the sum of the dissolved ion concentrations differs considerably from the TDS concentrations on January 16th.

Note that the pH and alkalinity of the Grand River are lower in January when the temperature is low. In natural waters, carbonate chemistry is the dominating factor in determining pH and alkalinity. When water temperature is low the equilibrium between carbon dioxide in the air and water shifts toward higher concentrations in the water. In water there is an equilibrium reaction between four forms of the carbonate species, dissolved carbon dioxide gas (CO_{2aq}), carbonic acid (H_2CO_3), bicarbonate ion (HCO_3^-) and carbonate ion (CO_3^{-2}), as indicated by

$$CO_{2aq} \rightleftarrows H_2CO_3 \rightleftarrows HCO_3^- \rightleftarrows CO_3^{-2} \qquad (9)$$

The CO_2 concentration is fixed by the gas–liquid equilibrium relationship. Thus the increased CO_2 concentration at lower temperatures results in greater quantities of carbonic acid and a drop in both pH and alkalinity.

The total Kjeldahl nitrogen is the sum of the organic and ammonia nitrogen present. In January, TKN is low and most of the nitrogen present is in the form of nitrate. In July, the TKN concentration has increased nearly sixfold and nearly all of it is organic. The organic nitrogen is

TABLE VII Characteristics of the Grand River near Harperville, Ohio, on January 16 and July 31, 1997

		Day	
Constituent	**Units**	**16 Jan.**	**31 Jul.**
Flow	m^3/sec	7.6	0.4
Total dissolved solids (TDS)	mg/L	128	228
Conductivity	μS/cm	291	384
Temperature (T)	°C	0	24
pH	—	7.5	8
Alkalinity (A)	meq/L	1.56	2.34
Magnesium, Mg^{+2}	mg/L	4	11
Sodium, Na^{+}	mg/L	8.4	19
Potassium, K^{+}	mg/L	1.5	2.9
Bicarbonate, HCO_3^{-}	mg/L	95	143
Sulfate, SO_4^{-2}	mg/L	22	27
Chloride, Cl^{-}	mg/L	19	28
Fluoride, Fl^{-}	mg/L	0.2	0.2
Ammonia, as N	mg/L	0.05	<0.02
Total Kjeldahl nitrogen (TKN)	mg/L	0.07	0.4
Nitrate and nitrite, as N	mg/L	0.41	0.06
Phosphorus, as P	mg/L	0.07	0.04
Iron, Fe^{+3}	mg/L	0.18	0.024
Manganese, Mn^{+2}	mg/L	0.033	0.073
Silica, as SiO_2	mg/L	3.1	1.7
Total organic carbon (TOC)	mg/L	5.7	7.7
Dissolved oxygen (DO)	mg/L	12	8.4

[Source: Water Resources Data, Water Year 1997, Vol. 2, United States Geological Survey.]

probably associated with increased phytoplankton concentrations, and this conclusion is supported by the increase in total organic carbon. Note also that both phosphorus and nitrate have decreased from January to July. Both results would be expected with increased phytoplankton growth.

Inorganic chemical characteristics of groundwaters differ significantly from surface waters, in part because of the longer time of contact with soil minerals (see Table III) and in part because of groundwater being isolated from the atmosphere. Characteristics of three shallow groundwaters from the southeastern coastal plain of the United States are given in Table VIII.

The three groundwaters described in Table VIII are typical of high-quality groundwaters with respect to low TDS concentrations and relatively low concentrations of the hardness ions, calcium and magnesium. The Ocala, FL, water would be classified as moderately hard and the Tampa, FL, and Virginia Beach, VA, waters would be classified as soft. Note the extremely low bicarbonate concentration in the Tampa, FL, water. Such low concentrations are very unusual. Alkalinity, the ability of a water to neutralize acid, is made up almost exclusively of bicarbonate in natural waters, and waters of low alkalinity tend to be acidic, as indicated by the pH of 5.4 for the Tampa water. Such waters are corrosive in pipes, and consequently the pH and alkalinity must be raised prior to discharge into distribution systems.

Hardness of the three waters of Table VIII is calculated in Table IX using Eq. (8). Waters are classified as soft, moderately hard, hard, and very hard for hardness values of <1, between 1 and 3, between 3 and 6, and greater than 6 meq/L, respectively. Thus the Ocala water is hard, the Tampa water is soft, and the Virginia Beach water is moderately hard.

2. Chemical Characteristics of Water—Organic Matter

Organic matter in water is derived from the decay of plant and animal matter, wastewater discharges, and storm and irrigation runoff containing agricultural chemicals. Natural organic matter (NOM) is composed largely of very complex, high molecular weight materials that fall under the general category of humic substances. Most natural waters contain between 1 and 10 mg/L of NOM. In some cases, the NOM gives water a brownish tint. Of greater importance are the reactions between NOM and the widely used disinfectant chlorine (Cl_2). Disinfection byproducts produced in these reactions include the carcinogenic trihalomethanes and haloacetic acids.

Organic matter in wastewater discharges can be classified as natural or synthetic. Natural organic compounds in wastewaters might be considered precursors of NOM in that they are composed of proteins, carbohydrates, and lipids derived from living organisms. Synthetic organic compounds include surfactants, pesticides, herbicides, solvents, and a variety of chemicals related to manufacturing of petrochemicals. Importance of organics in wastewaters is derived from their use as food sources by various organisms in the aquatic ecosystem and the toxicity or mutagenicity of many of the compounds. Wastewater discharges typically increase the supply of organics in receiving waters, with a resulting increase in microbial activity. The increase in activity at the lowest trophic level results in increased growth and activity at higher levels. Often species diversity decreases and chemical characteristics such as dissolved oxygen concentration that are important in overall water quality are modified. Toxic and mutagenic compounds that are discharged continuously often have impacts at extremely low concentrations. The MCLs placed on such compounds are often less than 5 μg/L and in some cases are less than 100 ng/L.

Organic molecules in wastewater discharges range widely in size with molecular weights from less than 20

TABLE VIII Characteristics of Three Groundwaters from the Southeastern Coastal Plain of the United States

Constituent	Units	Ocala, FL	Tampa, FL	Virginia Beach, VA
Temperature (T)	°C	23.9	23.5	19.5
Conductivity	μS/cm	380	195	321
Total dissolved solids (TDS)	mg/L	214	121	216
pH		6.9	5.4	7.1
Calcium, Ca^{+2}	mg/L	57.5	8.1	33
Magnesium, Mg^{+2}	mg/L	3.4	1.75	7.2
Sodium, Na^{+}	mg/L	8.1	10.1	21
Potassium, K^{+}	mg/L	1.15	2.80	2.3
Iron, Fe	mg/L	0.014	0.155	3.8
Manganese, Mn^{+2}	mg/L	0.0045	0.0065	0.110
Bicarbonate, HCO_3^-	mg/L	126	10	154
Sulfate, SO_4^{-2}	mg/L	10.3	17.5	10
Chloride, Cl^-	mg/L	4.15	18.5	22
Nitrate, NO_3^- as N	mg/L	1.2	0.06	0.05
Total phosphorus	mg/L	0.28	0.015	0.06
Bromide, Br^-	mg/L	0.075	0.030	0.14
Silica, as SiO_2	mg/L	7.5	5.0	42
Dissolved oxygen (DO)	mg/L	2.9	0.35	0.15
Total organic carbon (TOC)	mg/L	0.65	3.55	1.8

[Source: Berndt, M. P., *et al.* (1997). Water Resources Investigation Report 97-423m, USGS.]

to several million. Solubility of organic compounds varies greatly, as well. Hydrophilic compounds such as ethanol and acetic acid are completely miscible in water while polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) generally have solubilities of less than 1 mg/L. Many of the organic compounds that impact water quality are volatile, and this fact leads to the possibility of the air being a source as well as a sink for these materials. For example, partial combustion of gasoline results in the production of a number of compounds that are soluble in water. Thus surface waters near urban areas contain trace quantities of many compounds through exchange with the atmosphere.

Stormwater and agricultural runoff is generally low in organic matter concentration relative to municipal and industrial wastewater discharges. The organic materials present include NOM, pesticides, herbicides, and a wide variety of compounds that have been inadvertently or illegally applied to soil. Examples include waste oil and solvents from home shops, agricultural chemicals washed off soil and plants during irrigation or rains, and animal wastes from pets and livestock.

a. Measurement of organic concentration. Because of the wide range organic compounds present in most natural waters, two approaches to measurement are taken. Specific compounds may be monitored if they pose a particular hazard to the environment or in connection with a use such as drinking. Compounds on the USEPA list of priority pollutants are examples of specific compounds that may require monitoring. Measurement of individual compounds in the complex mixture of organics found in most natural waters is usually quite difficult and expensive. Often the compounds of interest are present at

TABLE IX Hardness of Groundwater Supplies in Ocala, FL, Tampa, FL, and Virginia Beach, VA

	Ca^{+2} concentration		Mg^{+2} concentration		
Water	mg/L	mmol/L	mg/L	mmol/L	Hardness meq/L
Ocala	57.5	1.44	3.7	0.15	3.18
Tampa	8.1	0.20	1.75	0.07	0.54
Virginia Beach	33	0.83	7.2	0.30	2.26

extremely low concentration, and sophisticated methods of extraction, concentration, identification, and quantification must be applied. An example is provided by gasoline which is often present in both groundwater and surface water at trace levels. Gasoline in groundwater is derived from leaking storage tanks while gasoline in surface waters is derived from boat engines and the atmosphere. Gasoline is composed of approximately 60 hydrocarbon compounds ranging in size from 5 to 12 carbons. Monitoring the individual compounds making up gasoline is difficult because of the generally low concentrations and the difficulty of separating out individual organic species among the wide range of compounds normally present.

In most cases, organic matter in water is measured using lumped parameters such as oxygen demand, total organic carbon, or total petroleum hydrocarbons. The most widely used measures are two forms of oxygen demand, the chemical oxygen demand (COD) and the biochemical oxygen demand (BOD) tests. Oxygen demand provides two types of information, a direct lumped measure of the transformable organic content based on stoichiometry and a measure of the quantity of oxygen that will be required in treatment or from the environment. Both types of information can be visualized by examination of Eq. (10) where the general conceptual form is given first and an example of sugar oxidation follows.

$$\eta_1[\text{Organics}] + \eta_2[\text{O}_2] \rightarrow \eta_3[\text{CO}_2] + \eta_4[\text{H}_2\text{O}] \tag{10a}$$

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O \tag{10b}$$

From Eq. (10), we can see that a stoichiometric relationship exists between the amount of organic material initially present and the amount of oxygen consumed. In the example of Eq. (10b), 6 mol of oxygen (192 g) are required to oxidize 1 mol of sugar (180 g). Thus 1.07 g of oxygen is equivalent to 1.0 g of organic material in this example. The amount of oxygen consumed in a reaction such as given in Eq. (10) provides a lumped measure of the amount of organic initially present. Because dissolved oxygen is the source of the oxygen used in aquatic reactions, the amount of oxygen consumed is of importance in estimating impacts on the aquatic ecosystem. Equation (10) is a model for abiotic oxidation of organics. Because we rarely know what compounds are present in water, a chemical procedure in which the amount of oxygen required to carry out the reaction of Eq. (10a) is determined. The oxygen consumed is called the COD and is reported in milligrams of O_2 required per liter.

Most organic materials found in water are stable. For example, sugar will not oxidize spontaneously when dissolved in water. However, many organic materials are metabolized by microorganisms, thus we are more often interested in the amount of oxygen consumed in biological oxidation of these materials, as depicted by

$$\begin{aligned}&\eta_1[\text{Organics}] + \eta_2[\text{O}_2] + \eta_3\left[\text{NH}_4^+\right] + \eta_4\left[\text{PO}_4^{-3}\right]\\ &\xrightarrow{\text{bacteria}} \eta_5[\text{NewCells}] + \eta_6[\text{CO}_2] + \eta_7[\text{H}_2\text{O}]\end{aligned} \tag{11}$$

The amount of oxygen required in biological reactions is less than in abiotic reactions such as Eq. (10) because a portion of the organic material is used for production of new cells. Additionally, nutrients such as nitrogen and phosphorus are required as for any process involving growth of living organisms. A BOD test has been developed to measure the oxygen consumed through biological oxidation. The BOD test environment is dilute and the reaction rates are relatively slow. The standard measurement of BOD is based on the amount of oxygen consumed in 5 days at 20°C. Thus the organic strength of a water is often given in mg BOD_5/L.

b. Recalcitrant organics. Many organic materials are either nonbiodegradable or very difficult to degrade biologically. Recalcitrance is caused by factors such as low solubility, branching of the molecular structure, characteristic bond structures such as ethers, and substitutions such as chloride and nitro radicals. Solvents such as trichloroethylene (TCE) and methylene chloride have become major water quality problems because of their resistance to biodegradation. TCE is a carcinogen and has a density of 1462 kg/m^3 and water solubility of 1100 mg/L. Methylene chloride is a suspected carcinogen with a density of 1326 kg/m^3 and a water solubility of 1320 mg/L. The compounds are found in groundwater as a result of improper handling of and leaking storage tanks. Both compounds tend to migrate to the water table as pure liquids, settle to the bottom of the aquifer because of their density, and gradually dissolve into the water column. Sorption onto soil organics retards the movement and results in residual soil contamination. TCE can be biologically degraded through a process known as co-metabolism which requires another organic compound to serve as both the energy and the carbon source for growth. Relatively few microorganisms have been found that will degrade TCE, as the compound is very stable in aquifers and soil. Methylene chloride does serve as a sole source of carbon and energy for a few bacteria. The fact that the compound has only one carbon is a limiting factor and the substitution of a chloride atom increases the recalcitrance. Thus methylene chloride is also quite stable in aquifers and soil.

3. Chemical Characteristics of Water—Gases

The principal gases of interest in defining water quality are oxygen (O_2), carbon dioxide (CO_2), and hydrogen sulfide (H_2S). The importance of oxygen relates to its low solubility and its requirement for the respiration of aerobic organisms from bacteria to fish. Carbon dioxide is important because the pH and alkalinity of natural waters is controlled by carbonate chemistry as noted above. Hydrogen sulfide results from biological processes in which sulfate is reduced to H_2S as discussed above. In addition to wastewater, swamps and wetlands often have considerable sulfate reduction.

Estimates of dissolved gas concentrations can be made using Henry's law (Eq. 12) which provides an accurate estimate of equilibrium conditions for dilute conditions.

$$mC_i = P_i \tag{12}$$

where m = Henry's law coefficient, L · atm/mg,
C_i = concentration of gas species i, mg/L, and
P_i = partial pressure of gas species i, atm.

The Henry's law constant is a function of temperature, as indicated in Table X.

Because oxygen is consumed during the biodegradation of organics and the oxidation of inorganics such as ammonia and sulfide, natural waters are generally not in equilibrium with the atmosphere. The rate of oxygen transfer is usually predicted using Eq. (12), which is based on the two-film model.

$$M_{O_2} = K_L a \left(C_S - C_{O_2} \right) \tag{13}$$

where M_{O_2} = mass transfer rate, $\text{mgO}_2/\text{L} \cdot \text{sec}$,
K_L = mass transfer rate coefficient, m/sec,
a = gas–liquid interface area per unit volume, m^{-1},
C_S = equilibrium concentration of oxygen with the atmosphere as predicted by Henry's law, mg/L, and
C_{O_2} = concentration of oxygen in the water, mg/L.

Determination of the mass transfer rate coefficient in natural water systems is very difficult. Reported values range from about 0.1 d^{-1} for small ponds and backwaters to about 1.1 d^{-1} for rapids and waterfalls. Equation (12) is generally used for estimating mass transfer rates between gas and water or volatilization rates from water to gas. The equation is not specific to oxygen.

TABLE X Henry's Law Coefficients of Several Gases ($\text{atm/L} \cdot \text{mg}^1$)

T, °C	O_2	CO_2	H_2S	N_2	Air[a]
0	0.0144	0.00030	0.00015	0.0340	0.0271
10	0.0171	0.00042	0.00021	0.0429	0.0344
20	0.0231	0.00058	0.00027	0.0519	0.0414
30	0.0267	0.00076	0.00034	0.0596	0.0590
40	0.0300	0.00099	0.00042	0.0669	0.0632

[Sources: Butler (1964), and Perry and Green (1984).]

[a] Based on molecular weight of 28.8 g/mol.

The Climate Monitoring and Diagnostics Laboratory (CMDL) of the National Oceanic and Atmospheric Administration (NOAA) has reported that the mean atmospheric concentrations of CO_2 increased from 330 to 365 ppm between 1975 and 2000. Using Henry's law, and assuming that raindrops come to equilibrium with the atmosphere (approximately correct), and an atmospheric temperature of 10°C, we would predict that the CO_{2aq} concentration of raindrops increased from 0.79 to 0.87 mg/L. The pH of rain is usually controlled by the equilibrium with atmospheric carbon dioxide. The corresponding pH values based on Eq. (9) are 5.60 and 5.58. Acid rain resulting from absorption of oxides of sulfur and nitrogen may have substantially lower pH values. In soils, acid rain may increase the solubility of minerals with the result that toxic ions are introduced into waterways or that mineral nutrients such as phosphorus are stripped from the soil. Lakes are often poorly buffered, and continued influx of acidic water results in pH decreases.

D. Biological Characteristics of Water

Biological characteristics of water are related to the resident population of aquatic organisms. Most of the properties of concern result from the presence of microorganisms. Larger organisms, such as fish, rooted plants, and amphibians such as frogs and alligators require certain types of water quality but do little to change water quality themselves. Transmission of disease by pathogenic organisms is the most important biological impact of water quality, at least from a short-term human perspective. Microoroganisms are important factors in defining water quality and in modifying both physical and chemical characteristics of water. Daily pH and dissolved oxygen cycles, seasonal heating and cooling, and long-term geological changes are impacted by microbial growth.

1. Classifications of Microorganisms in Water

The principal groups of microorganisms found in water are bacteria, algae, protozoa, and invertebrate animals. Bacteria are single-celled organisms, characterized by the procaryotic cell structure, which is simpler than that of the other microorganisms, and are characteristically smaller, ranging in size from about 0.2 to about 3 μm. Bacteria are often referred to as primary degraders because they carry out degradation of organic detritus in water and serve as food sources for larger organisms such as the protozoa and invertebrates. Algae, like bacteria, are single celled but they have eucaryotic cell structure similar to plants and animals. Most importantly, algae are photosynthetic and autotrophic, that is, they obtain

TABLE XI Common Water-Related Diseases

Category and method of contamination	Disease	Causative agent	Symptoms
Waterborne: ingesting contaminated water	Amebiasis	Protozoan (*Endameba histolytica*)	Prolonged diarrhea with bleeding, abscesses of the liver and small intestine.
	Shigellosis (dysentery)	Bacteria (*Shigella*, several species)	Severe diarrhea
	Cholera	Bacteria (*Vibrio cholerae*)	Extremely heavy diarrhea, dehydration
	Cryptosporidiosis	Protozoan	
	Gasteroenteritis	Virus (enteroviruses, parvovirus, rotovirus)	Mild to severe diarrhea
	Giardiasis	Protozoan (*Giardia lamblia*)	Mild to severe diarrhea, nausea, indigestion, flatulence
	Hepatitis	Virus (hepatitis A virus)	Jaundice, fever
	Leptospirosis	Bacteria (*Leptospira interrogans*)	Jaundice, fever, renal failure
	Salmonellosis	Bacteria (all species of *Salmonella*)	Fever, nausea, diarrhea
	Typhoid fever	Bacteria (*Salmonella typhosa*)	High fever, diarrhea, ulceration of the small intestine
Water-washed: washing with contaminated water	Shigellosis (dysentery)	Bacteria (*Shigella, several species*)	Severe diarrhea
	Scabies	Mite	Skin ulcers
	Trachoma	Virus	Eye inflammation, blindness
Water-based diseases	Filariasis	Worms (*Wuchereria bancrofti, Brugia malayi*)	Blocking of lymphnodes, swelling and permanent tissue damage
	Guinea worm	Worm (*Dracunculus medinensis*)	Larvae are ingested when water flea host is swallowed. Worm grows for one year and then exits through skin lesion.
	Onchocerciasis	Worm (*Onchocerca volvulus*)	Disease is transmitted by blackflies. Human hosts have chronic eye and skin conditions, blindness.
	Schistosomiasis	Worm (*Schistosoma mansoni, japonicum, haematobium*)	Intestinal and venus damage, blood loss in bladder.

[Sources: Hawkes (1971); Salvatto (1982). Madigan, *et al.* (2000).]

their energy from light and use CO_2 for a carbon source. Protozoa and invertebrate animals, such as worms, rotifers, and crustaceans, graze on the bacteria and algae and in turn serve as food for larger organisms.

Viruses are nonliving particles composed of a strand of either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) and a protective protein coat and are also found in water. Viruses are too small and too simple to contain the molecular machinery necessary to replicate on their own, which is the essential definition of life, but can infect an organism and force it to carry out replication of the viral particles. (Perhaps viruses were the inspiration for Jack Finney, author of *Invasion of the Body Snachers*.) Viruses are extremely host specific, often infecting only one strain in a species. Thus human viruses, which all cause a disease, will only be found in water as the result of a discharge from people. A few viruses cause diseases in more than one species. Examples are *Hantavirus* and *Ebola virus*, neither of which is a waterborne infection.

2. Pathogenic Organisms in Water

Pathogenic organisms in water include bacteria, protozoans, worms, and viruses. A list of common diseases associated with water is given in Table XI. Many of the diseases listed in Table XI are transmitted in other ways, such as by food preparation or through person to person contact.

3. Impacts of Microbial Growth on Water Quality

Environmentally, the largest impact of microorganisms is on the development of phytoplankton, the microscopic communities of bacteria, algae, and plants that grow near the water surface. Phytoplankton form the lowest level of the food chain and the largest mass of aquatic organisms. The initial state of most natural water systems is one of minimal mineral and organic content. Very little life can be supported in such waters. Heterotrophic organisms, such as bacteria, have no organic matter to supply energy and carbon, and autotrophic organisms have little in the way of energy sources other than sunlight. Thus photosynthetic organisms, the *Cyanobacteria* and the algae, are generally the first organisms to grow in these *oligotrophic* waters. Over long periods of time the photosynthetic organisms accumulate, many die and provide food sources for heterotrophic organisms, and the mineral components are recycled. Nutrients and organic matter are washed into the water, and the overall accumulation becomes a richer habitat. Diversity of life increases and species present may change due to competition. This process, called *eutrophication*, will eventually lead to heavy phytoplankton growth, floating *algal blooms*, and increased deposition of organic detritus. Species diversity will pass through a maximum. When extreme eutrophic conditions develop, oxygen cycling occurs due to net oxygen production during daylight hours and net oxygen consumption during nighttime hours. Consumption of CO_2 by photosynthetic organisms during the day is greater than production by heterotrophic organisms, and the pH tends to rise. During the night, when photosynthesis shuts down, a net CO_2 production occurs that results in pH decreases. The combined oxygen and pH cycling, illustrated for an extreme case in Fig. 9, results in a habitat supporting some types of macroscopic organisms and excluding others. Under extreme conditions, the dissolved oxygen concentration may fall to near zero. Usually such extreme conditions are local, that is, near the bottom or in a stagnant backwater area with little mixing. Only a few types of macroscopic organisms will survive in such a habitat. The high pH values that may occur in highly eutrophic environments limit the species diversity and also change the water chemistry considerably. Over millennia, naturally eutrophic waters gradually fill up with highly organic soils and become meadows. Human activities have caused rapid increases in eutrophication of most natural aquatic habitats on the earth. Well advertised cases such as Lake Tahoe on the California–Nevada border have received a great deal of public attention. However, few lakes can be considered safe from accelerated eutrophication due to societal activities and near-shore waters are also at risk. For example, the habitat around coral reefs throughout the world has been damaged in ways that are not entirely understood, but the damage is certainly related to discharges from urban and agricultural areas. Approximately 10% of coral reefs throughout the world were seriously damaged in the last half of the 20th century.

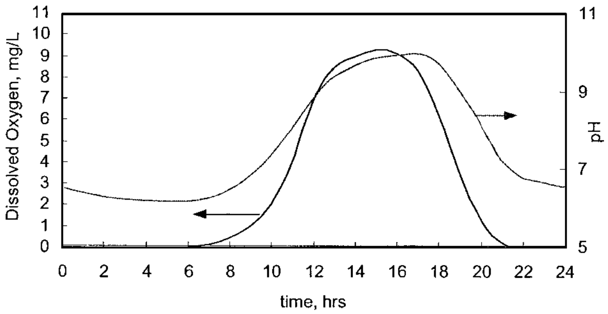


FIGURE 9 Dissolved oxygen and pH cycling in highly eutrophic waters. Note that the oxygen concentration may fall to zero under extreme conditions.

Microorganisms are also factors in the development of taste, odors, and colors in water, and the biofouling of heat transfer surfaces in cooling systems. Both *Cyanobacteria* and algae release organic compounds that cause tastes and odors in water. Problems usually occur in late summer when drinking water supplies are warmer. Microbial fouling of heat transfer surfaces is a very serious problem industrially. The principal related water quality problem results from the chemicals used for controlling fouling.

IV. MODIFICATION OF WATER QUALITY

Water is pure only at the moment of condensation. From that time forward materials in the environment interact with each drop of water. As shown in Table IV, rural and urban rainfall contain a surprising array of chemicals. The admonition *don't eat yellow snow* might be considered conservative in that water quality changes continuously as each droplet moves through the hydrologic cycle, and we never really know what may be present. Not all changes in water quality are deleterious, but the general result of contact with other components of the environment results in water becoming less useful. Municipal, industrial, and agricultural use of water results in severe deterioration of water quality that requires treatment to remove contaminants before discharge.

A. Deterioration of Water Quality

The deterioration of water quality during passage through the hydrologic cycle can be classified as natural or anthropogenic. Natural deterioration results from contact with mineral deposits, vegetation and animal wastes, atmospheric deposition, heating and cooling, and growth or addition of microorganisms. Anthropogenic deterioration results from the wide range of human activities that result in discharges to water courses and groundwater. Obviously there is some ambiguity and overlap in the definitions. However, two distinct differences between natural and anthropogenic deterioration should be noted: the ability or desirability of control and, in the case of anthropogenic deterioration, the involvement of synthetic or *xenobiotic* compounds.

Atmospheric deposition resulting from forest fires is natural, while deposition from agricultural burning is anthropogenic. The results are essentially the same, but the ability to control natural processes is far less than our ability to control anthropogenic processes. Natural deterioration of water quality would occur even if humans did not exist. We can modify natural deterioration in a number of ways, but the long-term impact is unclear. For example, managing forests to prevent fires may result in increasing the period between fires but larger conflagrations when fires do occur. Considering longer time scales, the resulting erosion and atmospheric deposition may result in greater rates of water quality deterioration than if fire prevention practices were minimal.

In evaluating natural water quality deterioration we suffer from the lack of a clear baseline. There are few pristine environments that can be used as a reference. For example, hikers must contend with *Giardia lamblia* in even the most remote places. How long this has been the case is unknown, but there is little reason to believe that the organism has not been historically resident in the wild animal population. Suspended solids in streams are a natural result of erosion caused by floods. Beaches are formed by sand carried downstream in floods. Flood control practices have both decreased erosion and flooding and trapped solids in reservoirs, and sand losses from beaches becomes a problem in places such as southern California.

Water quality of streams changes during passage from the source to the final discharge point, as indicated in Table XII. Each of the water quality parameters listed in Table XII changes considerably on a seasonal basis as well as on a spatial basis. Some knowledge of river geography is helpful in interpreting the data. For example, following the temperature of the Colorado River downstream from Cisco, Utah, we see a temperature of 22.5°C which is typical of a large, unshaded stream in late summer. The much lower temperature of 11°C at Lees Ferry, Arizona, would suggest that water releases from Lake Powell were from the lower depths. Similarly, the temperature of 13.5°C downstream of Lake Mead would have been from some depth. Finally, the temperature near the international border at Imperial Dam, Arizona, represents temperatures in the desert. The dissolved oxygen concentrations should be considered both in absolute terms and in terms of relative saturation. At Cisco, Utah, the river is 81% saturated with oxygen, while at Lees Ferry the water is at 67% saturation, which would be expected of water from some depth in a reservoir. We note that the water coming from Armistad Reservoir on the Rio Grande River is substantially depleted of oxygen and only 19% saturated. Total dissolved solids are expected to increase as water flows downstream, and this is generally the case for all for rivers in Table XII. Decreases shown result from the contributions of large tributaries or averaging of TDS values over time in reservoirs.

The importance of anthropogenic deterioration of water quality has been recognized for over 150 years. However, the accelerated industrial development of the past 50 years together with an increasing willingness to investigate and an increased ability to measure impacts of discharges has greatly changed society's views on water quality management. We find synthetic chemicals in significant concentrations throughout the natural water system. Additionally, we have learned that very low concentrations of chemicals may cause serious damage to the ecosystem. In the late 1980s, selenium was found to be the source of deformed birds and decreased birthrates of resident waterfowl in Kesterson Reservoir in the San Joaquin Valley of California. The source of the selenium was agricultural drainage from the Panoche Fan, a region on the west side of the San Joaquin Valley in which the soils had been washed from a high selenium region of the Coastal Range. At present, concentrations as low as 2 μg/L are believed to be a problem for some bird species. Analogously to selenium, heavy metals, such as copper, are accumulated in the food chain and very low concentrations can result in serious damage to organisms depending on the aquatic habitat. Sources of copper in streams include discharges from mines, automobile break linings, and pipes. The maximum contaminant limit goal for copper set by the USEPA for drinking water is 1.3 mg/L, while the recommended receiving water quality criterion for dissolved copper in freshwater is 13 μg/L because of the toxicity to certain aquatic species. The result is that many drinking waters may not meet discharge standards to sensitive streams.

More recently, there has been the discovery that many natural and synthetic compounds mimic the actions of hormones and can disrupt reproductive patterns. The most widely noted phenomenon is the *feminization* of male fish, but certainly the problem extends throughout the food chain. Initially, most endocrine disrupters were believed to be compounds with some structural similarity to natural hormones. More recent work has indicated that a wide range of materials, including heavy metals such as arsenic and mercury, may be involved.

B. Water Quality Modification

Water quality is modified to meet the requirements of specific uses. In most cases, we think of water quality modification in terms of treatment of surface or groundwaters for drinking or industrial purposes or of treatment of wastewaters to meet discharge requirements; and discussion of these topics is the principal purpose of this section. However, water quality can be selectively changed on a large scale through use of reservoirs and other constructed systems, as noted above. Temperature control is one of the most desirable potential benefits of large

TABLE XII Water Quality of Four Major U.S. Rivers in Late Summer 1999

	Colorado River				Rio Grande River			
	Cisco, UT	**Lees Ferry, AZ**	**Below Hoover Dam, AZ**	**Imperial Dam, AZ**	**El Paso, TX**	**Presidio, TX**	**Below Armistad Reservoir, TX**	**Brownsville, TX**
Date	8/30/99	8/17/99	8/25/99	8/25/99	8/24/99	8/31/99	9/10/99	8/24/99
Q, m^3/sec	170	560	462	250	34	6	21	6
T, °C	22.5	11	13.5	30	24.5	23	22.5	27
DO, mg/L	7	7.4	7.5	7.3	6.4	5.6	1.6	4.8
pH	8.3	8	8.1	8.2	8.3	7.9	7.2	7.7
SS	3303	2	8	146	724	582	622	39
TKN, mg/L	ND	0.2	0.2	0.3	0.8	0.7	0.2	0.4
P, mg/L	1.86	ND	0	0.02	0.37	0.08	0.01	0.15
DOC, mg/L	3.4	3.1	2.7	2.7	3.8	4.8	2.3	3.9
Turb, NTU	910	0.3	0.4	1.6	90	150	1.8	22
TDS, mg/L	636	438	563	688	547	1512	622	628

	Missouri River				Mississippi River			
	Culbertson, MT	**Pierre, SD**	**Omaha, NE**	**Hermann, MS**	**Jordan, MN**	**Clinton, IO**	**Thebes, IL**	**St. Francisville, LA**
Date	9/7/99	8/17/99	9/14/99	9/14/99	9/3/99	9/9/99	9/1/99	9/14/99
Q, m^3/sec	287	939	1492	2003	65	1149	4511	5050
T, °C	17	21	23.5	23.1	21.6	22	26.9	28.2
DO, mg/L	8.1	8	7.8	8.1	8.2	7.5	8.2	7.5
pH	8.4	8.4	8.4	7.8	8.2	8	8	7.8
SS	269	1	267	144	218	42	102	70
TKN, mg/L	0.3	0.3	0.5	0.5	1.3	0.7	0.8	0.5
P, mg/L	0.14	ND	0.19	0.19	0.23	0.16	0.24	0.18
DOC, mg/L	2.9	3.3	3.7	3.9	6.2	7.8	5.6	3.7
Turb, NTU	42	0.3	15	27	—	10	16	32
TDS, mg/L	352	509	545	514	23	221	383	311

[Source: US Geological Survey, National Stream Quality Accounting Network, *www.water.usgs.gov/nasqan*.]

impoundments. Drawing water off at selected depths allows improved support of fish migration, germination of seeds, and protection of ecosystems. Reservoirs also act as sediment traps. Downstream waters generally have low turbidity. A negative impact is that gravels trapped in reservoirs are required for breeding areas of many species and, as noted, loss of sand necessary for beach maintenance.

1. Water Treatment

Treatment of water to modify water quality nearly always requires a *treatment train*, that is, a sequence of steps in which various constituents are removed to produce a product water that meets specific requirements. Perhaps the simplest case is treatment of surface waters for municipal use. Municipal use requirements are dominated by the aesthetic and health requirements for drinking water. The most important contaminants in most surface waters are particulate matter and microorganisms. Other contaminants of concern include metal ions, surfactants, solvents, petroleum products, and agricultural chemicals.

a. Removal of suspended solids. Particulate matter in surface waters ranges in size from colloidal to large rocks. The rocks are removed quite easily, but colloid removal requires four distinct steps: coagulation to destabilize the particles, flocculation to promote particle growth, and sedimentation and filtration for the actual particle removal. Colloids are kept in suspension by Brownian motion. Their small size makes filtration extremely expensive. Coagulation is a process of particle destabilization by adding positive *counterions* to neutralize the net negative surface charge of the mostly clay colloids. A second approach is to form a precipitate in which the colloidal particles are enmeshed. The most common coagulant used is alum ($Al_2SO_4 \cdot 18H_2O$). Other coagulants in use include ferric chloride ($FeCl_3$), lime ($CaOH_2$), and organic polymers. In flocculation, the destabilized colloids are subjected to gentle mixing that results in particle collisions. On collision, the particles are bound by van der Waals forces and gradually grow to a size where they will settle from the water under quiescent conditions. Similar processes occur in natural waters and account for the accumulation of muds in many estuaries.

In sedimentation, particles settle to the bottom of quiescent tanks under gravitational forces, and treated water is removed at the surface. Typical hydraulic loading rates, based on tank surface area, are 35 to 45 $m^3/m^2 \cdot d$. Particles reaching the bottom are scraped to a sump and the resulting slurry, or sludge, is pumped to a thickening and drying facility. The product water from sedimentation usually has a turbidity of a few NTU. Granular media filtration is used to remove remaining particles down to a size of approximately 1 μm. Following filtration the turbidity should be less than 1 NTU and the bacterial counts are typically reduced by 99% (commonly referred to as two logs of removal, based on base 10 logarithms).

b. Disinfection.

Microorganisms in water are killed using a variety of disinfection processes. Because monitoring each human pathogen is logistically impossible, indicators of the possible presence of pathogens are monitored. The most commonly used indicators are fecal coliforms, bacteria known to make up a large fraction of the flora in the gut of warm-blooded animals. Presence of these organisms is taken as evidence of recent contamination by feces from warm-blooded animals and the possible presence of human pathogens. Fecal coliforms are commonly present in surface waters with typical *most probable numbers*, or MPNs, of 1000 to 10,000 per 100 mL of sample. Following disinfection, the MPN values should average less than one per 100 mL. One of the reasons that turbidity is an important parameter in drinking water quality is that bacteria attach to particles and are shielded from disinfection.

Chlorine gas is the most widely used disinfectant. Other forms of chlorine, such as chlorine dioxide, are also used. Monochloramine (NH_2Cl) and dichloramine ($NHCl_2$) are used to a limited extent. Ozone is used extensively in Europe but less so in the United States. Ultraviolet radiation is becoming the disinfection method of choice for filtered wastewaters because fewer problematic disinfection by-products are formed.

A disadvantage of chlorine is that reactions with organic matter, particularly humic materials, produce carcinogenic trihalomethanes such as chloroform and bromoform. Advantages of chlorine are that the compound is very effective for a wide range of organisms and that a residual concentration of chlorine can be maintained in the distribution system. The contact time necessary varies with the organism of concern and the chlorine concentration. Fecal coliforms are relatively sensitive to chlorine, while the protozoan *Cryptosporidium parvum* is quite resistant. Thus the use of fecal coliforms as indicator organisms has its limits.

New approaches to monitoring biological quality of drinking water will be developed using techniques of molecular biology. At present, the methods require too much time and are too expensive for regular monitoring. Additionally, there is not a generally accepted measure of probable presence of pathogens at the present time.

c. Metal ion removal.

Two types of metal ion problems exist in water: the presence of ions that cause aesthetic problems and the presence of ions that are toxic. Aesthetic problems include hardness and staining. Hardness is due to the presence of multivalent cations, principally calcium and magnesium. These ions react with soap to form a precipitate and cause scaling in hot-water pipes. Removal of hardness can be accomplished by precipitation of the metals as calcium carbonate and magnesium hydroxide or by ion exchange. Both methods are relatively inexpensive. The principal cause of staining is the presence of reduced iron and manganese in groundwater. Concentrations of iron and manganese are rarely more than a few milligrams per liter and both ions are rapidly oxidized and precipitated by contact with molecular oxygen.

Toxic metals in water include arsenic, barium, cadmium, hexavalent chromium (Cr^{+6}), lead, mercury, selenium, and silver. The most common sources are natural mineral formations, and consequently, toxic metals are most commonly associated with groundwater. Removal using ion exchange processes or by reverse osmosis is most common.

d. Organic matter removal.

Organic material is rarely present in large quantities in natural waters, either surface or ground. The principal problems associated with organic matter in drinking water sources is the presence of surfactants, solvents, petroleum products, and agricultural chemicals. Allowable concentrations of these materials in finished water is generally less than 100 μg/L, and in many cases, the maximum concentration limits are in the ng/L range. Sorption onto activated carbon is the most widely applied treatment method. Some contaminants, such as the components of gasoline, are volatile and can be removed using air stripping. Reverse osmosis is also used to remove organic matter from water in some cases.

2. Wastewater Treatment

Two principal objectives form the basis for wastewater treatment: protection of public health and protection of the aquatic environment. Wastewater treatment systems are similar to water treatment systems in many respects. In most cases, a treatment train must be used, and removal of suspended matter is very important. Toxic materials, including carcinogens, are sometimes present, and maximum concentration limits in the discharge for these materials are often very low. Two major differences between water and wastewater treatment are that biodegradable organic material in municipal and many industrial wastewaters is often present in relatively high concentrations and that nutrients such as nitrogen and phosphorus that may cause eutrophication of receiving waters may be present. A simplified schematic of a typical wastewater treatment system is shown in Fig. 10.

Because of the diverse nature of the organic matter in wastewaters, a lumped, surrogate parameter must be used to monitor the concentration. The parameter of choice, as noted above, is oxygen demand because oxygen is consumed when biodegradation takes place, and the relationship between organic matter initially present and oxygen consumed on degradation should be stoichiometric. Additionally, dissolved oxygen is very important in maintaining aquatic ecosystems. Using oxygen demand to measure organic concentration allows estimation of the damage that might be caused by discharge of a particular wastewater.

a. Particulate matter removal.

Particulate matter in wastewater is removed by sedimentation, and, if necessary, filtration. Because turbidity standards are rarely placed on discharges, coagulation and flocculation are usually not required. In municipal wastewaters the particulate matter is principally organic matter, although some sand and clay are normally present. Biodegradable particulate matter that is removed by physical processes undergoes anaerobic biological treatment in a separate process stream. The purposes of biological treatment of the solids are to reduce the volume of sludge and to stabilize the organic solids into nonputrescible materials. A benefit of biological solids treatment is the production of methane gas that can be used for heating or power generation.

b. Biological wastewater treatment.

In biological wastewater treatment organic material is oxidized by microbial communities maintained in either a *suspended growth* or an *attached growth* reactor. Both types of reactors make use of mixed cultures, that is, cultures including a number of microbial species. Such systems are self-optimizing in that the most competitive organisms for a particular set of environmental conditions dominate the culture. If the environmental conditions change (e.g., if temperature rises or falls) the population make-up will shift in species dominance. Bacteria are the dominant group of microorganisms involved in biological wastewater treatment. Higher organisms, such as protozoa, fungi, and invertebrates are present as components of the community but have mostly indirect impacts on the process performance.

The general concept of biological treatment processes can be understood by examination of Eq. (11), repeated below:

$$\eta_1[\text{Organics}] + \eta_2[\text{O}_2] + \eta_3\left[\text{NH}_4^+\right] + \eta_4\left[\text{PO}_4^{-3}\right] \xrightarrow{\text{bacteria}} \eta_5[\text{NewCells}] + \eta_6[\text{CO}_2] + \eta_7[\text{H}_2\text{O}] \quad (11)$$

Biodegradation of organic material in wastewater has two principal products, new cells and CO_2. The new cells are a waste product and must be removed from the treated water because they are a form of reduced carbon compound and will impose a load on the receiving water.

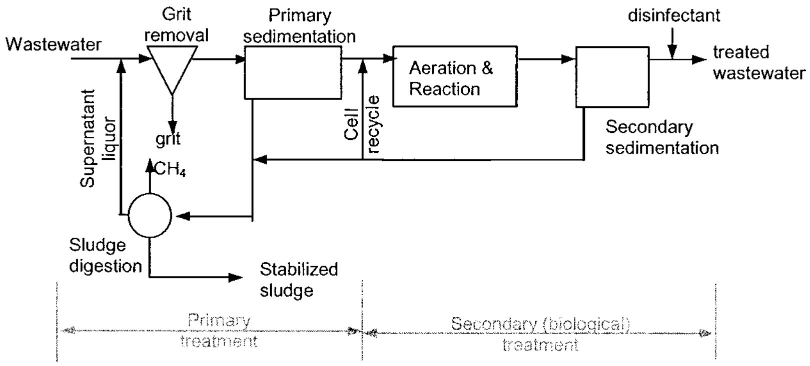


FIGURE 10 Schematic diagram of typical wastewater treatment system. The secondary system schematic diagram is for an activated sludge process, a type of suspended growth process. Attached growth systems recycle the effluent stream rather than the settled solids stream.

Carbon dioxide is the most oxidized form of carbon and imposes no load on receiving waters. Note that oxygen, nitrogen, and phosphorus are required by the bacteria carrying out the reaction. Oxygen is the terminal electron acceptor for the reaction process, and nitrogen and phosphorous are components of the bacterial cells produced. Other *trace* elements are required also, but these are normally present in wastewater in satisfactory quantities. Municipal wastewaters usually have excess concentrations of nitrogen and phosphorus and thus these materials need not be added during treatment. Many industrial wastes are nitrogen or phosphorus limited, and one or both must be added.

In suspended growth, or *activated sludge*, processes the microbial culture is grown as flocculent particles, or flocs, and kept in suspension by the action of the aeration system. Two types of aeration systems are in common use: diffused aeration and surface aeration. In diffused aeration, air, or in some cases pure oxygen, is injected at the bottom of the reactor as small bubbles. Oxygen is transferred from the bubbles into the liquid phase where it is used by the microorganisms. The rising bubbles mix the liquid and keep the organisms in suspension. The *mixed liquor* composed of wastewater and microbial community leaves the reactor and flows to a secondary sedimentation tank where the microbial flocs and the treated wastewater are separated. A portion of the settled floc equal to that grown in the cycle is discarded as waste sludge to the solids treatment stream, and the rest is returned to the aerated reactor for reuse.

In attached growth processes, typified by trickling filters, the microbial community is grown on inert packing such as rock or shaped polystyrene. Wastewater is sprinkled over the top of the packing and flows down and through the reactor as a thin liquid film. Air flows in the interstices of the packing, and oxygen is transferred from the air into the liquid film and then into the biofilm where the degradation reactions occur. Microbial growth, as indicated in Eq. (11), results in increasing biofilm depth. Eventually, the weight of the biofilm results in sloughing of small portions of the biofilm. The sloughed biofilm particles are removed in the secondary sedimentation tank and treated in the anaerobic sludge digester.

c. Performance of biological treatment systems. Biological treatment is expected to remove 85% of the entering organic matter measured as BOD and suspended solids. Most biological processes exceed this level of performance because virtually all of the biodegradable material entering is at least partially degraded. Additionally, most of the BOD leaving the secondary sedimentation tank consists of particulate material that has not settled. Influent organic concentrations are usually in the 200 to 350 mg BOD_5/L range while effluent BOD_5 and suspended solids concentrations from a well operating plant should be below 20 mg/L and are often below 10 mg/L.

Biological processes are subject to upset due to toxic contaminants, overloading, or limiting nutrients. Upset conditions are defined by a poor effluent quality and a change in the microbial culture that results in poor settling and high effluent suspended solids. Often the changes in the microbial culture are difficult to correct, and returning the process to good operation may take several weeks.

d. Nutrient removal. Un-ionized ammonia nitrogen is very toxic to fish and other forms of aquatic life. For this reason nitrification, the oxidation of ammonia nitrogen to nitrate, is a common requirement placed on wastewater discharges to surface waters. Excess nitrogen and phosphorus in wastewaters is available to organisms in receiving waters and often is the source of algal blooms in lakes, reservoirs, and rivers. Additionally, nitrate in water presents a health problem for infants, as described above. Thus, in some cases removal of nitrogen from wastewater discharges is required.

Nitrification can be achieved using a conventional biological treatment process operated at a low loading rate. Nitrate, the product of nitrification, serves as an electron acceptor, similar to the role of oxygen, but the reduced product is molecular nitrogen. By arranging the biological treatment sequences appropriately, nitrate nitrogen produced through nitrification can be reduced to molecular nitrogen through *denitrification*. This process sequence is in extensive use worldwide.

Phosphorus can be removed by precipitation with lime or iron. However, biological removal of phosphorus is increasingly utilized. In biological removal of phosphorus, the process sequence is set up to encourage the growth of polyphosphate accumulating organisms, bacteria that store polyphosphate in very high concentrations. These organisms are harvested, which effectively removes excess phosphorus from the treated wastewater.

e. Tertiary wastewater treatment. Primary and secondary treatments are focused on the removal of particles and biodegradable organic matter. Metal ions and many synthetic organic compounds pass through these processes virtually intact. Generally the concentrations of these *recalcitrant* materials are quite low. However, many of the materials are major environmental hazards even at extremely low concentrations, either directly or because they accumulate in the food chain. Appropriate methods of treatment to remove these materials include ion exchange, precipitation, adsorption on activated carbon, microfiltration, nanofiltration, and reverse osmosis.

C. Stormwater Treatment

Stormwater presents a difficult treatment problem because of the high volume flows generated over short time periods. In the Southwestern United States, a high fraction of the annual rainfall may occur in 10 to 15 storm events. Contaminants that collect on soil and paved surfaces over a periods of weeks and months are washed off in a short time frame, and concentrations in the initial flow can be quite high. Much of the contaminant load is soil related. However, animal droppings, litter, dust from break linings, herbicides, pesticides, solvents, spilled petroleum products, sanitary sewer overflows, and similar materials are also present. Dust from break linings includes copper which is hazardous to aquatic communities at very low concentrations. Animal droppings contain microorganisms and viruses that may cause human diseases. All of the spills and accidents that occur in urban life leave residues of materials that are potentially dangerous to humans or the aquatic environment. Much of the stormwater runoff is classified as non-point source flow. That is, the point of entry into water channels is distributed over the channel which makes treatment extremely difficult. Management of stormwater has until recently been focused on flood control and the focus has been on moving the runoff through urban areas as rapidly as possible. For these reasons, management of stormwater quality is a problem that has not been solved. Construction of facilities to treat large volumes of water that occur periodically to a very high standard will not be easy. At present the favored procedure in arid regions is to pond the water and percolate it to groundwater aquifers. However, this approach will deprive near-shore ecosystems of needed freshwater and potentially contaminate aquifers. Solutions to the stormwater quality management problem will need to be developed but the appropriate direction is not clear at this time.

V. REQUIREMENTS FOR WATER QUANTITY AND WATER QUALITY

Water use can be classified as municipal, environmental, agricultural, and industrial. Each use of water has both quantity and quality requirements. These requirements provide a framework for selection of municipal water supplies, setting recreational water, aquatic environment standards and wastewater discharge, operation of reservoirs, and selection of crops and other plantings.

A. Municipal Water Supplies

Municipal water is used for irrigation, car washing, dishwashing, toilet flushing, bathing, cooking, drinking, and a range of other uses. However, cooking and drinking are the dominant factors in setting standards for municipal water quality. Protection of human health will continue to be the dominating factor in selection of municipal water supplies and in the management of treatment and distribution systems. Water quantities required for municipal use are highly variable but average a little under 400 L/person · d in the United States. We are advised to drink about 2 L/d, and therefore the quality requirements for a tiny fraction of the municipal water supply set the standard for the entire volume.

1. Drinking Water Quality

Drinking water quality requirements have been focused on elimination of pathogenic organisms for over 100 years. Introduction of filtration in 1906 and chlorine disinfection in 1913 resulted in a very rapid drop in the incidence of waterborne diseases in the United States. The use of coliforms as an indicator of the possible presence of human contamination has clearly been successful in reducing the incidence of waterborne disease. At present, the standard for treated drinking water is that less than 1% of the samples will be positive for coliform organisms. In most years there are 20 to 30 incidents in which at least two people become ill from the same water source, and the total number of illnesses is annually 2000 to 3000. An exception occurred in March 1993 in Milwaukee, Wisconsin, when 400,000 people were infected with Cryptosporidium, an organism with considerably more resistance to chlorine than coliforms. Cryptosporidium was virtually unknown prior to 1990 and is completely absent from textbooks on water quality prior to that date. Thus the utility of the coliform standard must be questioned, although an adequate substitute is yet to be found.

Drinking water supplies must also be safe in terms of chemical constituents. Of particular concern are heavy metals, chlorinated hydrocarbons, and agricultural chemicals. Periodic monitoring of public water supplies is required to determine if contaminants determined by the USEPA to be priority pollutants are present above their respective MCLs.

Aesthetic quality of drinking water is extremely important also. Turbidity requirements for public health result in a clear, attractive water in most cases. However, high dissolved solids concentrations, while safe, result in water with a slightly alkaline taste. Water containing several milligrams per liter of humic material often has a brownish tint. Some groundwaters contain small concentrations of sulfides that result in objectionable odors. Such aesthetic characteristics usually result in large bottled water sales.

2. Quality Desired for Domestic Use

Desired characteristics of water used for nondrinking domestic use include low hardness, a slightly alkaline pH to prevent corrosion in the distribution system, and an absence of ions that are toxic to plants. An example of a toxic ion is boron which is toxic to citrus at concentrations below 1 mg/L.

3. Municipal Demand and Water Conservation

Municipal water consumption can be reduced considerably by introduction of low water use appliances and careful use of water for irrigation. Low-flow shower heads, low-flush volume toilets, low water use dish and clothes washers are extremely effective and have minimal if any detectable differences from conventional units for the user. Water consumption can be decreased by approximately 50% with little difficulty. Where irrigation is required for lawns, gardens, and plantings, there is a strong tendency to overwater, and equally large savings can often be achieved. Unfortunately water conservation efforts do not have a linear impact on water cost. A major factor in the cost of water to the consumer is the capital cost of the distribution system and maintenance costs over time. Pipe size is set by fire protection requirements and thus the distribution system costs are not a direct function of consumption rates. However, because of the cost of developing new water supplies and the limited supplies available in many locations, there is a need to practice conservation virtually everywhere.

B. Water Quality Standards for Recreation

Water quality standards for recreation are focused on prevention of waterborne infections. The standards in use are directly derived from drinking water standards. However, the levels of contamination are considerably less strict. In 1986 the USEPA recommended that *Escherichia coli* and enterococci be substituted for fecal coliform and fecal streptococci counts as measures of biological water quality for contact recreation such as swimming and water skiing. Current recommendations are that the steady-state geometric mean of freshwater samples not exceed 126 per 100 mL for *Escherichia coli* and not exceed 33 per 100 mL for enterococci. Marine samples, such as are taken in the surf, should not exceed 35 enterococci per 100 mL.

C. Water Quality of the Aquatic Environment

In the United States a water quality plan must exist for every drainage basin. The primary focus of basin plans is to set standards of water quality that protect the aquatic environment. The aquatic environment provides a baseline reference for water quality standards. Many aquatic organisms are very sensitive to heavy metals, pesticides, and other materials washed from the urban environment by storms or found in wastewater discharges. Insect larvae are often characterized by the need for high dissolved oxygen concentrations, and most organisms can thrive over a limited pH range. Additionally, many species have adapted to very specific environmental conditions and can be found in only a few habitats. Contaminants, such as the heavy metal ions arsenic, barium, cadmium, chromium, copper, lead, mercury, and zinc, bioaccumulate in the food chain. For example, bacteria or aquatic plants assimilate the metals and store them in higher concentration than in the water. Organisms that graze on the bacteria or plants are exposed to larger quantities because they feed over a period of time. The concentrations increase in each food chain step until toxic levels are attained. Bioaccumulation is a partial explanation for the fact that many aquatic ecosystem standards are much more stringent than drinking water standards.

Changes in the aquatic habitat result from changes in water availability and flow patterns. Aquatic plants grow in waters of specific depth ranges. These plants provide habitat for a wide range of organisms forming a major part of the ecosystem. Increases or decreases in flow or reservoir level resulting from flood control, power generation, or irrigation requirements can destroy an aquatic ecosystem very quickly, such as the example of the Nile river described above.

Urban drainage is now recognized as a serious ecological hazard. A complete understanding of the problems associated with urban drainage does not exist. However, impacts on the diversity of life in urban waters have been observed, and comparative measurements of the toxicity of urban and nonurban drainage indicate that there is reason to be concerned.

D. Wastewater Discharge Standards

Wastewater discharge standards are structured to protect public and ecological health and to maintain aesthetic qualities of the nation's waters. As for municipal water supplies and recreational waters, public health requirements are based on the bacterial quality, and coliforms, fecal coliforms, or *Escherichia coli* and enterococci counts are used as the principal parameters. Toxic ions and compounds are of concern for public health reasons but in most cases the sensitivity of the aquatic environment is considerably greater than that of humans, as noted above.

In most cases discharge requirements for lumped parameters are based on running averages of daily composite values. For example, the 30-day running average of the effluent BOD_5 value may not exceed 30 mg/L. Maximum values for any sample of about three times the 30-day average are typical. Average value requirements are based on the assumption that some variability in treatment process operations is inevitable. Discharge standards for specific pollutants tend to be based on single measurements. For example, copper concentration requirements as low as 2 μg that may not be exceeded in any sample are being imposed on many treatment facilities and on some stormwater discharges.

VI. WATER RESOURCES MANAGEMENT

The acceptance that water resources management and water quality management are tightly connected began in the last half of the 20th century. Prior to that time the disposal of urban runoff and wastewater discharges in rivers, lakes, and near-shore waters was considered to be a reasonable use of the resources. As evidence of the importance of aquatic ecosystems accumulated, pressure mounted both to stop pollution and to rehabilitate waters that had been damaged. Fortunately aquatic ecosystems have proven to be resilient and rivers such as the Ohio and the Willamette that were little more than open sewers in the 1950s were nearly recovered by 1990. It is now generally accepted that the water quality component is an essential part of any water resources management plan. Municipal, industrial, and agricultural water supply, flood control, recreation, and hydroelectric power generation are even more significant issues at the beginning of the 21st century than at the beginning of the 20th century. Adding the constraints imposed by water quality requirements to the increasing demand for water makes water resources management ever more challenging.

Water resources management programs are nearly always based on economic evaluation of alternative uses. Assigning value to specific water uses is a difficult process. People are willing to pay over one dollar per liter for bottled drinking water that has no more intrinsic value than tap water costing one cent. As municipal water prices increase, more lawns will go brown in the summer and a value becomes somewhat defined for that component of municipal water. The value of having water available for fire protection can be set using risk factors, insurance costs, and building values. None of these values will be precise and none will be constant in time. Estimation of the value of water to agriculture and industry is somewhat easier in that the cost of water is part of the overall balance sheet. Farmers pay for seed, fertilizer, equipment, fuel, and water and thus the cost of water is a factor in the price they must receive for their crops. The value for water is set by the ability to make a profit. Determining the value of flood protection and recreation is similar to determining the value of municipal water. The value of flood protection is based on the probability of floods of particular magnitudes and the estimated damage that would result from flooding. Generally, other alternatives such as preventing construction in the flood plain and moving communities currently in the flood plain are not investigated. The value of water for recreation is calculated by the amount of money spent on specific recreational activities: for example, the number of visitors to a facility, the prorated cost of their boats, fishing gear, and water skis, camping, hotel, transportation, and meal costs and everything else that goes into using a recreational site. Hydroelectric power generation has a clear market value and is perhaps the easiest of the classical water uses to evaluate.

Placing a value on water used for environmental purposes is perhaps the most difficult task in economic evaluation of water resources. Obtaining general agreement on the value of maintaining a marsh or wetland area is virtually impossible. Releasing water from a reservoir to protect an aquatic ecosystem has a clear cost in terms of alternative uses of the water. However, these costs should not be used to define the value of the ecosystem. Similarly the value of a mountain valley suitable for reservoir construction cannot be assessed simply by the recreational dollars spent by visitors or the number of cattle it will support. Some value must be placed on maintaining the ecosystem and some value must be placed on the availability of wild lands to the general public even if they are not used. Clearly, the value placed on environmental water use is subjective, but that is the case for most other uses as well.

The principal problems in setting the value of water occur in arid regions such as the Southwestern United States. Some help in developing management strategies is provided by regulatory agencies and legislation that require maintenance of water quality and aquatic ecosystems. Laws and regulations form a set of constraints on water use and set priorities among water uses and indirectly define the value of water for environmental uses.

A. Conservation as a Management Strategy

The most severe problems with the volume of municipal water use are in arid regions. In the United States the Southwestern states fall into this category. West Texas, New Mexico, Arizona, Utah, Nevada, and California have annual droughts of four to eight months duration, and the entire area is faced with serious long-term water supply deficits. Other parts of the United States experience periodic droughts that impact municipal water supplies, but these problems can be addressed by increasing storage or modest amounts of conservation. However, the situation in the Southwestern states has reached a point where water conservation has become a major priority. Competition exists between municipal, agricultural, and environmental uses of water. The value of water for municipal use is much greater than for agricultural use, and one solution is for municipalities to purchase water from agricultural users or those holding rights to agricultural water. There is no clear place for maintaining environmental uses except through legislation and regulation. Approximately 80 to 90% of water use in the Southwestern states is for irrigation. Thus competition for water may result in decreases in agricultural water use and corresponding decreases in agricultural production.

Considerable opportunities to conserve water exist in both municipal and agricultural applications. As noted above, introduction of water-saving appliances can decrease household water consumption by 50% or more. In arid regions, lawn and garden irrigation often accounts for the largest fraction of water use. Changing plantings to drought tolerant plants, decreasing or eliminating lawn area, and careful irrigation practice greatly decrease water consumption. Evaporation losses from swimming pools, ornamental lakes, and open reservoirs can also be significant sources of water consumption.

A significant reason for conservation in both arid and nonarid regions is the cost of wastewater disposal. Decreasing domestic use of water results in decreased hydraulic loading on wastewater treatment plants. The cost of wastewater treatment and disposal is very closely linked to the flow being treated. Thus communities with an abundant supply of water can also benefit from water conservation.

B. Wastewater Reclamation and Reuse

In arid regions there is increasing use of wastewater reclamation and reuse as part of the overall water resources management strategy. Wastewater reclamation involves the treatment or processing of wastewater to a quality level acceptable for reuse. Wastewater reuse may be direct or indirect. Direct wastewater reuse involves a direct link between reclamation and processing. Examples of direct reuse include application of reclaimed water for irrigation of crops or landscaping, industrial cooling water, and cleaning. Indirect reuse occurs when reclaimed water is discharged to a stream, impoundment, or aquifer where it is diluted by and mixed with freshwater prior to reuse. Indirect reuse may be planned or unplanned. At this time, most wastewater reuse is indirect and unplanned. Most rivers are used as receiving waters for wastewater and downstream users of all types are involved in indirect, unplanned wastewater reuse.

It is very important to understand that indirect, unplanned wastewater reuse is in widespread practice. Consider that the Mississippi River receives most of the municipal and industrial wastewater generated between the Allegheny and Rocky Mountains and the fact that New Orleans, Louisiana, produces potable water from the mixture. Planned reuse is focused on a particular application and the provision of water that would otherwise not be available. For example, the groundwater aquifers of the Los Angeles basin in California have been extensively overpumped. Treated wastewater from the Los Angeles County Sanitation Districts and the Orange County Sanitation Districts is discharged to spreading basins and used for recharge of the aquifers. Treated wastewater is directly reused extensively for landscape irrigation in many areas. Direct reuse of reclaimed wastewater for irrigation of nonfood crops is used throughout the world, and in many areas reclaimed wastewater is used for food crop irrigation. The technology is available to produce water of any desired quality, and reclamation of wastewater for potable use is feasible.

SEE ALSO THE FOLLOWING ARTICLES

DRINKING WATER QUALITY AND TREATMENT • ENVIRONMENTAL GEOCHEMISTRY • GREENHOUSE EFFECT AND CLIMATE DATA • HYDROGEOLOGY • OCEAN-ATMOSPHERIC EXCHANGE • POLLUTION, AIR • SOIL AND GROUNDWATER POLLUTION • STREAMFLOW • TRANSPORT AND FATE OF CHEMICALS IN THE ENVIRONMENT • WASTEWATER TREATMENT AND WATER RECLAMATION • WATER CONDITIONING, INDUSTRIAL • WATER POLLUTION

BIBLIOGRAPHY

Asano, T., and Levine, A. D. (1998). “Wastewater Reclamation, Recycling and Reuse: An Introduction,” In *Wastewater Reclamation and Reuse* (T. Asano, ed.), Technomic Publishing Co., Lancaster, PA.

Bedient, P. B., and Huber, W. C. (1992). *Hydrology and Flood Plain Analysis*, 2nd ed., Addison-Wesley, Reading, MA.

Berndt, M. P., Galeone, D. R., Spruill, T. B., and Crandall, C. A. (1997). *Groundwater Quality in Three Urban Areas in The South Costal Plain of the Southeastern United States, 1992*, Water Resources Investigation Report 97-423m, U.S. Geological Survey.

Bouchard, D. C., Williams, M. K., and Surampalli, R. Y. (1992). “Nitrate Contamination of Groundwater: Sources and Potential Health Effects,” *J. Am. Water Works Assoc.* **84**(9), 85–90.

Butler, J. N. (1964). *Solubility and pH Calculations*, Addison-Wesley, Reading, MA.

Davis, S. N., and DeWiest, R. J. M. (1966). *Hydrogeology*, John Wiley and Sons, New York.

Department of Health Services (1999). *Primary Drinking Water Standards*, California State Department of Health Services, Sacramento, CA.

Diazinon Review (2000). *National Registration Authority for Agricultural Chemicals*, Minister of Fisheries and Wildlife, Canberra, Australia.

Engman, E. T. (1986). "Roughness Coefficients for Routing Surface Runoff," *J. Irrigation Drainage Engineering*, ASCE, **112**(1), 39–53.

Ergas, S. J., Kinney, K., Fuller, M. E., and Scow, K. M. (1994). "Characterization of a Compost Biofiltration System Degrading Dichloromethane," *Biotechnol. Bioengineering* **44,** 1048–1054.

Eweis, J. B., Ergas, S. J., Chang, D. P. Y., and Schroeder, E. D. (1998). *Bioremediation Principles*, WCB McGraw-Hill, New York.

Freeze, R. A., and Cherry, J. A. (1979). *Groundwater*, Prentice-Hall, Englewood Cliffs, NJ.

Heath, R. C. (1982). *Basic Groundwater Hydrology*, U.S. Geological Survey, Water Supply Paper 2220.

Kaltreider, R. C., Davis, A. M., Laniviere, J. P., and Hamilton, J. W. (2001). "Arsenic Alters the Function of the Glucocorticoid Receptor as a Transcription Factor," *Environmental Health Perspect.* **109**(3).

Logan, B. E. (1999). *Environmental Transport Processes*, John Wiley & Sons, New York.

L'vovich, M. I. (1979). *World Water Resources and Their Future* (translated by R. L. Nase), American Geophysical Union.

Madigan, M. T., Martinko, J. M., and Parker, J. (2000). *Brock Biology of Microorganisms*, 9th ed., Prentice-Hall, Upper Saddle River, NJ.

Mansell, B. O. (2000). *Biological Denitrification of Nitrate Contaminated Groundwater in a Microporous Membrane Bioreactor*, Doctoral dissertation, Department of Civil and Environmental Engineering, University of California, Davis.

Metcalf and Eddy, Inc. (2001). *Wastewater Engineering*, 4th ed., McGraw-Hill Book Company, New York.

Nazaroff, W. W., and Alvarez-Cohen, L. (2000). *Environmental Engineering Science*, John Wiley & Sons, New York.

Pan, N. L., and Schroeder, E. D. (1997). *Impact of Storm Water Loading on Santa Monica Bay*, Report to the California Department of Transportation, Center for Environmental and Water Resources Engineering, University of California, Davis, CA.

Perry, R. H., and Green, D. W., eds. (1984). *Chemical Engineers Handbook*, 6th ed., McGraw-Hill Book Company, New York.

Public Health Service (2001). *9th Report on Carcinogens*, Revised January 2001, National Toxicology Program, U.S. Department of Health and Human Services, Washington, DC.

Tchobanoglous, G., and Schroeder, E. D. (1985). *Water Quality*, Addison-Wesley, Reading, MA.

U.S. Environmental Protection Agency (2000). *Draft Implementation Guidance for Ambient Water Quality Criteria for Bacteria-1986*, EPA 823-D-00-001, Office of Water, USEPA, Washington, DC.

U.S. Environmental Protection Agency (1999). *Enhanced Coagulation and Enhanced Precipitative Softening Guidance Manual*, EPA 815-R-99-012, Washington, DC.

Warrick, R. A., Le Provost, C., Meier, M. F., Oerlemans, J., and Woodworth, P. L. (1996). "Changes in Sea level," In *Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change* (J. T. Houghton, *et al.*, eds.), pp. 362–405.

Water Environment Federation (1998). *Standard Methods for the Examination of Water and Wastewater*, 20th edition, Alexandria, VA.

World Commission on Dams (2000). *Dams and Development*, London.