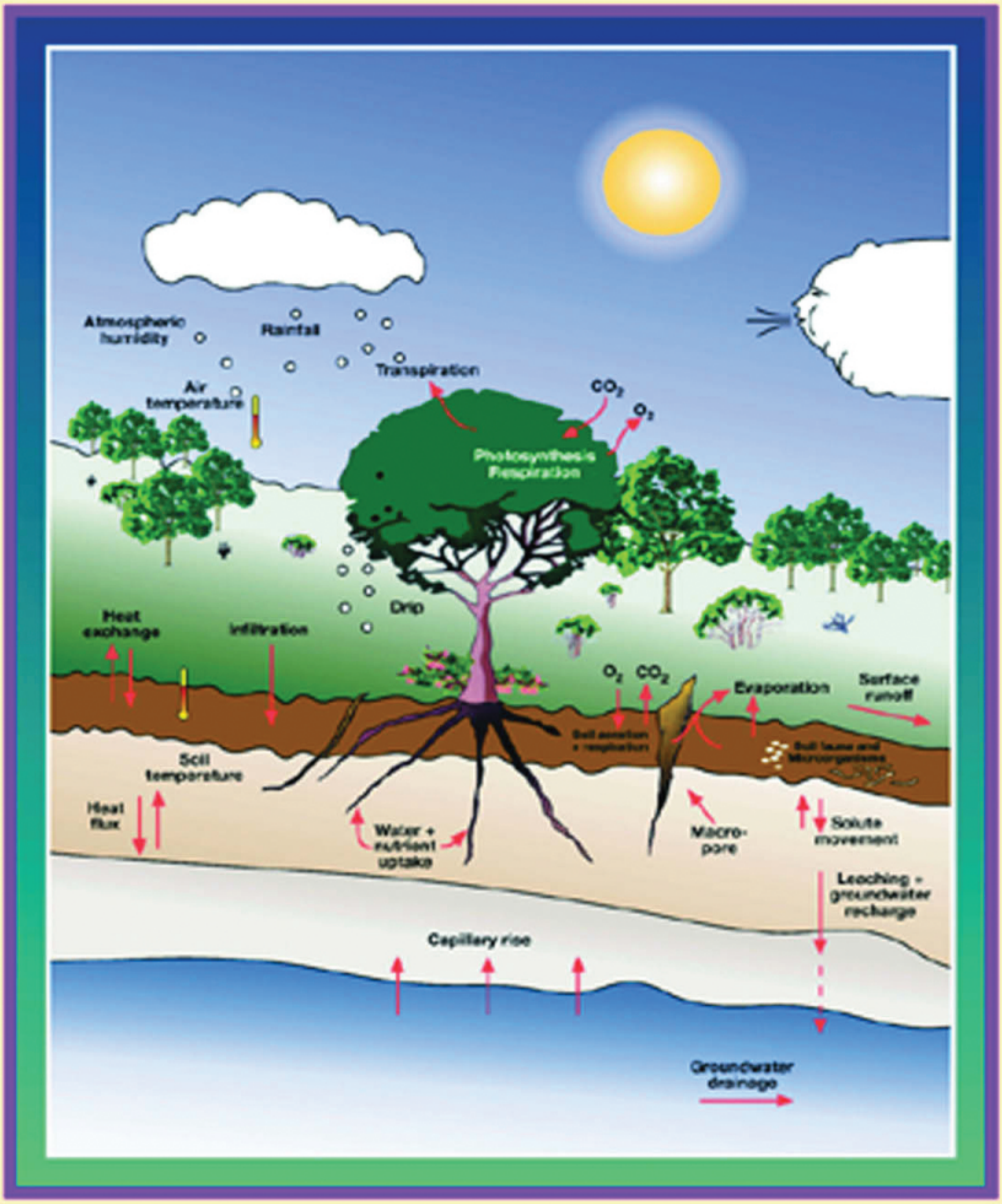
ELSEVIER
ACADEMIC
PRESS

Introduction to Environmental Soil Physics



Daniel Hillel

This book is but a clearing at the edge of the woods
where students might observe a few of the trees
as they prepare to set forth independently
to explore the great forest that yet lies beyond, –
and to care for it.

INTRODUCTION TO ENVIRONMENTAL SOIL PHYSICS

This Page Intentionally Left Blank

INTRODUCTION TO ENVIRONMENTAL SOIL PHYSICS

Daniel Hillel

Professor Emeritus
Plant, Soil, and Environmental Sciences
University of Massachusetts

ELSEVIER
ACADEMIC
PRESS

Amsterdam Boston Heidelberg London New York Oxford Paris
San Diego San Francisco Singapore Sydney Tokyo

This book is printed on acid-free paper. ♾

Academic Press
An imprint of Elsevier Science
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
http://www.academicpress.com

Academic Press
An imprint of Elsevier Science
84 Theobald's Road, London WC1X 8RR, UK
http://www.academicpress.com

Academic Press
An imprint of Elsevier Science
200 Wheeler Road, Burlington, Massachusetts 01803, USA
www.academicpressbooks.com

Library of Congress Cataloging-in-Publication Data: A catalog record for this book is available from the Library of Congress.

International Standard Book Number: 0-12-348655-6

PRINTED IN THE UNITED STATES OF AMERICA
03 04 05 06 9 8 7 6 5 4 3 2 1

CONTENTS

This Page Intentionally Left Blank

LIST OF TEXT BOXES

To see a world in a grain of sand
and heaven in a wild flower
Hold infinity in the palm of your hand
and eternity in an hour.
William Blake (1757–1827)
Songs of Innocence and Experience

PREFACE

What is special about this fragmented and loose outer layer of the earth's continental surface, which we call "soil"? What role does it fulfill in generating and sustaining life on earth, and why do we need to study it?

Considering the height of the atmosphere, the thickness of the earth's rock mantle, and the depth of the ocean, we note that the soil is an amazingly thin body — typically not much more than one meter thick and often less than that. Yet it is the crucible of terrestrial life, within which biological productivity is generated and sustained. It acts like a composite living entity, being home to a community of innumerable microscopic and macroscopic plants and animals. A mere fistful of soil typically contains millions of microorganisms, which perform the most vital functions of biochemistry. Another amazing attribute of the soil is its spongelike porosity and its enormous internal surface area. That same fistful of soil may actually consist of several acres of active surface upon which physicochemical processes take place continuously.

Realizing humanity's dependence on the soil, ancient peoples, who lived intimately with nature, actually revered the soil. It was their source of livelihood, as well as the material with which they built homes and that they learned to shape, heat, and fuse into household vessels and writing tablets (ceramic being the first synthetic material in the history of technology). In the Bible, the name assigned to the first human being was Adam, derived from the Hebrew word *adama*, meaning "soil." The name given to his mate was Hava (Eve, in transliteration), meaning "living" or "life-giving." Together, therefore, Adam and Eve signified quite literally "soil and life."

The same powerful metaphor is echoed in the Latin name for the human species — *Homo* — derived from *humus*, the material of the soil. Hence, the adjective *human* also implies "of the soil." Other ancient cultures evoked equally powerful associations. To the ancient Greeks, the earth was a manifestation of Gaea, the maternal goddess who, impregnated by Uranus (god of the sky), gave birth to all the disparate gods of the Greek pantheon.

Our civilization depends on the soil more crucially than ever, because our numbers have grown while available soil resources have diminished and deteriorated. Paradoxically, however, even as our dependence on the soil has increased, most of us have become physically and emotionally detached from it. The majority of the people in the so-called "developed" countries spend their lives in the artificial environment of a city, insulated from direct exposure to nature. Many children now assume as a matter of course that food originates in supermarkets.

Detachment has bred ignorance, and out of ignorance has come the arrogant delusion that our civilization has risen above nature and has set itself free of its constraints. Agriculture and food security, erosion and salination, degradation of natural ecosystems, depletion and pollution of surface waters and aquifers, and decimation of biodiversity — all these processes, which involve the soil directly or indirectly, have become mere abstractions to most people. The very language we use betrays disdain for that common material underfoot, often referred to as "dirt." Some fastidious parents prohibit their children from playing in the mud and rush to wash their "soiled" hands when the children nonetheless obey an innate instinct to do so. Thus is devalued and treated as unclean what is in fact the terrestrial realm's principal medium of purification, wherein wastes are decomposed and nature's productivity is continually rejuvenated.

Scientists who observe the soil discern a seething foundry in which matter and energy are in constant flux. Radiant energy from the sun streams onto the field and cascades through the soil and the plants growing in it. Heat is exchanged, rainwater percolates in the intricate passages of the soil, plant roots suck up that water and transmit it to their leaves, which transpire it back to the atmosphere. The leaves absorb carbon dioxide from the air and synthesize it with soil-derived water to form the primary compounds of life: carbohydrates, fats, proteins, and numerous other compounds (many of which provide medicinal as well as nutritional value). Oxygen emitted by the leaves makes the air breathable for animals, which feed on and in turn fertilize the plants.

The soil is thus a self-regulating biophysical factory, utilizing its own materials, water, and solar energy. It also determines the fate of rainfall and snowfall reaching the ground surface — whether the water thus received will flow over the land as runoff or seep downward to the subterranean reservoir called groundwater, which in turn maintains the steady flow of springs and streams. With its finite capacity to absorb and store moisture, the soil regulates all of these phenomena. Without the soil as a buffer, rain falling over the continents would run off immediately, producing violent floods rather than sustained stream flow.

The soil naturally acts as a living filter in which pathogens and toxins that might otherwise accumulate to foul the terrestrial environment are rendered harmless and transmuted into nutrients. Since time immemorial, humans and other animals have been dying of all manner of diseases and have then been buried in the soil, yet no major disease is transmitted by it. The term *antibiotic* was coined by soil microbiologists, who, as a consequence of their studies of soil bacteria and actinomycetes, discovered streptomycin (an important cure for tuberculosis and other infections). Ion exchange, a useful process of water purification, was also discovered by soil scientists studying the passage of solutes through beds of clay.

However unique in form and function, the soil is not an isolated body. It is, rather, a central link in the chain of interconnected domains comprising the terrestrial environment. The soil interacts with both the overlying atmosphere and the underlying strata, as well as with surface and underground bodies of water. Especially important is the interrelation between the soil and the climate. In addition to its function of regulating the cycle of water, soil regulates energy exchange and temperature. When virgin land is cleared of vegetation and turned into a cultivated field, the native biomass above the ground is often burned and the organic matter within the soil tends to decompose rapidly. These processes release carbon dioxide into the atmosphere, thus contributing to the earth's greenhouse effect and to global warming. On the other hand, the opposite action of soil enrichment with organic matter, such as can be achieved by means of reforestation and conservation farming, may help to absorb carbon dioxide from the atmosphere. To an extent, the soil's capacity to absorb and sequester carbon can thus help to mitigate the atmosphere's so-called greenhouse effect.

It takes nature thousands of years to create life-giving soil out of sterile bedrock. It takes but a few decades for unknowing or uncaring humans to destroy that wondrous work of nature. It is for us who do care for future generations to treat the soil with respect and humility, another word derived from *humus*. In the Book of Genesis, humans are said to have been placed in the Garden of Eden for a purpose, "to serve and preserve it." There is a profound truth in that perception. The earth and its soil can be a veritable Garden of Eden, but only if we do not despoil it and thereby banish ourselves from a life of harmony within it.

"To the wise man, the whole world's a soil," wrote Ben Johnson (1573–1637). His thought was echoed by William Butler Yeats (1865–1939): "All that we did, all that we said or sang, must come from contact with the soil."

This book is an abridged and updated version of my earlier book, published in 1998 under the title "Environmental Soil Physics." As such, this version is intended to serve as a basic text for introductory undergraduate courses in soil physics for students in the environmental, agricultural, and engineering sciences. The book is also meant to appeal to a broad range of students and professionals, as well as educated lay readers outside those formal categories, who may wish to acquire a fundamental understanding of the principles and processes governing the ways the soil functions in natural and in managed ecosystems. Those who seek a more detailed treatment of the various topics introduced herewith are invited to consult the larger book (1998), as well as the numerous references listed in the text and the Bibliography.

A textbook on so vital a subject ought by right to capture and convey the special sense of wonderment and excitement that impels the scientist's quest to comprehend the workings of nature, and hence should give some pleasure in the reading. It is my hope that this book might indeed be pondered, not merely studied, that its readers might find within it a few insights as well as facts, and that it will deepen their understandings as well as broaden their knowledge.

Acknowledgments

I am grateful to Dr. Charles Crumly, Editor-in-Chief of Applied Sciences at Academic Press/Elsevier, for initiating and encouraging the writing of this book, and to Ms. Angela Dooley, the Production Editor, for her empathy and competence in nursing the manuscript to publication.

Daniel Hillel
May, 2003

Part I

BASIC RELATIONSHIPS

1. SOIL PHYSICS AND SOIL PHYSICAL CHARACTERISTICS

SOIL SCIENCE

To begin, we define *soil* as the weathered and fragmented outer layer of the earth's terrestrial surface. It is formed initially through disintegration, decomposition, and recomposition of mineral material contained in exposed rocks by physical, chemical, and biological processes. The material thus modified is further conditioned by the activity and accumulated residues of numerous microscopic and macroscopic organisms (plants and animals). In a series of processes that may require hundreds or even thousands of years and that is called *soil genesis*, the loose debris of rock fragments is transmuted into a more or less stable, internally ordered, actively functioning natural body. Ultimately, this culminates in the formation of a characteristic *soil profile*, which resembles a layer cake. We can visualize the soil profile as a composite living body, in the same way that we think of the human body as a distinct organism, even though in reality it is an ensemble of numerous interdependent and symbiotically coordinated groups of organelles, cells, organs, and colonies of myriad organisms.

Soil science is the study of the soil in all its ramified manifestations and facets: as a central link in the biosphere, as a medium for the production of agricultural commodities, and as a raw material for industry and construction. As such, it shares interests with geology, sedimentology, terrestrial ecology, and geobotany as well as with such applied sciences as agronomy and engineering. Because of its varied interests and concerns, soil science itself is commonly divided into several subdivisions, including pedology (soil formation and classification), soil chemistry, soil mineralogy, soil biology, soil fertility, and soil mechanics. However, such distinctions are often arbitrary, because in fact all of the environmental sciences are inextricably interconnected.

SOIL PHYSICS

Soil physics is one of the major subdivisions of soil science. It seeks to define, measure, and predict the physical properties and behavior of the soil, both in its natural state and under the influence of human activity. As physics deals in general with the forms and interactions of matter and energy, so soil physics deals specifically with the state and movement of matter and with the fluxes and transformations of energy in the soil. On the one hand, the fundamental study of soil physics aims at understanding the mechanisms governing such processes as terrestrial energy exchange, the cycles of water and of transportable materials, and the growth of plants in the field. On the other hand, the practical application of soil physics aims at the proper management of the soil by means of cultivation, irrigation, drainage, aeration, improvement of soil structure, control of infiltration and evaporation, regulation of soil temperature, and prevention of erosion.

Soil physics is thus both a basic and an applied science, with a very wide range of interests. The study of soil science in general and of soil physics in particular is driven not only by the innate curiosity that is our species' main creative impulse, but also by urgent necessity. The intensifying pressure of population and development has diminished the soil resources of our small planet and has led to their unsustainable use and degradation in too many parts of the world.

Since the soil is not an isolated medium but is in constant dynamic interaction with the larger environment, soil physics is an aspect of the more encompassing field of environmental physics (sometimes called biospheric physics) and of the overall science of geophysics.

The early soil physicists were interested primarily in the engineering and the agricultural aspects of their discipline, hence their research focused on the soil as a material for construction or as a medium for the production of crops. Recent decades have witnessed an increasing emphasis on the environmental aspects and applications of soil physics. Consequently, research in soil physics has expanded its scope to include phenomena related to natural ecosystems and to processes affecting the quality of the environment. Processes occurring in the soil are now seen to affect the entire terrestrial environment, including local and regional climates, the natural food chain, biodiversity, and the fate of the voluminous waste products of our civilization (among which are many pathogenic and toxic agents).

Increasingly, the main concern of soil physics has shifted from the laboratory to the field and from a restricted one-dimensional view to an expansive three-dimensional view interfacing with the domains of sister disciplines such as meteorology and climatology, hydrology, ecology, and geochemistry. The larger domain of soil physics now encompasses greater complexity and variability in space and time, the treatment of which requires reliance on stochastic as well as deterministic methods. Consequently, the science is becoming ever more interesting and relevant.

The task of soil physics is made difficult by the enormous and baffling intricacy of a medium containing myriad mineral and organic components, all irregularly fragmented and variously associated in a geometric pattern that is so complex and labile as to challenge our imagination and descriptive powers. Some of the solid material consists of crystalline particles, while some is made up of amorphous gels that may coat the crystals and modify their behavior. The solid phase in the soil interacts with the fluids, water, and air that permeate soil pores. The entire soil is hardly ever in equilibrium as it alternately wets and dries, swells and shrinks, disperses and flocculates, hardens and softens, warms and cools, freezes and thaws, compacts and cracks, absorbs and emits gases, adsorbs and releases exchangeable ions, dissolves and precipitates salts, becomes acidic or alkaline, and exhibits aerobic or anaerobic conditions leading to chemical oxidation or reduction.

THE SOIL PROFILE

The most obvious part of any soil is its surface zone. Through it, matter and energy are transported between the soil and the atmosphere. The surface may be smooth or pitted, granular or crusted and cracked, friable or hard, level or sloping, vegetated or fallow, mulched or exposed. Such conditions affect the processes of radiation and heat exchange, water and solute movement, and gaseous diffusion.

Important though the surface is, however, it does not necessarily portray the character of the soil as a whole. To describe the latter, we must examine the soil in depth. We can do this by digging a trench and sectioning the soil from the surface downward, thus revealing what is commonly termed the *soil profile*.

The soil profile typically consists of a succession of more-or-less distinct strata. These may result from the pattern of deposition, or sedimentation, as can be observed in wind-deposited (*aeolian*) soils and particularly in water-deposited (*alluvial*) soils. If, however, the strata form in place by internal soil-forming (*pedogenic*) processes, they are called *horizons*.

The top layer, or *A horizon*, is the zone of major biological activity and is therefore generally enriched with organic matter and darker in color than the underlying soil. Here, plants and animals and their residues interact with an enormously diverse and labile multitude of microorganisms, such as bacteria, protozoa, and fungi, millions of which can be found in a mere handful of topsoil. In addition, there are usually varied forms of macroorganisms (including earthworms, arthropods, and rodents) that burrow in the soil. The A horizon is generally the most fertile zone of the soil, but it is also the zone most vulnerable to erosion by water and wind (especially if it is denuded of vegetative cover or its protective residues).

Underneath the A horizon is the *B horizon*, where some of the materials (e.g., clay or carbonates) that are leached from the A horizon by percolating water tend to accumulate. The B horizon is often thicker than the A horizon. The pressure of the overlying soil tends to reduce the porosity of the deeper layers. In some cases, an overly dense or indurated B horizon may inhibit gas exchange, water drainage, and root penetration.

Underlying the B horizon is the *C horizon*, which is the soil's parent material. In a soil formed of bedrock in situ (called a *residual soil*), the C horizon consists of the weathered and fragmented rock material. In other cases, the C horizon may consist of alluvial, aeolian, or glacial deposits.

The character of the profile depends primarily on the climate that prevailed during the process of soil formation. It also depends on the parent material, the vegetation, the topography, and time. These five variables are known as the *factors of soil formation* (Fanning and Fanning, 1989). Mature soils are such that have been subjected to those factors for a sufficient length of time so that full profile development has taken place. The A, B, C sequence of horizons is clearly recognizable in some cases, as, for example, in a typical zonal soil (i.e., a soil associated with a distinct climatic zone), such as a *podzol* (also known as a *spodosol*). In other cases, no clearly developed B horizon is discernible, and the soil is then characterized by an A,C profile. In a recent alluvium, hardly any profile differentiation is apparent.

The typical development of a soil and its profile, called *pedogenesis* (Buol et al., 2003), can be summarized: It begins with the physical disintegration of an exposed rock formation, which provides the soil's parent material. Gradually, the loosened material is colonized by living organisms. The consequent accumulation of organic residues at and below the surface brings about the development of a discernible A horizon. That horizon may acquire an aggregated structure, stabilized to some degree by gluelike components of the organic matter complex (known as *humus*) resulting from the decomposition of plant and animals residues. Continued weathering (decomposition and recomposition) of minerals may bring about the formation of clay. Some of the clay thus formed tends to migrate downward, along with other transportable materials (such as soluble salts), and to accumulate in an intermediate zone (the B horizon) between the surface zone of major biological activity and the deeper parent material of the so-called C horizon.

Important aspects of soil formation and profile development are the twin processes of *eluviation* and *illuviation* ("washing out" and "washing in," respectively), wherein clay and other substances emigrate from the overlying eluvial A horizon and accumulate in the underlying illuvial B horizon. The two horizons come to differ substantially in composition and structure.

Throughout these processes, the profile as a whole deepens as the upper part of the C horizon is gradually transformed, until eventually a quasi-stable condition is approached in which the counterprocesses of soil formation and soil erosion are more or less in balance. In the natural state, the A horizon may have a thickness of 0.1–0.5 meters. When stripped of vegetative cover and pulverized or compacted by tillage or traffic, this horizon may lose half or more of its original thickness within a few decades.

In arid regions, salts such as calcium sulfate and calcium carbonate, dissolved from the upper part of the soil, may precipitate at some depth to form a cemented *pan* (sometimes called *caliche*, from the Spanish word for "lime"). Erosion of the A horizon may bring the B horizon to the surface. In extreme cases, both the A and B horizons may be scoured off by natural or human-induced erosion so that the C horizon becomes exposed and a new cycle of soil formation may then begin. In other cases, a mature soil may be covered by a new layer of sediments (alluvial or aeolian) so that a new soil may form over a "buried" old soil. Where episodes of deposition occur repeatedly over a long period of time, a sequence of soils may be formed in succession, thus recording the pedological history — called the *paleopedology* — of the region (including evidence of the climate and vegetation that had prevailed at the time of each profile's formation).

Numerous variations of the processes described are possible, depending on local conditions. The characteristic depth of the soil, for instance, varies from location to location. Valley soils are typically deeper than hillslope soils, and the depth of the latter depends on slope steepness. In places, the depth of the profile can hardly be ascertained, because the soil blends into its parent material without any distinct boundary. However, the zone of

BOX 1.1 Soil Physics and the Environment

As human populations have grown and living standards have risen, the requirements for agricultural products have increased enormously. More land has been brought under cultivation, including marginal land that is particularly vulnerable to degradation by such processes as erosion, depletion of organic matter and nutrients, pollution, waterlogging, and salination. Other forms of land use — towns, roads, factories, airports, feedlots, waste disposal sites, and recreational facilities — usurp ever more land.

Consequently, the domain remaining for natural ecosystems has shrunk and been divided into smaller enclaves, to the detriment of numerous species. The mutual checks and balances that have long sustained the rich diversity of life on earth are now threatened by the human appetite for resources and the wanton way they are used and their waste products discarded. The task therefore is to supply human needs in ways that are sustainable locally and do not damage the larger environment.

Two alternative approaches have been proposed to prevent further destruction of the remaining natural ecosystems and to relieve pressure on fragile marginal lands. One way is to restrict human activities to choice areas, where production can be intensified. This calls for optimizing all production factors so as to achieve maximum efficiency in the utilization of soil, water, energy, and other necessary inputs (e.g., nutrients and pest control measures). The problems are that soil processes are difficult to control, and, because the soil is an open system in constant interaction with its surroundings, the time-delayed and space-removed consequences of soil processes are difficult to predict. Full control of agricultural production can ultimately be achieved only in enclosed spaces such as greenhouses or in confined fields.

Another approach is to devise more naturalistic modes of production that are compatible with the environment and do not require the drastic isolation of production from neighboring ecosystems. This agro-ecosystem approach is exemplified by the trends toward organic farming and agroforestry.

Either way, the physical attributes and processes of the soil are of prime importance. Physical factors strongly affect whether the soil is to be cool or warm, anaerobic or aerobic, wet or dry, compact or highly porous, hard or friable, dispersed or aggregated, impervious or permeable, eroded or conserved, saline or salt free, leached or nutrient rich. All these, in turn, determine whether the soil can be a favorable or unfavorable medium for various types of plants and other living organisms as well as for alternative modes of production; in short, whether the soil can be managed productively and sustainably while neutralizing — rather than transmitting — environmental pollutants.

biological activity seldom extends below 2–3 meters and in many cases is shallower than 1 meter.

SOILS OF DIFFERENT REGIONS

Each climatic zone exhibits a characteristic group of soils. In the humid tropics, there is a tendency to dissolve and leach away the silica and to accumulate iron and aluminum oxides. As a result, the soils are typically colored red, the hue of oxidized iron. Chunks or blocks excavated from such soils and dried in the sun may harden to form bricks; hence these soils are called *laterites*, from the Latin word *later*, meaning "brick." On the other hand, soils of the humid cool regions often exhibit an A horizon consisting of a thin surface layer darkened by organic matter and underlain by a bleached, ashlike layer; in turn, this overlies a clay-enriched B horizon. These soil were called *podzols* by the early Russian pedologists (from the Russian words *pod* = "ground" and *zola* = "ash") and are known as *spodosols* in the current American system. Soils that are poorly drained ("waterlogged") may exhibit conditions of chemical reduction in the profile, indicated by streaks of discoloration ("mottling").

In contrast with the soils of humid regions, from which nearly all readily soluble salts have been leached, the soils of arid regions tend to precipitate the moderately soluble salts of calcium and magnesium. (Especially prevalent are accumulations of calcium carbonate [lime] and calcium sulfate [gypsum].) Under certain conditions, arid-zone soils may even accumulate the more readily soluble salts of sodium ($NaCl$ and Na_2CO_3) and of potassium. Such soils are prone to excessive salinity, in extreme cases of which they become practically sterile. Irrigated soils in poorly drained river valleys of arid regions are particularly liable to undergo the process of salination. When such soils are leached of excess salts, they must be treated with soil amendments (such as gypsum) to replace the exchangeable sodium ions with calcium, lest the sodium ion cause dispersion of the clay and the breakdown of soil structure.

An outstanding soil formed naturally in some intermediate semihumid to semiarid regions (e.g., Ukraine, Argentina, and the prairie states of North America — the so-called "corn belt" of the United States) is the soil classically called *chernozem* (Russian for "black earth") with its unusually thick, humus-rich, fertile A horizon.

Whereas Russian pedologists were the first to develop a universal classification system of soils over a century ago, other schools of pedology have since offered alternative taxonomies, claimed to be more detailed and comprehensive. Notable among these is the systems offered by the U.N. Food and Agriculture Organization and the one developed by the U.S. Soil Survey. Each system recognizes hundreds of soil types and their variants.

By way of illustration, we present a hypothetical soil profile in Fig. 1.1. This is not a "typical" soil, for among the myriad of differing soil types recognized by pedologists no single type can be considered typical. Our figure is only meant to suggest the sort of contrasts in appearance and structure among different horizons that may be encountered in a soil profile. Pedologists classify soils by their mode of formation (genesis) and recognizable properties (see Fig. 1.2).

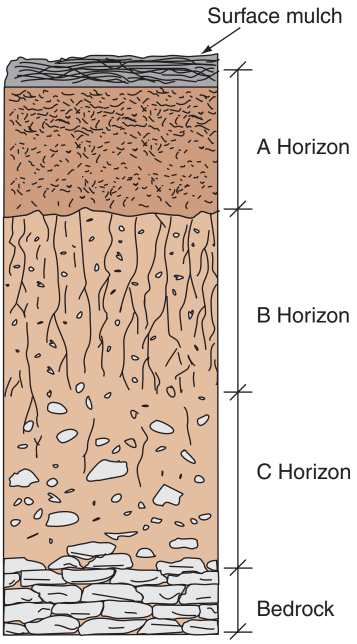


Fig. 1.1. Schematic representation of a hypothetical soil profile. The A horizon is shown with an aggregated crumblike structure, the B horizon with columnar structure, and the C horizon with incompletely weathered rock fragments.

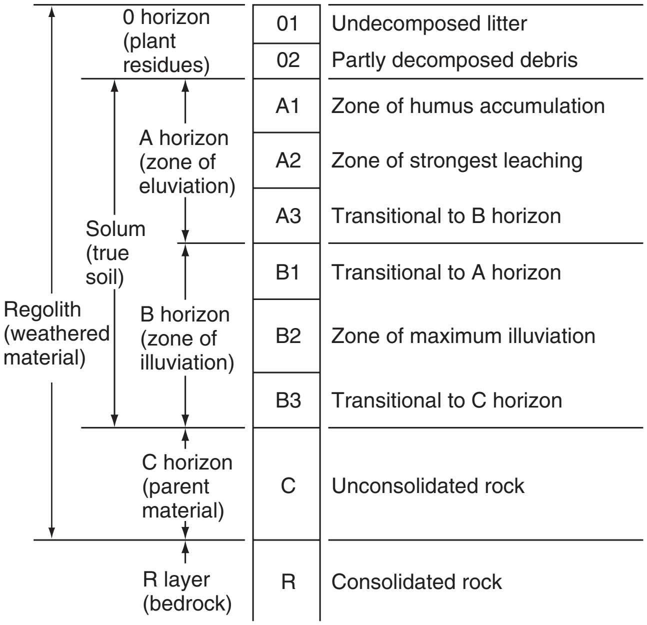


Fig. 1.2. Descriptive terminology for soil profile horizones.

SOIL AS A DISPERSE THREE-PHASE SYSTEM

The term *system* refers to a group of interacting, interrelated, or interdependent elements constituting an integrated entity. Since all of nature is in fact an integrated entity, defining any domain within nature to be a “system” is an admittedly arbitrary exercise. We choose to do so for reasons of convenience, since our own limitations prevent us from dealing with the entire complexity of nature all at once. However, the part of nature on which we may wish to focus our attention at any moment is necessarily a subsystem inside a larger system, with which our system interacts continuously. The modes of interaction generally include transfers or exchanges of materials and of energy.

Systems may vary widely in size, shape, and degree of complexity. They may consist of one or more substances and of one or more phases. The simplest system is one that is comprised of a single substance that has the same physical properties throughout. An example of such a system is a body of water consisting entirely of uniform ice. Such a system is homogeneous. A system comprised of a single chemical compound can also be heterogeneous if that substance exhibits different properties in different regions of the system. A region that is internally uniform physically is called a *phase*. A mixture of ice and liquid water, for instance, is chemically uniform but physically heterogeneous, because it includes two phases. The three ordinary phases in nature are the solid, liquid, and gaseous phases.

A system containing several substances may also be monophasic. A solution of salt and water, for example, is a homogeneous liquid. A system of several substances can also be heterogeneous. In a heterogeneous system the properties may differ not only between one phase and another, but also between the internal parts of each phase and the interfaces over which the phase comes into contact with one another. Interfaces exhibit specific phenomena resulting from the interaction of the phases. The importance of such phenomena (including adsorption, surface tension, and friction) in determining the behavior of the system as a whole depends on the magnitude of the interfacial area per unit volume of the system.

Systems in which at least one of the phases is subdivided into numerous minute particles, which together present a very large interfacial area per unit volume, are commonly called *disperse systems*. Colloidal sols (including aerosols), gels, emulsions, and aerosols are examples of disperse systems.

The soil is a heterogeneous, polyphasic, particulate, disperse, and porous system, with a large interfacial area per unit volume. (A handful of clay, for instance, may have an internal surface area of several hectares, a hectare being equal to 2.5 acres!). The disperse nature of soil and its consequent interfacial activity give rise to such phenomena as adsorption of water and chemicals, capillarity, ion exchange, swelling and shrinking, dispersion and flocculation.

The three phases of ordinary nature are represented in the soil as follows: the solid phase forms the *soil matrix*; the liquid phase is the water in the soil, which always contains dissolved substances, so it should properly be called the *soil solution*; and the gaseous phase is the *soil atmosphere*. The solid matrix of the soil consists of particles that vary in chemical and mineralogical composition as well as in size, shape, and orientation. It also contains amorphous substances, particularly organic matter, which is attached to the mineral grains

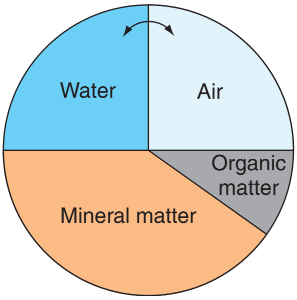


Fig. 1.3. Schematic composition (by volume) of a medium-textured soil at a condition considered optimal for plant growth. Note that the solid matter constitutes 50% and the pore space 50% of the soil volume, with the latter divided equally between water and air. the arrows indicate that water and air are related so that an increase in one is associated with a decrease in the other.

and may bind them together in assemblages called *aggregates*. The organization of the solid components of the soil determines the geometric characteristics of the pore spaces in which water and air are transmitted and retained. Finally, soil water and soil air vary in composition, both in time and in space.

The relative proportions of the three phases in the soil are not fixed but vary continuously, depending on such variables as weather, vegetation, and management. Figure 1.3 presents the hypothetical volume composition of a medium-textured soil at a condition considered approximately optimal for plant growth.

BOX 1.2 The Concept of Representative Elementary Volume

Some soil properties (for example, temperature) can be measured at a point, whereas other properties are volume dependent. Suppose we wish to measure some volume-dependent soil property, such as porosity. If our sample is very small, say, the size of a single particle or pore, the measured porosity may vary between zero and 100 percent, depending on the point at which we make our measurement (whether at a particle or at a pore). If we measure the porosity repeatedly at several adjacent points, the results will fluctuate widely. However, if we increase the scale or volume of the each sample so as to encompass within it both particles and pores, the fluctuations among repeated measurements at adjacent locations will diminish. If we keep enlarging the sample progressively, we will eventually obtain a consistent measurement of the soil's average porosity. The minimal volume of sample needed to obtain a consistent value of a measured parameter has been called the *representative elementary volume* (REV) (Bear, 1969). Obviously, the REV becomes larger in soils that are strongly aggregated (as well as in soils that are cracked or otherwise heterogeneous) than in more uniform soils.

The problem with the REV concept is that different parameters may exhibit different spatial or temporal patterns, so the REV for one parameter or property may differ from those for other parameters. That is to say, each property may have its own characteristic scale. Even more serious may be the failure of the REV concept in the case of "structured" fields, i.e., in fields that vary systematically in one direction or another. In such fields, increasing the size of the sample measured may not produce a consistent value at all.

VOLUME AND MASS RELATIONSHIPS OF SOIL CONSTITUENTS

Let us consider the volume and mass relationships among the three phases of the soil, and define some basic parameters that can help to characterize the soil physically. Figure 1.4 is a schematic depiction of a hypothetical soil in which the three phases have been separated and stacked one atop the other for the purpose of showing their relative volumes and masses.

In the figure, the masses of the phases are indicated on the right-hand side: the mass of air M_a, which is negligible compared to the masses of solids and water; the mass of water M_w; the mass of solids M_s; and the total mass M_t. (These masses can also be represented in terms of their weights, being the product of each mass and gravitational acceleration.) The volumes of the same components are indicated on the left-hand side of the diagram: volume of air V_a, volume of water V_w, volume of pores $V_f = V_a + V_w$, volume of solids V_s, and the total volume of the representative sample V_t.

On the basis of this diagram, we can now define terms that are generally used to express the quantitative interrelations of the three primary soil phases.

Density of Solids (Mean Particle Density) ρ_s

$$\rho_s = M_s / V_s \tag{1.1}$$

In most mineral soils, the mean mass per unit volume of solids is about 2600–2700 kg/m^3. This is close to the density of quartz, which is generally the most prevalent mineral in the coarsest fraction of the soil. Some of the minerals composing the finest fraction of the soil have a similar density. However, the presence of iron oxides and of various other "heavy" minerals (generally defined as those having a density exceeding 2900 kg/m^3) increases the average value of ρ_s, whereas the presence of low-density organic matter generally

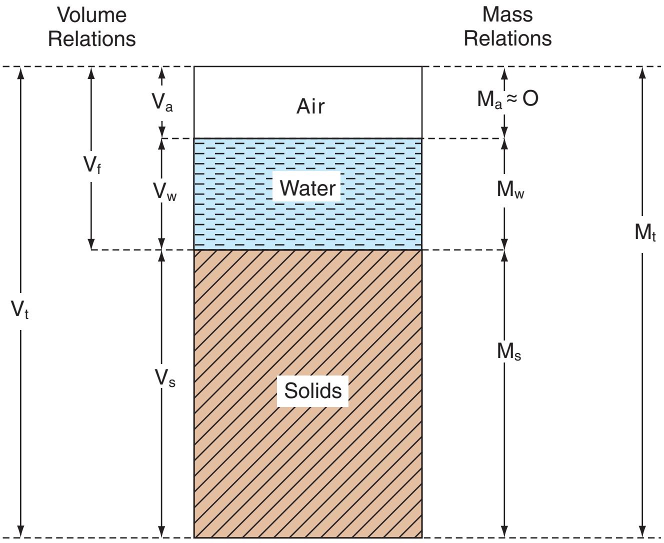


Fig. 1.4. Schematic diagram of the soil as a three-phase system.

lowers the mean density of the solid phase. Sometimes the density is expressed in terms of the *specific gravity*, σ_g, which is the ratio of the density of any material to that of water at 4°C and at atmospheric pressure. The latter density is about 1000 kg/m^3, so the specific gravity of the solid phase in a typical mineral soil is about 2.65, a value that is numerically (though not dimensionally) equal to the density expressed in the cgs. system of units (g/cm^3).

Dry Bulk Density ρ_b

$$\rho_b = M_s / V_t = M_s / (V_s + V_a + V_w) \tag{1.2}$$

The dry bulk density expresses the ratio of the mass of solids to the total soil volume (solids and pores together). Obviously, ρ_b is always smaller than ρ_s. If the pores constitute half the volume, then ρ_b is half of ρ_s, namely, about 1300–1350 kg/m^3, equivalent to a bulk specific gravity (i.e., the ratio of the soil's bulk density to the density of water at standard conditions) of 1.3–1.35. The bulk specific gravity of sandy soils with a relatively low volume of pores may be as high as 1.6, whereas that of aggregated loams and clay soils may be below 1.2. Whereas the mean particle density is typically constant, the bulk density is highly labile. It is affected by the structure of the soil, that is, its looseness or degree of compaction, as well as by its swelling and shrinkage characteristics. The latter depend both on clay content and water content. Even in extremely compacted soils, however, the bulk density remains appreciably lower than the density of the solid matter, since the particles can never interlock perfectly. However much the pore space can be reduced by compaction, it can never be eliminated.

Total (Wet) Bulk Density ρ_t

$$\rho_t = M_t / V_t = (M_s + M_w)/(V_s + V_a + V_w) \tag{1.3}$$

This is an expression of the total mass of a moist soil per unit volume. As such, this parameter depends more strongly than does the dry bulk density on soil wetness or water content.

Dry Specific Volume v_b

$$v_b = V_t / M_s = 1/\rho_b \tag{1.4}$$

The volume of a unit mass of a dry soil (the reciprocal of the dry bulk density) serves as another useful index of the degree of looseness or compaction of a soil body.

Porosity f

$$f = V_f / V_t = (V_a + V_w)/(V_s + V_a + V_w) \tag{1.5}$$

Porosity is an index of the relative pore space in a soil. Its value generally ranges from 0.3 to 0.6 (30–60%). Coarse-textured soils tend to be less porous than fine-textured soils, though the mean size of individual pores is greater in the former. In clayey soils, the porosity is highly variable because the soil alternately swells, shrinks, aggregates, disperses, compacts, and cracks. As generally defined, the term *porosity* refers to the volume fraction of pores, and this value should be equal, on average, to the areal porosity (the fraction of pores in a representative cross-sectional area) as well as to the average lineal porosity (the fractional length of pores along a straight line passing through the soil in any direction). However, the total porosity reveals nothing about the sizes or shapes of the various pores in the soil.

Void Ratio *e*

$$e = V_f/V_s = (V_a + V_w)/(V_t - V_f) \tag{1.6}$$

The void ratio is also an index of the fractional pore space, but it relates that space to the volume of solids rather than to the total volume of the soil. As such, it ranges between 0.3 and 2. The advantage of this index over the preceding one is that in the case of e any change in pore volume affects only the numerator of the defining equation, whereas in the case of f such a change affects both the numerator and the denominator. Void ratio is the index generally preferred by soil engineers, while porosity is more frequently used by agronomists.

Sample Problem

Prove the following relation of porosity to particle density and bulk density:

$$f = (\rho_s - \rho_b)/\rho_s = 1 - \rho_b/\rho_s$$

Substituting the definitions of f, ρ_s, and ρ_b, we can rewrite the equation as

$$V_f/V_t = 1 - (M_s/V_t)/(M_s/V_s)$$

Simplifying the right-hand side, we obtain

$$V_f/V_t = 1 - (V_s/V_t) = (V_t - V_s)/V_t$$

But since $V_t - V_s = V_f$, we have

$$V_f/V_t = V_f/V_t$$

Soil Wetness (Water Content)

The water content of a soil can be expressed in various ways: relative to the mass of solids, or to the total mass, or to the volume of solids, or to the total volume, or to the volume of pores. The various indexes are defined as follows.

Mass Wetness *w*

$$w = M_w/M_s \tag{1.7}$$

This is the mass of water relative to the mass of dry soil particles. The standard definition of *dry soil* refers to a mass of soil dried to equilibrium (in practice, over a 24-hour period) in an oven at 105°C, though a clay soil may still contain an appreciable amount of water at that state. Mass wetness is sometimes expressed as a decimal fraction but more often as a percentage. A sample of soil dried in “ordinary” air at ambient temperature (rather than in an oven) will generally retain several percent more water than if dried in the oven. Similarly, an oven-dry soil exposed to “ordinary” air will gradually gain appreciable moisture. This phenomenon results from the tendency of the soil’s clay fraction to adsorb moisture from the air, a property known as *hygroscopicity*. The amount thus adsorbed depends on the type and content of clay in the soil as well as on the humidity of the ambient atmosphere. The water content at saturation (when all pores are filled with water) is also higher in clayey than in sandy soils. In different soils, w can range between 25% and 60%, depending on bulk density. In the special case of organic soils, such as peat or muck soils, the saturation water content on the mass basis may exceed 100%.

Volume Wetness θ

$$\theta = V_w/V_t = V_w/(V_s + V_f) \tag{1.8}$$

The volume wetness (often termed *volumetric water content*) is generally computed as a percentage of the total soil volume. At saturation, therefore, it is equal to the porosity. In sandy soils, θ at saturation is on the order of 40%; in medium-textured soils it is approximately 50%; and in clayey soils it can approach 60%. In the last, in fact, the volume of water at saturation can exceed the porosity of the dry soil, since clayey soils swell upon wetting. The use of rather than θ to express water content is often more convenient because it is more directly applicable to the computation of fluxes and water volumes added to soil by rain or irrigation and to quantities extracted from the soil by evaporation and transpiration. The volume ratio is also equivalent to the depth ratio of soil water, that is, the depth of water per unit depth of soil.

Sample Problem

Prove the following relation between volume wetness, mass wetness, bulk density, and water density ($\rho_w = M_w/V_w$):

$$\theta = w\rho_b/\rho_w$$

Again, we start by substituting the respective definitions of w_v, w_m, ρ_b, and ρ_w:

$$V_w/V_t = [(M_w/M_s)(M_s/V_t)]/(M_w/V_w)$$

Rearranging the right-hand side, we get

$$V_w/V_t = (V_w M_w M_s)/(M_w M_s V_t) = V_w/V_t$$

Water Volume Ratio v_w

$$v_w = V_w/V_s \tag{1.9}$$

For swelling soils, in which porosity changes markedly with wetness, it may be preferable to refer the volume of water present in a sample to the invariant volume of particles rather than to total volume. At saturation, v_w is equal to the void ratio e.

Degree of Saturation s

$$s = V_w/V_f = V_w/(V_a + V_w) \tag{1.10}$$

This index expresses the water volume present in the soil relative to the pore volume. Index s ranges from zero in a completely dry soil to unity (100%) in a saturated soil. Complete saturation, however, is hardly ever attainable in field conditions, since some air is nearly always present. In a relatively dry soil the air phase occupies a continuous space, whereas in a very wet soil air may be occluded or encapsulated in the form of discontinuous bubbles.

Air-Filled Porosity (Fractional Air Content) f_a

$$f_a = V_a/V_t = V_a/(V_s + V_a + V_w) \tag{1.11}$$

This is a measure of the relative content of air in the soil and as such is an important criterion of soil aeration. It is related negatively to the degree of saturation s (i.e., $f_a = f - s$). The relative volume of air in the soil may also be expressed as a fraction, a, of the pore volume (rather than of the total soil volume). Thus,

$$a = V_a/V_f = V_a/(V_a + V_w) \tag{1.12}$$

Sample Problem

A sample of moist soil with a wet mass of 1.0 kg and a volume of 0.64 liters (6.4×10^{-4} m^3) was dried in the oven and found to have a dry mass of 0.8 kg. Assuming the typical value of particle density for a mineral soil (2650 kg/m^3), calculate the bulk density ρ_b, porosity f, void ration e, mass wetness w_m, volume wetness θ, water volume ratio v_w, degree of saturation s, and air-filled porosity f_a.

Bulk density: $\rho_b = M_s/V_t = 0.8 \text{ kg}/6.4 \times 10^{-4} \text{ m}^3 = 1250 \text{ kg/m}^3$

Porosity: $f = 1 - \rho_b/\rho_s = 1 - 1250/2650 = 1 - 0.472 = 0.528$

Alternatively, $f = V_f/V_t = (V_t - V_s)/V_t$

and since $V_s = M_s/\rho_s = 0.8 \text{ kg}/2650 \text{ kg/m}^3 = 3.02 \times 10^{-4} \text{ m}^3$

hence $f = (6.4 - 3.02) \times 10^{-4} \text{ m}^3/6.4 \times 10^{-4} \text{ m}^3 = 0.528 = 52.8\%$

Void ratio: $e = V_t/V_s = (V_t - V_s)/V_s = (6.4 - 3.02) \times 10^{-4} \text{ m}^3/3.02 \times 10^{-4} \text{ m}^3 = 1.12$

Mass wetness: $w = M_w/M_s = (M_t - M_s)/M_s = (1.0 - 0.8) \text{ kg}/0.8 \text{ kg} = 0.25 = 25\%$

Volume wetness: $\theta = V_w/V_t = 2.0 \times 10^{-4} \text{ m}^3/6.4 \times 10^{-4} \text{ m}^3 = 0.3125 = 31.25\%$

(*Note*: $V_w = M_w/\rho_w$, wherein $\rho_w \approx 1000 \text{ kg/m}^3$)

Alternatively, $\theta = w\rho_b/\rho_w = 0.25(1250 \text{ kg/m}^3/1000 \text{ kg/m}^3) = 0.3125$

Water volume ratio: $v_w = V_w/V_s = 2.0 \times 10^{-4} \text{ m}^3/3.02 \times 10^{-4} \text{ m}^3 = 0.662$

Degree of saturation: $s = V_w/(V_t - V_s) = 2.0 \times 10^{-4} \text{ m}^3/(6.4 - 3.02) \times 10^{-4} \text{ m}^3 = 0.592$

Air-filled porosity: $f_a = V_a/V_t = (6.4 - 2.0 - 3.02) \times 10^{-4} \text{ m}^3/6.4 \times 10^{-4} \text{ m}^3 = 0.216$

Sample Problem

What is the equivalent depth of water contained in a soil profile 1 m deep if the mass wetness of the upper 0.4 m is 15% and that of the lower 0.6 m is 25%? Assume a bulk density of 1200 kg/m^3 in the upper layer and 1400 in the lower layer. How much water does the soil contain, in cubic meters per hectare of land?

Recall that $\theta = w(\rho_b/\rho_w)$, where $\rho_w = 1000$ kg/m^3

Volume wetness in the upper layer: $\theta_1 = 0.15(1200/1000) = 0.18$

Equivalent depth in upper 0.4 m = 0.18 × 0.4 m = 0.072 m = 72 mm

Volume wetness in lower layer: $\theta_2 = 0.25(1400/1000) = 0.35$

Equivalent depth in lower 0.6 m = 0.35 × 0.6 m = 0.21 m = 210 mm

Total depth of water in 1-m profile = 0.072 m + 0.210 m = 0.282 m

Volume of water contained in 1-m profile per hectare
(1 ha = 10^4 m^2) = 0.282 m × 1000 m^2 = 2820 m^3

2. WATER PROPERTIES IN RELATION TO POROUS MEDIA

THE FLUID OF LIFE

Our planet is the planet of life primarily because it is blessed with the precise ranges of temperature and pressure that make possible the existence in the liquid state of a singular substance called water. So ubiquitous is water on our globe, covering nearly three-quarters of its surface, that the entire planet really should be called “water” rather than “earth.” However, as Coleridge’s Ancient Mariner complained, most of the water everywhere is unfit to drink. Less than 3% of the water on earth is “fresh” (i.e., nonsaline), and that amount is unevenly distributed. Humid regions are endowed with an abundance of it, even with a surfeit, so that often the problem is how to dispose of excess water. Arid and semiarid regions, on the other hand, are afflicted with a chronic shortage of it.

Life as we know it began in an aquatic medium, and water is still the principal constituent of living organisms. It is, literally, the essence of life. As Vladimir Vernadsky wrote a century ago: “Life is animated water.” Though we appear to be solid, we are really liquid bodies, similar to gelatin, which also seems solid but is in fact largely water, made consistent by the presence of organic material. The analogous material in our bodies is protoplasm. The body of a newborn infant contains nearly 90% water by mass, and even adults contain over 65% water. Growing herbaceous plants typically contain over 90% water. Far from being a bland, inert liquid, water is a highly reactive substance, a solvent, and a transporter of numerous substances.

The importance of water was recognized early in history, yet little was known about its real nature. In the Middle Ages, people believed that fresh water emanated magically from the bowels of the earth. They could not imagine that all the water flowing in innumerable springs and mighty rivers (such as the Nile, which appeared to the ancient Egyptians to come out of the driest desert!) could possibly result from so seemingly feeble a source as rain and snow. The first to conjecture this was Leonardo da Vinci, but only in the latter part of the 17th century did the English astronomer Edmond Halley and, separately, the Frenchman Claude Perrault, prove the principle by calculation and measurement. Water was long thought to be a single element, until early in the 18th century, when it was found to consist of hydrogen and oxygen in combination.

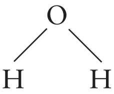
Notwithstanding its ubiquity, water remains something of an enigma, possessing unusual and anomalous attributes. Perhaps the first anomaly is that, being composed of two gases and having relatively low molecular weight, water is a liquid and not a gas at normal temperatures. (Its sister compound, hydrogen sulfide, H_2S, has a boiling point temperature of $-60.7°C$.) Compared to other common liquids, water has unusually high melting and boiling points, heats of fusion and vaporization, specific heat, dielectric constant, viscosity, and surface tension.

MOLECULAR STRUCTURE

One cubic meter of liquid water at 20°C contains some 3.4×10^{28} (34 billion billion billion) molecules, the diameter of which is about 3×10^{-10} meter (3×10^{-4} micrometers (μm), or 3 angstrom units). The chemical formula of water is H_2O, which signifies that each molecule consists of two atoms of hydrogen and one of oxygen. There are three isotopes of hydrogen (^1H, ^2H, ^3H) and three of oxygen (^{16}O, ^{17}O, ^{18}O), which can form 18 combinations. However, all isotopes but ^1H and ^{16}O are quite rare.

The hydrogen atom consists of a positively charged proton and a negatively charged electron. The oxygen atom consists of a nucleus having a positive charge of eight protons, surrounded by eight electrons, of which six are in the outer shell. Since the outer electron shell of hydrogen lacks one electron and that of oxygen lacks two electrons, one atom of oxygen can combine with two atoms of hydrogen in an electron-sharing molecule.

The strong intermolecular forces in liquid water are caused by the electrical polarity of the water molecule, which in turn is a consequence of the arrangement of electrons in its oxygen and hydrogen atoms (Fig. 2.1). The oxygen atom shares a pair of electrons with each of the two hydrogen atoms, through overlap of the 1s orbitals of the hydrogen atoms with two hybridized sp^3 orbitals of the oxygen atom. The H–O–H bond in water is not linear but bent at an angle of 104.5°:



That angle deviates slightly from a perfectly tetrahedral arrangement of the oxygen atom's four possible sp^3 orbitals, which would have an angle of 109.5°. The mean H–O interatomic distance is 9.65×10^{-5} μm. The arrangement of

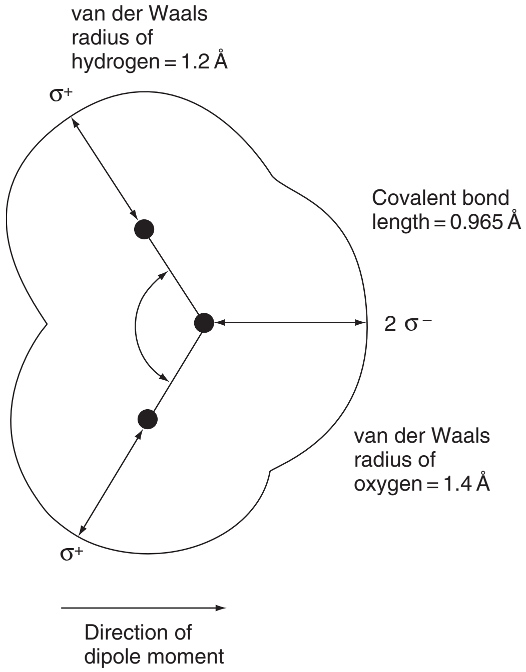


Fig. 2.1. Model of a water molecule. The curved lines represent the borders at which van der Waals attractions are counterbalanced by repulsive forces.

electrons in the molecule gives it electrical asymmetry. The electronegative oxygen atom attracts the electrons of the hydrogen atoms, leaving the hydrogen nuclei bare, so each of the two hydrogen atoms has a local partial positive charge. The oxygen atom, in turn, has a partial negative charge, located in the zone of the unshared orbitals. Though water molecules have no net charge, they form electrical dipoles.

HYDROGEN BONDING

Every hydrogen proton, though attached primarily to a particular molecule, is also attracted to the oxygen of a neighboring molecule, with which it forms a secondary link known as a *hydrogen bond*. Though the intermolecular link resulting from dipole attraction is not as strong as the primary link of the hydrogen to the oxygen of its own molecule, water can be regarded as a polymer of hydrogen-bonded molecules. This structure is most complete in ice crystals, in which each molecule is linked to four neighbors via four hydrogen bonds, thus forming a hexagonal lattice with a rather open structure (Fig. 2.2). When ice melts, this structure breaks partially, so additional molecules can enter the intermolecular spaces and each molecule can have more than four near neighbors. For this reason, liquid water can be more dense than ice at the same temperature, and thus lakes and ponds develop a surface ice sheet in winter rather than freeze solid from bottom to top as they would if ice were denser than liquid water.

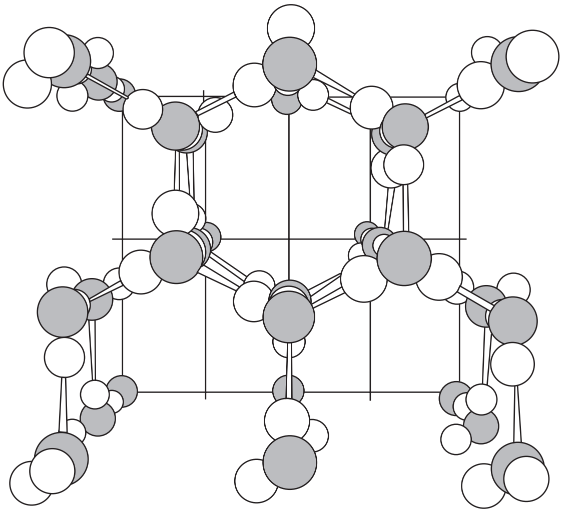


Fig. 2.2. Schematic structure of an ice crystal. The oxygen atoms are shown in gray and the hydrogen atoms in white. The pegs linking adjacent molecules represent hydrogen bonds.

STATES OF WATER

In the gaseous state, water molecules are largely independent of one another and occur mostly as monomers, designated $(H_20)_1$. Occasionally, colliding molecules fuse to form dimers $(H_20)_2$ or trimers $(H_20)_3$, but such combinations are rare. In the solid state, at the other extreme, a rigidly structured lattice forms with a tetrahedral configuration (Fig. 2.2) that can be depicted schematically as sheets of puckered hexagonal rings (Fig. 2.3). Nine alternative ice forms can occur when water freezes, depending on prevailing temperature and pressure. Figure 2.3 pertains to ice 1, the familiar form, which occurs and is stable at ordinary atmospheric pressure.

The orderly structure of ice does not totally disappear in the liquid state. The polarity and hydrogen bonds continue to bind water molecules together,

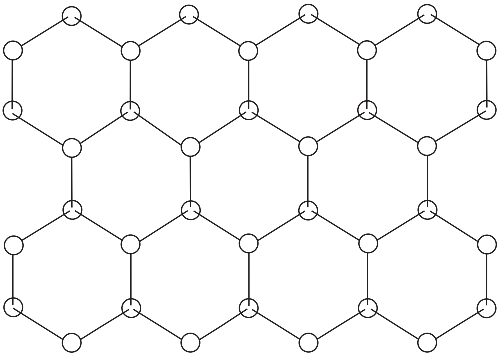


Fig. 2.3. The crystalline structure of ice.

though the structural forms that develop in the liquid state are much more flexible and transient than in the rigidly structured solid state. Hydrogen bonds in liquid water form an extensive three-dimensional network, the detailed features of which appear to be short-lived. According to the "flickering cluster" model, for instance, the molecules of liquid water associate and dissociate repeatedly in transitory or flickering polymer groups, designated $(H_20)_n$, having a quasi-crystalline internal structure. These microcrystals, as it were, form and melt so rapidly and randomly that, on a macroscopic scale, water appears to behave as a homogeneous liquid.

In transition from solid to liquid and from liquid to gas, hydrogen bonds must be broken (while in freezing and condensation they are reestablished). Hence relatively high temperatures and energies are required to achieve these transitions. To thaw 1 kg of ice, 3.35×10^5 J (80 cal/g) must be supplied. Conversely, the same energy (the *latent heat of fusion*) is released in freezing.

At the boiling point (100°C at atmospheric pressure), water passes from the liquid to the gaseous state and in so doing absorbs 2.26×10^6 J/kg (540 cal/g). This amount of heat is known as the *latent heat of vaporization*. Water can be vaporized at temperatures below 100°C, but such vaporization requires greater heat. At 30°C, the latent heat is about 2.43×10^6 J/kg (580 cal/g). *Sublimation* is the direct transition from the solid state to vapor, and the heat absorbed by it is equal to the sum of the latent heats of fusion and of vaporization.

OSMOTIC PRESSURE

Owing to the constant thermal motion of all molecules in a fluid (above a temperature of absolute zero), solute species spread throughout the solution in a spontaneous tendency toward a state of equal concentration throughout. This migration of solutes in response to spatial differences in concentration is called *diffusion*.

If a physical barrier is interposed across the path of diffusion, and if that barrier is permeable to molecules of the solvent but not to those of the solute, the former will diffuse through the barrier in a process called *osmosis* (from the Greek *osmos*, meaning "push"). As in the case of unhindered diffusion, this process tends toward a state of uniform concentration, even across the barrier. Barriers permeable to one substance in a solution but not to another are called *selective* or *semipermeable membranes*. Membranes surrounding cells in living organisms, for example, exhibit selective permeability to water while restricting the diffusion of solutes between the cells' interior and their exterior environment. Water molecules cross the membrane in both directions, but the net flow of water is from the more dilute solution to the more concentrated.

Figure 2.4a is a schematic representation of a pure solvent separated from a solution by a semipermeable membrane. Solvent will pass through the membrane and enter the solution compartment, driving the solution level up the left-hand tube until the hydrostatic pressure of the column of dilute solution on the left is sufficient to counter the diffusion pressure of

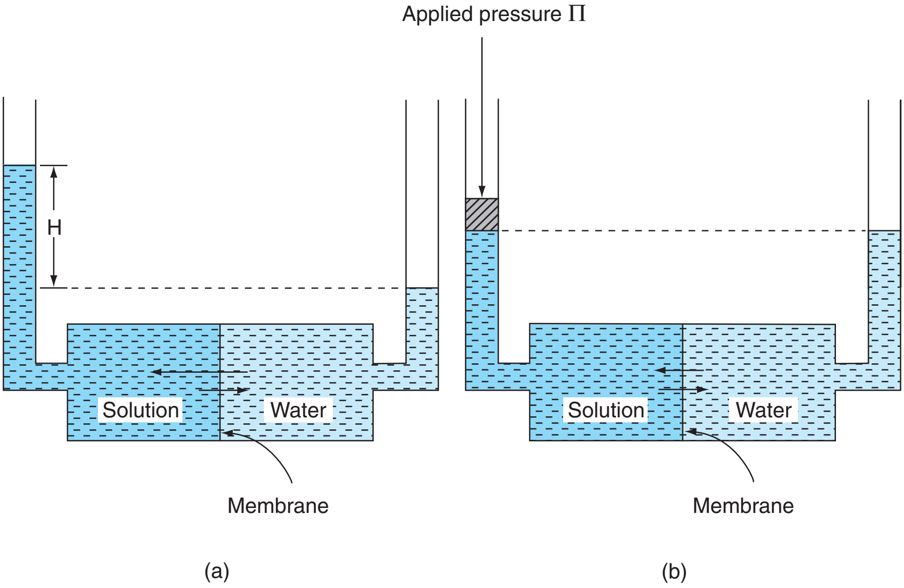


Fig. 2.4. Osmosis and osmotic pressure: (a) In osmosis, the flow of water molecules through the membrane into the solutions is at first greater than the reverse flow from the solution into the water compartment. The hydrostatic pressure due to the column of expanded solution increases the rate of water flow from the solution to the water compartment until, at equilibrium, the opposite flows are equal. (b) The osmotic pressure of the solution is equal to the hydrostatic pressure Π that must be applied to the solution to equalize the rate of flow to and from the solution and produce a net flow of zero.

the solvent molecules drawn into the solution through the membrane. The hydrostatic pressure at equilibrium, when solvent molecules are crossing the membranes in both directions at equal rates, is the osmotic pressure of the solution.

In dilute solutions, the osmotic pressure is proportional to the concentration of the solution and to its temperature, according to the following equation:

$$\Pi = MRT \tag{2.1}$$

Here Π is the osmotic pressure in atmospheres (to be multiplied by 0.101 to obtain megapascal units), M is the total molar concentration of the solute (whether molecules or dissociated ions), T is the temperature in degrees Kelvin, and R is the gas constant (0.08205 L atm/deg mole). The osmotic pressure increase with temperature is associated with the corresponding increase of the molecular diffusivity (self-diffusion coefficient) of water, D_w. According to the Einstein–Stokes equation,

$$D_w = kT/6\pi\eta r$$

where $k = R/N$, the Boltzmann constant (1.38×10^{-23} J/K), r is the rotation radius of the molecule ($\sim 1.5 \times 10^{-4}$), η is the viscosity, and N is Avagadro's number.

Sample Problem

Calculate the osmotic pressure of a 0.01 *M* solution of sodium chloride at 20°C, assuming complete dissociation into Na^+ and Cl^- ions.

$$\begin{aligned}\Pi = MRT &= (2 \times 0.01 \text{ mol/L})(0.08205 \text{ L atm/mole K})(293 \text{ K})\\ &= 0.48 \text{ atm} = 48.5 \text{ kPa}\end{aligned}$$

SOLUBILITY OF GASES

The concentration of dissolved gases in water in equilibrium with a gaseous phase generally increases with pressure and decreases with temperature. According to Henry's law, the mass concentration of a dissolved gas c_m is proportional to the partial pressure of the gas p_i in the ambient atmosphere:

$$c_m = s_c p_i / p_0 \tag{2.2}$$

where s_c is the solubility coefficient of the particular gas in water and p_0 is the total pressure of the atmosphere. The volume concentration is similarly proportional:

$$c_v = s_v p_i / p_0 \tag{2.3}$$

where s_v is the solubility coefficient expressed in terms of volume ratios (i.e., c_v is the volume of dissolved gas relative to the volume of water). Both s_c and s_v are determined experimentally. If the gas does not react chemically with the liquid, these properties should remain constant over a range of pressures, especially at low partial pressures of the dissolved gases. Solubility is, however, strongly influenced by temperature. Table 2.1 gives the s_v values of several atmospheric gases at various temperatures.

The solubilities of various gases (particularly oxygen) in varying conditions strongly influence such vital soil processes as oxidation and reduction, as well as respiration by roots and microorganisms.

TABLE 2.1 Solubility Coefficients of Gases in Water

Temperature (°C)	Nitrogen (N_2)	Oxygen (O_2)	Carbon dioxide (CO_2)	Air (without CO_2)
0	0.0235	0.0489	1.713	0.0292
10	0.0186	0.0380	1.194	0.0228
20	0.0154	0.0310	0.878	0.0187
30	0.0134	0.0261	0.665	0.0156
40	0.0118	0.0231	0.530	—

Sample Problem

A liter of water at 25°C dissolves 0.0283 L of oxygen when the pressure of oxygen in equilibrium with the solution is 1 atm (101 kPa). From this we can find the proportionality constant s_v in Henry's law with $P_{O_2} = 1$ atm, as follows:

$$\text{Restating Henry's law:} \quad c_v = s_v(P_{O_2}/P_{total})$$

(where c_v is the volume of gas dissolved at equilibrium with a partial pressure P_{O_2}/P_{total}) we can write

$$s_v = c_v/(P_{O_2}/P_{total}) = 0.0283 \text{ L/L atm}$$

With the prevailing pressure of oxygen in normal dry air equal to 159 mm Hg,

$$\begin{aligned} c_v = s_v(P_{O_2}/P_{total}) &= (0.0283)(159/760) = 0.00592 \text{ L/L } H_2O \\ &= 5.92 \text{ mL/L} \end{aligned}$$

ADSORPTION OF WATER ON SOLID SURFACES

Adsorption is an interfacial phenomenon resulting from the differential forces of attraction or repulsion occurring among molecules or ions of different phases at their exposed surfaces. As a result of cohesive and adhesive forces coming into play, the zones of contact among phases may exhibit a concentration or a density of material different from that inside the phases themselves. As different phases come in contact, various types of adsorption can occur: adsorption of gases on solids, of gases on liquid surfaces, and of liquids on solids.

The interfacial forces of attraction or repulsion may themselves be of different types, including electrostatic or ionic (coulombic) forces, intermolecular forces such as van der Waals and London forces, and short-range repulsive (Born) forces. The adsorption of water upon solid surfaces is generally of an electrostatic nature. The polar water molecules attach to the charged faces of the solids and to the ions adsorbed on them. This adsorption of water is the mechanism causing the strong retention of water by clay at high suctions.

The interaction of the charges of the solid with the polar water molecules may impart to the adsorbed water a distinct structure in which the water dipoles assume an orientation dictated by the charge sites on the solids. This adsorbed "phase" may have mechanical properties of strength and viscosity that differ from those of ordinary ("free") liquid water at the same temperature. The adsorption of water on clay surfaces is an exothermic process, resulting in the liberation of an amount of heat known as the *heat of wetting*.

VAPOR PRESSURE

According to the kinetic theory, molecules in a liquid are in constant motion due to their thermal energy. These molecules collide frequently, and occasionally one or another at the surface absorbs sufficient momentum to leap out of the liquid and into the atmosphere above it. Such a molecule, by virtue of its kinetic energy, thus changes from the liquid to the gaseous phase. This kinetic energy is then lost in overcoming the potential energy of intermolecular attraction while escaping from the liquid. At the same time, some of the randomly moving molecules in the gaseous phase may strike the surface of the liquid and be absorbed in it.

The relative rates of these two directions of movement depend upon the concentration of vapor in the atmosphere. An atmosphere that is at equilibrium with a body of pure water at standard atmospheric pressure is considered to be saturated with water vapor, and the partial pressure of vapor in such an atmosphere is called the *saturation vapor pressure*. The vapor pressure at equilibrium with any body of water depends on its physical condition (pressure and temperature) and on its chemical state (affected by the nature and concentration of solutes) but is independent of the absolute or relative quantity of liquid or gas in the system.

The saturation vapor pressure rises with temperature. As the kinetic energy of the molecules in the liquid increases, so does the evaporation rate. Consequently, a higher concentration of vapor in the atmosphere is required for the rate of return to the liquid to match the rate of escape from it. A liquid arrives at its boiling point when the vapor pressure becomes equal to the atmospheric pressure. If the temperature range is not too wide, the dependence of saturation vapor pressure on temperature is expressible by the following equation (see Table 2.2):

$$\ln p_0 = a - b/T \tag{2.4}$$

where $\ln p_0$ is the logarithm to the base e of the saturation vapor pressure p_0, T is the absolute temperature, and a and b are constants.

TABLE 2.2 Physical Properties of Water Vapor

Temperature (°C)	Saturation vapor pressure (torr)		Vapor density in saturated air (kg/m^3)		Diffusion coefficient (m^2/sec)
	Over liquid	Over ice	Over liquid	Over ice	
−10	2.15	1.95	2.36×10^{-3}	2.14×10^{-3}	0.211×10^{-4}
−5	3.16	3.01	3.41×10^{-3}	3.25×10^{-3}	—
0	4.58	4.58	4.85×10^{-3}	4.85×10^{-3}	0.226×10^{-4}
5	6.53	—	6.80×10^{-3}	—	
10	9.20	—	9.40×10^{-3}	—	0.241×10^{-4}
15	12.78	—	12.85×10^{-3}	—	
20	17.52	—	17.30×10^{-3}	—	0.257×10^{-4}
25	23.75	—	23.05×10^{-3}	—	
30	31.82	—	30.38×10^{-3}	—	0.273×10^{-4}
35	42.20	—	39.63×10^{-3}	—	
40	55.30	—	51.1×10^{-3}	—	0.289×10^{-4}
45	71.90	—	65.6×10^{-3}	—	
50	92.50	—	83.2×10^{-3}	—	

The vapor pressure also depends also on the hydrostatic pressure of the liquid water. At equilibrium with drops of water (which have a hydrostatic pressure greater than atmospheric), the vapor pressure is greater than in a state of equilibrium with free water (which has a flat interface with the atmosphere). On the other hand, in equilibrium with adsorbed or capillary water under a hydrostatic pressure smaller than atmospheric, the vapor pressure is smaller than in equilibrium with free water. The curvature of drops is considered to be positive, because these drops are convex toward the atmosphere, whereas the curvature of capillary water menisci is considered negative, because they are concave toward the atmosphere.

For water in capillaries with a concave air–water interface, the Kelvin equation applies:

$$-(\mu_l - \mu_l^\circ) = \mathrm{RT}\ \ln\ (p_l^\circ/p_l) = 2\gamma v_1 \cos\ \alpha/\mathrm{r_c}$$

in which $(\mu_l - \mu_l^\circ)$ is the change in potential of the water due to the curvature of the air–water interface, γ is the surface tension of water, α is the contact angle, v_1 is the partial molar volume of water, and r_c is the radius of the capillary.

Water present in the soil invariably contains solutes, generally electrolytic salts. The composition and concentration of the soil solution affect soil behavior. In humid regions the soil solution may have a concentration of only a few parts per million, but in arid regions it may be as high as several percent. The vapor pressure of electrolytic solutions is less than that of pure water:

$$v_1\Pi_o = RT\ \ln\ (p_1^\circ/p_l) = \mu_l - \mu_l^\circ$$

where Π_o is the osmotic pressure of the solution, μ_1° and p_1° are the chemical potential and vapor pressure of the pure liquid, and μ_l and p_l are the same for the solution. The soil solution has a lower vapor pressure than pure water, even when the soil is saturated. In unsaturated soil, capillary and adsorptive effects further lower the potential and hence also the vapor pressure.

SURFACE TENSION

Surface tension is a phenomenon occurring typically, though not exclusively, at the interface of a liquid and a gas. The liquid behaves as if it were covered by an elastic membrane in a constant state of tension that has the effect of causing the surface to contract. To be sure, no such membrane exists, yet the analogy is a useful one if not taken too literally.

If we draw an arbitrary line of length L on a liquid surface, there will be a force F pulling the surface to the right of the line and an equal force pulling the surface leftward. The ratio F/L is the surface tension, and its dimensions are those of force per unit length. The same phenomenon can also be described in terms of energy. Increasing the surface area of a liquid requires work, which remains stored as potential energy in the enlarged surface, just as energy can be stored in a stretched spring. That potential energy can perform work if the enlarged surface is allowed to contract again. Energy per unit area has the same dimensions as force per unit length.

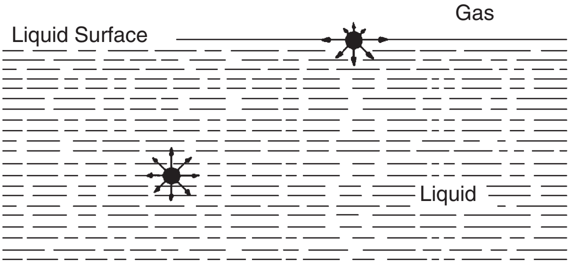


Fig. 2.5. Cohesive forces acting on a molecule inside the liquid and at its surface.

An explanation for the occurrence of surface tension is given in Fig. 2.5. A molecule inside the liquid is attracted to neighboring molecules in all directions equally by cohesive forces, while a molecule at the surface of the liquid is attracted into the relatively dense liquid phase by a net force greater than that attracting it toward the rarified gaseous phase. This unbalanced force draws the surface molecules inward into the liquid and results in the tendency for the surface to contract. This is why drops of a liquid in air (as well as bubbles of air in a liquid) assume the shape of a sphere, which is a body of minimal surface exposure relative to its volume.

Different liquids exhibit different surface tension values, as in the following list:

Water, 7.27×10^{-2} N/m (72.7 dyn/cm) at 20°C

Ethyl ether, 1.7×10^{-2} N/m (17 dyn/cm)

Ethyl alcohol, 2.2×10^{-2} N/m (22 dyn/cm)

Benzene, 2.9×10^{-2} N/m (29 dyn/cm)

Mercury, 0.43 N/m (430 dyn/cm)

CURVATURE OF WATER SURFACES AND HYDROSTATIC PRESSURE

Wherever an interface between fluids (say, between water and air) is not planar but curved, the resolution of forces due to surface tension creates a pressure differential across that interface. For a spherical interface (as in the case of a bubble of air immersed in a body of water), the pressure difference is proportional to the surface tension and inversely proportional to the curvature:

$$\Delta P = 2\gamma / R \tag{2.5}$$

Thus, the smaller the bubble is, the greater is its pressure.

If the bubble is not spherical, then instead of Eq. (2.5) we obtain

$$\Delta P = \gamma(1/R_1 + 1/R_2) \tag{2.6}$$

where R_1 and R_2 are the principal radii of curvature for any given point on the interface. Equation (2.6) reduces to (2.5) whenever $R_1 = R_2$.

Sample Problem

Calculate the hydrostatic pressure of water in raindrops of the following diameters: (a) 5; (b) 1; (c) 0.2 mm. Assume 4°C. What are the corresponding pressures at 25°C?

The relevant equation is: $\Delta P = 2\gamma/R$. At 4°C the surface tension γ is 7.5×10^{-2} N/m (75.0 dyn/cm).

(a) $\Delta P = (2 \times 7.5 \times 10^{-2}\ \text{N/m})/(2.5 \times 10^{-3}\ \text{m}) = 60\ \text{Pa} = 0.6\ \text{mbar}$
(b) $\Delta P = (2 \times 7.5 \times 10^{-2}\ \text{N/m})/(0.5 \times 10^{-3}\ \text{m}) = 300\ \text{Pa} = 3.0\ \text{mbar}$
(c) $\Delta P = (2 \times 7.5 \times 10^{-2}\ \text{N/m})/(0.1 \times 10^{-3}\ \text{m}) = 1500\ \text{Pa} = 15.0\ \text{mbar}$

At 25°C the surface tension of water is 7.19×10^{-2} N/m (71.9 dyn/cm).

(a) $\Delta P = (2 \times 7.19 \times 10^{-2}\ \text{N/m})/(2.5 \times 10^{-3}\ \text{m}) = 57.5\ \text{Pa} = 0.575\ \text{mbar}$
(b) $\Delta P = (2 \times 7.19 \times 10^{-2}\ \text{N/m})/(0.5 \times 10^{-3}\ \text{m}) = 287.6\ \text{Pa} = 2.876\ \text{mbar}$
(c) $\Delta P = (2 \times 7.19 \times 10^{-2}\ \text{N/m})/(0.1 \times 10^{-3}\ \text{m}) = 1438\ \text{Pa} = 14.38\ \text{mbar}$

We note that the hydrostatic pressure in drops is greatly influenced by drop size but not much by temperature in the range considered.

CONTACT ANGLE OF WATER ON SOLID SURFACES

If we place a drop of liquid on a dry solid surface, the liquid will usually displace the gas that covered the surface of the solid and it will spread over that surface to a certain extent. Where its spreading ceases and the edge of the drop comes to rest, it will form a typical angle with the surface of the solid. This angle, termed *contact angle*, is illustrated in Fig. 2.6.

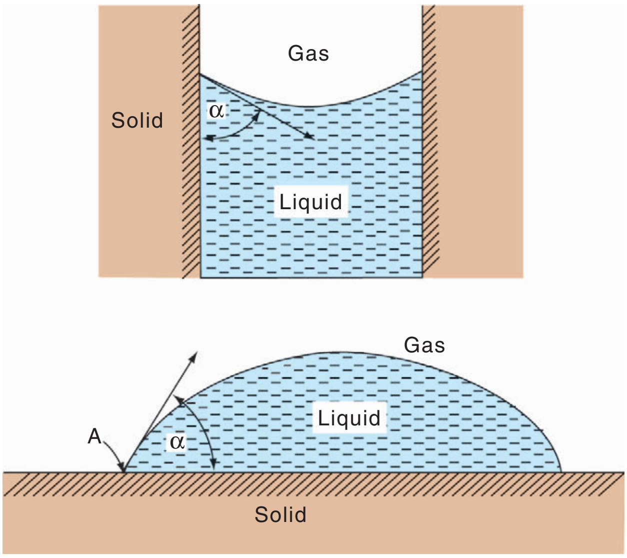


Fig. 2.6. The contact angle of a meniscus in a capillary tube and a drop resting on the surface of a solid.

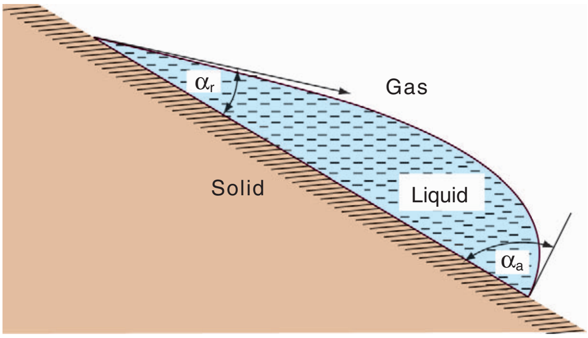


Fig. 2.7. Hypothetical representation of a drop resting on an inclined surface. The contact angle α_3 at the advancing edge of the drop is shown to be greater than the corresponding angle α_r at the receding edge.

We now consider what factors determine the magnitude of the angle α. We can expect that angle to be acute if the adhesive affinity between the solid and liquid is strong relative to the cohesive forces inside the liquid itself and to the affinity between the gas and the solid. We can then say that the liquid "wets" the solid. A contact angle of zero would mean the complete flattening of the drop and the perfect wetting of the solid surface by the liquid. On the other hand, a contact angle of 180° would imply a complete nonwetting or rejection of the liquid by the gas-covered solid. In that case the drop would retain its spherical shape without spreading over the surface at all (assuming no gravity effect). Surfaces on which water exhibits an obtuse contact angle are called *water repellent* or *hydrophobic* (Greek: "water-hating").

The contact angle of a given liquid on any particular solid is generally characteristic of their interaction under given physical conditions. This angle, however, may be different in the case of a liquid that is advancing over the surface than in the case of the same liquid receding over the surface. This phenomenon, where it occurs, is called *contact angle hysteresis*. The wetting angle of pure water on clean and smooth mineral surfaces is generally zero, but where the surface is rough or coated with adsorbed surfactants of a hydrophobic nature, the contact angle, and especially the wetting angle, can be considerably greater than zero and may even exceed 90°. This phenomenon is illustrated in given Fig. 2.7.

THE PHENOMENON OF CAPILLARITY

A capillary tube dipped in a body of free water will form a meniscus as the result of the contact angle of water with the walls of the tube. The curvature of this meniscus will be greater (i.e., the radius of curvature will be smaller) the narrower the tube. The occurrence of curvature causes a pressure difference to develop across the liquid–gas interface.

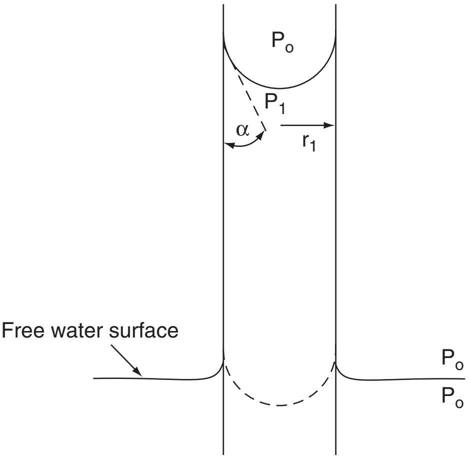


Fig. 2.8. Capillary rise.

A liquid with an acute contact angle (e.g., water on glass) will form a concave meniscus, and therefore the liquid pressure under the meniscus P_1 will be smaller than the atmospheric pressure P_0 (Fig. 2.8). Hence, water inside the tube will be driven up the tube from its initial location (shown as a dashed curve in Fig. 2.8) by the greater pressure of the free water (i.e., water at atmospheric pressure, under a horizontal air–water interface) at the same level.

The upward movement will stop when the pressure difference between the water inside the tube and the water under the flat surface outside the tube is countered by the hydrostatic pressure exerted by the water column in the capillary tube. If the capillary tube is cylindrical and if the contact angle of the liquid on the walls of the tube is zero, the meniscus will be a hemisphere (and in a two-dimensional drawing can be represented as a semicircle) with its radius of curvature equal to the radius of the capillary tube. If, however, the liquid contacts the tube at an angle greater than zero but smaller than 90°, then the diameter of the tube ($2r$) is the length of a chord cutting a section of a circle with an angle of $\pi - 2\alpha$, as shown in Fig. 2.9. Thus,

$$R = r/\cos\alpha \tag{2.7}$$

Here R is the radius of curvature of the meniscus, r is the radius of the capillary, and α is the contact angle.

The pressure difference ΔP between the capillary water (under the meniscus) and the atmosphere, therefore, is

$$\Delta P = (2\gamma\cos\alpha)/r \tag{2.8}$$

Recalling that hydrostatic pressure is proportional to the depth d below the free water surface (i.e., $P = \rho g d$, where ρ is liquid density and g is gravitational acceleration), we can infer that the hydrostatic tension (negative pressure) in

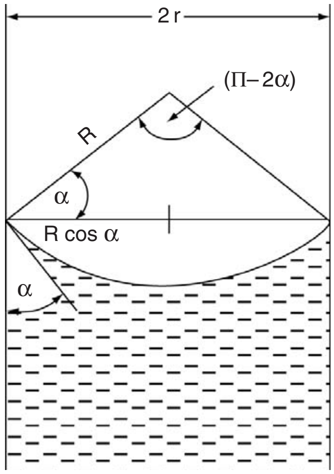


Fig. 2.9. The geometric relationship of the radius of curvature *R* to the radius of the capillary *r* and the contact angle α.

a capillary tube is proportional to the height *h* above the free water surface. Hence the height of capillary rise is

$$h_c = (2\gamma \cos \alpha)/g(\rho_1 - \rho_g)\gamma \tag{2.9}$$

where ρ_g is the density of the gas (generally neglected), ρ_l is the density of the liquid, g is the acceleration of gravity, r is the capillary radius, α is the contact angle, and γ is the surface tension.

When the liquid surface is concave, the center of curvature lies outside the liquid and the curvature, by convention, is regarded as negative. Thus, for a concave meniscus such as that of water in a clean glass capillary, ΔP is negative with reference to the atmosphere, indicating a capillary pressure deficit, or subpressure, commonly called *tension*. On the other hand, in a capillary tube that forms a convex meniscus (such as in the case of mercury in glass or of water in a waxy or otherwise water-repellent tube), ΔP would be positive, and hence capillary depression, rather than capillary rise, will result.

Sample Problems

Calculate the equilibrium capillary rise of water and mercury at 20°C in glass cylindrical capillary tubes of the following diameters: (a) 2 mm; (b) 0.5 mm; (c) 0.1 mm. Disregard the density of atmosphere.

$$\Delta h = (2\gamma \cos \alpha)/\ (\rho g r)$$

For water: $\gamma = 7.27 \times 10^{-2}$ kg/sec^2 (= N/m); $\alpha = 0$; $\rho = 998$ kg/m^3

(a) $r = 10^{-3}$ m:

$$\Delta h = (2 \times 7.27 \times 10^{-2}\ \text{kg/sec}^2)/(998\ \text{kg/m}^3 \times 9.81\ \text{m/sec}^2 \times 10^{-3}\text{m})$$
$$= 1.48 \times 10^{-2}\ \text{m} = 1.48\ \text{cm}$$

(b) $r = 2.5 \times 10^{-4}$ m:

$\Delta h = (2 \times 7.27 \times 10^{-2}$ kg/sec^2)/(998 kg/m$^3 \times 9.81$ m/sec$^2 \times 2.5 \times 10^{-4}$ m
$= 5.92 \times 10^{-2}$ m = 5.92 cm

(c) $r = 5.0 \times 10^{-5}$ m:

$\Delta h = (2 \times 7.27 \times 10^{-2}$ kg/sec^2)/(998 kg/m$^3 \times 9.81$ m/sec$^2 \times 5.0 \times 10^{-5}$ m
$= 29.6 \times 10^{-2}$ m = 29.6 cm

For mercury: $\gamma = 0.43$ kg/sec^2; $\alpha = 180°$; $\rho = 1{,}360$ kg/m^3

(a) $r = 10^{-3}$ m:

$\Delta h = [2 \times 0.43$ kg/sec$^2 \times (-1)]/(1360$ kg/m$^3 \times 9.81$ m/sec$^2 \times 10^{-3}$ m)
$= -0.64 \times 10^{-2}$ m = -0.64 cm (capillary depression)

(b) $r = 2.5 \times 10^{-4}$ m:

$\Delta h = [2 \times 0.43$ kg/sec$^2 \times (-1)]/(1360$ kg/m$^3 \times 9.81$ m/sec$^2 \times 2.5 \times 10^{-4}$ m)
$= -2.58 \times 10^{-2}$ m = -2.58 cm

(c) $r = 5.0 \times 10^{-5}$ m:

$\Delta h = [2 \times 0.43$ kg/sec$^2 \times (-1)]/(1360$ kg/m$^3 \times 9.81$ m/sec$^2 \times 5.0 \times 10^{-5}$ m
$= -12.9 \times 10^{-2}$ m = -12.9 cm

DYNAMIC AND KINEMATIC VISCOSITY

When a fluid is moved in shear (i.e., when adjacent layers of fluid are made to slide over each other), the force required is proportional to the velocity of shear. The proportionality factor is called the *viscosity*. As such, it is the property of the fluid to resist the rate of shearing and can be visualized as an internal friction. The coefficient of viscosity η is defined as the force per unit area necessary to maintain a velocity difference of 1 m/sec between two parallel layers of fluid that are 1 m apart. The viscosity equation is

$$\tau = F_s/A = \eta du/dx \qquad (2.10)$$

where τ is the shearing stress, consisting of a force F_s acting on an area A; η [dimensions: mass/ (length × time)] is the *coefficient of dynamic viscosity*; and du/dx is the velocity gradient perpendicular to the stressed area A.

The ratio of the dynamic viscosity of a fluid to its density is called the *kinematic viscosity*, designated ν. It expresses the shearing-rate resistance of a fluid independent of the density. While the dynamic viscosity of water exceeds that of air by a factor of about 50 (at room temperature), its kinematic viscosity is actually lower.

Liquids of lower viscosity flow more readily and are said to possess greater *fluidity* (the reciprocal of viscosity). The viscosity of water diminishes by over 2% per 1°C rise in temperature, and thus it decreases by more than half as the temperature increases from 5 to 35°C. The viscosity is also affected by the type and concentration of solutes present.

BOX 2.1 The Shape of a Raindrop

What is the shape of a free-falling raindrop? Most people answer that question unhesitatingly: A raindrop is shaped like a teardrop, flattened at the bottom and tapering to a point at the top. That nugget of conventional wisdom is so embedded in our collective consciousness that the "standard" teardrop is used not only in cartoons but also in respectable scientific publications. Irrigation companies emblazon it on their brochures as a logo, and so do agencies of the United Nations Organization dealing with the supply and utilization of water by the world's thirsty population.

Alas, the vertically elongated, top-pointed teardrop falling through the air is a physical impossibility. Although a drop may have that shape at the instant of detachment from the tip of a wire or a capillary, it will naturally "ball up" to form a sphere once it is suspended in air. Raindrops, in any case, do not form at the tip of a wire, but are condensed from vapor as small droplets, which later grow by coalescence. They are spherical from the start. Then, as they fall vertically through the air, other effects modify their shape somewhat.

Air flowing around the falling drop has a greater velocity near the horizontal "equator" of the drop than near its top and bottom "poles." According to Bernoulli's law, relating the variation of speed to the variation of pressure along streamlines ($p + \rho q^2/2 = \text{constant}$, where p is the pressure, q is the velocity, and ρ is the density of the fluid), the air around the sides of the drop will acquire a pressure smaller than the air above and below the drop, hence the drop will tend to compress vertically and will come to resemble an ellipsoid (with its horizontal axis longer than its vertical axis). If the drop were to continue accelerating, it might eventually come to resemble a pancake. However, Stokes' law enters the picture, with air resistance countering the acceleration and establishing a terminal velocity (which depends on drop size). Turbulence at its wake may cause the drop to resemble a hamburger bun, and random air currents might cause the drop to flutter, but its shape remains more or less the same. And so we can bid farewell to sad tears, as myriad happy raindrops hit the ground surface with their flattened faces and are absorbed into the soil, providing vital moisture to the roots of growing plants.

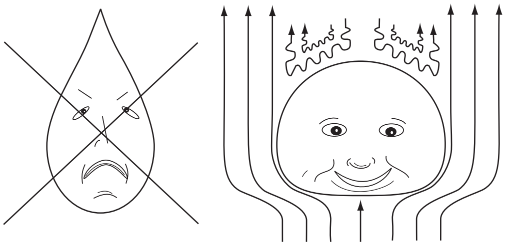


Fig. B2.1. Conventional and real raindrops.

Part II

THE SOLID PHASE

This Page Intentionally Left Blank

3. PARTICLE SIZES, SHAPES, AND SPECIFIC SURFACE

SOIL TEXTURE

Having defined the soil as a three-phase system, we now focus attention on the solid phase, which is the permanent component of the soil. We can readily imagine a soil without air or without water and, in a vacuum, without both (as is the case with the "soil" found on the moon), but there could hardly be a soil without the solid phase. The soil's solid phase consists of mineral particles having various shapes and sizes, as well as of amorphous compounds such as organic matter and hydrated iron oxides, which are generally attached to the particles. Because the content of the amorphous material is generally (though not invariably) small, we can in most cases represent the solid phase as consisting mainly of discrete particles. The largest soil particles are often visible to the naked eye, whereas the smallest are colloidal and can be observed only with the aid of an electron microscope.

In general, it is possible to classify or group soil particles according to their sizes and to characterize the soil as a whole in terms of the relative proportions of those size groups. The groups may differ from one another in mineral composition as well as in particle size. It is these attributes of the soil solid phase, particle size and mineral composition, that largely determine the nature and behavior of the soil: its internal geometry and porosity, its interactions with fluids and solutes, as well as its compressibility, strength, and thermal regime.

The term *soil texture* refers to the range of particle sizes in a soil, that is, whether a given soil contains a wide or relatively narrow array of particle sizes and whether the particles are mainly large, small, or of intermediate size. The term thus carries both qualitative and quantitative connotations. Qualitatively, it represents the "feel" of the soil material, whether coarse and gritty or fine and smooth to the touch. An experienced soil classifier can feel, by rubbing the moistened soil with his fingers, whether it is coarse textured or fine textured, and can even assess semiquantitatively to which of the textural "classes" the soil might belong.

In a more rigorously quantitative sense, however, the term *soil texture* denotes the precisely measured distribution of particle sizes and the proportions of the various size ranges of particles composing a given soil. As such, soil texture is an intrinsic attribute of the soil and the one most often used to characterize its physical makeup.

TEXTURAL FRACTIONS

The traditional method of characterizing particle sizes in soils is to divide the array of particle diameters into three conveniently separable size ranges known as *textural fractions* or *separates*, namely, *sand*, *silt*, and *clay*. The actual procedure of separating out these fractions and of measuring their proportions is called *mechanical analysis*, for which standard techniques have been devised. The results of this analysis yield the *mechanical composition* of the soil, a term often used interchangeably with *soil texture*.

Unfortunately, there is as yet no universally accepted scheme for classifying particle sizes. For instance, the classification standardized in America by the U.S. Department of Agriculture differs from that of the International Soil Science Society (ISSS), as well as from those promulgated by the American Society for Testing Materials (ASTM), the Massachusetts Institute of Technology (MIT), and various national institutes abroad. The classification followed by soil engineers often differs from that of agricultural soil scientists. The same terms are used to designate differing size ranges, an inconsistency that can be confusing indeed. Several of the often-used particle size classification schemes are compared in Fig. 3.1.

An essential criterion for determining soil texture is the upper limit of particle size that is to be included in the definition of *soil material*. Some soils contain large rocks that obviously do not behave like soil, but, if numerous, might affect the behavior of the soil in bulk. The conventional definition of soil material includes particles smaller than 2 mm in diameter. Larger particles are generally referred to as *gravel*, and still larger rock fragments, several centimeters in diameter, are variously called *stones*, *cobbles*, or — if very large — *boulders*.

The largest particles that are generally recognized as soil material are designated *sand*, defined as particles ranging in diameter from 2000 μm (2 mm) to 50 μm (USDA classification) or to 20 μm (ISSS classification). The sand fraction is often further divided into subfractions such as coarse, medium, and fine sand. Sand grains usually consist of quartz but may also be fragments of feldspar, mica, and, occasionally, heavy minerals such as zircon, tourmaline, and hornblende, though the latter are rather rare. In most cases, sand grains have more or less uniform dimensions and can be represented as spherical, though they are not necessarily smooth and may in fact have quite jagged surfaces (Fig. 3.2). That, together with their hardness, accounts for their abrasiveness.

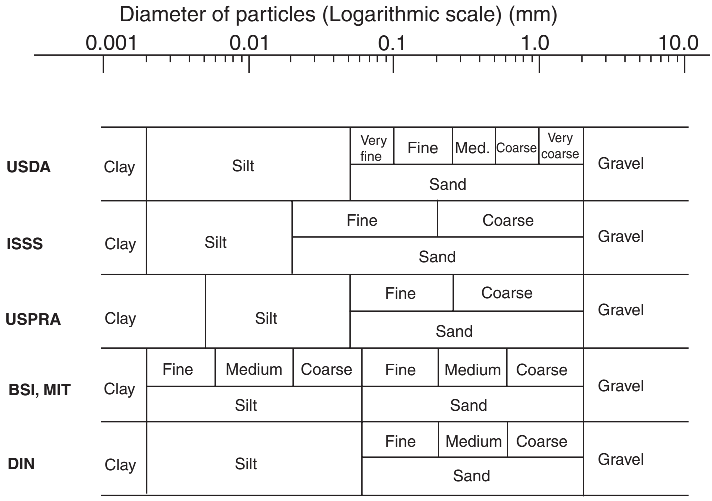


Fig. 3.1. Several conventional schemes for the classification of soil fractions according to particle diameter ranges; U.S. Department of Agriculture (USDA); International Soil Science Society (ISSS); U.S. Public Roads Administration (USPRA); British Standards Institute (BSI); Massachusetts Institute of Technology (MIT); German Standards (DIN).

The next fraction is silt, which consists of particles intermediate in size between sand and clay. Mineralogically and physically, silt particles generally resemble sand particles. However, since the silt are smaller, particles have a greater surface area per unit mass, and are often coated with strongly adherent clay, silt may exhibit, to a limited degree, some of the physicochemical characteristics generally attributed to clay.

The clay fraction, with particles ranging from 2 μm downwards, is the colloidal fraction. Clay particles are characteristically platelike or needlelike in shape and generally belong to a group of minerals known as the *aluminosilicates*. These are secondary minerals, formed in the soil itself in the course of its evolution from the primary minerals that were contained in the original rock. In some cases, however, the clay fraction may include particles (such as iron oxide and calcium carbonate) that do not belong to the aluminosilicate clay mineral category.

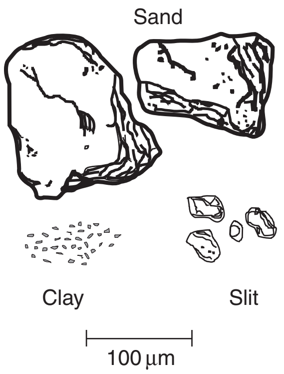


Fig. 3.2. A visual representation of the comparative sizes of sand, silt, and clay particles.

BOX 3.1 The Life Story of Sand

Sand is ubiquitous. We associate sand most commonly with beaches and deserts, but in fact it is an important constituent of nearly all soils. Where does sand come from?

Most sand originates in outcropping continental rocks, especially in mountainous areas that are exposed to weathering. Mechanical breakdown first dislodges rock fragments. Chemical assault follows mechanical disintegration in dissolving and recombining the more reactive constituent minerals, such as micas and feldspars, while the less reactive (more resistant) minerals — primarily quartz — remain as individual grains.

Sand grains are then carried downslope by surface runoff and are generally delivered into a stream. There they roll and bounce along the bottom of the streambed, resting temporarily in stagnant pools and then resuming their sporadic motion when a new flood arrives. Some of the river sand eventually reaches the seashore, though the journey may take hundreds of years. In the course of this journey, the grains are smoothed and polished by friction. Where masses of sand are deposited and dried, the winds take over, further sorting and redistributing the sand. The grains are propelled in spurts, alternately rising a few centimeters above the ground surface and bumping against other grains as they fall back to the surface at an oblique angle.

Because of its far greater surface area per unit mass and its resulting physicochemical activity, clay is the fraction with the most influence on soil behavior. Clay particles adsorb water and hydrate, thereby causing the soil to swell upon wetting and to shrink upon drying. Clay particles typically carry a negative electrostatic charge and, when hydrated, form an electrostatic double layer with exchangeable ions in the surrounding solution. Another expression of surface activity is the heat that evolves when dry clay is wetted, called the *heat of wetting*. A body of clay will typically exhibit plastic behavior and become sticky when moist and then cake up and crack to form cemented hard fragments when desiccated.

The relatively inert sand and silt fractions can be called the soil “skeleton,” while the clay, by analogy, can be thought of as the “flesh” of the soil. Together, all three fractions of the solid phase, as they are combined in various configurations, constitute the *matrix* of the soil.

TEXTURAL CLASSES

The overall textural designation of a soil, called *textural class*, is conventionally based on the mass ratios of the three fractions. Soils with different proportions of sand, silt, and clay are assigned to different classes, as shown in the triangular diagram of Fig. 3.3. To illustrate the use of the textural triangle, let us assume that a soil is composed of 50% sand, 20% silt, and 30% clay. Note that the lower left apex of the triangle represents 100% sand and the right side of the triangle represents 0% sand. Now find the point of 50% sand on the bottom edge of the triangle and follow the diagonally leftward line rising from that point and parallel to the zero line for sand. Next, identify the 20%

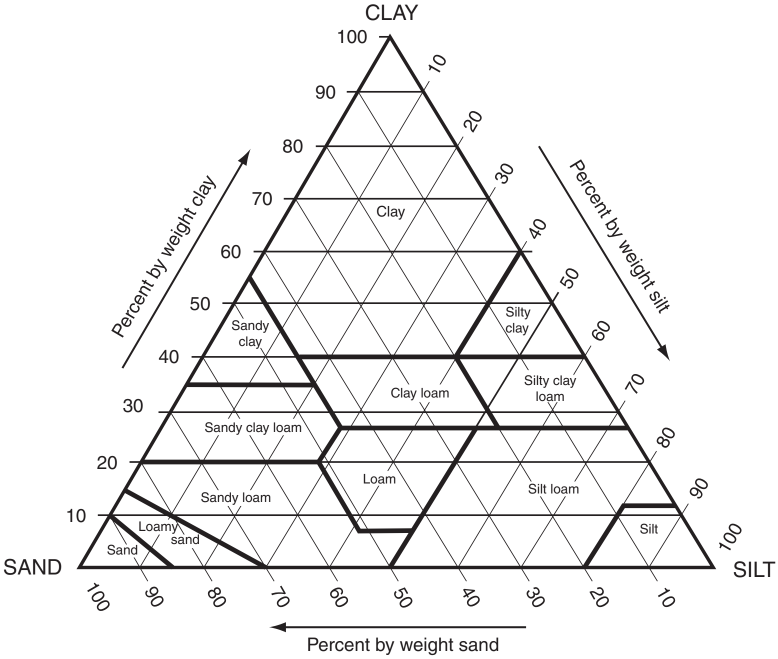


Fig. 3.3. Textural triangle, showing the percentages of clay (below 0.002 mm), silt (0.002–0.05 mm), and sand (0.05–2.0 mm) in the conventional soil textural classes.

line for silt, which is parallel to the zero line for silt, namely, the left edge of the triangle. Where the two lines intersect, they meet the 30% line for clay. The soil in this example happens to fit the category of "sandy clay loam."

Note that a class of soils called *loam* occupies a central location in the textural triangle. It refers to a soil that contains a ''balanced" mixture of coarse and fine particles, so its properties are intermediate among those of a sand, a silt, and a clay. As such, a loam is often considered to be the optimal soil for crop growth. Its capacity to retain water and nutrients is better than that of sand, while its drainage, aeration, and tillage properties are more favorable than those of clay. There are, however, exceptions to this generalization. Under different environmental conditions and for different plant species, a sand or a clay may be more suitable than a loam.

Sample Problem

Determine the textural class designations for soils with the distributions of particle sizes shown in Table 3.1.

Using the USDA classification (Fig. 3.1) and the textural triangle (Fig. 3.3), we obtain the following textural classes:

(a) % sand = 40, % silt = 45, % clay = 15; soil class: loam.
(b) % sand = 25, % silt = 60, % clay = 15; soil class: silt-loam.

(c) % sand = 20, % silt = 30, % clay = 50; soil class: clay.
(d) % sand = 60, % silt = 30, % clay = 10; soil class: sandy loam.

TABLE 3.1 Particle Size Distributions for Sample Problem 1

<0.0002 (%)	0.0002–0.002 (%)	0.002–0.01 (%)	0.01–0.05 (%)	0.05–0.25 (%)	0.25–2.0 (mm) (%)
(a) 5	10	20	25	20	20
(b) 6	9	30	30	15	10
(c) 10	30	30	10	10	10
(d) 4	6	10	20	30	30

PARTICLE SIZE DISTRIBUTION

Any attempt to divide into discrete fractions what is usually a continuous array of particle sizes is necessarily arbitrary, and the further classification of soils into distinct textural classes is doubly so. Although this approach is widely followed and in some ways useful, greater information is to be gained from measuring and displaying the complete array and distribution of particle sizes. Figure 3.4 presents typical *particle size distribution curves*. The ordinate of the graph indicates the percentage of soil particles with diameters smaller than the diameter denoted in the abscissa. The latter is drawn on a logarithmic scale to encompass several orders of magnitude of particle diameters while

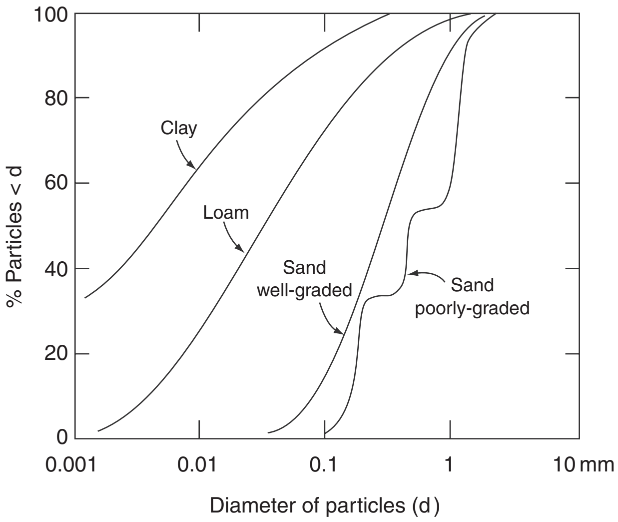


Fig. 3.4. Particle size distribution curves for various types of soil material (schematic).

allowing sufficient space for representing the fine particles. Note that this graph gives an integral, or cumulative, representation. In practice, the particle size distribution curve is constructed by connecting a series of n points, each expressing the cumulative mass fraction of particles finer than each of the n diameters measured ($F_l, F_2, \ldots, F_i, \ldots, F_n$). Thus,

$$F_i = (M_s - \Sigma M_i)/M_s \tag{3.1}$$

in which Ms is the total mass of the soil sample analyzed and ΣM_i is the cumulative mass of particles finer than the ith diameter measured.

The information obtainable from this representation of particle size distribution includes the diameter of the largest grains in the assemblage and the grading pattern, that is, whether the soil is composed of distinct groups of particles each of uniform size or whether it consists of a more or less uniform distribution of particle sizes. Soils having a preponderance of particles of one or several distinct sizes, indicating a steplike distribution curve, are called *poorly graded*. Soils with a flattened and smooth distribution curve (without apparent discontinuities) are called *well graded*.

This aspect of the particle size distribution can be expressed in terms of the so-called *uniformity index*, I_u, defined as the ratio of the diameter d_{60}, which includes 60% of the particles, to the smaller diameter d_{10}, which includes 10% of the particles (as shown in Fig. 3.4). This index, also called the *uniformity coefficient*, is used mostly with coarse-grained soils. For a soil material consisting entirely of equal-sized particles, if such were to exist, I_u would be unity. Some sand deposits may have uniformity indexes smaller than 10. Some well-graded soils, on the other hand, have I_u values greater than 1000.

The particle size distribution curve can also be differentiated graphically to yield a frequency distribution curve for grain sizes, with a peak indicating the most prevalent grain size. Attempts have been made to correlate this index, as well as the harmonic mean diameter of grains, with various soil properties, such as permeability (see Chapter 7).

MECHANICAL ANALYSIS

Mechanical analysis is the procedure for determining the particle size distribution of soil samples (Gee and Or, 2002). The first step in this procedure is to disperse the soil in an aqueous suspension. The primary soil particles, often naturally aggregated, must be separated and made discrete by removal of cementing agents (such as organic matter, calcium carbonate, and iron oxides) and by deflocculating the clay. Removal of organic matter is usually achieved by oxidation with hydrogen peroxide, and calcium carbonate can be dissolved by addition of hydrochloric acid. Deflocculation is carried out by means of a chemical dispersing agent (such as sodium metaphosphate) and by mechanical agitation (shaking, stirring, or ultrasonic vibration). The function of the dispersing agent is to replace the cations adsorbed to the clay, particularly divalent or trivalent cations, with sodium. This has the effect of increasing the hydration of the clay micelles, thus causing them to repel each other rather than coalesce, as they do in the flocculated state. Failure to disperse the soil completely will result in flocs of clay or aggregates settling as if they were silt-sized or sand-sized primary particles, thus biasing the results of mechanical analysis to indicate an apparent content of clay lower than the real value.

Actual separation of particles into size groups can be carried out by passing the suspension through graded sieves, down to a particle diameter of approximately 0.05 mm. To separate and classify still finer particles, the method of sedimentation is usually used, based on measuring the relative settling velocity of particles of various sizes in an aqueous suspension.

According to *Stokes' law*, the velocity of a spherical particle settling under the influence of gravity in a fluid of a given density and viscosity is proportional to the square of the particle's radius. We shall now derive this law, inasmuch as it governs the method of sedimentation analysis.

A particle falling in a vacuum will encounter no resistance as it is accelerated by gravity, so its velocity will increase as it falls. However, a particle falling in a fluid will encounter frictional resistance proportional to the product of its radius and velocity and to the viscosity of the fluid.

The resisting force due to friction F_r was shown by George Stokes in 1851 to be

$$F_r = 6\pi\eta r u \tag{3.2}$$

where η is the viscosity of the fluid, r is the radius of the particle, and u is its velocity. As the particle begins its fall, its velocity increases due to the acceleration of gravity. Eventually, the increasing resistance force equals the constant downward gravity force. Henceforth the particle continues to descend by inertia at a constant velocity known as the *terminal velocity* u_t.

The downward force due to gravity F_g is

$$F_g = (4/3)\pi r^3(\rho_s - \rho_f)g \tag{3.3}$$

Here $(4/3)\pi r^3$ is volume of the spherical particle, ρ_s is its density, ρ_f is the density of the fluid, and g is the acceleration of gravity. Subtracting fluid density from particle density accounts for the buoyancy that reduces the effective weight of particles in suspension (Archimedes' law).

Setting the two forces equal, we obtain Stokes' law:

$$u_t = (2/9)(r^2 g/\eta)(\rho_s - \rho_f) = (d^2 g/18\eta)(\rho_s - \rho_f) \tag{3.4}$$

where d is the diameter of the particle. Assuming that the terminal velocity is attained almost instantly, we can obtain the time t needed for the particle to fall through a vertical distance h:

$$t = 18h\eta/d^2 g(\rho_s - \rho_f)$$

Rearranging and solving for the particle diameter gives

$$d = [18h\eta/tg(\rho_s - \rho_f)]^{1/2} \tag{3.5}$$

Since all terms on the right-hand side of the equation, except t, are constants, we can combine them into a single constant A and write

$$d = A/t^{1/2} \qquad \text{or} \qquad t = B/d^2$$

where $B = A^2$.

One way of measuring particle size distribution is to use a pipette to draw samples of known volume from a given depth in the suspension at regular times during sedimentation. An alternative method is to use a *hydrometer* to monitor the density of the suspension at a given depth as a function of time. The density of the suspension diminishes as the largest particles and then progressively smaller ones settle out of the region of the suspension being monitored.

The use of Stokes' law to measure particle sizes is based on certain simplifying assumptions:

1. The particles are large enough to be unaffected by random motion of the fluid molecules.
2. The particles are rigid, spherical, and smooth.
3. All particles have the same density.
4. The suspension is dilute enough so particles settle independent.
5. Fluid flow around the particles is laminar (slow enough to avoid onset of turbulence).

In fact, we know that soil particles, though mostly rigid, are neither spherical nor smooth, and some may be platelike. Hence, the diameter calculated from the settlement velocity does not necessarily correspond to the actual dimensions of the particle. Rather, we should speak of an *effective* or *equivalent* settling diameter.

The results of a mechanical analysis based on sieving may differ from those of a sedimentation process for the same soil material. Moreover, soil particles are not all of the same density. Most silicates have specific gravity values of 2.6–2.7, whereas iron oxides and other heavy minerals may be twice as dense. For all these reasons, the mechanical analysis of soils yields only approximate results. Its greatest shortcoming, however, is that it only gives the *amount* of clay but not the *type* of clay, which can greatly affect soil behavior.

Sample Problem

Using Stokes' law, calculate the time needed for all sand particles (diameter >50 μm) to settle out of a depth of 0.2 m in an aqueous suspension at 30°C. How long would it take for all silt particles to settle out? How long for "coarse" clay (>1 μm)?

We use Eq. (3.5):

$$t = 18h\eta/d^2g(\rho_s - \rho_f).$$

Substituting the appropriate values for depth h (0.20 m), viscosity η (0.0008 kg/m sec), particle diameter d (50 μm, 2 μm, and 1 μm for the lower limits of sand, silt, and coarse clay, respectively), gravitational acceleration g (9.81 m/sec^2), average particle density ρs (2.65 × 10 km/m^3), and water density (10^3 kg/m^3), we obtain the following answers.

(a) For all sand to settle out, leaving only silt and clay in suspension:

$$t = \frac{18 \times 0.2 \times (8 \times 10^{-4})}{(50 \times 10^{-6})^2 \times 9.81 \times (2.65 - 1.0) \times 10^3} = 71 \text{ sec}$$

(b) For all silt to settle out, leaving only clay in suspension:

$$t = \frac{18 \times 0.2 \times (8 \times 10^{-4})}{(2 \times 10^{-6})^2 \times 9.81 \times 1.65 \times 10^3} = 44,500 \text{ sec} = 12.36 \text{ hr}$$

(c) For all coarse clay to settle out, leaving only fine clay in suspension:

$$t = \frac{18 \times 0.2 \times (8 \times 10^{-4})}{(1 \times 10^{-6})^2 \times 9.81 \times 1.65 \times 10^3} = 178,800 \text{ sec} = 49.44 \text{ hr}$$

PARTICLE SHAPES

Particle shapes, as well as particle sizes, affect the properties and behavior of the soil in bulk. Irregularly shaped particles expose a larger surface per unit mass than do regularly shaped particles. Whether particles are smooth or jagged has a bearing on the tendency of particles to fit closely together or interlock and, hence, on the resulting bulk density, the geometric configuration of the pore space, the interparticle bonding, and the resistance of the soil to deformation. This resistance determines the reaction of soil bodies to forces tending to cause compression as well as shearing. The jaggedness (as well as hardness) of particles also affects the frictional abrasiveness of soils toward traversing vehicles and tillage tools.

The techniques employed to examine particle shapes include direct observation via a microscope, image analysis, and the measurement of viscosity (Egashira and Matsumoto, 1981) and the light-scattering properties of suspensions

For particles that can be assigned the dimensions of length (L), breadth (B), and thickness (T), Heywood (1947) first defined the following shape parameters:

$$\text{Flatness ratio} = B/T \tag{3.6}$$

$$\text{Elongation ratio} = L/B \tag{3.7}$$

Other, dimensionless parameters that describe particle shapes are (Skopp, 2002):

$$\text{Sphericity} = A_s/A_a \tag{3.8}$$

$$\text{Circularity} = C_a/P_a \tag{3.9}$$

$$\text{Rugosity} = P_a/P_c \tag{3.10}$$

where A_a and A_s are, respectively, the surface area of an actual particle and the surface area of an "equivalent" sphere (i.e., a sphere that has the same apparent settling or sieve-passing properties as the particle C_a and P_a are, respectively, the circumference of a circle with the same area as the cross-sectional area of a particle and the particle's actual perimeter P_a; and P_c is the circumference of a circle circumscribing the particle.

These parameters can be used to describe the shapes of idealized particles and to model the characteristics of assemblages of such particles. However, the same criteria are of limited use in characterizing soils in which the particles are too irregular and variable to obey any simplistic criterion. Attempts have nonetheless been made to characterize the shapes of soil particles using empirically determined dimensionless shape factors.

SPECIFIC SURFACE

The specific surface of soil material can be defined as the surface area A_s of particles per unit mass (a_m) or per unit volume of particles (a_v) or per unit bulk volume of soil (a_b):

$$a_m = A_s/M_s \tag{3.11}$$

$$a_v = A_s/V_s \tag{3.12}$$

$$a_b = A_s/V_t \tag{3.13}$$

where M_s is the mass of particles of volume V_s contained in a bulk volume V_t of soil.

Specific surface has traditionally been expressed in terms of square meters per gram (a_m) or square meters per cubic centimeter (a_v). To convert from m^2/gm to m^2/kg, one need only multiply by 10^3. To convert from m^2/cm^3 to m^2/m^3, one needs to multiply by 10^6.

Specific surface obviously depends on the sizes of the soil particles. It also depends on their shapes. Flattened or elongated particles expose greater surface per unit mass or volume than do equidimensional (cubical or spherical) particles. Since clay particles are generally platy, they contribute more to the overall specific surface of a soil than is indicated by their small size alone. In addition to their external surfaces, some clay particles exhibit internal surface areas, such as are exposed when the accordion-like crystal lattice of montmorillonite expands on imbibing water. While the specific surface of sand may be no more than 1 or 2 square meters per gram, that of clay may be a hundred times higher. In fact, it is the clay fraction — its content and mineral composition — that largely determines the specific surface of a soil.

MEASURING SPECIFIC SURFACE BY ADSORPTION

The usual procedure for determining surface area is to measure the amount of gas or liquid needed to form a *monomolecular layer* over the entire surface in a process of adsorption (Pennell, 2002). The standard method is to use an inert gas such as nitrogen. Water vapor and organic liquids (e.g., glycerol and ethylene glycol) are also used. The adsorption phenomenon was described by de Boer (1953). At low gas pressures, the amount of a gas adsorbed per unit area, σ_a, is related to the gas pressure P, the temperature T, and the heat of adsorption Q by the equation

$$\sigma_a = k_i P \exp(Q_a/RT) \tag{3.14}$$

where R is the gas constant and k_i is also a constant. Thus, the amount of adsorption increases with pressure but decreases with temperature.

The equation of Langmuir (1918) relates the gas pressure P to the volume of gas adsorbed per unit mass of adsorbent m_a at constant temperature:

$$P/m_a = 1/kv_1 + P/v_1 \tag{3.15}$$

where k is a constant and v_1 is the volume of adsorbed gas forming a monomolecular layer over the adsorbent. Parameter v_1 is the slope of the curve of P/m_a versus P. The specific surface of the adsorbent can be calculated by determining the number of molecules in v_1 and multiplying this by the cross-sectional area of these molecules. Langmuir's equation assumes that only one layer of molecules is adsorbed and that the heat of adsorption is uniform during the process.

Brunauer, Emmett, and Teller (1938) derived what has come to be known as the BET equation, based on multilayer adsorption theory:

$$P/v(P_0 - P) = (1/v_1C) + (C - 1)P/v_1CP_0 \tag{3.16}$$

where v is the volume of gas adsorbed at pressure P, v_1 is the volume of a single molecular layer over the surface of the adsorbent, P_0 is the gas pressure required for monolayer saturation at ambient temperature, and C is a constant for the particular gas, adsorbent, and temperature. The volume v_1 can be obtained from the BET theory by plotting $P/v(P_0 - P)$ versus P/P_0. The density of the adsorbed gas is usually assumed to be that of the liquefied or solidified gas.

Polar adsorbents (such as water) may not obey the BET and Langmuir equations (which are similar at low pressures), because their molecules or ions tend to cluster at charged sites rather than spread evenly over the surface of the adsorbent. Techniques for measuring the specific surface of soils by using various adsorbed materials were described by Carter et al. (1986).

ESTIMATING SPECIFIC SURFACE BY CALCULATION

An estimation of the specific surface area can also be made by calculation based on the sizes, shapes, and relative abundance of different types of particles in the soil.

For a sphere of diameter d, the ratio of surface to volume is

$$a_v = (\pi d^2)/(\pi d^3/6) = 6/d \tag{3.17}$$

and the ratio of surface to mass is

$$a_m = a_v/\rho_s = 6/\rho_s d \tag{3.18}$$

For particles having a density ρ_s of about 2.65 g/cm^3 (2650 kg/m^3), Equation (3.18) can be approximated by

$$a_m \approx 2.3/d \tag{3.19}$$

Now suppose the particles are not spherical but rectangular: For a cube of edge L, the ratio of surface to volume is

$$a_v = 6L^2/L^3 = 6/L \tag{3.20}$$

and the ratio of surface to mass is

$$a_m = 6/\rho_s L \tag{3.21}$$

Thus, the expressions for particles of nearly equal dimensions, such as most sand and silt grains, are similar, and knowledge of the particle size distribution can allow us to calculate the approximate specific surface by the summation equation:

$$a_m = (6/\rho_s)\Sigma\, c_i(d_i^{\,2}/d_i^{\,3}) = (6/\rho_s)\Sigma\,(c_i/d_i) \tag{3.22}$$

in which c_i is the mass fraction of particles of average diameter d_i.

Now let's consider a platy particle. For the sake of simplicity, assume that a hypothetical plate is square shaped, with sides L and thickness z. The surface-to-volume ratio is

$$a_v = (2L^2 + 4Lz)/L^2z \tag{3.23}$$

and the surface-to-mass ratio is

$$a_m = 2(L + 2z)/\rho_s Lz \tag{3.24}$$

If the platelet is very thin, so that its thickness z is negligible compared to principal dimension L, and if $\rho_s = 2.65$ g/cm^3 (2650 kg/m^3), then

$$a_m \approx 2/\rho_s z \approx 0.75/z \text{ cm}^2/\text{g} \tag{3.25}$$

Thus, the specific surface of a clay can be estimated if the thickness of its platelets is known. The platelet thickness of dispersed montmorillonite is about 10 Å (or 10^{-7} cm = 10^{-9} m). Hence $a_m = 0.75/10^{-7}$ cm^2/g, or 750 m^2/g, which compares closely with the measured value of about 800 m^2/g. The average platelet thickness for illite clay is about 50 Å (or 5×10^{-11} m), and for kaolinite clays it is a few hundred angstroms.

Sample Problem

Estimate the approximate specific surface (m^2/g) of a soil composed of 10% coarse sand (average diameter 1 mm), 20% fine sand (average diameter 0.1 mm), 30% silt (average diameter 0.02 mm), 20% kaolinite clay (average platelet thickness 4×10^{-8} m), 10% illite clay (average thickness 0.50×10^{-8} m), and 10% montmorillonite (average thickness 10^{-9} m).

For the sand and silt fractions, we use Eq. (3.17):

$$a_m = (6/2.65)[(0.1/0.1) + (0.2/0.01) + (0.3/0.002)] = 0.0387 \text{ m}^2/\text{g} = 38.7 \text{ m}^2/\text{kg}$$

For the clay fraction, we use Eq. (3.22) in summation form to include the partial specific surface values for kaolinite, illite, and montmorillonite, respectively:

$$\begin{aligned} a_m &= (0.2 \times 0.75)/(400 \times 10^{-8}) + 0.1 \times 0.75/(50 \times 10^{-8}) + 0.1 \times 0.75/(10 \times 10^{-8}) \\ &= 3.78 \text{ m}^2/\text{g (kaol.)} + 15.09 \text{ m}^2/\text{g (ill.)} + 75.45 \text{ m}^2/\text{g (mont.)} \\ &= 94.32 \text{ m}^2/\text{g} = 94{,}320 \text{ m}^2/\text{kg} \end{aligned}$$

$$\text{Total for the soil} = 0.0387 + 94.32 = 94.36 \text{ m}^2/\text{g} = 94{,}360 \text{ m}^2/\text{kg}$$

Note: The clay fraction, only 40% of the soil mass, accounts for 99.96% of the specific surface. Montmorillonite alone (10% of the mass) accounts for nearly 80% of this soil's specific surface.

4. CLAY, THE COLLOIDAL COMPONENT

DEFINITIONS OF CLAY

The term *clay* carries several connotations, which are not necessarily mutually consistent. In daily language, it suggests a soil that tends to retain water and to become soft and sticky when wet. In the context of soil texture, it designates a range of particle sizes (namely, particles smaller than 2 micrometers) or a soil material with a high fractional content of this particle-size range. (Much of the clay therefore fits within the physicochemical category of *colloids*, which are finely divided materials that tend to disperse within gases and liquids and exhibit a very high degree of surface activity.) Finally, in the mineralogical sense, clay refers to a particular group of minerals, many of which occur in the clay fraction of the soil. That fraction thus differs from sand and silt not only in the range of particle sizes but also in mineralogical composition. Sand and silt consist mainly of weathering-resistant *primary minerals*, that is, minerals present in the original rock from which the soil was formed. Clay, however, includes *secondary minerals* formed in the soil itself by decomposition of the primary minerals and their recomposition into new ones.

The various *clay minerals* differ from one another in prevalence and properties and in the way they affect soil behavior. Rarely do any of these minerals occur in homogeneous deposits; in the soil they generally appear in mixtures, the specific composition of which depends in each case on the combination of conditions that governed soil formation. To understand why and how the clay fraction serves as the active constituent of soils, we must consider the structure and function of clay minerals.

STRUCTURE OF CLAY MINERALS

The forerunners of modern soil science assumed at first that clay consists of particles that are essentially similar to those of sand and silt, differing from them only in size. Later, they began to notice that when clays were dried from aqueous suspensions they tended to form thin flakes and also that moist clay could be skimmed and polished to form a smooth and shiny surface. These observations suggested that clay particles may be flat and capable of being oriented in different ways. But it was not until the advent of X-ray diffraction (described by Whittig and Allardice, 1986), of differential thermal analysis (Tan et al., 1986), and electron microscopy that the crystalline nature of clay minerals was proven and their structures were described.

The most prevalent minerals in the clay fraction of temperate region soils are silicate clays, whereas in tropical regions hydrated oxides of iron and aluminum are more prevalent. The typical aluminosilicate clay minerals appear as laminated microcrystals, composed mainly of two basic structural units: a tetrahedron of four oxygen atoms surrounding a central cation, usually Si^{4+}, and an octahedron of six oxygen atoms or hydroxyls surrounding a somewhat larger cation of lesser valency, usually Al^{3+} or Mg^{2+}. These basic building blocks are illustrated in Fig. 4.1.

The tetrahedra are joined together at their basal corners by means of shared oxygen atoms, in an hexagonal network that forms a flat sheet only 0.493 nanometers (4.93 Å) thick. This is illustrated in Fig. 4.2. The octahedra are similarly joined along their edges to form a triangular array as shown in Fig. 4.3. These sheets are about 0.505 nm (5.05 Å) thick.

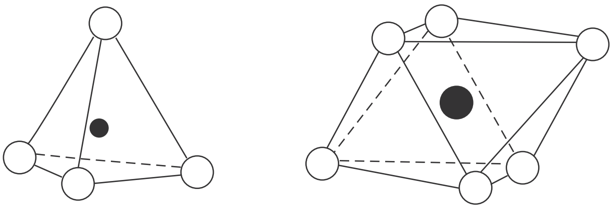


Fig. 4.1. The basic structural units of aluminosilicate clay minerals: a tetrahedron of oxygen atoms surrounding a silicon ion (left), and an octahedron of oxygens or hydroxyls enclosing an aluminum ion (right).

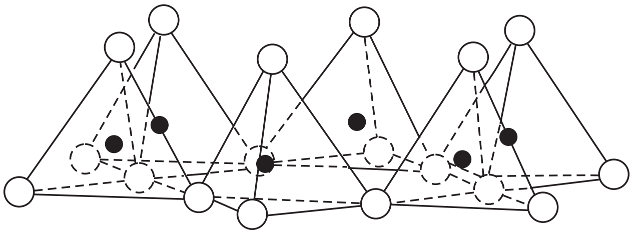


Fig. 4.2. Hexagonal network of tetrahedra forming a silica sheet.

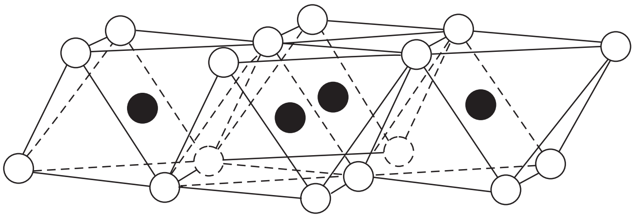


Fig. 4.3. Structural network of octahedra forming an alumina sheet.

The layered aluminosilicate clay minerals are of two main types, depending on the ratios of tetrahedral to octahedral sheets, whether 1:1 or 2:1 (2:2 minerals are also recognized). In 1:1 minerals like kaolinite, an octahedral sheet is attached by the sharing of oxygens to a single tetrahedral sheet. In 2:1 minerals like montmorillonite, it is attached in the same way to two tetrahedral sheets, one on each side. This is shown in Fig. 4.4. A clay particle is composed of multiply stacked composite layers (or unit cells) of this sort, called *lamellae*.

The structure described is an idealized one. Typically, some substitutions of ions of approximately equal radii, called *isomorphous replacements*, take place during crystallization. In the tetrahedral sheets, Al^{3+} may take the place of Si^{4+}, whereas in the octahedral layer Mg^{2+} may occasionally substitute for Al^{3+}. Consequently, internally unbalanced negative charges occur at different sites in the lamellae. Another source of unbalanced charge in clay crystals is the incomplete charge neutralization of terminal ions on lattice edges.

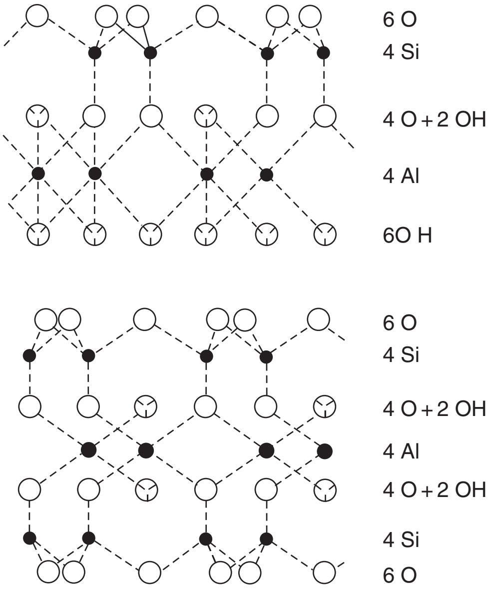


Fig. 4.4. Schematic representation of the structure of aluminosilicate minerals: kaolinite (top) and montmorillonite (bottom).

These unbalanced charges must be compensated externally by the adsorption of ions (mostly cations) from the solution surrounding the clay when it is wetted. These ions tend to concentrate near the external surfaces of the clay particle and occasionally penetrate into the interlamellar spaces. The adsorbed cations, including Na^+, K^+, H^+, Mg^{2+}, Ca^{2+}, and Al^{3+}, are not an integral part of the lattice structure and can be replaced, or exchanged, by other cations in the soil solution. The *cation exchange* phenomenon is of great importance in the soil, because it affects the retention and release of nutrients and other salts as well as the flocculation–dispersion processes of soil colloids.

PRINCIPAL CLAY MINERALS

Clay minerals are usually classified into two main groups, structured and amorphous. The structured clays are subclassified according to their internal crystal structure (i.e., the layered arrangement of tetrahedra and octahedra sheets) into two principal types, 1:1 and 2:1 minerals. The 2:1 clay minerals are further divided into expanding and nonexpanding types. Finally, each of these types includes a number of specific minerals, which can be identified on the basis of X-ray, electron microscope, or thermal analysis techniques.

The most common mineral of the 1:1 type is *kaolinite*. Other minerals in the same group are *halloysite* and *dickite*. The basic layer in the crystal structure is a pair of silica-alumina sheets, and these are stacked in alternating fashion and held together by hydrogen bonding in a rigid, multilayered lattice that often forms an hexagonal platelet. Since water and ions cannot enter between the basic layers, these cannot ordinarily be cleaved or separated. Moreover, since only the outer faces and edges of the platelets are exposed, kaolinite has a relatively low specific surface. Kaolinite crystals generally range in planar diameter from 0.1 to 2 μm, with a variable thickness in the range of about 0.02–0.05 μm. Owing to its relatively large particles and low specific surface, kaolinite exhibits less plasticity, cohesion, and swelling than most other clay minerals. The unit layer formula of kaolinite is $Al_4Si_4O_{10}(OH)_8$.

At the opposite end of the spectrum of aluminosilicate clay minerals is *montmorillonite*, a 2:1 mineral of the expanding type, which also includes *vermiculite* and *beidellite*. The lamellae of montmorillonite are stacked in loose assemblages called *tactoids*. Water and ions are drawn into the cleavage planes between the lamellae, and as the crystal expands like an accordion, it can readily be separated into flakelike thinner units and, ultimately, into individual lamellae, which are only 1 nm thick. As the montmorillonite crystals expand, their internal surfaces as well as external ones come into play, thus increasing the effective specific surface severalfold. Because of its tendency to expand and to disperse, montmorillonite exhibits pronounced swelling–shrinking behavior as well as high plasticity and cohesion. On drying, montmorillonitic soils, especially if dispersed, tend to crack and form hard clods. When tactoids are heated to several hundred degrees they tend to close irreversibly so that only their external areas act as adsorbing surfaces. The unit layer formula of montmorillonite is $Al_{3.5}Mg_{0.5}Si_8O_{20}(OH)_4$.

A clay mineral intermediate in properties between kaolinite and montmorillonite is *illite*. It belongs to a group of clay minerals called *hydrous micas*, which have a 2:1 silica:alumina ratio but are of the nonexpanding type. Isomorphous substitution of aluminum ions for silicon ions in the tetrahedral sheets (rather than substitutions of Mg^{2+} for Al^{3+} in the octahedral sheets, as is the case in montmorillonite), to the extent of about 15%, accounts for the relatively high density of negative charges in these sheets. This, in turn, attracts potassium ions and “fixes” them tightly between adjacent lamellae. As a result, the layers are bound together, so their separation and, hence, expansion of the entire lattice are effectively prevented. The unit layer formula of illite is $Al_4Si_7AlO_{20}(OH)_4K_{0.8}$, with the potassium occurring between the crystal units.

An example of a 2:2-type mineral is *chlorite*, in which magnesium rather than aluminum ions predominate in the octahedral sheets, which are in combination with the tetrahedral silica sheets. Its unit layer formula is $Mg_6Si_6Al_2O_{20}(OH)_4$, with $Mg_6(OH)_{12}$ occurring between the layers. In behavior, chlorite resembles illite. An additional group of silicate clays, in which the lattice is continuous in only one dimension, is known as *attapulgite* or *palygorskite*. The particles of this group are needlelike or tubelike, with microcavities providing internal surfaces.

Frequently, the various clay minerals occur not separately but in a complex mixture. At times, even the internal structure is mixed or interstratified, giving rise to composite minerals, which are somewhat loosely termed *bravaisite* (illite-montmorillonite, chlorite-illite, vermiculite-chlorite, etc.).

The clay fraction may also contain appreciable quantities of noncrystalline (amorphous) mineral colloids. *Allophanes*, for instance, are random combinations of poorly structured silica and alumina components expressible in the general formula $Al_2O_3{\cdot}2SiO_2{\cdot}H_2O$. The actual mole ratio of alumina to silica ranges in this group from 0.5 to 2.0. Phosphorus and iron oxides are frequently present in allophane. Notwithstanding its variable composition, allophane is sufficiently distinctive to be identified as a clay mineral. In behavior, this amorphous clay is similar to the crystalline clays with respect to adsorption, ion exchange, and plasticity.

Still another constituent of the clay traction is a group of hydrous oxides of iron and aluminum known as *sesquioxides*, which are prevalent mainly in the soils of tropical and subtropical regions and are responsible for the predominantly reddish or yellowish hue of these soils. Their composition can be formulated as $Fe_2O_3{\cdot}nH_2O$ and $Al_2O_3{\cdot}nH_2O$, in which the hydration ratio n is variable. Limonite and goethite are typical hydrated iron oxides, and gibbsite is a frequently encountered aluminum oxide. The sesquioxide clays may be partly crystallized but are often amorphous. Their electrostatic, adsorptive, and plasticity properties are less distinct than those of most of the silicate clays. Frequently, these oxides serve as cementing agents in the stabilization of soil aggregates, particularly in subtropical and tropical regions. Table 4.1 summarizes the properties of a few of the silicate clay minerals.

TABLE 4.1 **Typical Properties of Selected Clay Minerals (Approximate Values)**

Properties	Kaolinite	Illite	Montmorillonite	Chlorite	Allophane
Planar Diameter (μm)	0.1–4	0.1–2	0.01–1	0.1–2	
Basic layer thickness (Å)	7.2	10	10	14	
Particle thickness (Å)	500	50–300	10–100	100–1000	
Specific surface (m^2/g)	5–20	80–120	700–800	80	
Cation exchange capacity (mEq/100 g)	3–15	15–40	80–100	20–40	40–70
Area per charge ($Å^2$)	25	50	100	50	120

THE ELECTROSTATIC DOUBLE LAYER

When a colloidal particle is more or less dry, the neutralizing counterions are attached to its surface. Upon wetting, however, some of the ions dissociate from the surface and enter into solution. A hydrated colloidal particle of clay or humus thus forms a *micelle*, in which the adsorbed ions are spatially separated, to a greater or lesser degree, from the negatively charged particle (Fig. 4.5). Together, the particle surface, acting as a multiple anion, and the "swarm" of cations hovering about it form an *electrostatic double layer*.

The assemblage of adsorbed cations can be regarded as consisting of (a) a layer more or less fixed in position at the immediate proximity of the particle surface and known as the *Stern layer* and (b) a diffuse cloud of cations extending some distance away from the particle surface and gradually diminishing in concentration. This distribution is illustrated in Fig. 4.6. It results from an equilibrium between two opposing tendencies: (1) the electrostatic

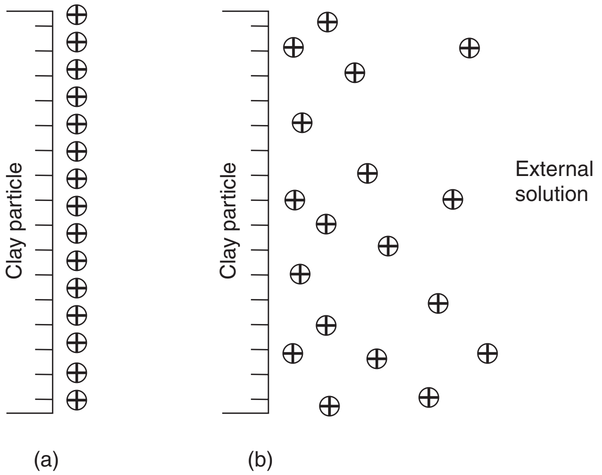


Fig. 4.5. Formation of a diffuse double layer in a dry (a) vs. hydrated (b) micelle.

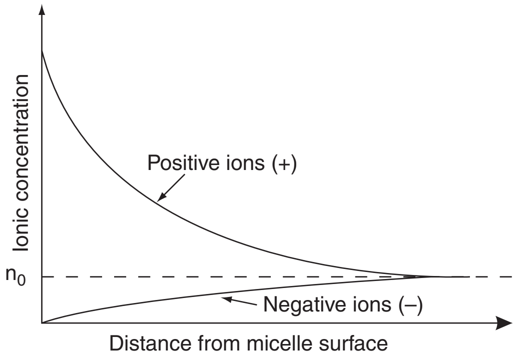


Fig. 4.6. The distribution of positive and negative ions in solution with distance from the surface of a clay micelle bearing a net negative charge. Here n_0 is the ionic concentration in the bulk solution outside the electrical double layer.

(coulombic) attraction of the negatively charged surface for the positively charged ions, which tends to pull the cations inward so as to attain the minimum potential energy level; and (2) the kinetic (Brownian) motion of the liquid molecules, inducing the outward diffusion of the adsorbed cations in a tendency toward equalizing the concentration throughout the solution phase, thus maximizing entropy. The equilibrium distribution of cations is such as to minimize the free energy of the system as a whole. The actual concentration of cations inside the double layer can be 100 or even 1000 times greater than in the external, or intermicellar, solution (i.e., outside the range of influence of the particle's electrostatic force field).

As cations are adsorbed positively by the colloidal particle, anions are generally repelled, or adsorbed negatively, and thus are relegated from the micellar to the intermicellar region of the solution surrounding the clay particles. In some special cases, however, anions may also be attracted to specific sites on colloidal surfaces, but this phenomenon of *anion adsorption* is much less prevalent than *cation adsorption* in soils.

Theoretical treatments of the electrostatic double layer have generally been based on the *Gouy–Chapman theory*, in which the negative charges are assumed to be constant and to be distributed uniformly over the particle surfaces (although in actual fact they may originate within the crystal lattice). The strength of the surface charge is proportional to the charge density (i.e., the number of charged sites per unit area).

The effective thickness of the diffuse double layer can be estimated by means of the following equation:

$$z = (1/ev)\ [(\varepsilon kT)/(8\pi n_0)]^{0.5} \tag{4.1}$$

in which z is the characteristic length, or extent, of the double layer, defined as the distance from the clay surface to where the ionic concentration is very nearly that of the external (intermicellar) solution; e is the elementary charge of an electron (4.77×10^{-10} esu); ε is the dielectric constant; k is the Boltzmann constant ($k = R/N$, where R is the gas constant and N is Avogadro's number); v is the valence of the ions in solution; n_0 is the concentration of the ions in the bulk solution; and T is the temperature in Kelvin.

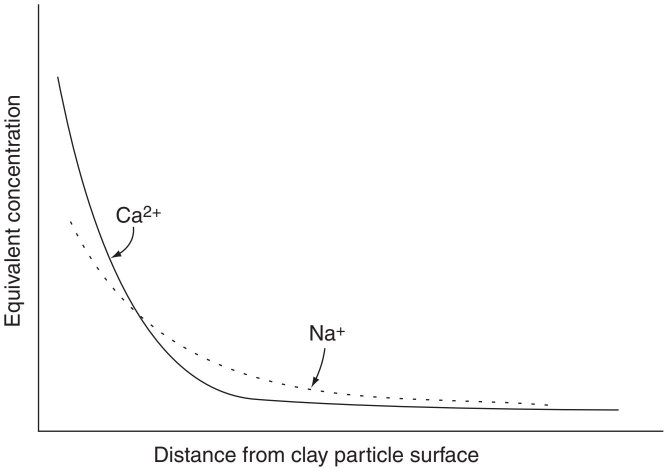


Fig. 4.7. Comparison of the distribution of a monovalent cation (Na^+) with a divalent cation (Ca^{2+}) in the diffuse double layer. Thanks to its greater charge, the divalent cation is attracted more strongly and held more closely to the particle surface. The area under each curve, signifying milliequivalents of cations per unit surface, is equal.

As the equation indicates, the diffuse double layer is compressed as the valence of the ions in solution is increased. For instance, if a solution of monovalent cations is replaced by a solution of divalent cations, the double layer becomes only half as thick (Fig. 4.7). The double-layer thickness is also affected by the solution's concentration (Fig. 4.8), because z of Eq. (4.1) is inversely proportional to the square root of n. Thus a tenfold increase of concentration will reduce the double layer to $1/10^{0.5}$, that is, to about a third of its previous thickness.

The foregoing considerations do not account for interparticle interaction, that is, where the diffuse ionic clouds of neighboring micelles intermingle. In

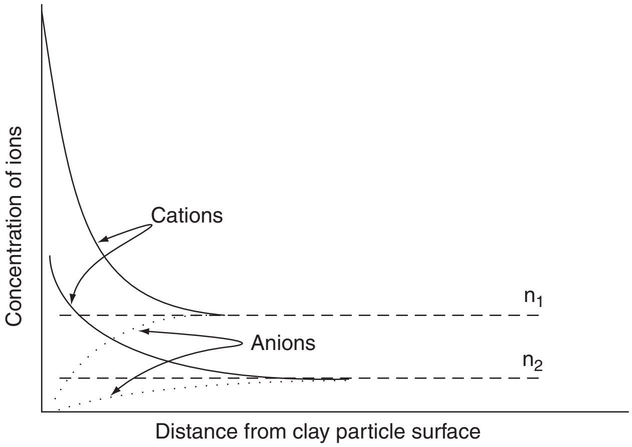


Fig. 4.8. Influence of ambient solution concentration (n_1, n_2) on thickness of the diffuse double layer. The solid curves represent cations and the dotted curves anions. Higher concentration of the external solution is seen to compress the double layer.

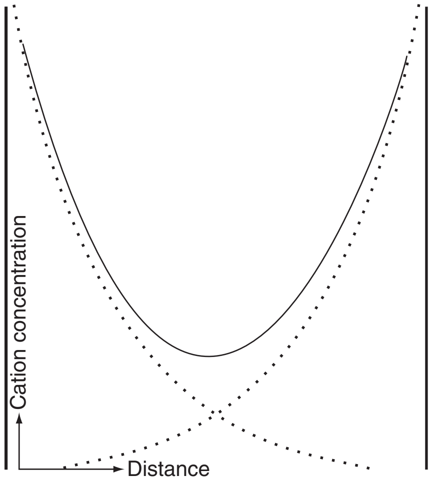


Fig. 4.9. Schematic illustration of the increased concentration of cations in the double layer due to interaction of adjacent particles. The dotted curves indicate the hypothetical distribution of cations for each of the particles if it were suspended alone in the same ambient solution.

this case, the concentration in the median plane, rather than in the external solution, is referred to (Langmuir, 1938):

$$n_c = \pi^2/[v^2 B(d + a)^2 \times 10^{-16}] \tag{4.2}$$

Here n_c is cation concentration at the median plane (moles/liter), v is the valence of the exchangeable cations, d is the distance from either plate to the median plane (angstroms), a is a correction factor (1–4 Å), and B is a constant related to the temperature and dielectric constant (10^{15} cm/mmol). The distribution of cations between two charged plates is illustrated in Fig. 4.9.

ION EXCHANGE

The cations in the double layer can be replaced or exchanged by other cations introduced into the solution. Under chemically neutral conditions, the total number of exchangeable cation charges, expressed in terms of chemical equivalents per unit mass of soil particles, is nearly constant and independent of the species of cations present. It is thus considered to be an intrinsic property of the soil material, generally called the *cation exchange capacity* (or the *cation adsorption capacity*), and has traditionally been reported in terms of milliequivalents of cations per 100 grams of soil. [The term *equivalent* is defined in chemistry as one gram atomic weight (mole) of hydrogen or the mass of any ion that can combine with or replace it in chemical reactions.] Thus expressed, the cation exchange capacity ranges from nil in sands to 100 meq/100 g, or even more, in clays and in organic soils.

The cation exchange capacity depends not only on clay content but also on clay type (i.e., on specific surface and charge density), as indicated in

Table 4.1. In the top layer of soils rich in organic matter, humus can account for as much as 1/4 or 1/3 of the exchange capacity. The cation exchange phenomenon affects the movement and retention of ions in the soil, which can be important in plant nutrition and in environmental processes involving the transport of pollutants. Cation exchange also affects the flocculation–dispersion processes of soil colloids and hence the development and degradation of soil structure. The total number of electrostatic charges of clay particles divided by the total exposed surface area is called the *surface charge density*.

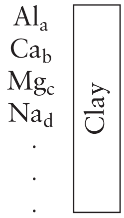
Because of differences in their valences, radii, and hydration properties, different cations are adsorbed with varying degrees of tenacity or preference and hence are either more readily or less readily exchangeable. In general, the smaller the ionic radius and the greater the valence, the more closely and strongly is the ion adsorbed. However, the greater the ion's hydration, the farther it is from the adsorbing surface and the weaker its adsorption. Sodium ion, for example, has an atomic radius of only 9.8×10^{-11} m when in the "naked" state, but it tends to be strongly hydrated and its effective radius increases eightfold when it is enveloped by water molecules. Monovalent cations, attracted by only a single charge, can be replaced more easily than divalent or trivalent cations. The order of preference in exchange reactions is as follows:

$$Al^{3+} > Ca^{2+} > Mg^{2+} > NH^{+} > K^{+} > H^{+} > Na^{+} > Li^{+} \tag{4.3}$$

An example of an exchange reaction is the following:

$$Na_2\ [\text{clay}] + Ca^{2+} \leftrightarrow Ca\ [\text{clay}] + 2Na^{+} \tag{4.4}$$

In nature, soil material is seldom adsorbed homogeneously with a single ionic species. Typically, the exchange capacity is taken up by several cations in varying proportions, all of which together constitute the so-called *exchange complex*. A typically heterogeneous exchange complex can be represented in the following way (with subscripts a, b, c, d, etc., indicating equivalent fractions of the cation exchange capacity, adding up to 100%):



In arid regions, calcium, magnesium, and sometimes sodium tend to predominate in the exchange complex. In humid regions, where soils are highly leached and often acid, hydrogen and aluminum ions, the latter often released from the clay crystal lattice under low pH conditions, play an important role. The presence of exchangeable hydrogen is usually referred to as *base unsaturation* of the exchange complex, and its measure in equivalent terms for the soil mass of an agricultural field is taken to be an indication of the amount of lime needed to neutralize the soil's *reserve acidity*.

The composition of the exchange complex depends on the concentration and ionic composition of the ambient solution. This is expressed in the *Gapon equation*:

$$A_e/B_e = c([A_s]^{1/a}/[B_s]^{1/b}) \tag{4.5}$$

in which A and B are fractional concentrations of cations with valences of a and b, respectively, and subscripts e and s refer to concentrations in the exchange complex and in the ambient solution, respectively. The coefficient c depends on the nature of the charged surface and the nature of the adsorbed cation. This equation indicates that the adsorption mechanism favors cations of higher valence and that this preference increases as the solution becomes increasingly dilute.

For the important case of calcium–sodium exchange, the *selectivity coefficient* (Shainberg, 1973) is given by

$$\sigma_{\mathrm{Ca-Na}} = n_0([\mathrm{Ca_e}][\mathrm{Na_s}]^2/[\mathrm{Na_e}]^2[\mathrm{Ca_s}]) \tag{4.6}$$

Here, subscript e denotes the ionic equivalent fractions of Ca and Na in the exchange phase, subscript s denotes the mole fractions of these ions in the solution phase, and n_0 is the total molar concentration of the solution. The selectivity coefficient σ for a range of exchangeable sodium fractions has been reported to be about 4 in typical soils in Israel (Levy and Hillel, 1968).

For ordinarily encountered solution concentrations in nonsaline soils, divalent cations predominate in the exchange complex if present in appreciable concentrations. In a mixed calcium–sodium system most of the calcium ions are usually adsorbed tightly in the Stern layer, whereas the sodium ions are relegated to the diffuse region of the double layer.

Cation exchange reactions are rapid and reversible; so the composition of the exchange complex responds to frequent changes in the composition and concentration of the soil solution. The composition of the soil's exchange complex in turn governs the soil's pH as well as swelling and flocculation–dispersion tendencies.

In addition to the *structural charges* due to isomorphous substitutions within the crystal lattice of the clay, there are *surface charges* due to the imbalance of proton and hydroxyl charges on the exposed peripheries or edges (as opposed to the faces) of the clay particles, where the lattice bonds are broken (Sposito, 1984). The former charges are permanent and are generally unaffected by the ambient solution surrounding the particles, whereas the latter charges are affected by the pH of the ambient solution. The structural charges are generally negative, but the surface charges may become positive at low values of pH. Consequently, pH-sensitive clay minerals may display a capacity for anion adsorption. This has been noticed specifically in kaolinite. Anion adsorption is selective. Among the common anions, silicate and phosphate ions are more strongly adsorbed than are sulfate, nitrate, or chloride ions. Anion adsorption can be important in the retention of phosphate in soils. Various nonionic substances (including organic compounds) can also be adsorbed by clay, generally by means of hydrogen and van der Waals bonds. Organic matter (humus) can also exhibit anion exchange, particularly at low pH values.

BOX 4.1 The Uses of Clay

Some eight millennia ago, humans discovered the use of clay for pottery making. In Mesopotamia, an important use of clay was for writing. Clay tablets were fashioned and then while still moist imprinted with a sharp reed stylus. If then fired, such tablets could become permanent records of political and commercial deeds, legal codes, myths, and poetry.

The value of clay for ceramics arises from its plasticity and its tendency to harden when heated. Hence it can be shaped at will and the desired shape can then be made durable, strong, thermally and electrically resistant, and either pervious or impervious (depending on the type of clay and on the temperature to which it is heated). When the temperature is high enough, the clay particles tend to fuse so the shaped object becomes practically impervious. Wall tiles, ceramic pipes, and stoneware pots are among the objects made in this way. Sanitary faiences are made with kaolinite fired at very high temperatures. When the kaolinite is quite pure and fusing is complete, a translucent glass is obtained, called *porcelain*, which is used for the manufacture of fine chinaware as well as electric insulators.

An important use of clay is in construction. Even today, an estimated third of all humans live in houses made of clay. Clay mixed with sand or straw is the plasterlike material called *daub*. Packed into molds and then dried, it is made into mudbricks. When fired, it is hardened into stonelike bricks. Cement is a mixture of limestone and clay (about 25%) fired at a temperature of 1400°C.

The fact that clay can be made into impermeable pastes has been utilized since antiquity. Cisterns and channels were coated with clay so that they would hold and convey water with minimum leakage. Porous soil can be made impermeable by the injection of mixtures of water and clay. Injection of clayey mud is a common practice in drilling for oil. Pumped into a drill pipe, the mud cools the drill bit and rises along the pipe, carrying away the drill cuttings.

The adsorptive capacities of clay are important in agriculture and in industry. Adsorbed ions are loosely bound and therefore are easily exchanged with hydrogen from the roots of plants. Hence the clay minerals act to provide a source for the nutrient mineral required by plants. The ability of smectite clays to form stable colloidal suspensions is useful in the preparation of polishes, cosmetics, and pesticides, among many products. Kaolin, or white clay, serves widely for coating paper. (The paper on which this book is printed may contain over 20% kaolin.) Without it the paper would be much less opaque and would absorb the ink so that the lines of illustrations would become blurred.

HYDRATION AND SWELLING

Clay particles are hardly ever completely dry. Even after oven-drying at 105°C for 24 hr, which is the standard for drying soil material, clay particles still retain appreciable amounts of adsorbed water, as shown in Fig. 4.10. The affinity of clay surfaces for water is demonstrated by the hygroscopic nature of clay soils, that is, their ability to sorb and condense water vapor from the air. So-called “air-dry” soil commonly contains several percent water, the exact percentage depending of course on the kind and quantity of clay as well as on the humidity of the air.

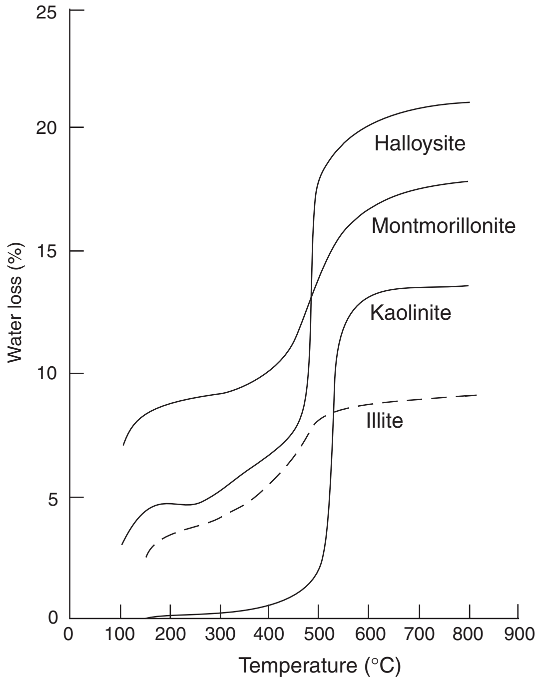


Fig. 4.10. Dehydration curves of clay minerals. (After Marshall, C. E., 1964)

In the oven-dry state, the water associated with clay is so tightly held that it can be considered a part of the clay itself. When water is added to initially dry clay, the water films surrounding each particle thicken and the water is then more loosely held. The entire physical behavior of a clay-containing soil mass (including strength, consistency, plasticity, and the conduction of water and heat) is strongly influenced by the degree of hydration.

Water is attached to clay surfaces by several mechanisms, including the electrostatic attraction of the dipolar, oriented water molecules to charged sites as well as their hydrogen bonding to exposed oxygen atoms on the clay crystal. Still another mechanism of hydration arises from the presence of adsorbed cations. Because the cations associated with the clay also tend to hydrate, they contribute to the overall hydration of the clay system. Quantitatively, this effect depends on the type of cations present and on the cation adsorption capacity of the clay.

The strength of clay–water adsorption is clearly greatest for the first layer of water molecules. The second layer is attached to the first by hydrogen bonding, and the third to the second, and so forth, so the influence of the attractive force field of the clay surface diminishes rapidly with distance, to become vanishingly small beyond a few molecular layers. The effective thickness and the physical properties of the adsorbed water in the vicinity of the clay surface have been a subject of controversy, with some investigators contending that adsorbed water is quasi-crystalline and hence differs significantly from bulk water in viscosity, diffusivity to ions, dielectric constant, and density. Others claim that the difference in properties between adsorbed water and capillary water in the soil is scarcely detectable and is in any case inconsequential.

Sample Problem

The diameter of a water molecule is about 3×10^{-8} cm (3×10^{-10} m). One cubic centimeter of water at standard temperature and pressure has a mass of about 1 gram and contains about 3.4×10^{22} molecules. The mass per molecule M_m is therefore about 3×10^{-23} g.

Calculate the mass of water adsorbed by 100 g of montmorillonite, illite, kaolinite, and fine quartz sand if the specific surface areas a_s are 800, 100, 10, and 1 m²/g, respectively. Assume monomolecular adsorption.

For montmorillonite:

Surface area of sample: $A_s = ma_s = 100 \text{ g} \times 800 \text{ m}^2/\text{g}$
$= 8 \times 10^4 \text{ m}^2 = 8 \times 10^8 \text{ cm}^2$

Area of one molecule $A_m \approx (3 \times 10^{-8} \text{ cm})^2 \approx 9 \times 10^{-16} \text{ cm}^2$

Number of molecules adsorbed: $n = A_s/A_m = 8 \times 10^8 \text{ cm}^2 / 9 \times 10^{-16} \text{ cm}^2$
$\approx 9 \times 10^{23}$

Mass of water adsorbed: $M_w = nM_m \approx (9 \times 10^{23})(3 \times 10^{-23} \text{ g}) = 27 \text{ g}$

For illite: $M_w \approx 3.375$ g; for kaolinite ≈ 0.338 g; for fine quartz sand ≈ 0.034 g.

When a confined body of clay sorbs water, *swelling pressures* (defined as the pressure that must be applied to a body of soil that is allowed to imbibe water, to prevent it from expanding) develop, related to the osmotic pressure difference between the adsorbed water and the external solution. In a partially hydrated micelle, the thickness of the enveloping water is less than the potential thickness of the fully expanded diffuse double layer. The double layer, thus truncated, will tend to expand to its full potential thickness and dilution by the osmotic absorption of additional water, if available. As each micelle expands, its swarm of positively charged cations repels those of the adjacent micelles. Thus the micelles tend to push each other apart. Though this causes the system to swell and increase its overall porosity, it may have the internal effect of closing the soil's larger pores, thus reducing its permeability.

We have already shown that the concentration of ions between associated clay micelles is greater than in the external solution. The actual concentration difference depends on the interparticle distance (i.e., on the degree of hydration of the clay) and on the potential extent of the diffuse double layer (which, in turn, depends upon the valences and concentrations of the adsorbed cations). The osmotic attraction of a clay assemblage for "external" water may be roughly twice as high with monovalent than with divalent cations, since there are normally twice as many of the former than of the latter. Hence swelling is greatest with monovalent cations, such as sodium, and with distilled water as the external solution. With calcium as the dominant cation in the exchange complex, swelling is reduced. A similar restraining effect is caused at low pH values by the presence of trivalent aluminum. High salinity

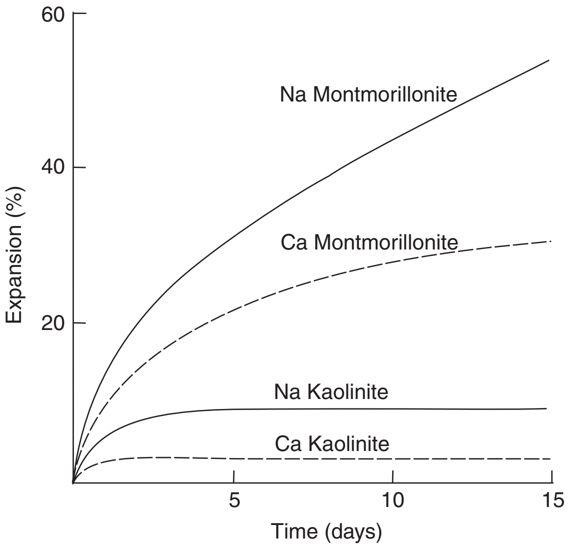


Fig. 4.11. Volume changes of montmorillonite and kaolinite clays during hydration.

of the soil solution will also suppress swelling. However, if a saline soil in which sodium salts predominate is leached of excess salts with freshwater without concurrent addition of calcium (e.g., in the form of gypsum), strong swelling may result from the predominance of sodium ions in the adsorbed phase.

The time-dependent volume increase of clays and clay–sand mixtures in the process of hydration is illustrated in Fig. 4.11. This time dependence is due to the low permeability of clay systems. The eventual swelling is seen to depend on the amount and nature of the clay present. In general, swelling increases with increasing specific surface area. It is also affected by the arrangement or orientation of soil particles and by the possible occurrence of interparticle cementation by such materials as iron or aluminum oxides, carbonates, and humus, which may constrain the expansion of the soil matrix.

When a hydrated body of clay is dried, a process opposite to swelling occurs, namely, shrinkage. In the field, the shrinkage that typically begins at the surface often causes the formation of numerous cracks, which break the soil mass into fragments of various sizes, from small aggregates to large blocks. An extreme example of this can be seen in *vertisols*, which are soils rich in expansive clay (e.g., montmorillonite). When subject to alternating wetting and drying, as in a semiarid region, such soils tend to heave and then to settle and to form wide, deep cracks and slanted sheer planes extending deep into the soil profile. Vertisols can be problematic both in agriculture (where they are exceedingly difficult to till because they tend to puddle and then to harden upon drying) and in engineering (where their successive heaving and subsidence tends to warp road pavements and to damage buildings).

BOX 4.2 Clay Soils That Till Themselves

Vertisols behave strangely: They perform a kind of tillage on their own. The part of the profile where we would normally expect to find more or less distinct differentiation between an A horizon and a B horizon in fact exhibits very little in the way of profile development. The entire soil, sometimes to a depth of 1 m or more, actually churns and mixes itself repeatedly.

Here is how it happens: During the wet season, these clayey soils swell markedly. The differential swelling of zones within the profile causes parts of the soil to move relative to adjacent parts, thus creating typically oblique shear planes (called *slickensides*) in the subsoil. During the dry season, the soil shrinks from the top downward and forms an extensive network of deep, wide cracks. The masses of soil between cracks form separated, column-like blocks. The tops of these blocks desiccate and break into numerous small clods, which in turn dribble into the cracks, often to a depth of several decimeters. When the soil is next wetted and in turn desiccated again, new material is exposed at the surface, just as some of the material formerly at the surface is now mixed deep in the profile.

The same process also causes stones (where they are present, as in volcanic areas) to rise in the soil. The small clods that drop into the cracks sometimes dribble past the stones that are embedded in the soil. When the soil is next wetted, these clods swell and push the stones upwards. Farmers in such areas are likely to find a fresh crop of stones on the soil surface every year.

FLOCCULATION AND DISPERSION

As clay particles interact with one another, forces both of repulsion and of attraction can come into play. The one or the other type of force may predominate, depending on physicochemical conditions. When the repulsive forces are dominant, the particles separate and remain apart from each other — and the clay is said to be *dispersed*. On the other hand, when the attractive forces prevail, the clay becomes *flocculated*, a phenomenon analogous to the coagulation of organic colloids, as the particles associate in packets or flocs.

These phenomena can be observed quite readily in dilute aqueous suspensions of clay. For the clay particles to enter into and remain in a state of suspension, they must be dispersed. This can be induced by the addition of a sodic dispersing agent and by mechanical agitation, in the manner described in connection with mechanical analysis. The dispersed suspension is typically turbid and remains so as long as the suspension is stable. This state can be changed rather dramatically by the addition of polyvalent cations or by the addition of salt to increase the overall electrolyte concentration. The turbid suspension suddenly clarifies as the clay particles flocculate and the flocs (behaving like large composite particles) settle to the bottom.

The main repulsive force between clay micelles derives from the like charges of the ionic swarms surrounding the particles, manifested by the swelling tendency of clay–water systems. An attractive force may result, however, if two clay platelets are brought close enough together (within about 1.5 nm) so that their counterions intermingle to form a unified layer of positive charges, which then attracts the negatively charged particles on both sides. This process,

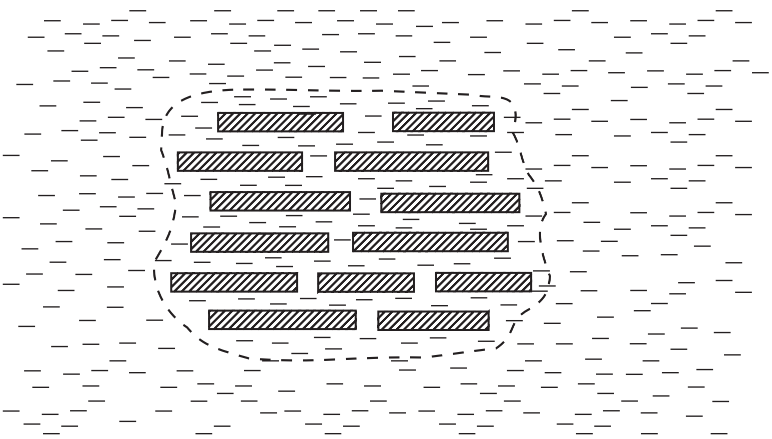


Fig. 4.12. Schematic representation of montmorillonite tactoid.

called *plate condensation*, brings about the formation of *tactoids*, or packets of parallel-oriented platelets (like packs of cards) that form a more or less stable association (Fig. 4.12).

Another type of electrostatic attraction between particles can occur when the plate edges develop positive charges, as they tend to do at low pH values. If the repulsion normally due to the diffuse double layer is not so great as to prevent clay particles from coming together, the positive charges on the edge of one particle may form bonds with the negative charges on the face of another. A “card-house” structure of flocs can then develop, as shown in Fig. 4.13.

The important fact about these various attractive and repulsive forces is that they operate with different intensities and over different ranges. The combined force field, shown schematically in Fig. 4.14, consists of regions within which net attraction prevails and regions over which repulsion predominates. For example, coulombic (electrostatic) forces are inversely proportional to distance squared, whereas London–van der Waals forces are inversely proportional to the seventh power of the distance, so the latter are effective within a narrow space of only a nanometer or so, whereas the former extend to distances 10 times as great.

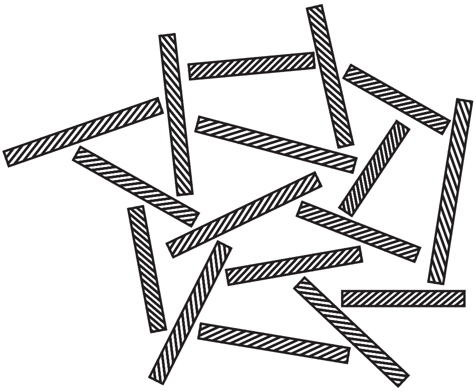


Fig. 4.13. Edge-to-face electrostatic bonding resulting in card-house structure.

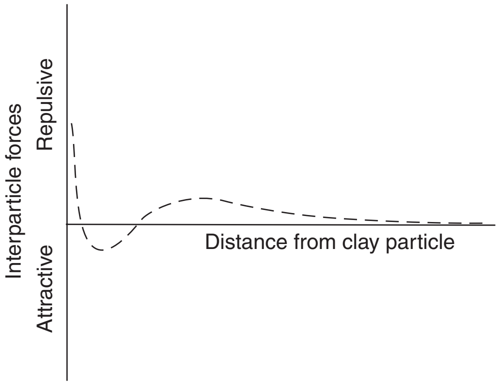


Fig. 4.14. Schematic representation of the combined force field surrounding a hydrated clay particle.

Starting from a dilute stable suspension in which the micelles are far apart so that the repulsive forces prevail, flocculation can only be achieved if these repulsive forces are suppressed sufficiently to allow colliding particles to clump together, rather than bounce apart, as they approach within the range of the attractive forces. Repulsion is maximized, favoring dispersion, when the double layer is fully extended, as when the ambient solution is very dilute and of high pH (thus preventing positive charges from forming at the particle edges), when the dominant cation is monovalent, and when the soil is fully hydrated. Mechanical agitation also inhibits flocculation. In contrast, repulsion is minimized, allowing flocculation, whenever the solution concentration is high and the monovalent cations are replaced by polyvalent ones.

The balance of repulsive versus attractive forces is reversible and, under certain circumstances, can easily shift either way. A soil can thus disperse, flocculate, redisperse, reflocculate, and so forth, several times over, as, for example, when it is irrigated with water of varying salinity and ionic composition.

Sample Problem

If the cation exchange capacity of a soil is 40 meq/100 g (400 mEq/kg), and if sodium ions constitute 25% of the exchange complex, what is the minimum amount of gypsum required per hectare to replace sodium with calcium in the upper 0.20-m layer of the soil? Assume a bulk density of 1.2×10^3 kg/m^3.

The chemical formula for gypsum is $CaSO_4 \cdot 2H_2O$. Hence the molecular weight is 172 and the equivalent weight is 86.

Mass of soil in upper 0.2-m layer $= 10^4 \text{m}^2/\text{hectare} \times (2 \times 10^{-1}$ m depth) $\times 1.2 \times 10^3 \text{ kg/m}^3 = 2.4 \times 10^6$ kg.

Equivalents of sodium to be removed from upper 0.2 m layer $= 2.4 \times 10^6$ kg $\times (400 \times 0.25$ mEq/kg$) = 2.4 \times 10^5$ Eq.

Mass of gypsum needed $= 2.4 \times 10^5 \text{ Eq} \times 86 \text{ g/Eq} \approx 2 \times 10^7 \text{ gm} \approx 20$ metric tons.

HUMUS: THE ORGANIC CONSTITUENT OF SOIL COLLOIDS

Although the main topic of this chapter is the mineral colloidal matter known as clay, we must emphasize the important fact that an entirely different kind of colloidal matter exists in soils, called *humus*. This generally dark-colored material, found mostly in the surface zone (the A horizon) of soils, is defined in the *Glossary of Soil Science Terms* (Soil Science Society of America, 1996) as "the more or less stable fraction of the soil organic matter remaining after the major portion of added plant and animal residues have decomposed." So defined, humus does not include undecomposed or partially decomposed organic residues, such as recent stubble or dead roots.

Like clay, humus particles are negatively charged. During hydration, each particle of humus forms a micelle and acts like a giant, composite anion, capable of adsorbing various organic and inorganic constituents, including cations. The cation exchange capacity of humus is much greater, per unit mass, than that of clay. Unlike most clay, moreover, humus is generally not crystalline but amorphous. Because it is composed mostly of carbon, oxygen, and hydrogen, its charges are due not to the isomorphous substitutions of cations but to the dissociation of carboxylic (–COOH) and phenolic (OH) groups. Since the cation exchange process depends on replacement of the hydrogen in these groups, it is pH dependent, with the cation exchange capacity generally increasing at higher pH values.

Humus is not a single compound, nor does it have the same composition in different locations. Rather, it is a complex mixture of numerous compounds, including lignoproteins, polysaccharides, polyuronides, and other compounds too varied to list. Furthermore, the organic colloids of humus, although "more or less" stable, are in fact amenable to bacterial action, particularly if the soil's temperature, moisture, or aeration regimes are modified.

The content of humus in mineral soils varies from as high as 10% or even more in the top layer of *chernozem* (the typically black soil that occurs in the American prairie and in the plains of Ukraine) down to nil in desert soils and is of the order of 1–3% in many intermediate soils. The humus content generally diminishes in depth through the B horizon and becomes negligible at the bottom of the normal root zone, unless the soil is a deposit of alluvial material with a high original content of humus. Organic soils such as peat and muck may contain well over 50% organic matter, though not all of that would fit the accepted definition of humus.

The importance of humus goes beyond its effect on cation adsorption or even plant nutrition. Humus often coagulates in association with clay and serves as a cementing agent, binding and stabilizing soil aggregates and thus improving soil structure. This aspect is discussed in Chapter 5.

5. SOIL STRUCTURE AND AGGREGATION

SOIL STRUCTURE DEFINED

Bricks thrown haphazardly atop one another become an unsightly heap. The same bricks, only differently arranged and mutually bonded, can give rise to a home or a school. Similarly, a soil can be merely a loose and unstable assemblage of random particles, or it can consist of a distinctly structured body of interbonded particles associated into aggregates having regular sizes and shapes. Hence it is not enough to study the properties of individual soil particles. To understand how the soil behaves as a composite body, we must consider the manner in which the various particles are packed and held together to form a continuous spatial network that constitutes the *soil matrix* or the *soil fabric* (Gregorich et al., 2002).

The arrangement or organization of the particles in the soil (i.e., the internal configuration of the soil matrix) is called *soil structure*. Since soil particles differ in shape, size, and orientation and can be variously associated and interlinked, the mass of them can form complex and irregular patterns that are difficult to characterize in exact geometric terms. A further complication is the inherently unstable nature of soil structure and its nonuniformity in space. Soil structure is affected by changes in climate, biological activity, and soil management practices, and it is vulnerable to destructive forces of a mechanical and physicochemical nature.

For these various reasons, we have no truly objective or universally applicable way to measure soil structure per se, and the term *soil structure* therefore expresses a qualitative concept rather than a directly quantifiable property. The numerous methods proposed for characterizing soil structure are designed not to measure soil structure itself but soil attributes that are supposed to depend on structure. Many of these methods are specific to the purpose for which they were devised and some are completely arbitrary.

The difficulty of defining soil structure notwithstanding, we can readily perceive its importance, inasmuch as it determines the total porosity as well as the shapes and sizes of the pores in the soil. Soil structure affects the retention and transmission of fluids in the soil, including infiltration and aeration. Moreover, as soil structure influences the mechanical properties of the soil, it may also affect such disparate phenomena as germination, root growth, tillage, overland traffic, and erosion. Agriculturists are usually interested in having the soil, at least in its surface zone, in a loose and highly porous and permeable condition. Engineers, on the other hand, often desire a dense and rigid structure so as to provide maximal stability and resistance to subsequent deformation as well as minimal permeability. In either case, knowledge of basic soil structure relationships is essential for efficient management of the soil.

TYPES OF SOIL STRUCTURE

In general, we can recognize three broad categories of soil structure — *single grained*, *massive*, and *aggregated*. When particles are entirely unattached to one another, the structure is completely loose, as it is in the case of a coarse granular soil or an unconsolidated deposit of desert dust. Such soils were labeled *structureless* in the older literature of soil physics, but, since even a haphazard arrangement is a structure of sorts, we prefer the designation *single-grained* structure. On the other hand, when the soil is tightly packed in large cohesive blocks, as is sometimes the case with dried clay, the structure can be called *massive*. Between these two extremes, we can recognize an intermediate condition in which the soil particles are associated in quasi-stable small clods known as *aggregates* or *peds*.

This last type of structure, called *aggregated*, is usually the most desirable condition for plant growth, especially in the critical early stages of germination and seedling establishment. The presence and maintenance of stable aggregates is the essential feature of soil *tilth*, a qualitative term used by agronomists to describe the highly desirable, yet elusive, physical condition in which the soil is an optimally loose, friable, and porous assemblage of stable aggregates, permitting free entry and movement of water and air, easy cultivation and planting, and unobstructed germination and emergence of seedlings as well as the growth of roots.

STRUCTURE OF GRANULAR SOILS

The structure of most coarse-textured soils is single grained, because there is little tendency for the grains to adhere and form aggregates. The arrangement and internal mode of packing of the grains depends on their distribution of sizes and shapes as well as the manner in which the material had been deposited or formed in place. The two extreme cases of possible packing arrangements are, on the one hand, a system of uniform spherical grains in a state of open packing (and hence minimal density) and, on the other hand, a gradual distribution of grain sizes in which progressively smaller grains fill the voids between larger ones in an "ideal" succession that provides maximal density. An assemblage of grains of uniform size is called *monodisperse*, whereas an assemblage of widely varying grain sizes is called *polydisperse*.

Let us consider a hypothetical system consisting entirely of uniform spheres. Although not realistic, this is a useful exercise, inasmuch as it can help us to establish theoretical limits of porosity by which we may later evaluate real systems. With monodisperse spheres, the minimal density and hence maximal porosity is obtained in the case of cubic (open) packing. In this mode of packing, each grain touches six neighbors on opposite sides of three orthogonal axes and hence is said to have a *coordination number* of 6. The bulk density is then $\pi\rho_s/6$ (where ρ_s is the particle density), and the porosity is 47.6%, regardless of the diameter of the spheres. If $\rho_s = 2.65 \times 10^3$ kg/m^3, we obtain a bulk density of about 1.39×10^3 kg/m^3.

In contrast with the cubic arrangement, we may have the tetrahedral or octahedral arrrangement, both of which provide the densest possible packing of uniform spheres. Both of these conditions have a coordination number of 12, a bulk density of $\pi\rho_s/3(2^{1/2})$ (about 1.97×10^3 kg/m^3 = 1.97 g/cm^3), and a porosity of 25.9%. A schematic representation of modes of packing is given in Fig. 5.1. A quantitative summary is given in Table 5.1.

The analysis of polydisperse systems is obviously more complex than that of monodisperse systems, even if we continue to assume spherical shape. A mixture of two grain sizes depends on the relative concentration of the components. For any particular ratio of grain sizes, there exists an optimal composition that will result in minimal porosity. In principle, the density of a polydisperse system consisting of many particle sizes can be greater than that of a monodisperse system, because smaller particles can fit into the spaces between larger ones (Fig. 5.2). Depending on the array of relative particle radii, coordination numbers as high as 30 or more and porosities lower than 20% are possible. The actual porosities of natural sediments generally lie between the theoretically derivable limits for the ideal packings of monodisperse and polydisperse spheres; that is, they range between 25% and 50%.

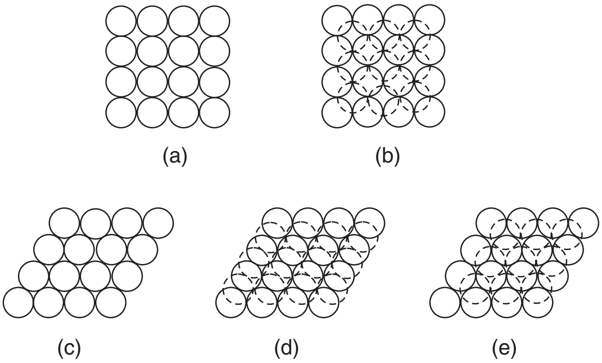


Fig. 5.1. Models of packing of equal spheres. (a) simple cubic, (b) cubical tetrahedral, (c) tetragonal sphenoidal, (d) pyramidal, and (e) tetrahedral. (Deresiewicz, 1958).

TABLE 5.1 **Packing of Spheres**[a]

Type of packing	Coordination number	Spacing of layers	Volume of unit prism	Density	Porosity (%)
Simple cubic	6	$2R$	$8R^3$	$\pi/6$ (0.5236)	47.64
Cubic tetrahedral	8	$2R$	$4\sqrt{3}\,R^3$	$\pi/3\sqrt{3}$ (0.6046)	39.54
Tetragonal sphenoidal	10	$R\sqrt{3}$	$6R^3$	$2\pi/9$ (0.6981)	30.19
Pyramidal	12	$R\sqrt{2}$	$4\sqrt{2}\,R^3$	$\pi/3\sqrt{2}$ (0.7405)	25.95
Tetrahedral	12	$2R\sqrt{2/3}$	$4\sqrt{2}\,R^3$	$\pi/3\sqrt{2}$	25.95

[a] From Deresiewiez (1958, Table 1).

Loose deposits of granular material cannot be compacted very effectively by application of static pressure (which, however, can be very effective in compacting unsaturated clayey soils). Static pressure is merely transmitted along the soil skeleton and borne by the high frictional resistance to the mutual sliding of the grains. Granular material can, however, be compacted by the application of vibratory action. The vibration pulsates the grains and allows smaller ones to enter between larger ones, thus increasing the packing density. The vibration of granular materials is effective in dry and in saturated states, whereas at unsaturated moisture contents surface tension forces of the water menisci that are wedged between the particles lend cohesiveness and hence greater rigidity to the matrix. The structure of nonideal granular materials (in which the grains are not spherical but oblong or angular) is much more difficult to formulate theoretically, because the orientations of particles must be taken into account.

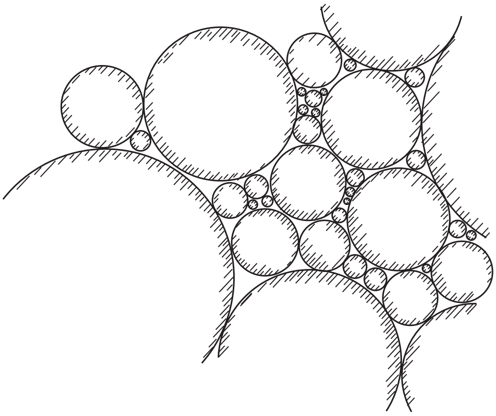


Fig. 5.2. Packing of polydisperse particles (hypothetical).

Sample Problem

Calculate the overall porosity and bulk density values for an assemblage of uniform spherical aggregates in open (cubic) and dense (tetrahedral) packing. Assume that each aggregate has a bulk density ρ_b of 1.8×10^3 kg/m^3 (1.8 g/cm^3) and that particle density ρ_s is 2.65×10^3 kg/m^3. How much of the total porosity is due to macro-(between-aggregate) pores and how much to micro-(within-aggregate) pores in each packing mode?

To determine the microporosity of the aggregates themselves:

Recall that porosity $f = 1 - (\rho_b/\rho_s) = 1 - (1.8/2.65) = 0.32$ and that the volume of a sphere $= (4/3)\pi r^3 = (\pi/6)d^3$, where r is radius and d is diameter.

In cubic packing: Assuming the diameter to be of unit length, each such sphere occupies a cube of unit volume ($d^3 = 1 \times 1 \times 1 = 1$). Therefore the fractional volume of each sphere in its cube $= \pi/6 = 0.5236$.

Hence the macro-(interaggregate) porosity $= 1 - 0.5236 = 0.4764$.
As a fraction of a unit cube, the microporosity $= 0.5236 \times 0.32 = 0.1676$.
Therefore, total porosity $= 0.4764 + 0.1676 = 0.644 = 64.4\%$.

For dense (tetrahedral) packing:

The fractional volume of the spheres $= 0.7405$ (see Table 5.1).
The macroporosity $= 1 - 0.7405 = 0.2595$.
The porosity of each aggregate (as before) $= 0.32$.
As a fraction of the total volume, microporosity $= 0.7405 \times 0.32 = 0.237$.
Therefore, total porosity $= 0.2595 + 0.237 = 0.496 = 49.6\%$.

Of the total porosity, macropores constituted $0.2595/0.496 = 0.523$ and micropores $= 0.478$.

To compute bulk densities (ρ_b):

From $f = 1 - (\rho_b/\rho_s)$, we obtain: $\rho_b = \rho_s (1 - f)$.
For open packing, $\rho_b = 2.65(1 - 0.644) = 0.934$ g/cm^3 $= 0.934 \times 10^3$ kg/m^3.
For dense packing, $\rho_b = 2.65(1 - 0.496) = 1.3396$ g/cm^3 $= 1.34 \times 10^3$ kg/m^3.

Note: Even though the porosity of the aggregates themselves did not change, the overall macroporosity and microporosity (and hence the total porosity and bulk density) changed greatly with the change in packing arrangement.

BOX 5.1 Soil Physics at the Beach

Those who go to the beach merely to swim miss the fun of observing soil physics in action on the sandy shore. The labor of trudging through the soft dry sand is lightened by the realization of why it is so. Sand has such low specific surface that its particles do not bond to one another, so their assemblage has hardly any firmness.

Approaching the shore, however, we notice a zone of firm sand, where walking is much easier. It is the same sand, so what makes it rigid? Aha, it must be the water in the sand. But we know that liquid water is itself a rather “soft” substance, so how can it impart rigidity to the sand? The answer is that water in a moist but unsaturated sand tends to cluster in pendular rings at points where grains touch one another. These rings, also called *menisci*, exhibit surface-tension forces that resist any deformation tending to increase the water-to-air interfacial area.

Now we come to the water's edge and find ourselves walking on saturated sand. Again, it seems soft: In a saturated medium, the water-to-air interfaces disappear and no longer resist deformation. Here, yet another interesting phenomenon comes into play. As each wave of the sea recedes, it leaves behind a strip of glistening, saturated sand. Now if we step on the sand, the area around our foot seems to dry up instantly. That's puzzling: If this were like a saturated sponge, it would spurt water rather than turn dry. So why does saturated sand appear to turn dry when trodden?

The answer lies in a phenomenon called *dilatancy* (pronounced "die late and see" — a good idea in any case). It is the tendency of some materials to expand when subjected to shearing. In the case of sand, the sliding of one plane over another (when a differential force is applied to adjacent areas) causes the rolling of grains over one another. The porosity in the zone of shearing thereby increases. Where this "puffing up" occurs, the volume of water in the sand is no longer sufficient for saturation. Therefore, the sand appears to "dry up."

Now we notice the ripples in the sand under the shallows near the water's edge and the appearance of dark streaks in the troughs between ridgelets. That banding is caused by the segregation of "heavier" minerals (e.g., magnetite, hornblende, zircon) from the "lighter" minerals (e.g., quartz), due to differences in density. The dark minerals are often deposited at the water's edge as the waves — and then the winds — sweep the lighter particles higher up the beach.

Is that all? Certainly not. Notice the frictional sound of the dry sand as it is tread on and the gurgling sound of the sand as it releases entrapped air after each wave's inundation. There is more, much more, to observe at the beach (and everywhere else!).

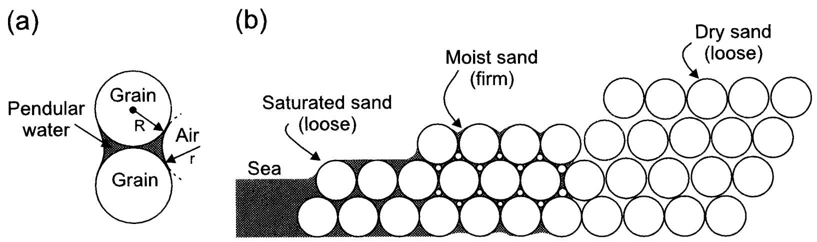


Fig. B5.1. Water in sand: (a) Pendular ring between adjacent grains. (b) The rings disappear in both dry sand and saturated sand.

STRUCTURE OF AGGREGATED SOILS

In soils with an appreciable content of clay, the primary particles tend, under favorable circumstances, to group themselves into composite structural units known as *aggregates*. Such aggregates are not characterized by any universally fixed size, nor are they necessarily stable. The visible aggregates, which are generally of the order of several millimeters to several centimeters in diameter, are often called *peds* or *macroaggregates*. These are usually assemblages of smaller groupings, or *microaggregates*, which themselves are associations of the ultimate structural units, that is, the flocs, clusters, or packets of clay particles. Bundles of these last units attach themselves to, and sometimes engulf, the much larger primary particles of sand and silt. The internal organization of these various groupings can be studied by means of electron microscopy, using both scanning and transmission methods (the latter with thin sections of soil samples embedded in congealed, transparent plastic materials).

A prerequisite for aggregation is that the clay be flocculated. However, flocculation is a necessary but not sufficient condition for aggregation. As stated by Richard Bradfield of Cornell University a half century ago, "Aggregation is flocculation — plus!" That "plus" is *cementation*.

A complex interrelationship of physical, biological, and chemical reactions is involved in the formation and degradation of soil aggregates (Kay and Angers, 2002). An important role is played by the extensive networks of roots that permeate the soil and tend to enmesh soil aggregates. Roots exert pressures that compress aggregates and separate between adjacent ones. Water uptake by roots causes differential dehydration, shrinkage, and the opening of numerous small cracks. Moreover, root exudations and the continual death of roots and particularly of root hairs promote microbial activity, which results in the production of humic cements. Since these binding substances are transitory, being susceptible to further microbial decomposition, organic matter must be replenished and supplied continually if aggregate stability is to be maintained in the long run.

ROLE OF ORGANIC MATTER AND MICROBIAL ACTIVITY

Active humus is accumulated and soil aggregates are stabilized most effectively under perennial sod-forming herbage. Annual cropping systems, on the other hand, hasten the decomposition of humus and the destruction of soil aggregates. The soil surface is especially vulnerable if exposed and desiccated in the absence of a protective cover. The foliage of close-growing vegetation, and its residues, protects surface soil aggregates against slaking by water, particularly under raindrop impact.

Soil structure is affected by the time-variable activity of countless microorganisms, including species of protozoa, bacteria, fungi, actinomycetes, etc. Especially important are rhizospheric bacteria, which flourish in direct association with roots of specific plants, as well as fungi, which often form extensive adhesive networks of fine filaments known as *mycelia* or *hyphae*. The composition of the soil microfauna and microflora depends on the thermal and moisture regimes, on soil pH and oxidation–reduction potential, the nutrient status of the soil substrate, and the type and quantity of organic matter present.

Soil microorganisms bind aggregates by a complex of mechanisms, such as adsorption, physical entanglement and envelopment, and cementation by excreted mucilaginous products. Prominent among the many microbial products capable of binding soil aggregates are polysaccharides, hemicelluloses or uronides, levans, as well as numerous other natural polymers. Such materials are attached to clay surfaces by means of cation bridges, hydrogen bonding, van der Waals forces, and anion adsorption mechanisms. Polysaccharides, in particular, consist of large, linear, and flexible molecules capable of forming multiple bonds with several particles at once. In some cases, organic polymers hardly penetrate between the individual clay particles but form a protective capsule around soil aggregates. In other cases, solutions of active organic agents penetrate into soil aggregates and then precipitate more or less irreversibly as insoluble (though still biologically decomposable) cements.

In addition to increasing the strength and stability of intra-aggregate bonding, organic products may further promote aggregate stability by reducing wettability and swelling. Some of the organic materials are inherently hydrophobic or become so as they dehydrate, so the organo-clay complex may have a reduced affinity for water.

Some inorganic materials can also serve as cementing agents. The importance of clay and of its state of flocculation should be obvious by now. Adhesion between clay particles is the ultimate internal binding force within microaggregates. Calcium carbonate as well as iron and aluminum oxides can impart considerable stability to soil aggregates. The latter are the reason for the remarkable stability of aggregates in tropical soils that may contain little organic matter.

A model of the internal bonding forms that can constitute a soil aggregate is shown in Fig. 5.3. The model is composed of sand or silt-size quartz particles and of "domains" of oriented clay bonded by electrostatic forces. Stability of the aggregate is enhanced by linkage of organic polymers between the quartz particles and the faces or edges of clay crystals.

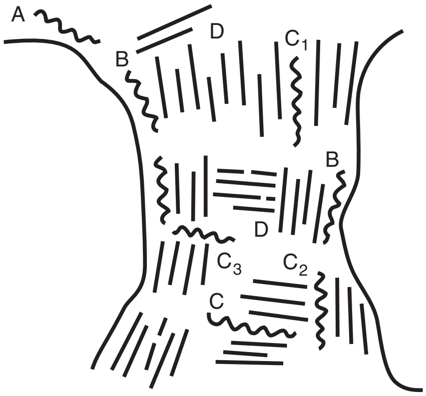
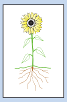


Fig. 5.3. Possible arrangements of quartz particles, clay domains, and organic matter in a soil aggregate: (A) quartz–organic colloid–quartz, (B) quartz–organic colloid–clay domain, (C) clay domain–organic colloid–clay domain, (C_1) face–face, (C_2) edge–face, (C_3) edge–edge, and (D) clay domain edge–clay domain face. (After Emerson, 1959.)



BOX 5.2 The Role of Earthworms

Prominent among the macrobiological fauna affecting soil structure are earthworms, which have aroused the interest of such observers as Aristotle and Darwin. Their role in the soil was described vividly and aptly two centuries ago by George White (quoted by Russel, 1973):

> Worms seem to be the great promoters of vegetation, which would proceed but lamely without them, by boring, perforating, and loosening the soil, and rendering it pervious to rains and fibers of plants, by drawing straws and stalks of leaves and twigs into it, and, most of all, by throwing up such infinite numbers of lumps of earth called wormcasts, which being their excrement, is a fine manure of grain and grass . . . . The earth without worms would soon become cold, hardbound, and void of fermentation, and consequently sterile.

A population of several million earthworms (of various species) per acre is not uncommon in regions where the supply of moisture and fresh organic matter is adequate. Such a population of earthworms can digest and expel as "casts" many tons of soil per acre per year. The earthworm activity may penetrate to a depth of half a meter or more (Blanchart, 1992).

BREAKDOWN OF AGGREGATES

The state of soil aggregation at any time results from the balance between the forces or processes promoting aggregation and those tending to cause its breakdown. A particularly destructive condition may result when thoroughly desiccated aggregates are suddenly submerged in water. The water drawn into each aggregate over its entire periphery may trap and compress the air originally present in the dry aggregate. Because the cohesive strength of the outer part of the clod is reduced by swelling and the pressure of the entrapped air builds up in proportion to its compression, sooner or later the latter exceeds the former and the clod may actually explode. More typically, however, a series of small explosions, each marked by the escape of a bubble of air, shatters the clod into fragments. This destructive process is known as *air slaking* (Fig. 5.4).

Another process by which water can destroy aggregates is the hammering impact of falling raindrops along with the scouring action of flowing water in surface runoff. On the other hand, drying processes, causing shrinkage, can (in themselves) increase aggregate stability by making the aggregates more dense and cohesive as well as by causing gluelike organic gums and gels to "set" irreversibly so as to serve as stable cementing agents. Freezing, incidentally, is itself analogous to drying, in that it causes the extraction of liquid water from regions of the soil toward the freezing sites (*ice lenses*).

With so many factors active simultaneously, aggregate stability is obviously not a simple property but a complex attribute, which depends on the relative strength of the intra-aggregate bonds versus the stresses induced by swelling, scouring, and air entrapment. Accordingly, the task of maintaining aggregation involves strengthening the internal bonding mechanisms while at the same time

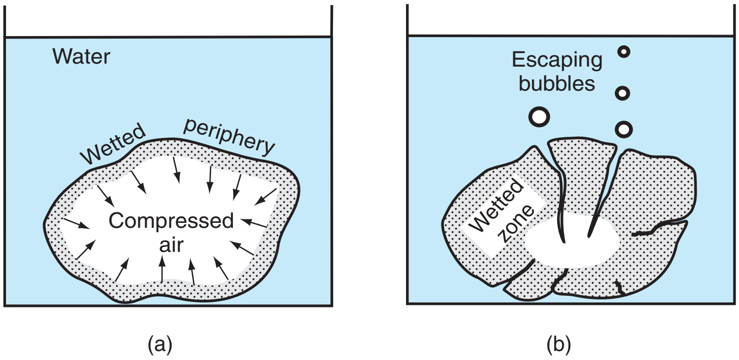


Fig. 5.4. Air slaking of an initially dry aggregate suddenly submerged in water. (a) *Early stage*: The periphery of the aggregate is wetted and water moves into the aggregate, compressing the air inside. (b) *Bursting stage*: Because the wetted zone is weakened by swelling and the pressure of entrapped air increases in proportion to its compression, eventually the aggregate is shattered and air bubbies out. This point may be quite abrupt and result in the collapse of the shattered aggregate.

reducing the destructive forces, for example, preventing sudden submergence and direct exposure of surface aggregates to desiccation, raindrop impact, and running water, as well as avoiding compressive traffic and excessively pulverizing tillage.

CHARACTERIZATION OF SOIL STRUCTURE

The structure of the soil can be studied directly by microscopic observation of thin slices under polarized light. The arrangements of minute clay particles can be examined by means of electron microscopy, using either transmission or scanning techniques. The structure of single-grained soils, as well as of aggregated soils, can be expressed quantitatively in terms of the total porosity and of the pore size distribution. The structure of aggregated soils can, in addition, be characterized qualitatively by specifying the typical shapes of aggregates found in various horizons within the soil profile or quantitatively by measuring their sizes. Additional methods of characterizing soil structure are based on measuring mechanical properties (e.g., resistance to penetration, compression, or shearing) and permeability to fluids. None of these methods has been accepted universally. In each case, the choice of the method to be used depends on the problem, the soil, the equipment available, and, not the least, the soil physicist.

Total porosity f of a soil sample is usually computed from the measured bulk density ρ_b, using the following equation (see Chapter 1):

$$f = 1 - \rho_b/\rho_s \qquad (5.1)$$

where ρ_s is the average particle density.

Bulk density is generally measured by means of a *core sampler* designed to extract “undisturbed” samples of known volume from various depths in the profile. An alternative is to measure the volumes and masses of individual clods (*not* including interclod cavities) by *immersion in mercury* or by coating with paraffin wax prior to *immersion in water*. Still other methods are the *sand-funnel* and *balloon technique*, used in engineering, and *gamma-ray attenuation densitometry* (Grossman and Reinsch, 2002).

Pore size distribution measurements can be made in coarse-grained soils by means of the *pressure-intrusion method* (Flint and Flint, 2002), in which a nonwetting liquid, generally mercury, is forced into the pores of a predried sample. The pressure is applied incrementally, and the volume penetrated by the liquid is measured for each pressure step, equivalent (by the capillary theory) to a range of pore diameters. In the case of fine-grained soils, capillary condensation methods or, more commonly, *desorption methods* are used. In the latter method, a presaturated sample is subjected to a stepwise series of incremental suctions, and the capillary theory is used to obtain the equivalent pore size distribution. Water is commonly used as the permeating liquid, though nonpolar liquids have also been tried in an attempt to assess, by comparison, the possible effect of water saturation and desorption in modifying soil structure.

Where the aggregates are fairly distinct, it is sometimes possible to divide pore size distribution into two distinguishable ranges, *macropores* and *micropores*. The macropores are mostly the interaggregate spaces, which serve as the principal avenues for the infiltration and drainage of water and for aeration. The micropores are the intraaggregate capillaries responsible for the retention of water and solutes. However, the demarcation is seldom truly distinct, so differentiating between macropores and micropores is often arbitrary.

SHAPES OF AGGREGATES

The shapes of aggregates observable in the field (Fig. 5.5) can be classified as follows:

1. *Platy*: Horizontally layered, thin and flat aggregates resembling wafers. Such structures occur, for example, in recently deposited clay soils.

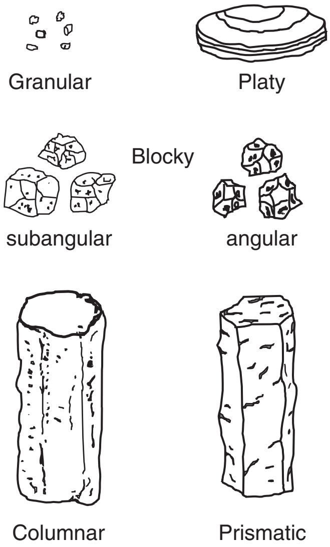


Fig. 5.5. Observable forms of soil aggregation.

TABLE 5.2 **Classification of Soil Structure According to Soil Survey Staff (1951; 1993)**

	A Type: Shape and arrangement of peds						
				Blocklike–polyhedral–spheroidal. Three approximately equal dimensions arranged around a point			
	Platelike. Horizontal axes longer than vertical; arranged around a horizontal plane	**_Prismlike_. Horizontal axes shorter than vertical; arranged around vertical line. Vertices angular**		**_Blocklike–polyhedral_. Plane or curved surfaces accommodated to faces of surrounding peds**		**_Spheroidal–polyhedral_. Plane or curved surfaces not accommodated to faces of surrounding peds**	
		Without rounded caps	**With rounded caps**	**Faces flattened; vertices sharply angular**	**Mixed rounded, flattened faces; many rounded vertices**	**Relatively nonporous peds**	**Porous peds**
B Class: Size of peds	**Platy**	**Prismatic**	**Columnar**	**Blocky**	**Subangular blocky**	**Granular**	**Crumb**
1. Very fine or very thin	<1 mm	<10 mm	<10 mm	<5 mm	<5 mm	<1 mm	1 mm
2. Fine or thin	1–2 mm	10–20 mm	10–20 mm	5–10 mm	5–10 mm	1–2 mm	1–2 mm
3. Medium	2–5 mm	20–50 mm	10–20 mm	10–20 mm	10–20 mm	2–5 mm	2–5 mm
4. Coarse or thick	5–10 mm	50–100 mm	50–100 mm	20–50 mm	20–50 mm	5–10 mm	
5. Very coarse (very thick)	>10 mm	>100 mm	>100 mm	>50 mm	>50 mm	>10 mm	

C Grade: Durability of peds		
	0. Structureless	No aggregation or orderly arrangement.
	1. Weak	Poorly formed, nondurable, indistinct peds that break into a mixture of a few entire and many broken peds and much unaggregated material.
	2. Moderate	Well-formed, moderately durable peds, indistinct in undisturbed soil, that break into many entire and some broken peds but little unaggregated material.
	3. Strong	Well-formed, durable, distinct peds, weakly attached to each other, that break almost completely into entire peds.

2. *Prismatic* or *columnar*: Vertically oriented pillars, up to 15 cm in diameter. Such structures are common in the B horizon of clayey soils, particularly in semiarid regions. Where the tops are flat, these vertical aggregates are called *prismatic*, and where rounded, *columnar*.
3. *Blocky*: Cubelike blocks of soil, up to 10 cm in size, sometimes angular with well-defined planar faces. These structures occur most commonly in the upper part of the B horizon.
4. *Spherical*: Rounded aggregates, generally not much larger than 2 cm in diameter, often found in a loose condition in the A horizon. Such units are called *granules* and, where particularly porous, *crumbs*.

The shapes of aggregates, as well as their sizes and densities, generally vary within the profile. Because the overburden pressure increases with depth, and inasmuch as the deeper layers do not experience such extreme fluctuations in moisture content as does the alternately saturated and desiccated surface layer, the decrease of swelling and shrinkage activity causes the deeper aggregates to be larger. A typical structural profile in semiarid regions consists of a granulated A horizon underlain by a prismatic B horizon, whereas in humid temperate regions a granulated A horizon may occur with a platy or blocky B horizon. The number of variations found in nature are, however, legion. A detailed classification of aggregate shapes is given in Table 5.2.

AGGREGATE SIZE DISTRIBUTION

Aggregate size distribution is an important determinant of the soil's pore size distribution and has a bearing on the erodibility of the soil surface, particularly by wind. In the field, adjacent aggregates often adhere to one another, though of course not as tenaciously as do the particles within each aggregate. Separating and classifying soil aggregates necessarily involves a disruption of the original, in situ structural arrangement. The application of too great a force may break up the aggregates themselves. Hence the determination of aggregate size distribution depends on the mechanical means employed to separate the aggregates.

Screening through flat sieves is difficult to standardize and entails frequent clogging of the sieve openings. A more practical approach is to use a rotary sieve machine with concentrically nested sieves of different aperture sizes. The operation of such a device can be standardized, thus minimizing the arbitrary personal factor, and clogging can be practically eliminated.

Various indexes have been proposed for the distribution of aggregate sizes (Nimmo and Perkins, 2002). If a single characteristic parameter is desired (so as to allow correlation with such factors as erosion, infiltration, evaporation, and aeration), a method must be adopted for assigning an appropriate weighting factor to each size range of aggregates. One of the most widely used indexes is the *mean weight diameter*, based on weighting the masses of aggregates of the various size classes according to their respective sizes. The mean weight diameter X is thus defined by the following equation:

$$X = \sum_{i=1}^{n} x_i w_i \qquad (5.2)$$

Here x_i is the mean diameter of any particular size range of aggregates separated by sieving and w_i is the weight of the aggregates in that size range as a fraction of the total dry weight of the sample. The summation accounts for all size ranges, including the group of aggregates smaller than the openings of the finest sieve.

An alternative index of aggregate size distribution is the *geometric mean diameter* Y, calculated according to the following equation:

$$X = \exp\left[\left(\sum_{i=1}^{n} w_i \log x_i\right) \middle/ \left(\sum_{i=1}^{n} w_i\right)\right] \tag{5.3}$$

where w_i is the weight of aggregates in a size class of average diameter x_i and the denominator $\Sigma\, w_i$ (for i values from 1 to n) is the total weight of the sample.

Sample Problem

Calculate the mean weight diameters of the assemblages of aggregates given in Table 5.3. The percentages refer to the mass fractions of dry soil in each diameter range.

First we determine the mean diameters of the seven aggregate diameter ranges:

Range: 0–0.5, 0.5–1, 1–2, 2–5, 5–10, 10–20, 20–50 mm
Mean: 0.25, 0.75, 1.5, 3.5, 7.5, 15, 35 mm

Recall that the mean weight diameter X is defined by Eq. (5.2):

$$X = \sum_{i=1}^{n} x_i w_i \tag{5.2}$$

Hence, for the dry-sieved virgin soil,

$$Y = (0.25 \times 0.1) + (0.75 \times 0.1) + (1.5 \times 0.15) + (3.5 \times 0.15) + (7.5 \times 0.2) + (15 \times 0.2) + (35 \times 0.1) = 8.85 \text{ mm}$$

For the dry-sieved cultivated soil,

$$X = (0.25 \times 0.25) + (0.75 \times 0.25) + (1.5 \times 0.15) + (3.5 \times 0.15) + (7.5 \times 0.1) + (15 \times 0.07) + (35 \times 0.03) = 4.30 \text{ mm}$$

For the wet-sieved virgin soil,

$$X = (0.25 \times 0.3) + (0.75 \times 0.15) + (1.5 \times 0.15) + (3.5 \times 0.15) + (7.5 \times 0.15) + (15 \times 0.05) + (35 \times 0.05) = 4.56 \text{ mm}$$

For the wet-sieved cultivated soil,

$$X = (0.25 \times 0.5) + (0.75 \times 0.25) + (1.5 \times 0.15) + (3.5 \times 0.05) + (7.5 \times 0.04) + (15 \times 0.01) + (35 \times 0.0) = 1.16 \text{ mm}$$

Note: Wet sieving reduced the mean weight diameter from 8.85 to 4.56 mm in the virgin soil and from 4.30 to 1.16 mm in the cultivated soil. This indicates the degree of instability of the various aggregates under the slaking effect of immersion in water. The influence of cultivation is generally to reduce the water stability of soil aggregates and hence to render the soil more vulnerable to crusting and erosion processes.

TABLE 5.3 **Aggregate Data for Sample Problem**

Aggregate diameter range (mm)	Dry sieving		Wet sieving	
	Virgin soil (%)	Cultivated soil (%)	Virgin soil (%)	Cultivated soil (%)
0.0–0.5	10	25	30	50
0.5–1.0	10	25	15	25
1–2	15	15	15	15
2–5	15	15	15	5
5–10	20	10	15	4
10–20	20	7	5	1
20–50	10	3	5	0

AGGREGATE STABILITY

Determining the state of aggregation of a soil at any particular moment might not suffice to portray the dynamic nature of soil structure. By whatever measure, the state of aggregation varies over time and space as aggregates form, disintegrate, and reform periodically. For instance, a newly cultivated field may for a time exhibit a nearly optimal array of aggregate sizes, with large interaggregate pores favoring high infiltration rates and unrestricted aeration. However, soil structure may begin to deteriorate quite visibly and rapidly, because the soil is subject to destructive forces resulting from intermittent rainfall (causing swelling and shrinking, slaking and erosion) followed by dry spells (exposing the soil to deflation by wind). Repeated traffic, especially by heavy machinery, further tends to crush the aggregates remaining at the surface and to compact the soil to some depth.

Soils vary, of course, in the degree of their vulnerability to externally imposed destructive forces. *Aggregate stability* is a measure of this vulnerability. More specifically, it expresses the resistance of aggregates to breakdown when subjected to potentially disruptive processes. The reaction of a soil to forces acting on it depends not only on the soil itself but also on the nature of the forces and the manner they are applied.

To test aggregate stability, soil physicists generally subject samples of aggregates to artificially induced forces designed to simulate processes that are likely to occur in the field. The nature of the forces applied during such testing depends on the investigator's perception of the phenomenon to be simulated. The degree of stability is then assessed by determining the fraction of the original sample's mass that has withstood destruction (Nimmo and Perkins, 2002).

If an indication of mechanical stability is sought, measurements can be made of the resistance of aggregates to prolonged dry sieving or to crushing forces. Most frequently, however, the concept of aggregate stability is applied in relation to the destructive action of water. Although mentioned before, it bears repeating that the very wetting of aggregates may cause their collapse, because the bonding substances dissolve or weaken and as the clay swells and possibly disperses. Aggregates are more vulnerable to sudden than to gradual wetting, owing to the air occlusion effect. Raindrops and flowing water provide the energy to detach particles and transport them away. Abrasion by particles carried as suspended matter in runoff water may scour the surface and contributes to the overall breakdown of the aggregated structure at the soil surface.

The classical and still most prevalent procedure for testing the *water stability* of soil aggregates is the *wet sieving* method. A representative sample of air-dry aggregates is placed on the uppermost of a set of graduated sieves and immersed in water to simulate flooding. The sieves are then oscillated vertically so that water is made to flow up and down through the screens and the assemblage of aggregates. At the end of a specified period of sieving (e.g., 20 min) the nest of sieves is removed from the water and the oven-dry weight of material left on each sieve is determined. The results should be corrected for the coarse primary particles retained on each sieve to avoid designating them falsely as aggregates. This is done by dispersing the material collected from each sieve, using a mechanical stirrer and a sodic dispersing agent, and then washing the material back through the same sieve. The weight of sand retained after the second sieving is then subtracted from the total weight of undispersed material retained after the first sieving, and the percentage of stable aggregates %SA is given by

$$\%\mathrm{SA} = 100 \times \frac{(\text{weight retained}) - (\text{weight of sand})}{(\text{total sample weight}) - (\text{weight of sand})} \tag{5.4}$$

An alternative approach is to subject soil aggregates to simulated rain. In the *drop method*, aggregates are bombarded with drops of water in a standardized manner. The number of drops needed for total dissipation of the aggregates or the fractional mass of the aggregates remaining after a given time, can be determined. Another way is to subject a soil in the field to simulated rainfall of controllable raindrop sizes and velocities. The condition of the soil surface can then be compared to the initial condition.

To determine the stability of microaggregates, comparative *sedimentation analysis* can be carried out with dispersed versus undispersed samples. An index indicating the fractional amount of clay associated in microaggregates can then be calculated. Still another index of soil structural stability is obtained by comparing the *permeability* of the soil to an inert fluid with its permeability to water. The permeability of an appropriately packed soil sample is first measured by using air, a fluid that is presumed to have no effect on structure, and then by using water. A ratio of unity, if such were to occur, would indicate perfect stability. Values greater than unity indicate a scale of increasing instability.

SOIL CRUSTING

The aggregates at the soil surface are the most vulnerable to destructive forces. The aggregates that collapse during wetting may form a layer of dispersed mud, typically several millimeters thick, which clogs the macropores of the top layer and thus tends to inhibit the infiltration of water and the exchange of gases between the soil and the atmosphere. Such a layer is often called a *surface seal*. As it dries, this dispersed layer shrinks to become a dense, hard *crust*, which impedes seedling emergence by its hardness and tears seedling roots as it cracks, forming a characteristic polygonal pattern. The effect of soil crusting on seedlings depends on crust thickness and strength as well as on the size and vigor of the seedlings. Most sensitive are small-seeded vegetables and grasses. In soils prone to crusting, seedling emergence may occur only through the crust's cracks.

Attempts have been made to characterize soil crusting, particularly with respect to its effect on seedling emergence, in terms of the resistance of the dry crust to the penetration of a probe (Parker and Taylor, 1965) as well as in terms of its mechanical strength as measured by the *modulus of rupture* test or the *tensile strength* test (see Chapter 13). These tests were designed to imitate the process by which a seedling forces its way upward by penetrating and rupturing the crust. However, the critical crust strength that prevents emergence obviously depends on crust thickness and soil wetness as well as on plant species and depth of seed placement (Hillel, 1972b).

Crust strength increases as the degree of colloidal dispersion increases. The evaporation process charges the soil surface zone with relatively high concentrations of sodic salts and, consequently, with a high exchangeable sodium percentage. With subsequent infiltration of rain or irrigation water, the salts are leached but the exchangeable sodium percentage (ESP) remains high. The resulting combination of high ESP and low salt concentration induces colloidal dispersion, which further contributes to the formation of a hard crust.

SOIL CONDITIONERS

We have already described the beneficial effect on soil aggregation of various natural polymers, products of the microbial decay of organic matter in the soil. Such substances as polysaccharides and polyuronides promote aggregate stability by gluing particles together within aggregates as well as by coating aggregate surfaces. As the soil dries, the gel-like glues undergo practically irreversible dehydration, becoming more or less stable cementing agents that bind flocs of clay to one another as well as to silt and sand grains.

Considerable work has been done to develop and apply synthetic compounds capable of duplicating the effect of natural polymers. Where natural aggregation or aggregate stability is lacking, such synthetic polymers, called *soil conditioners*, can help in the stabilization of aggregates (Wallace and Terry, 1997). Materials have indeed been produced that are effective in relatively small quantities (e.g., 0.1% of the treated soil mass) and can produce a dramatic improvement of soil structure, especially in the soil's upper layer, with consequent beneficial effects upon infiltration, aeration, and the prevention of crusting and erosion. The practical application of such products in the field, however, is still constrained by their high cost.

This Page Intentionally Left Blank

Part III

THE LIQUID PHASE

6. WATER CONTENT AND POTENTIAL

SOIL WETNESS

The variable amount of water contained in a unit mass or volume of soil and the energy state of water in the soil are important factors affecting the growth of plants. Numerous other soil properties depend on water content. Among these are mechanical properties, such as consistency, plasticity, strength, compactibility, penetrability, stickiness, and trafficability. In clayey soils, swelling and shrinking associated with addition or extraction of water change the bulk density, porosity, and pore size distribution. Soil water content also governs the air content and gas exchange of the soil, thus affecting the respiration of roots, the activity of microorganisms, and the chemical state of the soil (e.g., oxidation–reduction potential).

The per-mass or per-volume fraction of water in the soil is expressed in terms of *soil wetness*. The physicochemical condition or state of soil water is characterized in terms of its *free energy* per unit mass, called the *potential* (a term that alludes to the "potential energy" of water in the soil). Of the various components of this potential, the *matric potential* manifests the tenacity with which soil water is held by the soil matrix.

Wetness and matric potential are functionally related, and the graphical representation of this relationship is termed the *soil-moisture characteristic curve*. The relationship is not unique, however, because it is affected by the direction and rate of change of soil moisture and is sensitive to changes in soil structure. Wetness and matric potential vary in space and time as the soil is repeatedly wetted by rain, drained by gravity, and dried by evaporation and root extraction.

The lowest wetness we are likely to encounter in nature is a variable state called *air dryness*; in the laboratory it is an arbitrary state known as the *oven-dry* condition. On the other hand, the wettest possible condition of a soil is that of *saturation*, defined as a condition in which all soil pores are filled with water. Saturation is relatively easy to define in the case of nonswelling (e.g., sandy) soils. It can be difficult to define in the case of swelling soils, because such soils may continue to imbibe water and swell even after all pores have been filled with water.

In the field, however, the soil seldom attains complete saturation, because bubbles of air may form and remain occluded within the matrix even when the soil is flooded with excess water. Moreover, air bubbles may effervesce within the soil due to microbial action or whenever the temperature rises and the solubility of gases is exceeded. For this reason, some investigators prefer the term *satiation* to describe the condition in the field in which the soil, though not completely saturated, is as wet as it can get under the circumstances.

Here, we prefer the term *soil wetness* to the older (and still prevalent) term *soil water content*, not only for reasons of verbal economy but also because "wetness" implies intensity, whereas "content" implies extensity. One can speak of the water content when referring to the total amount of water in a bucket. In contrast, the wetness of the soil pertains to the relative amount rather than the absolute amount of water in a soil body, independent of its size.

MASS AND VOLUME RATIOS

The fractional content of water in the soil can be expressed in terms of either mass or volume ratios. As given in Chapter 1,

$$w = M_w/M_s \tag{6.1}$$

$$\theta = V_w/V_t = V_w/(V_s + V_w + V_a) \tag{6.2}$$

where w, the mass wetness, is the dimensionless ratio of water mass m_w to dry soil mass M_s and θ, the volume wetness, is the ratio of water volume V_w to total (bulk) soil volume V_t. The latter is equal to the sum of the volumes of solids (V_s), water (V_w), and air (V_a).

Both θ and w are usually multiplied by 100 and reported as percentages by volume or mass, respectively. The two expressions can be related to each other from knowledge of the bulk density ρ_b of the soil and the density of water:

$$\theta = w(\rho_b/\rho_w) = w\Gamma_b \tag{6.3}$$

where Γ_b is the bulk specific gravity of the soil (a dimensionless ratio between the soil bulk density and the water density, Γ_b generally varies between 1.1 and 1.7). The conversion is simple for nonswelling soils in which bulk density, hence bulk specific gravity, are constant regardless of wetness. However, the conversion can be difficult in the case of swelling soils, for which the bulk density is a function of mass wetness. The relation of bulk density (or of its reciprocal, the *bulk specific volume*) to mass wetness is illustrated schematically in Fig. 6.1.

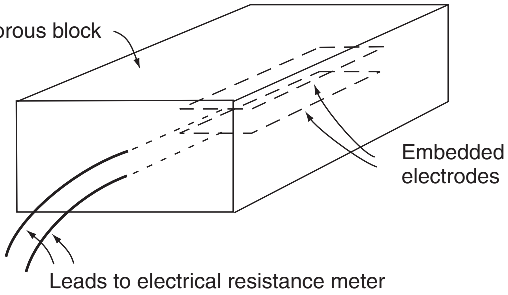


Fig. 6.1. An electrical resistance block. The embedded electrodes may be plates, screens, or wires in a parallel or concentric arrangement.

In many cases, it is useful to express the water content of a soil profile in units of *depth*, that is, as the volume of water contained in a specified total depth of soil d_t per unit area of land. This indicates the equivalent depth d_w that soil water would have if it were extracted and then ponded over the soil surface. Thus

$$d_w = \theta d_t = w\Gamma_b d_t \tag{6.4}$$

Usually d_w is given in millimeters, as are rainfall and evaporation.

Another expression of soil wetness is the liquid ratio θ_r, defined as the volume of water per unit volume of the solid phase:

$$\theta_r = w(\rho_s/\rho_w) = \theta(\rho_w/\rho_b)(\rho_s/\rho_w) = \theta(\rho_s/\rho_b) \tag{6.5}$$

This expression, along with *void ratio* (defined in Chapter 1), is useful mainly with soils having a nonrigid (swelling and shrinking) matrix.

Sample Problem

The data in Table 6.1 were obtained before and after an irrigation. From these data, calculate the mass and volume wetness values of each layer before and after the irrigation, and determine the amount of water (in millimeters) added to each layer and to the profile as a whole.

Using Eq. (6.6), we obtained the following mass wetness values:

$$w_1 = (0.160 - 0.150)/(0.150 - 0.050) = 0.1$$
$$w_2 = (0.146 - 0.130)/(0.130 - 0.050) = 0.2$$
$$w_3 = (0.230 - 0.200)/(0.200 - 0.050) = 0.2$$
$$w_4 = (0.206 - 0.170)/(0.170 - 0.050) = 0.3$$

Using Eq. (6.3), we obtain the following volume wetness values:

$$\theta_1 = 1.2 \times 0.1 = 0.12$$
$$\theta_2 = 1.5 \times 0.2 = 0.30$$
$$\theta_3 = 1.2 \times 0.2 = 0.24$$
$$\theta_4 = 1.5 \times 0.3 = 0.45$$

TABLE 6.1 Data for Sample Problem

Sampling time	Sample number	Depth (m)	Bulk density (kg/m^3)	Wet sample + container (kg)	Dry sample + container (kg)	Container (kg)
Before irrigation	1	0.0–0.4	1.2×10^3	0.160	0.150	0.05
	2	0.6–1.0	1.5×10^3	0.146	0.130	0.05
After irrigation	3	0.0–0.4	1.2×10^3	0.230	0.200	0.05
	4	0.6–1.0	1.5×10^3	0.206	0.170	0.05

Using Eq. (6.4), we obtain the following water depths per layer:

$$d_{w_1} = 0.12 \times 400 \text{ mm} = 48 \text{ mm}$$
$$d_{w_2} = 0.30 \times 600 \text{ mm} = 180 \text{ mm}$$
$$d_{w_3} = 0.24 \times 400 \text{ mm} = 96 \text{ mm}$$
$$d_{w_4} = 0.45 \times 600 \text{ mm} = 270 \text{ mm}$$

Depth of water in profile before irrigation = 48 + 180 = 228 mm.
Depth of water in profile after irrigation = 96 + 270 = 366 mm.
Depth of water added to top layer = 96 – 48 = 48 mm.
Depth of water added to bottom layer = 270 – 180 = 90 mm.
Depth of water added to entire profile = 48 + 90 = 138 mm.

MEASUREMENT OF SOIL WETNESS

The need to determine the amount of water contained in the soil arises frequently in many agronomic, ecological, and hydrological investigations aimed at understanding the soil's chemical, mechanical, hydrological, and biological relationships. There are direct and indirect methods to measure soil moisture (Gardner, 1986; Topp and Ferré, 2002), and, as we have already pointed out, there are several alternative ways to express it quantitatively. As yet there is no universally recognized standard method of measurement and no uniform way to compute and present the results of soil-moisture measurements. We next describe, briefly, some of the most prevalent methods for this determination.

Sampling and Drying

The traditional (gravimetric) method of measuring mass wetness consists of removing a sample by augering into the soil and then determining its moist and dry weights. The *moist weight* is determined by weighing the sample as it is at the time of sampling; the *dry weight* is obtained after drying the sample to a constant weight in an oven. The more or less standard method of drying is to place the sample in an oven at 105°C for 24 hours. An alternative method of drying, suitable for field use, is to impregnate the sample in a heat-resistant container with alcohol, which is then burned off, thus vaporizing the water. The *mass wetness*, also called *gravimetric wetness*, is the ratio of the weight loss in drying to the dry weight of the sample (mass and weight being proportional):

$$w = \frac{(\text{wet weight}) - (\text{dry weight})}{\text{dry weight}} = \frac{\text{weight loss in drying}}{\text{weight of dried sample}} \tag{6.6}$$

The gravimetric method, depending as it does on sampling, transporting, and repeated weighings, entails practically inevitable errors. It is also laborious and time consuming, since the samples must be transported from the field to the laboratory and a period of at least 24 hr is usually considered necessary for complete oven drying. The standard method of oven drying is itself arbitrary. Some clays may still contain appreciable amounts of adsorbed water even at 105°C. On the other hand, some organic matter may oxidize and decompose at this temperature, so the weight loss may not be due entirely to the evaporation of water.

The errors of the gravimetric method can be reduced by increasing the sizes and number of samples. However, the extraction of samples from the field is an invasive and destructive exercise, which may disturb an observation or experimental plot sufficiently to distort the results. Hence many workers prefer indirect methods, which, once installed and calibrated, permit repeated or continuous measurements at the same points with much less time, labor, and soil disturbance.

Electrical Resistance

The electrical resistance of a soil body depends not only on its water content, but also on its composition, texture, and soluble-salt concentration. On the other hand, the electrical resistance of porous bodies placed in the soil and left to equilibrate with soil moisture can sometimes be calibrated against soil wetness. Such units (called *electrical resistance blocks*) generally contain a pair of electrodes embedded in a porous material such as gypsum, nylon, or fiberglass. (See Fig. 6.1.)

Porous blocks placed in the soil tend to equilibrate with the matric suction (tension) of soil water (see later in this chapter) rather than with the water content per se. Different soils can have greatly differing wetness versus suction relationships (e.g., a sandy soil may retain less than 5% moisture at, say, 15-bar suctions, whereas a clayey soil may retain three times as much). Calibration of porous blocks against suction is therefore preferable to calibration against soil wetness, particularly when the soil used for the purpose is a disturbed sample differing in structure from the soil in situ.

The equilibrium of porous blocks with soil moisture may be affected by hysteresis, that is, by the direction of change of soil moisture (whether increasing or decreasing) prior to the equilibration. Moreover, the hydraulic properties of the blocks themselves or inadequate contact with the soil may prevent the rapid attainment of equilibrium and cause a time lag between the state of water in the soil and the state of water being measured in the block. This effect, as well as the sensitivity of the block, may not be constant over the entire range of variation in soil wetness. Gypsum blocks are generally more responsive in the dry range, whereas porous nylon blocks, because of their larger pore sizes, are more sensitive in the wet range of soil-moisture variation.

The electrical conductivity of a porous block made of inert material is due primarily to the permeating fluid rather than to the block's solid matrix. Thus it depends on the electrolytic solutes present in the fluid as well as on the volume content of the fluid. Blocks made of such materials as fiberglass, for instance, are highly sensitive to even small variations in salinity of the soil solution. On the other hand, blocks made of plaster of Paris (gypsum) maintain a nearly constant electrolyte concentration corresponding primarily to that of a saturated solution of calcium sulfate. This tends to mask, or buffer, the effect of small or even moderate variations in the soil solution (such as those due to fertilization or low levels of salinity). However, an undesirable consequence of the solubility of gypsum is that these blocks eventually deteriorate in the soil. Hence the relationship between electrical resistance and moisture suction varies not only from block to block but also for each block over time, because the gradual dissolution of the gypsum changes the internal porosity.

For these and other reasons (e.g., temperature sensitivity), the evaluation of soil wetness by means of electrical resistance blocks is likely to be of limited accuracy. On the other hand, an advantage of such blocks is that they can be connected to a recorder to obtain a continuous indication of soil moisture changes in situ.

Neutron Scattering

This method has gained widespread acceptance as an efficient and reliable technique for monitoring soil moisture in the field. Its principal advantages over the gravimetric method are that it allows less laborious, more rapid, nondestructive (after initial installation), and periodically repeatable measurements, in the same locations and depths, of the volumetric wetness of a representative volume of soil. The method is practically independent of temperature and pressure. Its main disadvantages, however, are the high initial cost of the instrument, low degree of spatial resolution, difficulty of measuring moisture in the soil surface zone, and the health hazard associated with exposure to neutron and gamma radiation.

The instrument, known as a *neutron moisture meter* (Fig. 6.2), consists of two main components: (a) a *probe* (containing a *source of fast neutrons* and a *detector of slow neutrons*), which is lowered into an *access tube* inserted vertically into the soil, and (b) a *scaler* or *rate meter* (usually battery powered and portable) to monitor the flux of the slow neutrons that are scattered and attenuated in the soil.

The purpose of the access tube is both to maintain the bore hole into which the probe is lowered and to standardize measuring conditions. Aluminum tubing is usually the preferred material for access tubes since it is nearly transparent to a neutron flux.

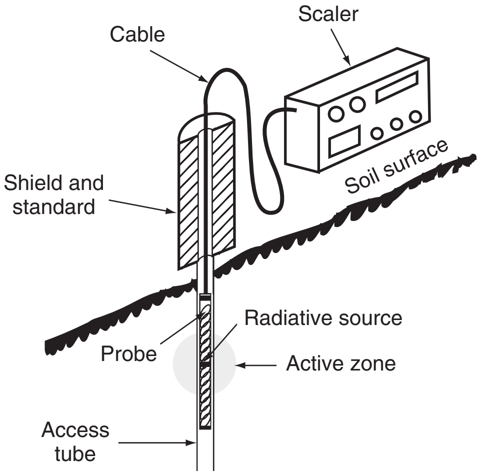


Fig. 6.2. Components of a portable neutron soil-moisture meter, including a probe (with a source of fast neutrons and a detector of slow neutrons) lowered from a shield containing hydrogenous material (e.g., paraffin, polyethylene) into the soil via an access tube. A scaler-rate meter is shown alongside the probe. Recent models incorporate the scaler into the shield body, and the integrated unit is lightweight for easy portability.

A source of fast neutrons is generally obtained by mixing a radioactive emitter of *alpha particles* (helium nuclei) with beryllium. Frequently used is a 2- to 5-millicurie pelletized mixture of radium and beryllium. An Ra-Be source emits about 16,000 neutrons per second per milligram (or millicurie) of radium. The energies of the neutrons emitted by this source vary from 1 to 15 MeV (million electron volts), with a preponderant energy range of 2–4 MeV and an average speed of about 1600 km/sec. Hence, they are called *fast neutrons*. An alternative source of fast neutrons is a mixture of americium and beryllium. Both radium and americium incidentally also emit gamma radiation, but that of the americium is lower in energy and hence less hazardous than that of the radium. The source materials are chosen for their longevity (e.g., radium-beryllium has a half-life of about 1620 yr) so that they can be used for a number of years without appreciable change in radiation flux.

The fast neutrons are emitted radially into the soil, where they encounter and collide elastically (as do billiard balls) with various atomic nuclei. Through repeated collisions, the neutrons are deflected and “scattered,” and they gradually lose some of their kinetic energy. As the speed of the initially fast neutrons diminishes, it approaches a speed that is characteristic for particles at the ambient temperature. For neutrons this is about 2.7 km/sec, equivalent to an energy of about 0.03 eV. Neutrons slowed to such a speed are said to be *thermalized* and are called *slow neutrons*. Such neutrons continue to interact with the soil and are eventually absorbed by the nuclei present.

The effectiveness of various nuclei present in the soil in moderating or thermalizing fast neutrons varies widely. The average loss of energy is maximal for collisions between particles of approximately the same mass. Of all nuclei encountered in the soil, the ones most nearly equal in mass to neutrons are the nuclei of hydrogen (protons), which are therefore the most effective moderators of fast neutrons in the soil. The average number of collisions required to slow a neutron from 2 MeV to thermal energies is 18 for hydrogen, 114 for carbon, 150 for oxygen, and $9N + 6$ for nuclei of larger mass number N. If the soil contains an appreciable concentration of hydrogen, the emitted fast neutrons are thermalized within close proximity of the source.

The slow neutrons thus produced scatter randomly in the soil, quickly forming a swarm or cloud of constant density around the probe. The equilibrium density of the slow neutron cloud is determined by the rate of emission by the source and the rates of thermalization and absorption by the medium (i.e., soil) and is established within a small fraction of a second. Certain elements that might be present in the soil (e.g., boron, cadmium, and chlorine) exhibit a high absorption capacity for slow neutrons, and their presence in nonnegligible concentrations might tend to reduce the density of slow neutrons. By and large, however, the density of slow neutrons formed around the probe is nearly proportional to the concentration of hydrogen in the medium surrounding the probe and therefore is approximately proportional to the volume fraction of water present in the soil. Thus

$$N_w = m\theta + b \qquad \text{and} \qquad N_w/N_s = y\theta \tag{6.7}$$

in which N_w is slow-neutron count rate in wet soil, N_s is the count rate in water or in a standard absorber (i.e., the shield of the probe), θ is volumetric wetness, y is a constant, and m and b are the slope and intercept, respectively, of the line indicating N_w as a function of θ.

As the thermalized neutrons repeatedly collide and bounce about randomly, a number of them (proportional to the density of neutrons thus thermalized and scattered and, therefore, approximately linearly related to the concentration of soil moisture) return to the probe. Here they are counted by the detector of slow neutrons. The detector cell is usually filled with BF_3 gas. When a thermalized neutron encounters a ^{10}B nucleus and is absorbed, an alpha particle is emitted, creating an electrical pulse on a charged wire. The number of pulses over a measured time interval is counted by a scaler or indicated by a rate meter (Hignett and Evett, 2002).

The effective volume of soil in which the water content is measured depends on the energy of emitted neutrons as well as on the concentration of hydrogen nuclei; that is, for a given source and soil, it tracks the volume concentration of soil moisture. If the soil is relatively dry, the cloud of slow neutrons surrounding the probe will be less dense and extend farther from the source, and vice versa for wet soil. With commonly used radium-beryllium and americium-beryllium sources, the so-called *sphere of influence*, or effective volume of measurement, varies with a radius of less than 0.10 m in a wet soil to 0.25 m or more in a dry soil. The low and variable degree of spatial resolution makes the neutron moisture meter unsuitable for the detection of moisture profile discontinuities (e.g., wetting fronts or boundaries between layers). Measurements made within 0.20 m of the surface are unreliable because of the possible escape of fast neutrons through the surface. However, a special surface probe is available commercially to allow measurement of the average moisture in the soil's top layer.

For the sake of safety, and also to provide a convenient means of making standard readings, the probe containing the radiation source is normally carried inside a protective shield designed to prevent the escape of gamma-rays as well as of fast neutrons. The shield is usually a cylindrical container lined with lead and filled with some hydrogenous material, such as paraffin or polyethylene. Improper or excessive use of the equipment can be hazardous. The danger from exposure to radiation depends on the strength of the source, the quality of the shield, the distance from source to operator, and the duration of contact. With strict observance of safety rules, however, the equipment can be used without undue risk.

Sample Problem

Calibration of a neutron probe shows that when a soil's volumetric wetness is 15%, we get a reading of 24,000 cpm (counts per minute), and at a wetness of 40% we get 44,000 cpm. Find the equation of the straight line defining the calibration ($Y = mX + b$, where Y is counts per minute, X is volumetric wetness, m is the slope of the line, and b is the intercept on the Y axis). Using the equation derived, find the wetness value corresponding to a count rate of 30,000 cpm.

We first obtain the slope m:

$$m = (Y_2 - Y_1)/(X_2 - X_1) = (44{,}000 - 24{,}000)/(40 - 15) = 800 \text{ cpm per 1\% wetness}$$

We next obtain the Y intercept b:

$$Y = 800X + b, \qquad b = Y - 800X = 44{,}000 - 800 \times 40 = 12{,}000 \text{ cpm}$$
$$(\text{or: } 24{,}000 - 800 \times 15 = 12{,}000 \text{ cpm})$$

The complete equation is therefore

$$Y = 800X + 12{,}000$$

Now, to find the wetness value corresponding to 30,000 cpm, we set

$$30{,}000 = 800X + 12{,}000, \qquad X = (30{,}000 - 12{,}000)/800 = 22.5\%$$

Gamma-Ray Absorption

The gamma-ray scanner for measuring soil moisture generally consists of two spatially separated units, or probes: (1) a source, usually containing a pellet of radioactive cesium (^{137}Cs emitting gamma radiation with an energy of 0.661 MeV), and (2) a detector, normally consisting of a scintillation counter (e.g., a sodium iodide crystal or a synthetic scintillator) connected to a photomultiplier and preamplifier. If the emission of radiation is monoenergetic and radial and if the space between the source and the detector is empty and the two units are a constant distance apart, then the fraction of the emitted radiation received by the detector will depend only on the angular section intercepted, that is, on the distance of separation and the size of the scintillation unit. On the other hand, should the space between the units be filled with some material, a fraction of the original radiation that would otherwise be detected will be absorbed, depending on the interposing mass, that is, on the thickness and density of the intervening material. In the event the material placed between source and detector is a body of soil of constant bulk density, the intensity of the transmitted radiation will vary only with changes in water content. In fact, it will be an exponential function of soil wetness, as follows (Gurr, 1962; Ferguson and Gardner, 1962):

$$N_w/N_d = \exp(-\theta_m \mu_w x) \tag{6.8}$$

where N_w/N_d is the ratio of the monoenergetic radiation flux transmitted through wet soil (N_w) to that transmitted through dry soil (N_d), μ_w is the mass attenuation coefficient for water, x is the thickness of the transmitting soil, and θ_m, is the water mass per unit bulk volume of soil. Equation (6.8) can be transformed to give wetness as a function of the relative transmission rate:

$$\theta_m = \frac{\ln(N_w/N_d)}{\mu_w x} = -\frac{\log_{10}(N_w/N_d)}{0.4343\,\mu_w x} \tag{6.9}$$

where $\ln(N_w/N_d)$ is the natural logarithm of the count ratio for wet to dry soil and $\log_{10}$ is the common logarithm for the same ratio.

The gamma-ray absorption method is used mostly in the laboratory, where the dimensions and density of the soil sample as well as the ambient temperature can be precisely controlled. A high degree of spatial resolution (e.g., 2 mm or so) can be obtained by collimation of the radiation. This is done by drilling a narrow hole or slot into the lead (Pb) wall shielding the source (as well, perhaps, as into a second shield placed in front of the detector), thus allowing passage of only a very narrow beam.

Since the absorption of radiation depends on the entire mass between source and detector, the readings can be related uniquely to soil moisture only if soil bulk density is constant or if its change is monitored simultaneously. To permit concurrent measurement of soil bulk density and moisture changes in swelling or shrinking soils, dual-source scanners have been developed, in which both cesium 137 and americium 241 are used. Analysis of the concurrent transmission of the two beams can allow separation of the change in attenuating mass between that due to bulk density and that due to soil wetness (Gardner, 1986).

The *double-probe gamma-ray method* has also been adapted to field use (Fig. 6.3). In principle, this technique offers several advantages over the neutron moisture meter, in that it allows better depth resolution in the measurement of soil-moisture profiles (i.e., about 1 cm in effective measurement width), sufficient to detect discontinuities between profile layers as well as movement of wetting fronts and conditions near the soil surface. However, the field device is still too cumbersome for general usage. Not the least of the problems is the accurate installation and alignment of two access tubes that must be strictly parallel along with the accurate determination of soil bulk density, which might vary in depth and time.

The health hazard associated with use of gamma-ray equipment is similar in principle to that discussed in connection with the neutron moisture meter. The equipment is considered safe only if strict attention is paid to all the safety rules.

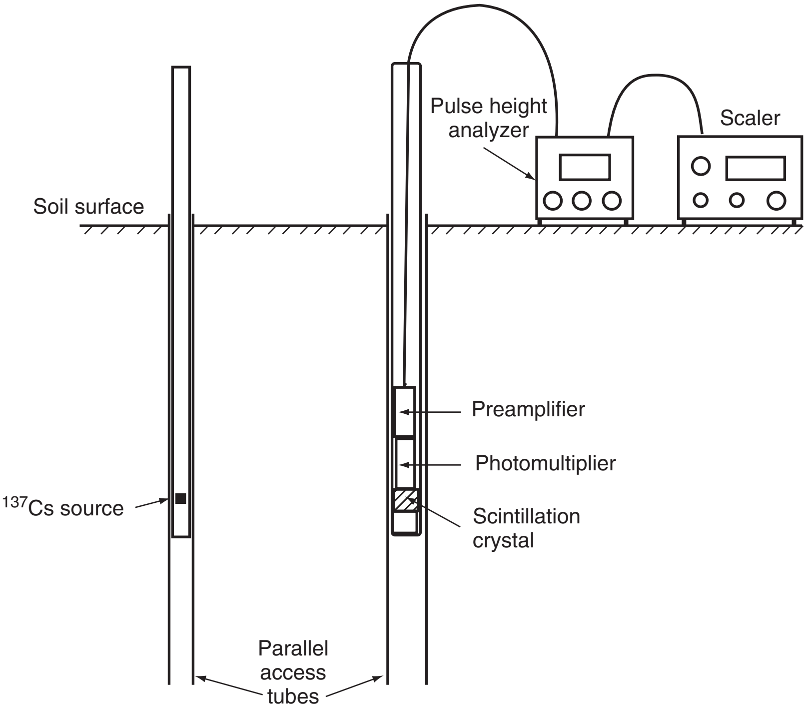


Fig. 6.3. Double-probe gamma-ray apparatus for monitoring soil moisture or density.

Time-Domain Reflectometry (TDR)

This is a relatively new method of measuring soil wetness, based on the unusually high *dielectric constant* of water. A *dielectric*, in general, is a non conductor of electricity, that is, a substance that, when placed between two charged surfaces (a capacitor), allows no net flow of electric charge but only a displacement of charge. (The simplest form of a capacitor consists of two parallel metal plates separated by a layer of air or some other insulating material, such as mica.) The dielectric constant, also called *relative permittivity* (or specific inductive capacity), is defined as the ratio of the *capacitance* of a capacitor with the given substance as dielectric to the capacitance of the same capacitor with air (or a vacuum) as the dielectric.

The value of the dielectric constant (ε_r) depends on the nature of the material. The ε_r value for dry air is usually taken to be 1, for paraffin wax 2.0–2.5, for rubber 2.8–3.0, for porcelain and mica 6.0–8.0. The value for dry soil is roughly 4 (Jackson and Schmugge, 1989). In contrast, the value for water is about 80. Hence the water content of a soil determines its dielectric constant.

In the TDR technique (Topp and Davis, 1985), a pair of parallel metal rods connected to a signal receiver is inserted into the soil. The rods serve as conductors, while the soil between and around the rods is the dielectric medium. When a step voltage pulse or signal is propagated along the parallel transmission lines, the signal is reflected from the ends of those rods and returns to the

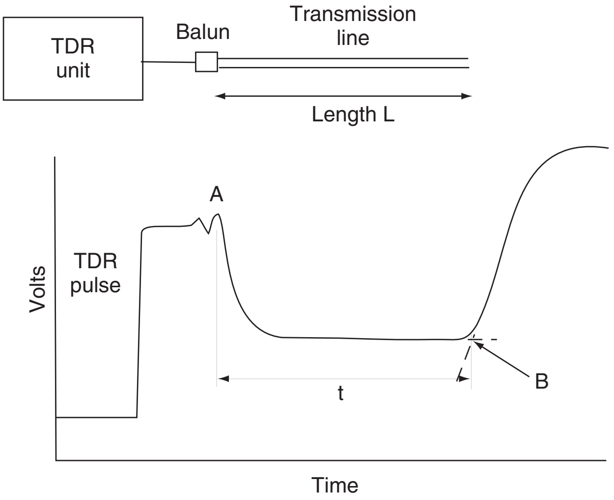


Fig. 6.4. The essential components of a TDR system (above) and an idealized TDR output trace (obtainable with an oscilloscope) showing how the propagation time is determined. (After Topp and Davis, 1985.)

receiver. A device then measures the time between sending and receiving the reflected signal, as shown in Fig. 6.4. For a fixed line length, the time interval relates inversely to the propagation velocity v of the signal in the soil (i.e., the propagation velocity diminishes as the amount of water present increases, so the time interval increases with soil wetness).

The approximate equation given by Topp and Davis (1985) is:

$$v = c/(\varepsilon_r)^{0.5} \tag{6.10}$$

where c is the propagation velocity of an electromagnetic wave in free space (3×10^8 m/sec). The velocity can be determined from the length L of the transmission lines (with the travel distance, forth and back, being $2L$) and the measured signal time t in the soil, using the time-domain reflectometer. The dielectric constant results from inserting $v = 2L/t$ into Eq. (6.10):

$$\varepsilon_r = (ct/2L)^2 \tag{6.11}$$

Topp and Davis (1985) found that the dielectric constant as a function of the soil's volume wetness is only weakly dependent on soil type, bulk density, temperature, or the electrical conductivity of pore water. They offered the following empirical equation:

$$\theta = -5.3 \times 10^{-2} + 2.9 \times 10^{-2}\varepsilon_r - 5.5 \times 10^{-4}\varepsilon_r^2 + 4.3 \times 10^{-6}\varepsilon_r^3 \tag{6.12}$$

The same investigators claimed that the volume wetness of soils can be thus determined with an accuracy of ±2% and a precision (or repeatability) of ±1%. They deemed this accuracy to be sufficient for using the TDR technique for irrigation applications without having to carry out a calibration for each soil or field. They recommended that the transmission rods be spaced 50 mm apart.

The soil "sampled" or measured by TDR is essentially a cylinder whose axis lies midway between the rods and whose diameter is 1.4 times the spacing between the rods. This gives a cross section of 3800 mm^2 for rods spaced 50 mm apart. The volume of the measurement varies directly with the length of the rods. Lengths of 0.1–1.0 m are evidently practical. Davis et al. (1983) even used transmission lines that were 20 m long, placed horizontally in a sandy soil to detect pillars or columns of wet soil that resulted from leaks out of a "sealed" lagoon.

Although the velocity of propagation of the TDR pulse as it travels in the soil is evidently unaffected by the soil solution's electrical conductivity, the intensity of the transmitted signal is affected. The attenuation of the signal amplitude can therefore serve to indicate the soil's salinity (Dalton et al., 1984). However, conductivity measurements using TDR are difficult to carry out with a high level of accuracy in the field.

Other Methods

Additional approaches to the measurement of soil wetness include techniques based on the dependence of soil thermal properties on water content and the use of ultrasonic waves, radar waves, and microwaves (Topp and Ferré, 2002). Some of these and other methods have been tried for the remote sensing of land areas from aircraft or satellites (Jackson and Schmugge, 1989; Engman, 1995; Engman and Chauhan, 1995.

ENERGY STATE OF SOIL WATER

Classical physics recognizes two principal forms of energy, kinetic and potential. Since the movement of water in the soil is quite slow, its kinetic energy, which is proportional to the velocity squared, is generally considered to be negligible. On the other hand, the potential energy, which is due to position or internal condition, is of primary importance in determining the state and movement of water in the soil.

The potential energy of soil water varies over a wide range. Differences in potential energy between one point and another induce water to flow within the soil. The spontaneous and universal tendency of all matter in nature is to move from where the potential energy is higher to where it is lower and for each parcel of matter to equilibrate with its surroundings. Soil water obeys the same universal tendency toward that elusive state known as *equilibrium*, definable as a condition of uniform potential energy throughout. In the soil, water moves constantly in the direction of decreasing potential energy. The rate of decrease of potential energy with distance is in fact the moving force causing flow.

Knowledge of the relative potential energy state of soil water at each point within the soil can allow us to evaluate the forces acting on soil water in all directions and to determine how far the water in a soil system is from equilibrium. This is analogous to the well-known fact that an object will tend to fall spontaneously from a higher to a lower elevation and that in so doing it performs work but that lifting it requires an investment of work. Since potential energy is a measure of the amount of work a body can perform by virtue of the energy stored in it, knowing the potential energy state of water in the soil and in the plant growing in that soil can help us to estimate how much work the plant must expend to extract a unit amount of water.

Clearly, it is not the absolute amount of potential energy "contained" in the water that is important in itself, but rather the relative level of that energy in different regions within the soil. The concept of soil-water potential is a criterion, or yardstick, for this energy. It expresses the specific potential energy of soil water relative to that of water in a standard reference state. The standard state generally assumed is a hypothetical reservoir of pure water, at atmospheric pressure, at the same temperature as that of soil water (or at any other specified temperature), and at a given and constant elevation. Since the elevation of this hypothetical reservoir can be set at will, it follows that the potential determined by comparison with this standard is not absolute, but by employing even so arbitrary a criterion we can determine the relative magnitude of the specific potential energy of water at different locations or times within the soil.

Just as an energy increment can be viewed as the product of a force and a distance increment, so the ratio of a potential energy increment to a distance increment can be viewed as constituting a force. Accordingly, the force acting on soil water, directed from a zone of higher to a zone of lower potential, is equal to the negative *potential gradient* ($-d\phi/dx$), which is the change of energy potential ϕ with distance x. The negative sign indicates that the force acts in the direction of decreasing potential.

The concept of soil-water potential is of great fundamental importance. This concept replaces the arbitrary categorizations that prevailed in the early stages of soil physics and that purported to recognize and classify different forms of soil water, for example, "gravitational water," "capillary water," "hygroscopic water". The possible values of soil-water potential are continuous and do not exhibit any abrupt transitions from one energy level to another.

When the soil is saturated and its water is at a hydrostatic pressure greater than the atmospheric pressure (as, for instance, under a water table), the potential energy of that water may be greater than that of the reference-state reservoir described, and is therefore considered positive. Water then tends to move spontaneously from the soil into such a reservoir. If, on the other hand, the soil is unsaturated, its water is no longer free to flow out toward a reservoir at atmospheric pressure. On the contrary, the spontaneous tendency is for the soil to draw water from such a reservoir if placed in contact with it, much as a blotter draws ink (or as a napkin absorbs spilled coffee). The energy potential of soil moisture is taken to be negative.

The magnitude of the potential at any point depends not only on hydrostatic pressure, however, but also on such additional physical factors as elevation (relative to that of the reference elevation), temperature, and concentration of solutes. When solutes are present that are not uniformly distributed, osmotic effects occur. The potential energy of soil water tends to be lower where the concentration of solutes is higher.

TOTAL SOIL-WATER POTENTIAL

Thus far, we have described the energy potential of soil water in a qualitative way. More precisely, this energy potential can be regarded in terms of the difference in *partial specific free energy* between soil water and standard water. According to a soil physics terminology committee of the International Soil Science Society (Aslyng, 1963), the *total potential of soil water* is "the amount of work that must be done per unit quantity of pure water to transport reversibly and isothermally an infinitesimal quantity of water from a pool of pure water at a specified elevation at atmospheric pressure to the soil water (at the point under consideration)."

This is merely a formal definition, since in actual practice the potential is not measured by transporting water per the definition, but by measuring some other property or properties related to the potential in some known way (e.g., hydrostatic pressure, vapor pressure, elevation). The definition specifies transporting an infinitesimal quantity, in any case, to ensure that the determination procedure does not change either the reference state (i.e., the pool of pure, free water) or the soil-water potential being measured. In any case this definition provides a conceptual rather than an actual working tool.

Soil water is subject to a number of possible forces, each of which may cause its potential to differ from that of pure, free water at a reference elevation. Such force fields result from the attraction of the solid matrix for water as well as from the presence of solutes and the action of external gas pressure and gravitation. Accordingly, the total potential of soil water can be thought of as the sum of the separate contributions of these various factors, as follows:

$$\phi_t = \phi_g + \phi_p + \phi_o + \ldots \tag{6.13}$$

where ϕ_t is the total potential, ϕ_g is the gravitational potential, ϕ_p is the pressure (or matric) potential, ϕ_o is the osmotic potential, and the ellipsis signifies that additional terms are theoretically possible.

Not all of the separate potentials given in Eq. (6.13) act in the same way, and their separate gradients may not always be equally effective in causing flow (e.g., the osmotic potential gradient requires a semipermeable membrane to induce liquid flow). The gravity and pressure potentials generally pertain to the soil solution, whereas the osmotic potential pertains, strictly speaking, to the pure water substance. (The latter difference can be significant whenever the salt content of the soil solution is appreciable.) Despite these differences among the component potentials, the advantage of the total-potential concept is that it provides a unified measure by which the state of water can be evaluated in most cases everywhere within the soil–plant–atmosphere continuum.

GRAVITATIONAL POTENTIAL

Every body on the earth's surface is attracted toward the earth's center by a gravitational force equal to the weight of the body, that weight being the product of the mass of the body and the gravitational acceleration. To raise a body against this attraction, work must be expended, and this work is stored by the raised body in the form of *gravitational potential energy*. The amount of this energy depends on the body's position in the gravitational force field.

The gravitational potential of soil water at each point is determined by the elevation of the point relative to some arbitrary reference level. For the sake of convenience, it is customary to set the reference level at the elevation of a pertinent point within the soil or below the soil profile being considered (e.g., at the water table, where such exists at some definable depth) so that the gravitational potential can always be taken as positive or zero. On the other hand, if the soil surface is chosen as the reference level, as is sometimes done, then the gravitational potential for all points below the surface is negative with respect to that reference level.

At a height z above a reference, the gravitational potential energy E_g of a mass M of water occupying a volume V is

$$E_g = Mgz = \rho_w Vgz \tag{6.14}$$

where ρ_w is the density of water and g is the acceleration of gravity. Accordingly, the gravitational potential in terms of the potential energy per unit mass is

$$\phi_{g,m} = gz \tag{6.15}$$

and in terms of potential energy per unit volume is

$$\phi_{g,v} = \rho_w gz \tag{6.16}$$

The gravitational potential is independent of the chemical and pressure conditions of soil water and dependent only on relative elevation.

PRESSURE POTENTIAL

When soil water is at hydrostatic pressure greater than atmospheric, its pressure potential is considered positive. When at a pressure lower than atmospheric (a subpressure commonly known as *tension* or *suction*), the pressure potential is considered negative. Thus, water at a free-water surface is at a zero pressure potential, water below that surface is at a positive pressure potential, while water that has risen in a capillary tube (or in the pores of the soil) above that surface is at a negative pressure potential. This principle is illustrated in Fig. 6.5.

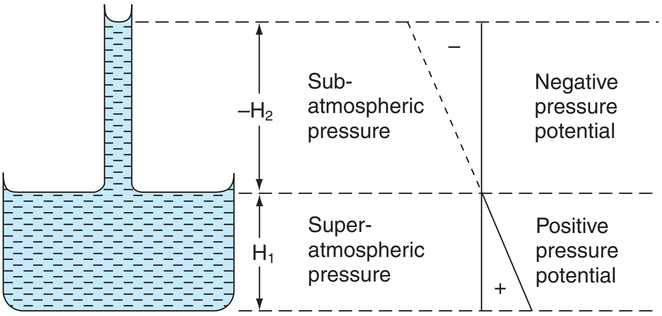


Fig. 6.5. Superatmospheric and subatmospheric pressures below and above a free-water level.

The pressure potential below the groundwater level has been termed the *submergence potential*. The hydrostatic pressure P of water in reference to atmospheric pressure is

$$P = \rho g h \tag{6.17}$$

Here h is the piezometric head (submergence depth below the free-water surface). The potential energy of this water is then

$$E = P dV \tag{6.18}$$

and thus the submergence potential, taken as the potential energy per unit volume, is equal to the hydrostatic pressure, P:

$$\phi_{\mathrm{ps}} = P \tag{6.19}$$

The negative pressure potential of soil moisture was termed *capillary potential* in the older literature of soil physics and is now called *matric potential*. It results from the interactive capillary and adsorptive forces between water and the soil matrix. These forces bind water in the soil and lower its potential energy below that of bulk water.

The terms *matric potential, matric suction*, and *soil-water suction* have been used interchangeably. According to the lSSS. committee cited earlier, the *matric suction* is defined as the negative gauge pressure, relative to the external gas pressure on soil water, to which a solution identical in composition with the soil solution must be subjected in order to be in equilibrium through a porous membrane wall with the water in the soil. This definition implies the use of either a *tensiometer* relative to the prevailing atmospheric pressure (conventionally taken to be zero) of the gas phase or a pressure plate extraction apparatus in which the gas phase is pressurized sufficiently to bring the water phase to atmospheric pressure.

As shown in Chapter 2, capillarity results from the surface tension of water and its contact angle with the solid particles. In an unsaturated (three-phase) soil system, curved menisci form, which obey the equation of capillarity:

$$P_0 - P_c = \Delta P = \gamma(1/R_1 + 1/R_2) \tag{6.20}$$

where P_0 is the reference atmospheric pressure (taken as zero), P_c is the pressure of soil water (which can be smaller than atmospheric), ΔP is the pressure deficit (subpressure) of soil water, γ is the surface tension of water, and R_1, R_2 are the principal radii of curvature of the meniscus.

If the soil were like a simple bundle of capillary tubes, the equations of capillarity might suffice to describe the relation of the negative pressure potential (tension, or suction) to the radii of the soil pores in which the menisci are contained. However, in addition to the capillary phenomenon, the soil also exhibits adsorption, which forms hydration envelopes over the particle surfaces. These two mechanisms of soil-water affinity are illustrated in Fig. 6.6.

The presence of water in films as well as under concave menisci is most important in clayey soil and at high suctions, and it is influenced by the electric double layer and the exchangeable cations present. In sandy soils, adsorption is relatively unimportant and the capillary effect predominates. In general, however, the negative pressure potential results from the combined

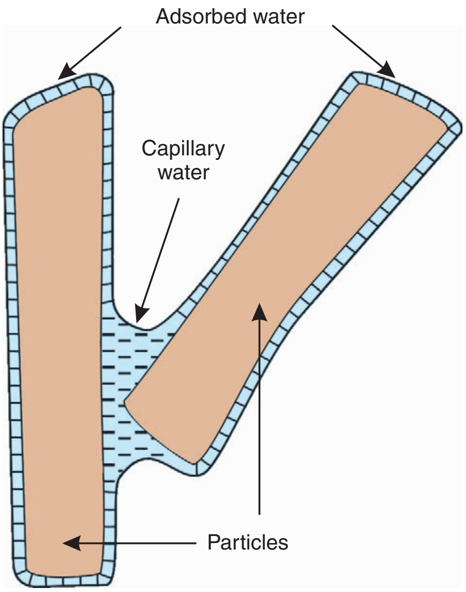


Fig. 6.6. Water in an unsaturated soil is subject to capillarity and adsorption, which combine to produce a "negative" matric potential, or a matric suction.

effect of the two mechanisms, which cannot easily be separated. The capillary "wedges" are at a state of internal equilibrium with the adsorption "films," and the ones cannot be changed without affecting the others. Hence, the older term, *capillary potential*, is inadequate and the better term is *matric potential*, because it denotes the total effect resulting from the affinity of the water to the whole matrix of the soil, including its pores and particle surfaces together.

In the absence of solute effects, the liquid and vapor phases in an unsaturated porous medium are related at equilibrium by

$$\phi_m = RT \ln(p/p_0) \tag{6.21}$$

where R is the gas constant for water vapor, T is the absolute temperature, and p/p_0 is the relative humidity (i.e., the ratio of the atmosphere's vapor pressure at equilibrium with the unsaturated medium to the "saturated" vapor pressure at equilibrium with a body of pure free water).

OSMOTIC POTENTIAL

The presence of solutes affects the thermodynamic properties of water and lowers its potential energy. In particular, solutes lower the vapor pressure of soil water. Thus,

$$\phi_m + \phi_o = RT \ln(p/p_0) \tag{6.22}$$

where ϕ_o is the osmotic potential. While this phenomenon may not affect liquid flow in the soil significantly, it does come into play whenever a membrane or diffusion barrier is present that transmits water more readily than salts. The osmotic effect is important in the interaction between plant roots and soil as well as in processes involving vapor diffusion.

There is a difference in principle between the osmotic potential and the other potential terms defined. Whereas the pressure and gravitational potentials refer to the *soil solution* (i.e., soil water along with its dissolved constituents), the osmotic potential applies to the water substance alone. Strictly speaking, therefore, ϕ_o should not simply be added to ϕ_m and ϕ_g as if those terms were similarly applicable and mutually independent. This fundamental difference can, however, be ignored in practice as long as the soil solution is dilute enough and the solutes it contains do not affect the matrix (and hence the matric potential) significantly.

QUANTITATIVE EXPRESSION OF SOIL-WATER POTENTIAL

The soil-water potential (Table 6.2) is expressible physically in at least three ways:

1. *Energy per unit mass*: This is the fundamental expression of potential, using units of ergs per gram or joules per kilogram. The dimensions of energy per unit mass are length squared per time squared: L^2T^{-2}.

2. *Energy per unit volume*: Since liquid water is only very slightly compressible in the pressure range ordinarily encountered at or near the earth's

TABLE 6.2 Energy Levels of Soil Water

Soil-water potential				Soil-water suction[a]		Vapor pressure (torr) 20°C (%)	Relative humidity[b] at 20°C (%)
Per unit mass		Per unit volume		Pressure	Head		
(erg/g)	(joule/kg)	(bar)	(mm H_2O)	(bar)	(mm H_2O)		
0	0	0	0	0	0	17.5350	100.00
-1×10^4	−1	−0.01	−102	0.01	102	17.5349	100.00
-5×10^4	−5	−0.05	−510	0.05	510	17.5344	99.997
-1×10^5	−10	−0.1	−1,020	0.1	1,020	17.5337	99.993
-2×10^5	−20	−0.02	−2,040	0.2	2,040	17.5324	99.985
-3×10^5	−30	−0.3	−3,060	0.3	3,060	17.5312	99.978
-4×10^5	−40	−0.4	−4,080	0.4	4,080	17.5299	99.971
-5×10^5	−50	−0.5	−5,100	0.5	5,100	17.5286	99.964
-6×10^5	−60	−0.6	−6,120	0.6	6,120	17.5273	99.965
-7×10^5	−70	−0.7	−7,140	0.7	7,140	17.5260	99.949
-8×10^5	−80	−0.8	−8,160	0.8	8,160	17.5247	99.941
-9×10^5	−90	−0.9	−9,180	0.9	9,180	17.5234	99.934
-1×10^6	−100	−1.0	−10,200	1.0	10,200	17.5222	99.927
-2×10^6	−200	−2	−20,400	2	20,400	17.5089	99.851
-3×10^6	−300	−3	−30,600	3	30,600	17.4961	99.778
-4×10^6	−400	−4	−40,800	4	40,800	17.4833	99.705
-5×10^6	−500	−5	−51,000	5	51,000	17.4704	99.637
-6×10^6	−600	−6	−61,200	6	61,200	17.4572	99.556

[a] In the absence of osmotic effects (soluble salts), soil-water suction equals matric suction; otherwise, it is the sum of matric and osmotic suctions.

[b] Relative humidity of air in equilibrium with the soil at different suction values.

surface, its density is practically independent of potential. Hence, there is a direct proportion between the expression of the potential as energy per unit mass and its expression as energy per unit volume. The latter expression yields the dimensions of pressure (for just as energy can be expressed as the product of pressure and volume, so the ratio of energy to volume is equivalent dimensionally to pressure). This equivalent pressure can be expressed as dynes per square centimeter, newtons per square meter, bars, atmospheres, or pascals. The basic dimensions are those of force per unit area (or mass per length per time squared): $ML^{-1}T^{-2}$. This mode of expression is convenient for the osmotic and pressure potentials, but it is seldom used for the gravitational potential.

3. *Energy per unit weight* (hydraulic head): Whatever can be expressed in units of hydrostatic pressure can also be expressed in terms of an equivalent head of water, which is the height of a liquid column corresponding to the given pressure. For example, a pressure of 1 atm is equivalent to a vertical water column, or hydraulic head, of 10.33 m and to a mercury head of 0.76 m. A pressure of 1 MPa is equivalent to a hydraulic head of 102.27 m, and 1 kPa to about 0.1 m. This mode of expression is certainly simpler, and often more convenient, than the preceding mode. Therefore, it is common to characterize the state of soil water in terms of the *total potential head*, the *gravitational potential head*, and the *pressure potential head*, which are usually expressible in meters. Accordingly, instead of

$$\phi = \phi_g + \phi_p \tag{6.23}$$

we could write

$$H = H_g + H_p \tag{6.24}$$

which reads: The total potential head of soil water (H) is the sum of the gravitational (H_g) and pressure (H_p) potential heads. H is commonly called the *hydraulic head*.

The use of various expressions for the soil-water potential can be perplexing to the uninitiated. However, these alternative expressions are in fact equivalent, and each of them can readily be translated into any of the others. If we use ϕ to designate the potential in terms of energy per mass, P for the potential in terms of pressure, and H for the potential head, then

$$\phi = P/\rho_w \tag{6.25}$$

$$H = P/\rho_w g = \phi/g \tag{6.26}$$

where ρ_w is the density of water and g is the acceleration of gravity.

A remark is in order concerning the use of the synonymous terms *tension* and *suction* in lieu of "negative" (or "subatmospheric") pressure. Tension and suction are merely semantic devices to avoid the use of the unesthetic negative sign, which generally characterizes the pressure of water in an unsaturated soil. These terms allow us to speak of the osmotic and matric potentials in positive terms. The two potentials are illustrated schematically in Fig. 6.7.

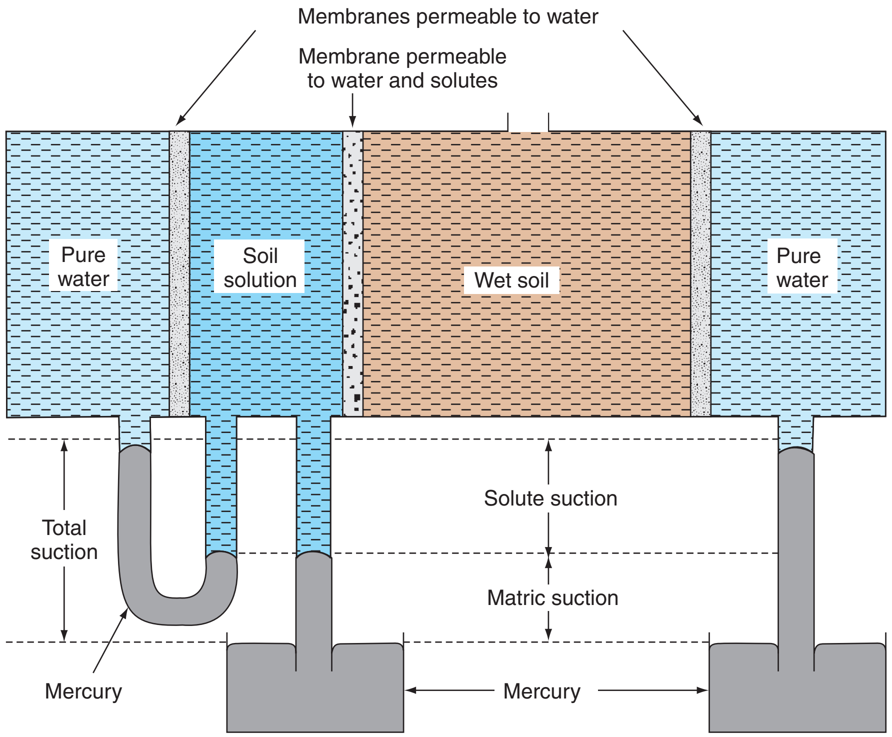


Fig. 6.7. In an isothermal system at equilibrium, matric suction is the pressure difference across a membrane separating the soil solution in situ from the same solution in bulk, with the membrane being permeable to the solution but not to solid particles or air. Osmotic (solute) suction is the pressure difference across a semipermeable membrane separating the bulk phases of pure water and the soil solution. Total suction is the sum of the matric and osmotic suction values and is thus the pressure difference across a semipermeable membrane separating pure water from a soil that contains a solution (After Richards, S. J., 1965.)

SOIL-MOISTURE CHARACTERISTIC CURVE

Water in a saturated soil is at atmospheric pressure if it is at equilibrium with free water at the same elevation. That implies zero hydrostatic pressure and hence zero suction. If a slight suction (i.e., a water pressure slightly subatmospheric) is applied to water in a saturated soil, no outflow may occur until, as suction is increased, a critical value is exceeded at which the largest surface pore begins to empty and its water content is displaced by air. This critical suction is called the *air-entry suction*. Being the threshold of desaturation, air-entry suction is generally small in coarse-textured and in well-aggregated soils having large pores; but it tends to be larger in dense, poorly aggregated, medium-textured or fine-textured soils. Since in coarse-textured soils the pores tend to be fairly uniform in size, these soils typically exhibit critical air-entry phenomena more distinctly than do soils with a wider array of pore sizes.

As suction is applied incrementally, the first pores to be emptied are the relatively large ones that cannot retain water against the suction applied.

Recalling the capillary equation ($-P = \psi = 2\gamma/r$), we can readily predict that a gradual increase in suction will result in the emptying of progressively smaller pores until, at high suction values, only the very narrow pores retain water. Similarly, an increase in soil-water suction is associated with decreasing thickness of the hydration envelopes adsorbed to the soil-particle surfaces. Increasing suction is thus associated with decreasing soil wetness. The amount of water remaining in the soil at equilibrium is a function of the sizes and volumes of the water-filled pores and of the amount of water adsorbed to the particles; hence it is a function of the matric suction. This function is usually measured experimentally, and it is represented graphically by a curve called the *soil-moisture retention curve*, also known as the *soil-moisture release curve* or the *soil-moisture characteristic*.

As yet, no universally applicable theory exists for the prediction of the matric suction versus wetness relationship from basic soil properties (i.e., texture and structure). The adsorption and pore geometry effects are generally too complex to be described by a simple model. Several empirical equations have been proposed that describe the soil-moisture characteristic for some soils and within limited suction ranges. One such equation was advanced by Visser (1966):

$$\psi = a(f - \theta)^b/\theta^c \tag{6.27}$$

Here ψ is matric suction, f is porosity, θ is volumetric wetness, and *a, b, c* are constants. Use of this equation is hampered by the difficulty of evaluating its constants. Visser found that *b* varied from 0 to 10, *a* from 0 to 3, and *f* from 0.4 to 0.6.

Alternative equations to describe the relationship between wetness and matric suction have been proposed by Laliberte (1969), White et al. (1970), Su and Brooks (1975), and van Genuchten (1978). An equation presented by Brooks and Corey (1966a) is

$$(\theta - \theta_r)/(\theta_m - \theta_r) = (\psi_e/\psi)^\lambda \tag{6.28}$$

for suction values greater than the air-entry suction ψ_e. The exponent λ has been termed the *pore size distribution index*. In this equation, θ is the volume wetness (a function of the suction ψ), θ_m is the maximum wetness (saturation), and θ_r is the "residual" wetness remaining at high suction in the small pores that do not form a continuous network (the intra-aggregate pores).

Physically based water retention prediction models have recently been described by Haverkamp and Reggiami (2002) and by Nimmo (2002).

The amount of water retained at low values of matric suction (say, between 0 and 100 kPa) depends on capillarity and the pore size distribution. Hence it is strongly affected by the soil's structure. At higher suctions, water retention is due increasingly to adsorption, so it is influenced less by the structure and more by the soil's texture and specific surface. The greater the clay content, in general, the greater the water retention at any particular suction and the more gradual the slope of the curve. In a sandy soil, most of the pores are relatively large, and once these large pores are emptied at a given suction, only a small amount of water remains. In a clayey soil, more of the water is adsorbed, so increasing the matric suction causes a more gradual decrease in wetness (Fig. 6.8).

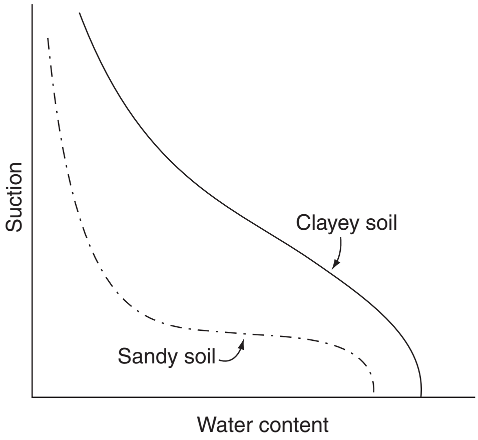


Fig. 6.8. The effect of texture on soil-water retention.

Because soil structure influences the shape of the soil-moisture characteristic curve primarily in the low-suction range, we may expect that the effect of compaction (which destroys the aggregated structure) is to reduce the total porosity and especially the volume of the large interaggregate pores. As a result of compaction, the saturation water content as well as the initial decrease of water content with the application of low suction are diminished. On the other hand, the volume of intermediate-size pores is likely to be greater in a compact soil (because some of the originally large pores have been squeezed into intermediate size by compaction), while the micropores remain unaffected, and thus the curves for the compacted and uncompacted soil tend to converge in the high-suction range (Fig. 6.9).

If two soil bodies differing in texture, structure, and initial wetness are brought into direct physical contact, water from one will normally move into

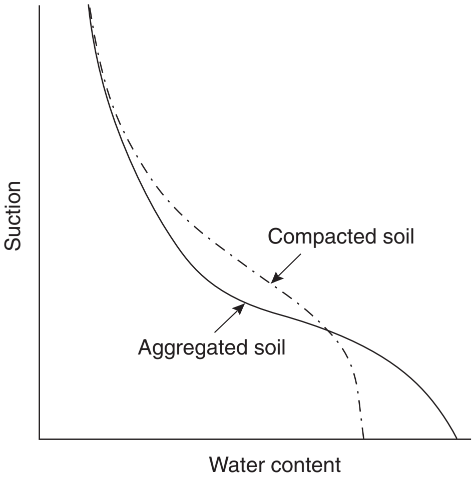


Fig. 6.9. The effect of soil structure on soil-water retention.

the other in a tendency toward potential energy equilibrium. Even at equilibrium, however, the two different soils will not generally be equal in wetness. Rather, each soil will end up with a content of water determined by its own soil-moisture characteristic. Two soil bodies or layers can thus attain equilibrium and yet exhibit a marked difference in wetness. We can easily envision a situation in which a drier soil layer will contribute water to a wetter one if the water potential of the former is higher than that of the latter, owing to textural, structural, elevational, osmotic, or thermal differences.

The slope of the soil-moisture characteristic curve, which is the change of water content per unit change of matric potential, is generally termed the *differential* (or *specific*) *water capacity* c_θ:

$$c_\theta = d\theta/d\phi_p \qquad \text{or} \qquad c_\theta = -d\theta/d\psi \tag{6.29}$$

The c_θ term is analogous to the well-known *differential heat capacity*, which is the change in the heat content of a body per unit change in the thermal potential (temperature). However, while the differential heat capacity is fairly constant with temperature for many materials, the differential water capacity in soils is strongly dependent on the matric potential.

HYSTERESIS

The relation between matric potential and soil wetness can be obtained in two ways: (1) in desorption, by taking an initially saturated sample and applying increasing suction, in stepwise manner, to gradually dry the soil while taking successive measurements of wetness versus suction; and (2) in sorption, by gradually wetting up an initially dry soil sample while reducing the suction incrementally. Each of these two methods yields a continuous curve, but the two curves will in general not be identical. The equilibrium soil wetness at a given suction is greater in desorption (drying) than in sorption (wetting). This dependence of the equilibrium content and state of soil water upon the direction of the process leading up to it is called *hysteresis* (Haines, 1930; Miller and Miller, 1956; Mualem, 1984).

Figure 6.10 shows a typical soil-moisture characteristic curve and illustrates the hysteresis effect in the relationship between soil wetness and matric suction at equilibrium.

The hysteresis effect may be attributed to several causes:

1. The geometric nonuniformity of the individual pores (which are generally irregularly shaped voids interconnected by smaller passages), resulting in the "ink bottle" effect, illustrated in Fig. 6.11.
2. The contact-angle effect, mentioned in Chapter 2, by which the contact angle is greater and hence the radius of curvature greater in the case of an advancing meniscus than in the case of a receding one. A given water content will tend therefore to exhibit greater suction in desorption than in sorption. (Contact-angle hysteresis can arise because of surface roughness, the presence of adsorbed impurities on the solid surface, and the mechanism by which liquid molecules adsorb or desorb when the interface is displaced.)

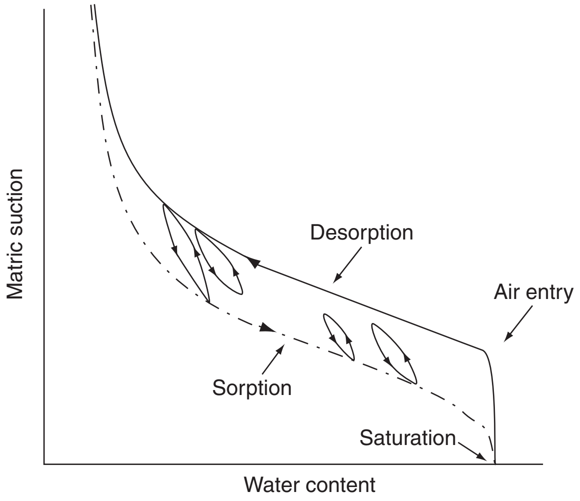


Fig. 6.10. Suction vs. water content curves in sorption and desorption. The intermediate loops are scanning curves, representing complete or partial transitions between the main branches.

3. Entrapped air, which further decreases the water content of newly wetted soil. Failure to attain true equilibrium (although not, strictly speaking, true hysteresis) can accentuate the hysteresis effect.
4. Swelling, shrinking, or aging phenomena, which result in differential changes of soil structure, depending on the wetting and drying history of the sample (Hillel and Mottes, 1966). The gradual solution of air or the release of dissolved air from soil water can also have a differential effect upon the suction–wetness relationship in wetting and drying systems.

Of particular interest is the ink bottle effect. Consider the hypothetical pore shown in Fig. 6.11. This pore consists of a relatively wide void of radius R, bounded by narrow channels of radius r. If initially saturated, this pore drains abruptly the moment the suction exceeds ψ_r, where $\psi_r = 2\gamma/\mathrm{r}$. For this pore to rewet, however, the suction must decrease to below ψ_R, where $\psi_R = 2\gamma/R$, whereupon the pore abruptly fills. Since $R > r$, it follows that $\psi_\mathrm{r} > \psi_R$.

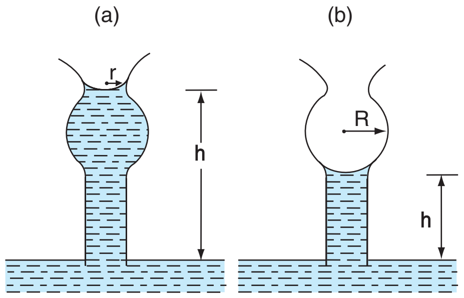


Fig. 6.11. The "ink bottle" effect determines the equilibrium height of water in a variable-width pore: (a) in capillary drainage (desorption) and (b) in capillary rise (sorption).

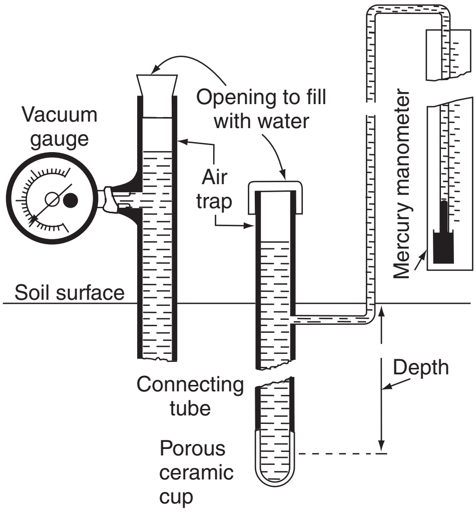


Fig. 6.12. The essential parts of a tensiometer. (After Richards, S. J., 1965.)

Desorption depends on the narrow radii of the connecting channels, whereas sorption depends on the diameters of the large pores. These discontinuous spurts of water, called *Haines jumps* (after W. B. Haines, who first noted the phenomenon in 1930), can be observed readily in coarse sands. The hysteresis effect is generally more pronounced in coarse-textured soils in the low-suction range, where pores empty at an appreciably larger suction than that at which they fill.

The two complete characteristic curves, the one describing the gradual transition from saturation to dryness and the other portraying the transition from dryness to saturation, are called the *main branches* of the hysteretic soil-moisture characteristic. When a partially wetted soil commences to drain or when a partially desorbed soil is rewetted, the relation of suction to moisture content follows some intermediate curve as it “scans” from one main branch to the other. Such intermediate spurs are called *scanning curves*. Cyclic changes often entail wetting and drying scanning curves, which may form loops between the main branches (Fig. 6.12). The ψ–θ relationship can thus become very complicated. Indeed, because of its complexity, the hysteresis phenomenon is too often ignored. The soil-moisture characteristic commonly reported in the literature is the *desorption curve*, also known as the *soil-moisture release curve*. The *sorption curve* is equally important but more difficult to determine, so it is seldom reported.

MEASUREMENT OF SOIL-MOISTURE POTENTIAL

The measurement of soil wetness, described earlier in this chapter, though essential in many soil physical and engineering investigations, is obviously not sufficient to provide a description of the state of soil water. To obtain such a description, evaluation of the energy status of soil water (soil-moisture potential or suction) is necessary. In general, the twin variables of wetness and potential should each be measured directly, because the translation of one to the other on the basis of calibration curves of soil samples is too often unreliable.

Total soil-moisture potential is often thought of as the sum of matric and osmotic (solute) potentials and is a useful index for characterizing the energy status of soil water with respect to plant water uptake. The sum of the matric and gravitational (elevation) heads is generally called the *hydraulic head* (or hydraulic potential) and is useful in evaluating the directions and magnitudes of the water-moving forces throughout the soil profile. Methods are available for measuring matric potential as well as total soil-moisture potential, separately or together (Dane and Hopmans, 2002). A schematic representation of the components of the soil-moisture potential is shown in Fig. 6.7. To measure matric potential in the field, an instrument known as a *tensiometer* is used; in the laboratory, use is often made of tension plates and of air-pressure extraction cells. Total soil-moisture potential can be obtained by measuring the equilibrium vapor pressure of soil water by means of thermocouple psychrometers.

We next describe the tensiometer, which has won acceptance as a practical device for in situ measurement of matric suction, hydraulic head, and hydraulic gradients.

The Tensiometer

The tensiometer is an instrument designed to provide a continuous indication of the soil's matric suction (also called *soil-moisture tension*) in situ. The essential parts of a tensiometer are shown in Fig. 6.12. The instrument consists of a porous cup, generally of ceramic material, connected through a tube to a manometer, with all parts filled with water. When the cup is placed in the soil where the suction measurement is to be made, the bulk water inside the cup comes into hydraulic contact and tends to equilibrate with soil water through the pores in the ceramic walls. When the tensiometer is initially placed in the soil, the water contained in it is generally at atmospheric pressure. Soil water, being generally at subatmospheric pressure, exercises a suction, which draws out a certain amount of water from the rigid and airtight tensiometer, thereby causing the pressure inside the tensiometer to fall below atmospheric pressure. This subpressure is indicated by a manometer, which may be a simple water- or mercury-filled U tube, a vacuum gauge, or an electrical transducer (Young and Sisson, 2002).

A tensiometer kept in place tends to track the changes in the soil's matric suction. As soil moisture is depleted by drainage or plant uptake or as it is replenished by rain or irrigation, corresponding readings on the tensiometer gauge occur. Owing to the hydraulic resistance of the cup and the surrounding soil and of the contact zone between the cup and the soil, the tensiometer response may lag behind suction changes in the soil. The lag time can be minimized by the use of a null-type device or a transducer-type manometer with rigid tubing, so practically no flow of water need take place as the instrument responds to changes in the soil.

Since the walls of the tensiometer's porous cup are permeable to both water and solutes, the solutes in the soil solution diffuse freely into the porous cup, so water inside the tensiometer tends to assume the same solute composition and concentration as soil water. Therefore the instrument does not indicate the osmotic suction of soil water (unless it is equipped with some type of an auxiliary salt sensor).

Suction measurements by tensiometry are limited to matric suction values below 1 atm (about 1 bar, or 100 kPa). This is due first of all to the fact that the vacuum gauge or manometer measures a partial vacuum relative to the external atmospheric pressure. Even if the gauging of pressure is independent of atmospheric pressure (e.g., if it is done by means of a strain-gauge manometer), we are still faced with the general failure of water columns in macroscopic systems to withstand tensions exceeding 1 atm. (Water in narrow capillaries, however, can maintain continuity at much higher tensions; witness, for example, the continuity of liquid water in the xylem vessels of tall trees.) Furthermore, because the ceramic cup is generally made of rather permeable and porous material, in the interest of promoting rapid equilibrium with soil moisture, higher suction may cause the entry of air from the soil into the cup. Such air entry equalizes the internal tensiometric pressure to the atmospheric pressure. Consequently, soil suction may continue to increase even though the tensiometer fails to show it.

In practice, the useful limit of most tensiometers is a maximal tension of about 0.8 atm (80 kPa). However, the limited range of suction measurable by the tensiometer is not as serious a problem as it may seem. Though the suction range of 0–0.8 atm is a small part of the total range of suction variation encountered in the field, it actually encompasses the greater part of the soil wetness range. In many agricultural soils the tensiometer range accounts for more than 50% (and in coarse-textured soils 75% or more) of the amount of soil water available to plants. Thus, where soil management (as in irrigation) is aimed at maintaining the low-suction conditions that are most favorable for plant growth, tensiometers are definitely useful.

The Thermocouple Psychrometer

At equilibrium, the potential of soil moisture is equal to the potential of the water vapor in the ambient air. If thermal equilibrium is ensured and the gravitational effect is neglected, the vapor potential can be taken to be equal to the sum of the matric and osmotic potentials, since air acts as an ideal semipermeable membrane in allowing only water molecules to pass (provided that the solutes are nonvolatile).

A *psychrometer* is an instrument designed to indicate the *relative humidity*, that is, the ratio of the partial pressure of water vapor in the air to the equilibrium partial pressure of vapor in vapor-saturated air at the same temperature. The instrument generally measures the difference between the temperatures registered by a wet bulb and a dry bulb thermometer. The *dry bulb thermometer* indicates the temperature of a nonevaporating surface in thermal equilibrium with the ambient air. The *wet bulb thermometer* indicates the generally lower temperature of an evaporating surface, where latent heat is absorbed in proportion to the rate of evaporation.

The relative humidity of a body of air in equilibrium with a moist soil will depend on the temperature as well as on the state of water in the soil — that is, on the constraining effects of adsorption, capillarity, and solutes, all of which act to reduce the evaporability of soil water relative to that of pure, free water at the same temperature. Hence the relative humidity of an unsaturated soil's atmosphere will generally be under 100%, and the deficit to "saturation" will depend on the soil moisture potential, or suction, due to the combined effects of the osmotic and matric potentials.

Recent decades have witnessed the development of highly precise, miniaturized thermocouple psychrometers that make possible the in situ measurement of soil moisture potential (Rawlins and Campbell, 1986; Andraski and Scanlon, 2002). A *thermocouple* is a double junction of two dissimilar metals. If the two junctions are subjected to different temperatures, they will generate a voltage difference. If, on the other hand, an electromotive force (emf) is applied between the junctions, a difference in temperature will result. Depending on which way a direct current is applied, one junction can be heated while the other is cooled, and vice versa. The soil psychrometer (Fig. 6.13) consists of a fine wire thermocouple, one junction of which is equilibrated with the soil atmosphere by placing it inside a hollow porous cup embedded in the soil while the other junction is kept in an insulated medium to provide a temperature lag.

During operation, an electromotive force is applied so the junction exposed to the soil atmosphere is cooled to a temperature below the dew point of that atmosphere. At this point a droplet of water condenses on the junction, allowing it to serve as a wet bulb thermometer. This is a consequence of the so-called *Peltier effect* (Yavorsky and Detlaf, 1972). The cooling is then stopped; as the water from the droplet reevaporates, the junction attains a wet bulb temperature. That temperature remains nearly constant until the junction dries out, after which it returns to the ambient soil temperature. While evaporation takes place, the difference in temperature between the "wet bulb" and the insulated

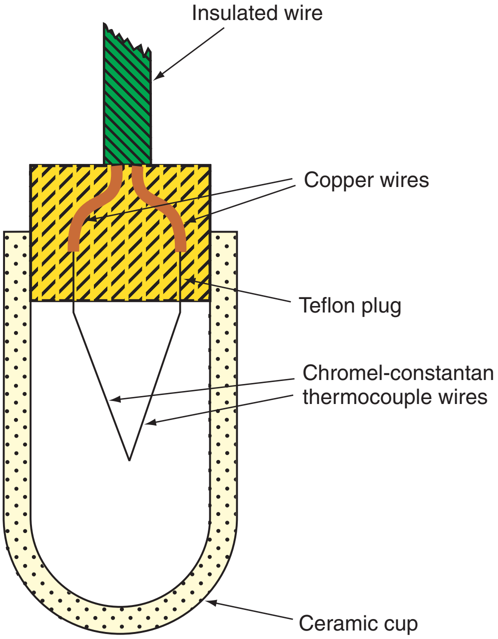


Fig. 6.13. Cross section of a thermocouple psychrometer contained in an air-filled ceramic cup.

junction serving as "dry bulb" generates an emf, which is indicative of the soil moisture potential. The relative humidity (i.e., the vapor pressure relative to that of pure, free water) is related to the soil-water potential according to

$$\phi = RT \ln(p/p_0) \tag{6.30}$$

where p is the vapor pressure of soil water, p_0 is the vapor pressure of pure, free water at the same temperature and air pressure, and R is the specific gas constant for water vapor.

Thermocouple psychrometry can be quite useful considerably beyond the suction range of the tensiometer (i.e., from 0.2 to 5 MPa, or 2 to 50 bar), where the change of wetness per unit change of suction is quite small. This instrument is used mostly in research and is now manufactured commercially. It is also applied to the measurement of plant-water potential.

Measurement of the Soil-Moisture Characteristic Curve

The fundamental relation between soil wetness and matric suction is often determined by means of a tension-plate assembly (Fig. 6.14) in the

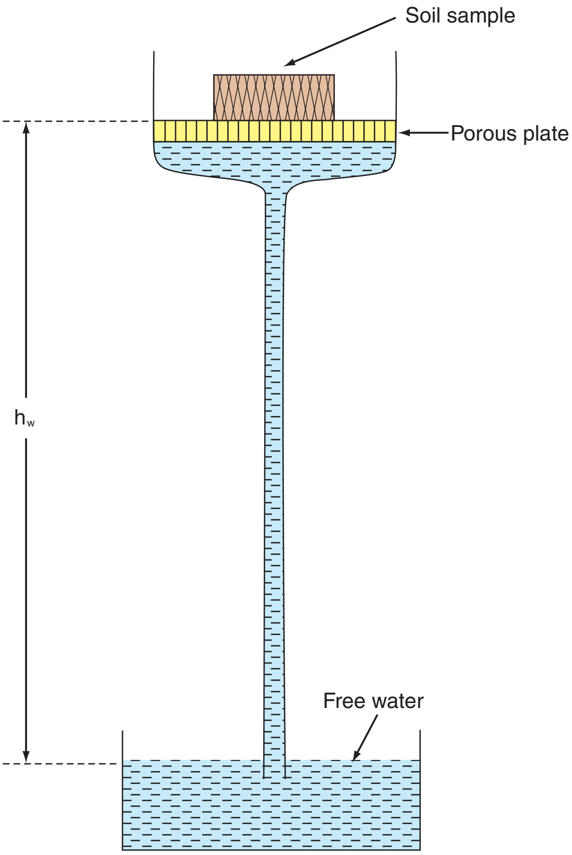


Fig. 6.14. Tension-plate assembly for equilibrating a soil sample with a known matric suction value. This assembly is applicable in the range of 0–1 bar only.

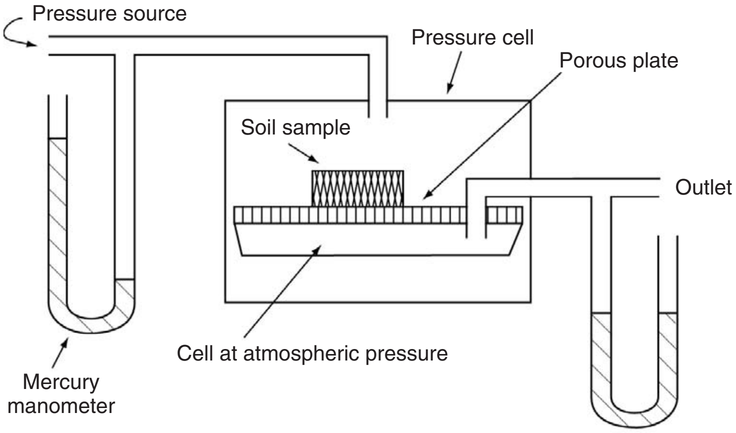


Fig. 6.15. Pressure-plate apparatus for moisture characteristic measurements in the high-suction range. The lower side of the porous plate is in contact with water at atmospheric pressure. Pressurized air is used to extract water from initially saturated soil samples.

low-suction (<1 bar) range and by means of a pressure-plate or pressure-membrane apparatus (Fig. 6.15) in the higher-suction range. These instruments allow the application of successive suction values and the repeated measurement of the equilibrium soil wetness at each suction (Dane and Hopmans, 2002).

The suction value obtainable by tension-plate devices is limited to 1 atm (1 bar, or 100 kPa) if the soil air is kept at atmospheric pressure and the pressure difference across the plate is controlled by vacuum or by a hanging water column. Matric suction values considerably greater than I bar (say, 20 bar or even more) can be obtained by increasing the pressure of the air phase. This requires placing the porous-plate assembly inside a pressure chamber, as shown in Fig. 6.15. The limit of matric suction obtainable with such a device is determined by the design of the chamber (i.e., its safe working pressure) and by the maximal air pressure difference the saturated porous plate can bear without allowing air to bubble through its pores. Ceramic plates generally do not hold pressures greater than about 2 MPa (20 bar), but cellulose acetate membranes can be used with pressures as high as 10 MPa (100 bar).

Soil-moisture retention in the low-suction range (0–100 kPa, equivalent to 0–1 bar) is strongly influenced by soil structure and pore size distribution. Hence, measurements made with disturbed samples (e.g., dried, screened, and artificially packed samples) cannot be expected to represent field conditions. The use of undisturbed soil cores is therefore preferable. Even better, in principle, is the in situ determination of the soil moisture characteristic curve by making simultaneous measurements of wetness (e.g., with the neutron moisture meter or with TDR) and of suction (using tensiometers) in the field (Bruce and Luxmoore, 1986). However, this approach is often frustrated by soil heterogeneity and by uncertainties over equilibrium as well as hysteretic phenomena as they might occur in the field.

REMOTE SENSING OF SURFACE-ZONE SOIL MOISTURE

Remote sensing is the collection of information regarding an object of interest, conducted from some distance without actual contact with that object. It is usually accomplished by detecting and measuring various portions (or bands) of the electromagnetic spectrum, using airborne or satellite-borne electronic scanning devices. Remote sensing of the earth's surface includes aerial photography, multispectral imagery, infrared imagery, radar, and microwave scanning. These techniques may be passive or active. Passive techniques measure signals emitted spontaneously from the ground. Active sensing techniques consist of generating a signal that is sent to the ground and of measuring its response.

Remote sensing is a promising technique in environmental soil physics and hydrology (Engman, 1995). It permits measurements that are not generally possible with traditional techniques, especially in areas where on-site stations are sparse and where data are difficult to obtain otherwise.

Information regarding the varying moisture content of the soil's surface zone is of particular interest. However insignificant this thin layer, with the small amount of water contained in it, may seem in comparison with the total amount of water on earth, it is indicative of the entire interaction between the land surface and the atmosphere as it regulates such processes as the exchange of energy and the allocation of precipitated water among infiltration, runoff, evaporation, and transpiration. Moreover, topsoil moisture affects seed germination, subsequent plant growth, and — ultimately — the success or failure of agriculture.

Of the various techniques suggested for measuring soil moisture, microwave technology appears to be the most promising at present. It can be used from a space platform (as well as from aircraft and truck-mounted devices) and can provide quantitative data of moisture in the soil's top layer (of the order of 5 cm) under a variety of topographic and vegetative conditions. The theoretical basis for measuring soil moisture by microwave techniques is based on the contrast between the dielectric constant of liquid water and of dry soil particles (namely, 80 compared to 3–5). That contrast is due to the tendency of the polarity of water molecules (i.e., their ability to align themselves in an applied electromagnetic field). As soil moisture increases, so does the soil's dielectric constant.

Passive microwave remote sensing of top-layer moisture has been shown to be practical. In the case of a bare surface, a radiometer can be used directly to measure the intensity of emission from the soil. According to Schmugge (1990), this emission is proportional to the product of surface temperature and surface emissivity, commonly referred to as the *microwave brightness temperature* (T_b):

$$T_b = \tau(H)[rT_{sky} + (1 - r)T_{soil}] + T_{atm} \tag{6.31}$$

where $\tau(H)$ is the atmospheric transmissivity for a radiometer at height H above the ground, r is the smooth-surface reflectivity, T_{soil} is the temperature of the soil, T_{atm} is the average temperature of the atmosphere, and T_{sky} is the contribution from the reflected sky brightness. For microwave wavelengths greater than 5 cm, the atmospheric transmissivity approaches 99% and the values of T_{sky} and T_{atm} are both less than 5 K (i.e., small compared to the soil contribution). Assuming the last terms to be negligible, we can simplify the equation to

$$T_b = (1 - r)T_{soil} = \varepsilon T_{soil} \tag{6.32}$$

where $\varepsilon = (1 - r)$ is the emissivity, which depends primarily on the dielectric constant of the soil and secondarily on the surface roughness.

The reflection coefficient of the ground is related nonlinearly to the dielectric constant and hence also to soil moisture. Therefore, most algorithms used for this type of remote sensing of soil moisture are calibrated with reference to empirical ground data (Engman, 1995). There is also some uncertainty over the exact thickness of the soil layer that is effectively monitored. The *penetration depth*, defined as the distance in the soil over which microwave radiated power is attenuated by a factor of e (= 2.718), also depends on the wavelength as well as on the dielectric constant (and hence on soil moisture). For a wavelength of 5 cm, the penetration depth varies from 10 cm for a soil with 1% volumetric wetness to less than 1 cm for a soil with 30% moisture. For a wavelength of 20 cm, the penetration depth varies from about 40 cm with 1% moisture to 5 cm with 30% moisture by volume (Njoku and Kong, 1977).

For the active microwave approach, the measured radar backscatter, σ_t, is made up of the backscatter from vegetation, σ_v, and from the soil, σ_s, and the attenuation caused by the vegetation canopy L:

$$\sigma_t = \sigma_v + \sigma_s/L \tag{6.33}$$

The soil backscatter can be related directly to soil moisture by

$$\sigma_s = R\alpha\theta \tag{6.34}$$

where R is a surface roughness term, α is a soil moisture sensitivity term, and θ is the volumetric moisture content. Because R and α are known to vary with wavelength, polarizataion, and angle of incidence, it is difficult to estimate these terms independently. Hence, as in the case of the passive microwave approach, the relationship between measured backscatter and soil moisture requires an empirical correlation with ground data, even for bare soils.

The effect of vegetation is, on the one hand, to attenuate the microwave emission from the soil and, on the other hand, to contribute its own emission to the total radiative flux. The net effect depends on the structure and density of the vegetative cover as well as on its degree of hydration. The interference of vegetation with the measurement of soil moisture can be minimized by choosing the optimal spectral band, because plant canopies are evidently more transparent to longer wavelengths than to shorter wavelengths in the range of microwave studies (Jackson and Schmugge, 1991).

7. WATER FLOW IN SATURATED SOIL

LAMINAR FLOW IN NARROW TUBES

Before beginning a discussion of flow in so complex a medium as soil, let us first consider the basics of fluid flow in narrow tubes. Early theories of fluid dynamics were based on the hypothetical concept of a “perfect” fluid, both frictionless and incompressible. In such a fluid, contacting layers exhibit no *tangential forces* (shearing stresses), only *normal forces* (pressures). Such fluids do not in fact exist. In the flow of real fluids, adjacent layers do transmit tangential stresses (drag), and the existence of intermolecular attraction causes the molecules in contact with a solid wall to adhere to it rather than slip over it. The flow of a real fluid is associated with the property of *viscosity*, defined in Chapter 2.

We can visualize the nature of viscosity by considering the motion of a fluid between two parallel plates, the lower one at rest, the upper one moving at a constant velocity (Fig. 7.1). Experience shows that the fluid adheres to both walls, so its velocity at the lower plate is zero, and that at the upper plate is equal to the velocity of the plate. Furthermore, the velocity distribution in the fluid between the plates is linear, so the fluid velocity at any plane between the plates is proportional to the distance y from the lower plate.

Maintaining the relative motion of the plates at a constant velocity requires applying a tangential force to overcome the frictional resistance in the fluid. This resistance, per unit area of the plate, is proportional to the velocity of the upper plate U and inversely proportional to the distance h between the plates. The shearing stress τ_s at any point is proportional to the velocity gradient du/dy. The viscosity η is the proportionality factor between τ_s and du/dy:

$$\tau_s = \eta \, du/dy \tag{7.1}$$

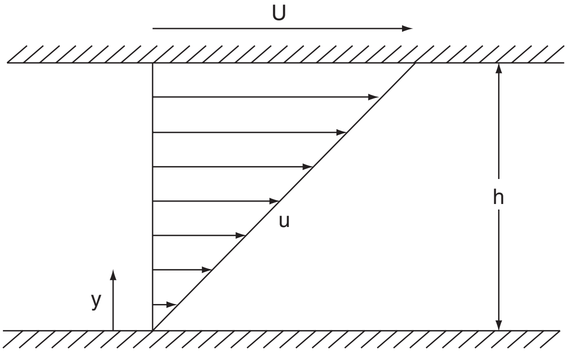


Fig. 7.1. Velocity distribution in a viscous fluid between two parallel plates, the upper one moving at a velocity *U* relative to the lower one.

Equation (7.1) bears a formal similarity to Hooke's law of elasticity. However, in an elastic solid the shearing stress is proportional to the strain, whereas in a viscous fluid the shearing stress is proportional to the time rate of the strain (i.e., the velocity gradient).

Now let us consider flow through a straight, cylindrical tube of uniform diameter $D = 2R$ (Fig. 7.2). The velocity is zero at the wall (because of adhesion), maximal at the axis, and constant on cylindrical surfaces that are concentric about the axis. Adjacent cylindrical laminae, moving at different velocities, slide over each other. A parallel motion of this kind is called *laminar*. Fluid flow in a horizontal tube is normally caused by a pressure gradient acting in the axial direction. A fluid "particle" is accelerated by the pressure gradient and retarded by the frictional resistance.

Imagine a coaxial fluid cylinder of length L and radius y. For flow velocity to be constant (no net force, hence no acceleration), the pressure force acting on the face of the cylinder $\Delta p\ \pi y^2$ (in which $\Delta p = p_1 - p_2$) must be equal to the frictional resistance due to the shear force $2\pi y L \tau_s$ acting on the circumferential area. Thus,

$$\tau_s = (\Delta p/L)(y/2)$$

Recalling Eq. (7.1), $\tau_s = -\eta\ (du/dy)$ (the negative sign arises because in this case u decreases with increasing y), we obtain

$$du/dy = -(\Delta p/\eta L)(y/2)$$

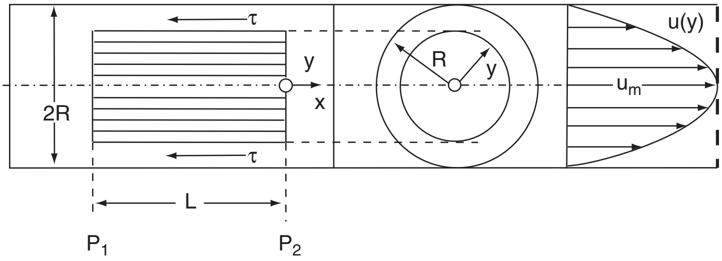


Fig. 7.2. Laminar flow through a cylindrical tube.

which, on integration, gives

$$u(y) = -(\Delta p/\eta L)(c - y^2/4)$$

The constant of integration c is evaluated from the boundary condition of no slip at the wall ($u = 0$ at $y = R$), so $c = R^2/4$. Therefore,

$$u(y) = (\Delta p/4\eta L)(R^2 - y^2) \tag{7.2}$$

Equation (7.2) indicates that the velocity is distributed parabolically over the radius, with maximum velocity $u_{\max}$ on the axis ($y = 0$):

$$u_{\max} = \Delta p\ R^2/4\eta L$$

The discharge Q, that is, the volume flowing through a section of length L per unit time, can now be evaluated. The volume of a paraboloid of revolution is (1/2)(base × height). Hence

$$Q = (1/2)\pi R^2 u_{\max} = \pi R^4 \Delta p/8\eta L \tag{7.3}$$

This equation, known as *Poiseuille's law*, indicates that the volume flow rate is proportional to the pressure drop per unit distance ($\Delta p/L$) and the fourth power of the radius of the tube.

The mean velocity over the cross section is

$$u_{\text{mean}} = \Delta p R^2/8\eta L = (R^2/a\eta)\nabla \mathrm{p} \tag{7.4}$$

where ∇p is the pressure gradient ($\Delta p/L$). Parameter a, equal to 8 in a circular tube, varies with the shape of the conducting passage.

Laminar flow prevails only at relatively low flow velocities and in narrow tubes. As the radius of the tube and the flow velocity are increased, the point is reached at which the mean flow velocity is no longer proportional to the pressure drop, and the parallel *laminar flow* changes into a *turbulent flow* with fluctuating eddies. Conveniently, however, laminar flow is the rule rather than the exception in most water flow processes taking place in soils, because of the narrowness of soil pores.

Sample Problem

The water in an irrigation hose is at a hydrostatic pressure of 100 kPa (1 bar). Five drip-irrigation emitters are inserted into the wall of the hose. Calculate the drip rate (L/hr) from the emitters if each contained a coiled capillary tube 1 m long and the capillary diameters are 0.2, 0.4, 0.6, 0.8, and 1.0 mm. Assuming laminar flow, what fraction of the total discharge is due to the single largest emitter?

We use Poiseuille's law to calculate the discharge: $Q = \pi R^4\ \Delta P/8\eta L$. Substituting the values for π (3.14), the pressure differential ΔP (10^5 N/m^2 = 10^5 kg/m sec^2), the viscosity η (10^{-3} kg/m sec, at 20°C), the capillary tube length (1 m), and the appropriate tube radii (0.1, 0.2, 0.3, 0.4, and 0.05 mm), we obtain

$$\begin{aligned} Q_1 &= \frac{3.14 \times (10^{-4}\ \text{m})^4 \times 10^5\ \text{kg/m sec}^2}{8 \times 10^{-3}\ \text{kg/m sec} \times 1\ \text{m}} = 3.9 \times 10^{-9}\ \text{m}^3/\text{sec} \\ &= 0.014\ \text{L/hr} \end{aligned}$$

$$Q_2 = \frac{3.14 \times (2 \times 10^{-4}\ \text{m})^4 \times 10^5\ \text{kg/m sec}^2}{8 \times 10^{-3}\ \text{kg/m sec} \times 1\ \text{m}} = 2^4\ Q_1 = 16 \times 1.4 \times 10^{-2}$$

$$= 0.226\ \text{L /hr}$$

$$Q_3 = 3^4 Q_1 = 81 Q_1 = 1.14\ \text{L/hr}$$

$$Q_4 = 4^4 Q_1 = 256 Q_1 = 3.61\ \text{L/hr}$$

$$Q_5 = 5^4 Q_1 = 625 Q_1 = 8.81\ \text{L/hr}$$

Total discharge from all five emitters:

$$Q_{\text{total}} = 0.014 + 0.226 + 1.14 + 3.61 + 8.81 = 13.8\ \text{L/hr}$$

Fractional contribution of the single largest emitter:

$$Q_5/Q_{\text{total}} = 8.81/13.8 = 0.639 = 63.8\%$$

The single largest emitter thus accounts for nearly two-thirds of the total discharge, while the smallest emitter accounts for only 0.1% (though its diameter is only one-fifth that of the largest emitter).

Note: Modern drip-irrigation emitters generally depend on partially turbulent (rather than completely laminar) flow, to reduce sensitivity to pressure fluctuations and vulnerability to clogging by particles.

DARCY'S LAW

If the soil were merely a bundle of straight and smooth tubes, each uniform in radius, we could assume the overall flow rate to equal the sum of the separate flows through the individual tubes. Knowledge of the size distribution of the tube radii would then enable us to calculate the total flow, caused by a known pressure difference, using Poiseuille's equation.

Unfortunately from the standpoint of physical simplicity, however, soil pores do not resemble uniform, smooth, cylindrical tubes but are irregularly shaped, tortuous, and intricately interconnected. Flow through soil pores is limited by numerous constrictions, or ''necks,'' with occasional ''dead-end'' spaces. The fluid velocity varies drastically from point to point, even along the same passage, and varies even more among different passages. Hence, the actual geometry and flow pattern in a typical soil specimen are too complicated to be described in microscopic detail. For this reason, flow through a complex porous medium such as a soil is generally described in terms of a *macroscopic flow-velocity vector*, which is the overall average of the microscopic velocities over the total volume considered. The detailed flow pattern is thus ignored, and the conducting body is treated as though it were a uniform medium, with the flow spread out over the entire cross section, solid and pore space alike. (An implicit assumption here is that the soil volume taken is sufficiently large relative to the pore sizes and microscopic heterogeneities to permit the averaging of velocity and potential over the cross section.)

We now examine the flow of water in a macroscopically uniform, saturated soil body, and attempt to describe the quantitative relations connecting the rate of flow, the dimensions of the body, and the hydraulic conditions at the inflow

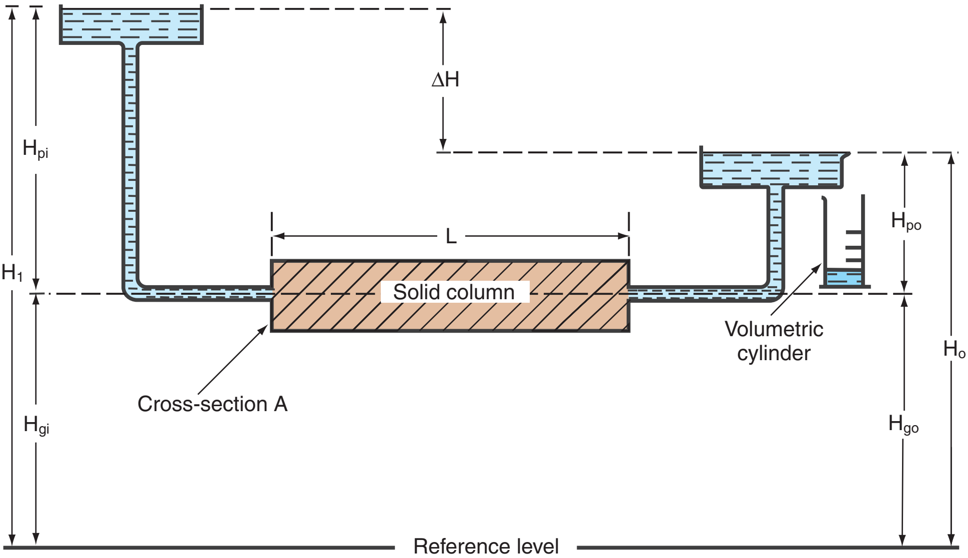


Fig. 7.3. Flow in a horizontal saturated column.

and outflow boundaries. Figure 7.3 shows a horizontal column of soil through which a steady flow of water is occurring from left to right, from an upper reservoir to a lower one, in each of which the water level is maintained constant.

Experience shows that the discharge rate Q, being the volume V flowing through the column per unit time, is directly proportional to the cross-sectional area and to the hydraulic head drop ΔH and inversely proportional to the length of the column L:

$$Q = V/t \propto A\ \Delta H/L$$

The usual way to determine the hydraulic head drop across the system is to measure the head at the inflow boundary H_i and at the outflow boundary H_o relative to some reference level. ΔH is the difference between these two heads:

$$\Delta H = H_i - H_o$$

Obviously, no flow occurs in the absence of a hydraulic head difference, that is, when $\Delta H = 0$.

The head drop per unit distance in the direction of flow ($\Delta H/L$) is the *hydraulic gradient*, which is, in fact, the driving force. The specific discharge rate Q/A (i.e., the volume of water flowing through a unit cross-sectional area per unit time t) is called the *flux density* (or simply the flux) and is indicated by q. Thus, the flux is proportional to the hydraulic gradient:

$$q = Q/A = V/At \propto \Delta H/L$$

The proportionality factor K is termed the *hydraulic conductivity*:

$$q = K\ \Delta H/L \tag{7.5}$$

This equation is known as *Darcy's law*, after Henry Darcy, the French engineer who discovered it over a century ago in the course of his classic investigation of seepage rates through sand filters in the city of Dijon (Darcy, 1856; Hubbert, 1956; Philip, 1995).

Where flow is unsteady (the flux changes with time) or the soil nonuniform, the hydraulic head may not decrease linearly along the direction of flow. Where the hydraulic head gradient or the conductivity is variable over time and space, we must consider the instantaneous localized gradient, flux, and conductivity values rather than average values for the soil system as a whole. A more exact expression of Darcy's law is, therefore, in differential form. Slichter (1899) generalized Darcy's law into a three-dimensional macroscopic differential equation of the form

$$q = -K\, \nabla H \tag{7.6}$$

where ∇H is the three-dimensional gradient of the hydraulic head. Since $K\, \nabla H$ is the product of a scalar K and a vector ∇H, the flux q is a vector, the direction of which is determined by ∇H. In an isotropic medium it is orthogonal to surfaces of equal hydraulic potential H.

Stated verbally, Darcy's law postulates that the flow of a viscous liquid through a porous medium is in the direction of, and proportional to, the *driving force* acting on the liquid (i.e., the *hydraulic gradient*) and is also proportional to a transmitting property of the conducting medium (the *hydraulic conductivity*). In a one-dimensional system, Eq. (7.6) takes the form

$$q = -K\, dH/dx \tag{7.7}$$

Mathematically, Darcy's law is similar to the linear transport equations of classical physics, such as *Ohm's law* (stating that the current, or flow rate of electricity, is proportional to the electrical potential gradient), *Fourier's law* (the rate of heat conduction is proportional to the temperature gradient), and *Fick's law* (the rate of diffusion is proportional to the concentration gradient).

BOX 7.1 Henry Cut the Mustard

The city of Dijon had long been known for its production of mustard, an important ingredient of France's celebrated *haute cuisine*. The pungent wastes of that industry, however, contributed to the fouling of the city's water supply. Even though the denizens of Dijon, being stalwart Frenchmen and Frenchwomen, did not drink much water (preferring, as they still do, the taste of *le vin*), they still needed water for cooking and sanitation. So they invited an engineer named Henry Darcy to design a filtration system. Being an engineer, Darcy must have looked for a soil physics handbook on filtration, but found none, so he had to experiment from scratch. He ended up formulating a new law and achieving immortal, if posthumous, fame. The use of filters packed with inert sand was but a first step in the greatly needed improvement of municipal water supplies. (It led eventually to the use of active materials such as clay, charcoal, diatomaceous earth, and ion exchange resins, along with chlorination, oxidation, flocculation, and — lately — reverse osmosis.) Henry Darcy was a pioneer in the task of purifying drinking water. His ultimate contribution to fundamental soil physics and hydrology was an unintended byproduct of that direct goal.

Sample Problem

Let us suppose we were given the task of purifying Dijon's water, with its 10,000 denizens. Since they drink mostly wine, their daily water requirements are modest, say, no more than 20 liters per day per person, on average. Let us suppose further that we knew (with the benefits of hindsight) that a column thickness of 0.30 m was adequate for filtration and that the hydraulic conductivity of the available sand was 2×10^{-5} m/sec. Could we calculate the area of the filter bed needed under a hydrostatic pressure head of 0.7 m? Consider the flow to be vertically downward to a fixed drainage plane.

We begin by calculating the discharge Q needed:

$$Q = \frac{10^4 \text{ persons} \times 20 \text{ L/person day} \times 10^{-3} \text{m}^3/\text{L}}{8.64 \times 10^4 \text{ sec/day}} = 2.31 \times 10^{-3} \text{m}^3/\text{sec}$$

We recall Darcy's law:

$$Q = AK\,\Delta H/L$$

Hence, the area needed is:

$$A = QL/K\,\Delta H$$

The hydraulic head drop ΔH equals the sum of the pressure head and gravitational head drops:

$$\Delta H = 0.7 + 0.3 = 1.0 \text{ m}$$

Substituting these values for L (0.3 m), ΔH (1 m), and K (2×10^{-5} m/sec), we obtain

$$A = \frac{2.31 \times 10^{-3} \text{m}^3/\text{sec} \times 0.3 \text{ m}}{2 \times 10^{-5} \text{m/sec} \times 1 \text{m}} = 34.7 \text{ m}^2$$

Note: Since populations and per capita water use tend to increase and filter beds tend to clog, it might be wise to apply a factor of safety to our calculations and increase the filtration capacity severalfold (especially to accommodate peak demand hours). Incidentally, per capita water use in the U.S. (with running toilets, showers, and dishwashers) ranges from 100 to 400 L/day.

GRAVITATIONAL, PRESSURE, AND TOTAL HYDRAULIC HEADS

The water entering the column of Fig. 7.3 is under a pressure, which is the sum of the hydrostatic pressure and the atmospheric pressure acting on the surface of the water in the reservoir. Since the atmospheric pressure is the same at both ends of the system, we can disregard it and consider only the hydrostatic pressure. Accordingly, the water pressure at the inflow boundary $P_i = \rho_w g H_{pi}$. Since ρ_w and g are both nearly constant, we can express this pressure in terms of the pressure head H_{pi}.

Water flow in a horizontal column occurs in response to a pressure head gradient. Flow in a vertical column may be caused by gravitation as well as pressure. The *gravitational head* H_g at any point is determined by the height of the point relative to some reference plane, while the pressure head is determined by the height of the water column resting on that point. The total hydraulic head H is the sum of these two heads:

$$H = H_p + H_g \tag{7.8}$$

To apply Darcy's law to vertical flow, we must consider the total hydraulic head at the inflow and at the outflow boundaries (H_I and H_o, respectively):

$$H_i = H_{pi} + H_{gi} \qquad \text{and} \qquad H_o = H_{po} + H_{go}$$

Darcy's law thus becomes

$$q = K[(H_{pi} + H_{gi}) - (H_{po} + H_{go})]/L$$

The gravitational head is often designated as z, which is the vertical distance in the rectangular coordinate system x, y, z. It is convenient to set the reference level as the point $z = 0$ at the bottom of a vertical column or at the center of a horizontal column. However, the exact elevation of this hypothetical level is unimportant, since the absolute values of the hydraulic heads determined in reference to it are immaterial and only their differences from one point in the soil to another affect flow.

The pressure and gravity heads can be represented graphically in a simple way. To illustrate this, we shall immerse and equilibrate a vertical soil column in a water reservoir so that the upper surface of the column will be level with the water surface, as shown in Fig. 7.4. The coordinates of Fig. 7.4 are arranged so that the height above the bottom of the column is indicated by the vertical axis z; and the pressure, gravity, and hydraulic heads are indicated on the horizontal axis. The gravity head is determined with respect to the reference level $z = 0$, and it increases with height at the ratio of 1:1. The pressure head is determined with reference to the free-water surface, at which the hydrostatic pressure is zero. Accordingly, the hydrostatic pressure head at the top of the column is zero and at the bottom of the column is equal to L, the column length. Just as the gravity head diminishes from top to bottom, so the pressure head increases.

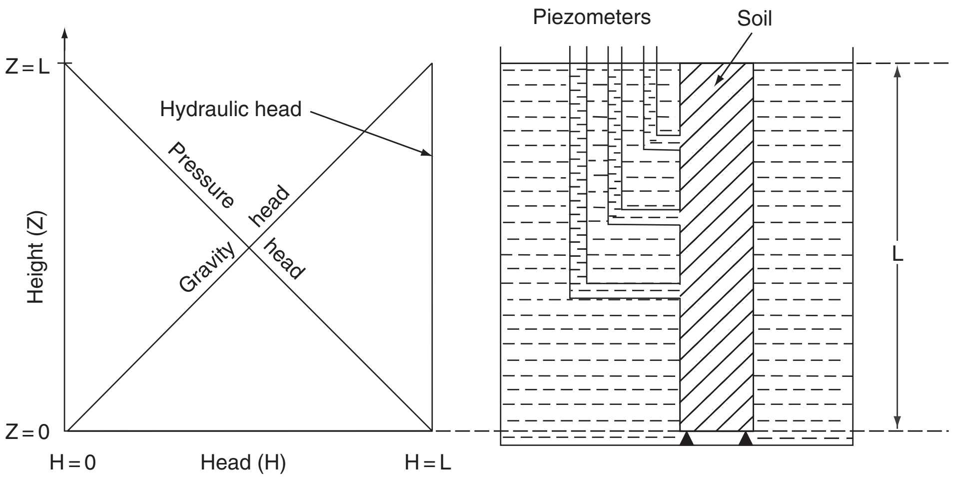


Fig. 7.4. Distribution of pressure, gravity, and total hydraulic heads in a vertical column immersed in water, at equilibrium.

Consequently their sum, which is the hydraulic head, remains constant all along the column. This is a state of equilibrium, at which no flow occurs.

This statement deserves to be further elaborated. The water pressure is not equal along the column, being greater at the bottom than at the top of the column. Why, then, will the water not flow from a zone of higher pressure to one of lower pressure? If the pressure gradient were the only force causing flow (as it is, in fact, in a horizontal column), the water would tend to flow upward. However, opposing the pressure gradient is a gravitational gradient of equal magnitude, resulting from the fact that the water at the top is at a higher gravitational potential than that at the bottom. In the illustration given, the two opposing gradients in effect cancel each other, so the total hydraulic head is constant throughout the column, as indicated by the standpipes (*piezometers*) connected to the column at the left.

As we already pointed out, it is convenient to set the reference level at the bottom of the column so that the gravitational potential can always positive. On the other hand, the pressure head of water can be either positive or negative: It is positive under a free-water surface (i.e., a water table) and negative above it.

FLOW IN A VERTICAL COLUMN

Figure 7.5 shows a uniform, saturated vertical column, the upper surface of which is ponded under a constant head of water H_1, and the bottom surface of which is set in a lower, constant-level reservoir. Flow is thus taking place from the higher to the lower reservoir through a column of length L.

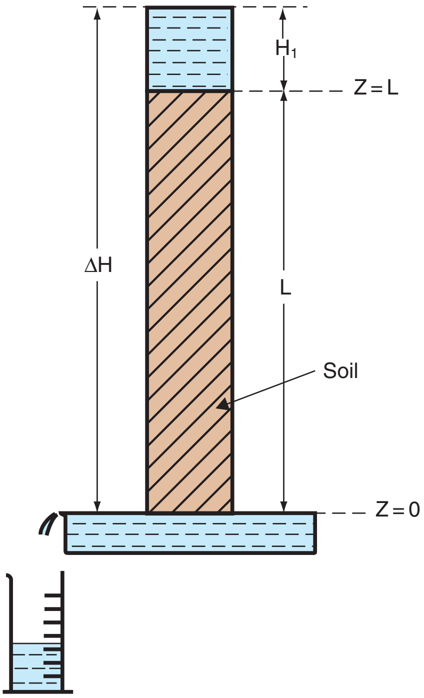


Fig. 7.5. Downward flow of water in a vertical saturated column.

In order to calculate the flux according to Darcy's law, we must know the hydraulic head gradient, which is the ratio of the hydraulic head drop (between the inflow and outflow boundaries) to the column length, as shown here:

		Pressure head		**Gravity head**
Hydraulic head at inflow boundary H_i	=	H_1	+	L
Hydraulic head at outflow boundary H_o	=	0	+	0
Hydraulic head difference $\Delta H = H_i - H_o$	=	H_1	+	L

The Darcy equation for this case is

$$q = K\,\Delta H/L = K(H_1 + L)/L = KH_1/L + K \tag{7.9}$$

Comparison of this case with the horizontal one shows that the rate of downward flow of water in a vertical column is greater than in a horizontal column by the magnitude of the hydraulic conductivity. It is also apparent that, if the ponding depth H_1 is negligible, the flux is equal to the hydraulic conductivity. This is due to the fact that, in the absence of a pressure gradient, the only driving force is the gravitational head gradient, which, in a vertical column, has the value of unity (since this head varies with height at the ratio of 1:1).

We now examine the case of upward flow in a vertical column, as shown in Fig. 7.6. In this case, the direction of flow is opposite to the direction of the gravitational gradient, and the hydraulic gradient becomes

		Pressure head		**Gravity head**
Hydraulic head at inflow boundary H_i	=	H_1	+	0
Hydraulic head at outflow boundary H_o	=	0	+	L
Hydraulic head difference $\Delta H = H_i - H_o$	=	H_1	−	L

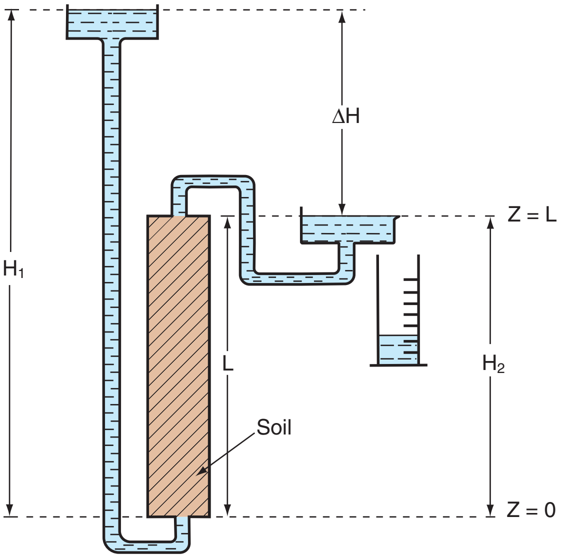


Fig. 7.6. Steady upward flow in a saturated vertical column.

The Darcy equation is thus:

$$q = K(H_1 - L)/L = KH_1/L - K = K\ \Delta H/L$$

FLOW IN A COMPOSITE COLUMN

Figure 7.7 depicts steady flow through a soil column consisting of two distinct layers, each homogeneous within itself and differing from the other layer in thickness and hydraulic conductivity. Layer 1 is at the inlet and layer 2 is at the outlet side of the column. The hydraulic head values at the inlet surface, at the interlayer boundary, and at the outlet end are designated H_1, H_2, and H_3, respectively. At steady flow, the flux through both layers must be equal:

$$q = K_1(H_1 - H_2)/L_1 = K_2(H_2 - H_3)/L_2 \tag{7.10}$$

where q is the flux, K_1 and L_1 are the conductivity and thickness (respectively) of the first layer, and K_2 and L_2 are the same for the second layer. Here we have disregarded any possible contact resistance between the layers. Thus,

$$H_2 = H_1 - qL_1/K_1 \quad \text{and} \quad qL_2/K_2 = H_2 - H_3$$

Therefore,

$$qL_2/K_2 = H_1 - qL_1/K_1 - H_3 \quad \text{and} \quad q = (H_1 - H_3)/(L_2/K_2 + L_1/K_1) \tag{7.11}$$

The reciprocal of the conductivity has been called the *hydraulic resistivity*, and the ratio of the thickness to the conductivity ($R_h = L/K$) has been called the hydraulic resistance per unit area. Hence,

$$q = \Delta H/(R_{h1} + R_{h2}) \tag{7.12}$$

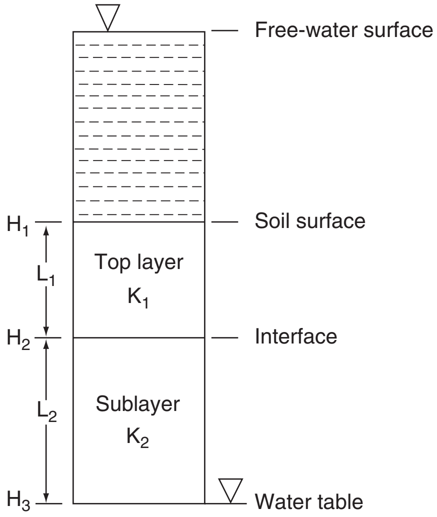


Fig. 7.7. Downward flow through a composite column.

where ΔH is the total hydraulic head drop across the entire system and R_{h1}, R_{h2} are the hydraulic resistances of layers 1 and 2. Equation (7.12) is in complete analogy to Ohm's law for an electric current through two constant resistances in series.

Sample Problem

Let us consider two cases of steady downward percolation through a two-layer soil profile, the top of which is submerged under a 1-m head of water and the bottom of which is defined by a water table. Each of the two layers is 0.50 m thick. In the one case, the conductivity of the top layer is 10^{-6} m/sec and that of the sublayer is 10^{-7} m/sec. In the second case, the same layers are reversed.

To calculate the flux and the hydraulic and pressure heads at the interface between the layers, we use the Ohm's law analogy for steady flow through two resistors in series:

$$q = \Delta H/(R_1 + R_2)$$

where q is the flux, ΔH is the total hydraulic head drop across the profile, and R_1, R_2 are the hydraulic resistances of the top layer and sublayer, respectively.

Each resistance is proportional directly to the layer's thickness and inversely to its conductivity (i.e., $R = L/K$). The pressure head at the soil surface is 1 m, and the gravity head (with reference to the soil's bottom) is also 1 m. Both the pressure and gravity heads at the bottom boundary are zero. Hence,

$$q = (1\text{ m} + 1\text{ m})/(0.5\text{ m}/10^{-6}\text{ m/sec} + 0.5\text{ m}/10^{-7}\text{ m/sec}) = 3.64 \times 10^{-7}\text{ m/sec}$$

We can now apply Darcy's equation to the top layer alone to obtain the hydraulic head at the interlayer interface:

$$q = K_1\ \Delta H_1/L_1 = K_1(H_{\text{surface}} - H_{\text{interface}})/L_1$$

Hence:

$$\begin{aligned} H_{\text{interface}} &= H_{\text{surface}} - qL_1/K_1 \\ &= 2\text{ m} - (3.64 \times 10^{-7}\text{ m/sec})(0.5\text{ m})/10^{-6}\text{ m/sec} = 1.818\text{ m} \end{aligned}$$

Since the gravity head at the interface is 0.5 m (above our reference datum at the bottom of the profile), the pressure head H_p is

$$H_p = H - H_g = 1.818 - 0.50 = 1.318\text{ m}$$

We now reverse the order of the layers so that the less conductive layer overlies the more conductive. The total head drop remains the same and so does the total resistance. Therefore the flux remains the same (assuming that both layers are still saturated and the conductivity of each does not change). Applying Darcy's equation to the top layer, we have, as previously,

$$H_{\text{interface}} = 2.0 - (3.64 \times 10^{-7})(0.5)/10^{-7} = 0.18\text{ m}$$

In this case the pressure head at the interface is

$$H_p = H - H_g = 0.18 - 0.50 = -0.32\text{ m}$$

Note: Comparison between the interface pressures of the two cases illustrates an important principle regarding flow in layered profiles. With the more conductive layer on top, flow is impeded at the interface and there is a pressure buildup, which in our case increased the pressure head from 1 m at the surface to 1.218 m at the interface. The opposite occurs when the upper layer is less conductive. In this case the pressure is dissipated through the top layer, often to the extent that a negative pressure develops at the interface.

FLUX, FLOW VELOCITY, AND TORTUOSITY

As stated earlier, the *flux density* (herein, simply *flux*) is the volume of water V passing through a unit cross-sectional area A (perpendicular to the flow direction) per unit time t, the dimensions of which are equivalent to length per time:

$$q = V/At = L^3/L^2T = LT^{-1}$$

(In SI units, the flux is expressed in cubic meters per square meter per second, equivalent to meters per second). These are the dimensions of velocity, yet we prefer the term *flux* to *flow velocity*, the latter being too ambiguous. Since soil pores vary in shape, width, and direction, the actual flow velocity in the soil is highly variable (e.g., wider pores conduct water more rapidly, and the liquid in the center of each pore moves faster than the liquid in close proximity to the particles). Strictly speaking, therefore, one cannot refer to a single velocity of liquid flow, but at best to an average velocity.

Even the average velocity of the flowing liquid differs from the flux as we have defined it. Flow does not in fact take place through the entire cross-sectional area A, because part of this area is plugged by particles and only the porosity fraction permits flow. Since the real area through which flow takes place is smaller than A, the average velocity of the liquid must be greater than the flux q. Furthermore, the actual length of the path traversed by an average parcel of liquid is greater than the soil column length L, owing to the labyrinthine, or tortuous, nature of the pore passages, as shown in Fig. 7.8.

Tortuosity can be defined as the ratio of the average roundabout path to the apparent, or straight, flow path; that is, the ratio of the average length of the pore passages (as if they were stretched out in the manner one might stretch out a coiled or tangled telephone wire) to the length of the soil specimen. Tortuosity is thus a dimensionless geometric parameter of porous media, which, though difficult to measure precisely, is always greater than 1 and may exceed 2. The *tortuosity factor* is sometimes defined as the inverse of what we defined as the tortuosity.

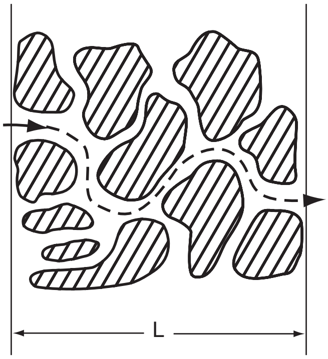


Fig. 7.8. Flow path tortuosity in the soil.

HYDRAULIC CONDUCTIVITY, PERMEABILITY, AND FLUIDITY

Recalling that the hydraulic conductivity K is the ratio of the flux to the potential gradient (i.e., the slope of flux vs. gradient, Fig. 7.9) and that the dimensions of flux are LT^{-1}, we note that the dimensions of hydraulic conductivity depend on those assigned to the potential gradient. In Chapter 6, we showed that the simplest way to express the potential is by use of length, or head, units (though, strictly speaking, H is not a true length but a pressure equivalent in terms of a water column height $H = P/\rho g$). Therefore, the hydraulic head gradient H/L, being the ratio of length to length, is dimensionless. Accordingly, the dimensions of hydraulic conductivity are the same as those of flux, namely, LT^{-1}. However, If the gradient is expressed in terms of the variation of hydraulic pressure with length, then the hydraulic conductivity assumes the dimensions of $M^{-1}L^3T$. Since the latter is cumbersome, the use of head units is generally preferred.

In a saturated soil of stable structure, as well as in a rigid porous medium such as sandstone, the hydraulic conductivity is usually found to be nearly constant. Its order of magnitude is about 10^{-4}–10^{-5} m/sec in a sandy soil and 10^{-6}–10^{-9} m/sec in a clayey soil.

To appreciate the practical significance of these values in more familiar terms, consider the hypothetical case of an unlined (earth-bottom) reservoir or pond in which one wishes to retain water against losses caused by downward seepage. Assuming, for the sake of simplicity, that seepage into and through the underlying soil is by gravity alone (i.e., no pressure or suction gradients in the soil), it will take place at a rate approximately equal to the hydraulic conductivity. A coarse sandy soil might have a K value of, say, 10^{-4} m/sec and would therefore lose water at the enormous rate of nearly 9 m/day (there being 8.64×10^4 sec/day). A loam soil with a K value of 10^{-6} m/sec would lose "only" about 0.1 m/day (100 mm/day). Finally, and in contrast, a bed of clay with a conductivity of 10^{-8} m/sec would allow the seepage of no more than 1 mm/day, less than the generally expectable rate of evaporation.

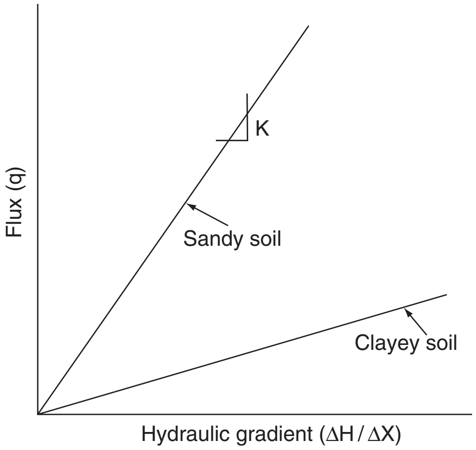


Fig. 7.9. The linear dependence of flux on hydraulic gradient, the hydraulic conductivity being the slope (i.e., the flux per unit gradient).

The hydraulic conductivity is obviously affected by structure as well as by texture, being greater if the soil is highly porous, fractured, or aggregated than if it is tightly compacted and dense. Hydraulic conductivity depends not only on total porosity but also, and primarily, on the sizes of the conducting pores.

Contrary to the convenient assumption (just stated) that the hydraulic conductivity of a saturated soil remains constant, it is often found to vary (generally to diminish) over time. Because of various chemical, physical, and biological processes, the hydraulic conductivity may change as water permeates the soil and flows in it. Changes occurring in the composition of the exchange complex (as when water entering the soil differs in composition or concentration of solutes from the original soil solution) can greatly affect the hydraulic conductivity. In general, the conductivity decreases with decreasing concentration of electrolytic solutes. This is due to swelling and dispersion phenomena, which are also affected by the species of cations present. Detachment and migration of clay particles during flow may result in the clogging of pores.

In practice, it is extremely difficult to saturate a soil with water without trapping some air. Encapsulated air bubbles may block pore passages, as shown in Fig. 7.10. Temperature changes may cause the flowing water to dissolve or to release gas bubbles and will also cause a change in their volume, thus affecting conductivity.

The hydraulic conductivity K is not a property of the soil alone. Rather, it depends jointly on the attributes of the soil and of the fluid. The soil characteristics that affect K are the total porosity, the distribution of pore sizes, and tortuosity — in short, the soil's pore geometry. The fluid attributes that affect conductivity are density and viscosity.

It is possible in theory, and sometimes in practice, to separate K into two factors: *intrinsic permeability* of the soil k and *fluidity* of the permeating liquid (or gas) f:

$$K = \mathrm{k}f \tag{7.13}$$

If K is given in terms of m/sec (LT^{-1}), then k is in units of m^2 (L^2) and f is in units of m^{-1} sec^{-1} ($L^{-1}T^{-1}$).

Fluidity is related directly to density and inversely to viscosity:

$$f = \rho g/\eta \tag{7.14}$$

Hence,

$$k = K\eta/\rho g \tag{7.15}$$

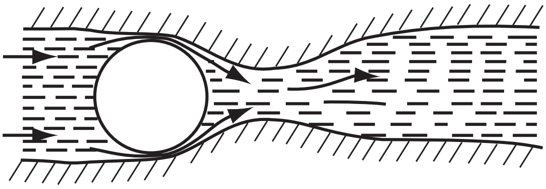


Fig. 7.10. An entrapped air bubble plugging flow.

where η is the dynamic viscosity (N sec/m^2, or Pa sec), ρ is the fluid density (kg/m^3), and g is the gravitational acceleration (m/sec^2).

While fluidity varies with temperature and composition of the fluid, permeability is ideally a property of the porous medium's pore geometry alone, provided the fluid and the solid matrix do not interact in such a way as to change the properties of either. In a stable porous body, the same permeability will be obtained with different fluids (water, air, or oil). However, in many cases water does interact with the solid matrix to modify its permeability, so hydraulic conductivity cannot be resolved into separate and independent properties of water and of soil.

LIMITATIONS OF DARCY'S LAW

Darcy's law is not universally valid for all conditions of liquid flow in porous media. It has long been recognized that the linearity of the flux versus hydraulic gradient relationship fails at high flow velocities, where inertial forces are no longer negligible compared to viscous forces. Darcy's law applies only as long as flow is laminar (i.e., nonturbulent movement of adjacent layers of the fluid relative to one another) and where soil–water interaction does not result in a change of fluidity or of permeability with a change in gradient. Laminar flow prevails in silts and finer materials for most commonly occurring hydraulic gradients found in nature. In coarse sands and gravels, however, hydraulic gradients much in excess of unity may result in nonlaminar flow conditions, so in such cases Darcy's law may not be applicable.

The quantitative criterion for the onset of turbulent flow is the *Reynolds number* N_{Re}:

$$N_{\mathrm{Re}} = du\rho/\eta \tag{7.16}$$

where u is the mean flow velocity, d is the effective pore diameter, ρ is the liquid density, and η is its viscosity. In straight tubes, the critical value of N_{Re} beyond which turbulence sets in has variously been reported to be of the order of 1000–2200 (Childs, 1969). However, the critical Reynolds number at which water flowing in a tube becomes turbulent is apparently reduced greatly when the tube is curved and its diameter varies. For porous media, therefore, it is safe to assume that flux remains linear with hydraulic gradient only as long as N_{Re} is smaller than unity. As flow velocity increases, especially in systems of large pores, the occurrence of turbulent eddies or nonlinear laminar flow results in "waste" of effective energy; that is, some energy is dissipated by the internal churning of the liquid so that the hydraulic potential gradient becomes less effective in inducing flow . This is illustrated in Fig. 7.11.

Deviations from Darcy's law may also occur at the opposite end of the flow-velocity range, namely, at low gradients and in narrow pores. Some investigators (Swartzendruber, 1962; Miller, R. J. and Low, 1963; Nerpin et al., 1966) have claimed that, in clayey soils, small hydraulic gradients may cause no flow or only low flow rates that are less than proportional to the gradient.

A possible reason for this anomaly is that the water in close proximity to the particle surfaces and subject to their adsorptive force fields may be more

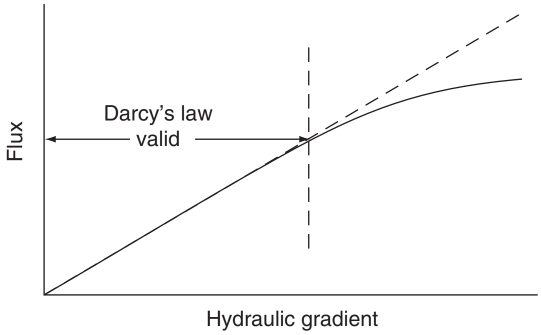


Fig. 7.11. The deviation from Darcy's law at high flux, where flow becomes turbulent.

rigid than ordinary bulk water and may exhibit the properties of a *Bingham liquid* (having a *yield value*) rather than a *Newtonian liquid* (Hillel, 1980a). The adsorbed (or "bound") water may even have a quasi-crystalline structure similar to that of ice. Some soils may exhibit a *threshold gradient* (Miller, R. J. and Low, 1963), below which the flux is either zero (the water remaining apparently immobile) or at least lower than predicted by the Darcy relation, and only at gradients exceeding the threshold value does the flux become proportional to the gradient (Fig. 7.12). These phenomena and their possible explanations, though highly interesting, are generally of little importance in practice, so Darcy's law can be employed in the vast majority of cases pertaining to the flow of water in soil.

Another possible cause for apparent flow anomalies in clay soils is their tendency to swell or compress (Smiles, 1976). As commonly formulated, Darcy's law applies to flow relative to a geometrically fixed porous matrix, and it may seem to fail when the particles composing the matrix are themselves moving relative to a fixed frame of reference.

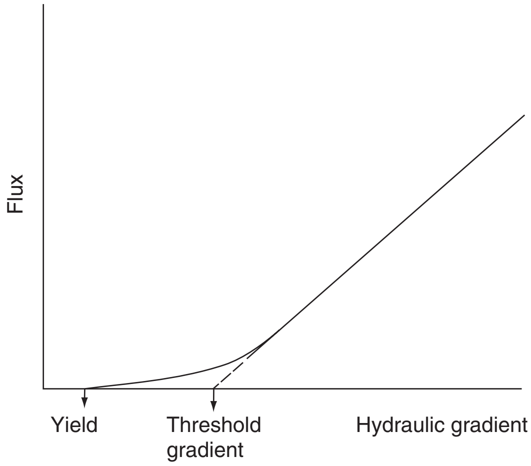


Fig. 7.12. Possible deviations from Darcy's law at low gradients.

HOMOGENEITY AND ISOTROPY

The hydraulic conductivity may be uniform throughout a soil, or it may vary from point to point, in which case the soil is said to be hydraulically *inhomogeneous*. If the conductivity is the same in all directions, the soil is *isotropic*. However, the conductivity at each point may differ for different directions (e.g., the horizontal conductivity may be greater, or smaller, than the vertical), a condition known as *anisotropy*.

A soil may be homogeneous and nonetheless anisotropic, or it may be inhomogeneous (e.g., layered) and yet isotropic at each point. Some soils exhibit both inhomogeneity and anisotropy. In certain cases, K may also be asymmetrical (or directional); that is to say, K may have a different value for opposite directions of flow along a given line. A basic review of anisotropy and layering was given by Bear et al. (1968). Anisotropy is generally due to the structure of the soil, which may be laminar, or platy, or columnar, etc., thus exhibiting a pattern of micropores or macropores with a distinct directional bias.

To illustrate the meaning of the concepts of homogeneity and isotropy, let us consider the hypothetical situation of two continuous bodies of soil or, better yet, two realms within the same body of soil, as shown in Fig. 7.13. The hydraulic conductivities for the three principal axes of the x, y, z coordinate system are designated K_x, K_y, K_z, respectively, for the soil body on the left; and the corresponding conductivities are designated $K_x{}^*$, $K_y{}^*$, $K_z{}^*$ for the body on the right. Four possible cases exist:

Case I. The soil is homogeneous and isotropic throughout both realms.

$$\left.\begin{array}{ll}\text{Homogeneous:} & K_x = K_x{}^*, K_y = K_y{}^*, K_z = K_z{}^* \\ \text{Isotropic:} & K_x = K_y = K_z, K_x{}^* = K_y{}^* = K_z{}^*\end{array}\right\}\text{ single value of } K$$

Case II. The soil is homogeneous but anisotropic.

$$\left.\begin{array}{ll}\text{Homogeneous:} & K_x = K_x{}^*, K_y = K_y{}^*, K_z = K_z{}^* \\ \text{Anisotropic:} & K_x \neq K_y \neq K_z, K_x{}^* \neq K_y{}^* \neq K_z{}^*\end{array}\right\}\text{ 3 values of } K$$

Case III. The soil is inhomogeneous but isotropic.

$$\left.\begin{array}{ll}\text{Inhomogeneous:} & K_x \neq K_x{}^*, K_y \neq K_y{}^*, K_z \neq K_z{}^* \\ \text{Isotropic:} & K_x = K_y = K_z, K_x{}^* = K_y{}^* = K_z{}^*\end{array}\right\}\text{ 2 values of } K$$

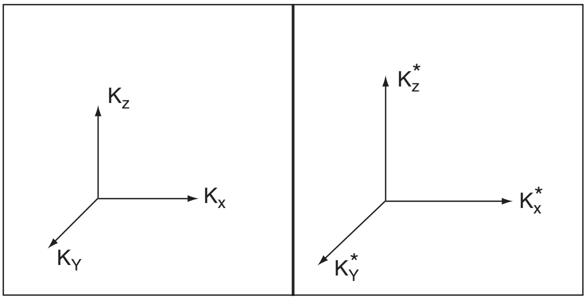


Fig. 7.13. Schematic illustration of homogeneity and isotropy for two adjacent realms in the soil.

Case IV. The soil is both inhomogeneous and anisotropic.

$$\left.\begin{array}{ll}\text{Inhomogeneous:} & K_x \neq K_x^*, K_y \neq K_y^*, K_z \neq K_z^* \\ \text{Anisotropic:} & K_x \neq K_y \neq K_z, K_x^* \neq K_y^* \neq K_z^*\end{array}\right\} \text{6 values of } K$$

Analysis of anisotropic flow systems is also complicated by the fact that the flow direction may not be orthogonal to the equipotential lines or planes; that is to say, the flow does not necessarily take place in the direction of the steepest potential gradient.

MEASURING HYDRAULIC CONDUCTIVITY OF SATURATED SOILS

Methods for measuring hydraulic conductivity in the laboratory were reviewed by Klute and Dirksen (1986) and by Reynolds et al. (2002), and for measurement in the field, by Amoozegar and Warrick (1986). The use of *laboratory permeameters* is illustrated in Figs. 7.14 and 7.15. Measurements are made either with dried and fragmented specimens, packed into the flow cells in a standardized manner, or, preferably, with *undisturbed core samples* taken directly from the field. In either case, provision must be made to avoid boundary flow along the walls of the container.

Field measurements can be made below the water table, as by the *auger-hole method* or by the *piezometer method* (Reynolds et al., 2002). Techniques have also been proposed for measurements above the water table, as by the *double-tube method* (Bouwer, 1961, 1962), the *shallow-well pump-in method*, and the *field-permeameter* method (Winger, 1960). Topp and Sattlecker (1983) described a rapid measurement of horizontal and vertical components of a soil's saturated hydraulic conductivity. Mallants et al.

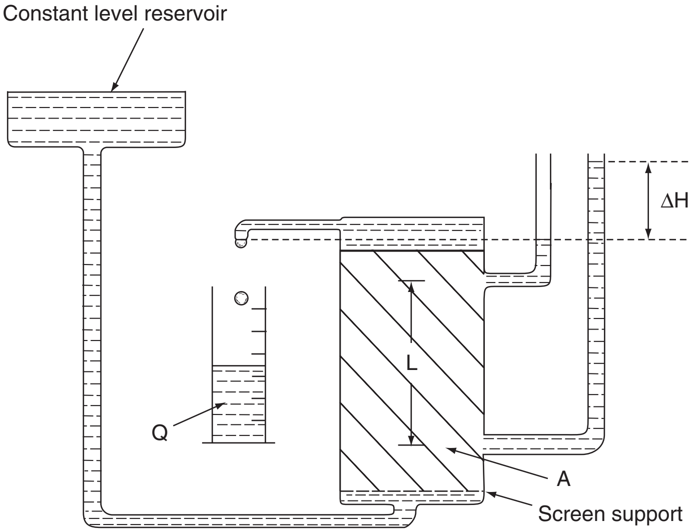


Fig. 7.14. The measurement of saturated hydraulic conductivity with a constant-head permeameter: $K = VL/At\,\Delta H$.

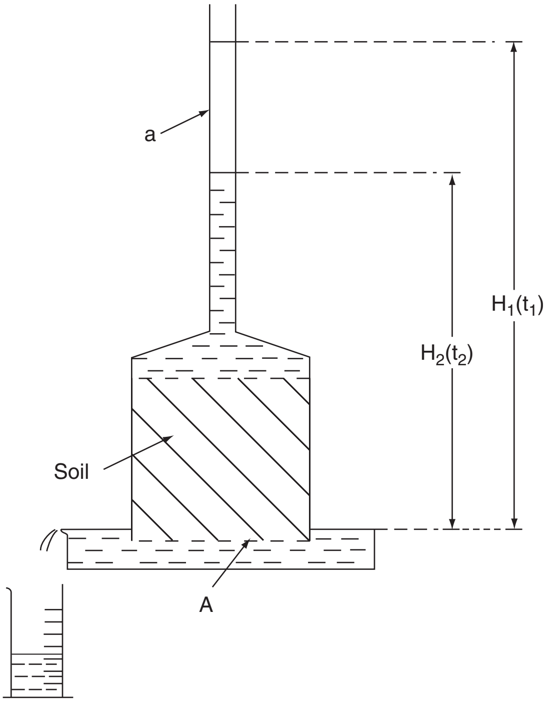


Fig. 7.15. The measurement of saturated hydraulic conductivity with a falling head permeameter: $K = [2.3aL/A(t_2 - t_1)](\log H_1 - \log H_2)$. Note that H_1, H_2 are the values of hydraulic head at times t_1 and t_2, respectively.

(1997) analyzed the spatial dependence of saturated hydraulic conductivity in a soil with macropores.

Typical values of saturated hydraulic conductivity for soils of different texture are listed in Table 7.1. Owing to soil heterogeneity, the apparent hydraulic conductivity measured often depends on the scale of the measurement. Thus, the *K* value measured on a cubic centimeter or decimeter may differ from the average value measured on a cubic meter. Too often, this is ignored and *K* values are reported without specifying the scale of the measurement.

TABLE 7.1 Hydraulic Conductivity at Saturation (K_s), Particle Diameters (d), and Height of Capillary Rise (h_c) for some Solids[a]

	K_s (m/sec)	d (mm)	h_c (m)
Clay	10^{-10}–10^{-8}	<0.002	2–4
Silt	10^{-8}–10^{-6}	0.002–0.05	0.7–1.5
Sand	10^{-5}–10^{-3}	0.05–2	0.12–0.35
Gravel	10^{-2}–10^{-1}	>2	Nil

[a] After Boeker and van Grandelle (1995).

EQUATIONS OF SATURATED FLOW

Darcy's law, by itself, is sufficient only to describe *steady*, or *stationary*, flow processes, in which the flux remains constant and equal throughout the conducting medium (and hence the potential and gradient at each point remain constant in time). *Unsteady*, or *transient*, flow processes, in which the magnitude and possibly even the direction of the flux and potential gradient vary in time, require the inclusion of an additional law, namely, the *law of conservation of matter*. To understand how this law applies, consider a small volume element (say, a cube) of soil, into and out of which flow takes place at possibly differing rates. The mass conservation law, expressed in the *equation of continuity*, states that if the rate of inflow into the volume element exceeds the rate of outflow, then the volume element must be storing the excess and increasing its water content. Conversely, if outflow exceeds inflow, storage must be decreasing.

Consider first the simplest case, one-dimensional flow, with q_x as the flux in the direction of x. Any increase of q_x with x must equal the decrease of water content w with time t:

$$\partial\theta/\partial t = -\partial q_x/\partial x \tag{7.18}$$

which in multidimensional systems becomes

$$\partial\theta/\partial \mathrm{t} = -\nabla \cdot q \tag{7.19}$$

We recall Darcy's law,

$$q = -K\ \nabla H \tag{7.20}$$

which in one dimension is

$$q_x = -K\ (dH/dx) \tag{7.21}$$

(where H is the hydraulic head and K is the hydraulic conductivity). Now we combine equation (7.20) with the continuity equation (7.19) to obtain the *general flow equation*:

$$\partial\theta/\partial t = \nabla \cdot K\ \nabla H \tag{7.22}$$

In applying this equation, the assumptions are usually made that inertial forces are negligible in comparison with viscous forces, that water is continuously connected throughout the flow region, that isothermal conditions prevail, and that no chemical or biological phenomena change the fluid or the porous medium during the flow process.

In one dimension, Eq. (7.22) becomes

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x}\left(K\,\frac{\partial H}{\partial x}\right) \tag{7.23}$$

Note that Eq. (7.23) is a partial differential equation with two independent variables, t and x.

Since the hydraulic head can be resolved into a pressure head H_p and a gravitational head (an elevation z above some reference datum), we can rewrite Eq. (7.22) as

$$\frac{\partial\theta}{\partial t} = \nabla \cdot [K(\nabla H_p + \nabla z)] \tag{7.24}$$

In horizontal flow, $\nabla z = 0$, so for this case,

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x}\left[K\left(\frac{\partial H_p}{\partial x}\right)\right] \tag{7.25}$$

while in vertical flow, $\delta z/\delta z = 1$, and therefore, for this case,

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z}\left[K\left(\frac{\partial H_p}{\partial z} + 1\right)\right] \tag{7.26}$$

In a saturated soil with an incompressible matrix, $\delta\theta/\delta t = 0$, the conductivity is usually assumed to remain constant; hence Eq. (7.23) becomes

$$K_s = \partial^2 H/\partial x^2 = 0 \tag{7.27}$$

where K_s is the hydraulic conductivity of the saturated soil (the "saturated conductivity"). For three-dimensional flow conditions, and allowing for anisotropy, the equation is

$$K_x \frac{\partial^2 H}{\partial x^2} + K_y \frac{\partial^2 H}{\partial y^2} + K_z \frac{\partial^2 H}{\partial z^2} = 0 \tag{7.28}$$

where K_x, K_y, and K_z represent the hydraulic conductivity values in the three principal directions x, y, z.

In an isotropic soil (where $K_x = K_y = K_z$ at each point) that is also homogeneous (the K values of all points are equal), we obtain the well-known *Laplace equation*:

$$\partial^2 H/\partial x^2 + \partial^2 H/\partial y^2 + \partial^2 H/\partial z^2 = 0 \tag{7.29}$$

This second-order partial differential equation of the elliptical type can be solved in some cases to obtain a quantitative description of steady-state flow in various geometric configurations.

8. WATER FLOW IN UNSATURATED SOIL

FLOW IN UNSATURATED VERSUS SATURATED SOIL

Most of the processes involving soil–water interactions in the field, including the supply of moisture and nutrients to plant roots as well as the transport of water and solutes beyond the root zone, occur while the soil is in an unsaturated condition. Unsaturated flow processes are in general complicated and difficult to describe quantitatively, since they often entail changes in the state and content of soil water during flow. Such changes involve complex relations among such variables as soil wetness, suction, and conductivity, whose interrelations are further complicated by hysteresis as well as by spatial variability.

In the preceding chapter, we stated that the flow of water in the soil is driven by a hydraulic potential gradient, that it occurs in the direction of decreasing hydraulic potential, and that its rate (flux) is proportional to the potential gradient and is affected by the geometric properties of the pore channels through which flow takes place. These principles apply in unsaturated as well as saturated soils. However, the nature of the moving force and the effective geometry of the conducting pores may be very different.

Apart from the gravitational force (which is completely independent of soil wetness), the primary moving force in a saturated soil is the gradient of a positive pressure potential. On the other hand, water in an unsaturated soil is subject to a subatmospheric pressure, or matric suction, which is equivalent to a negative pressure potential. The gradient of this potential likewise constitutes a moving force.

Matric suction is due, as we have noted, to the physical affinity between water and the matrix of the soil, including both the soil-particle surfaces and the capillary pores. When suction is uniform all along a horizontal column of soil, the column is at equilibrium and there is no moving force. Not so when a suction gradient exists. In that case, water will be drawn from a zone where the hydration envelopes surrounding the particles are thicker to where they are thinner, and from a zone where the capillary menisci are less curved to where they are more strongly curved. In other words, water will flow spontaneously from where matric suction is lower to where it is higher. It will flow in the pores that are water filled at the existing suction and creep along the hydration films over the particle surfaces, in a tendency to equilibrate the potential. (The ideal state of equilibrium, like human happiness, may never be achieved in practice, but its natural pursuit is a universal rule.)

Perhaps the most important difference between unsaturated and saturated flow is in the hydraulic conductivity. When the soil is saturated, all of the pores are water filled and conducting. The water phase is then continuous and the conductivity is maximal. When the soil desaturates, some of the pores become air filled, so the conductive portion of the soil's cross-sectional area diminishes. Furthermore, as suction develops, the first pores to empty are the largest ones, which are the most conductive (remember Poiseuille's law!), thus relegating flow to the smaller pores. At the same time, the large empty pores must be circumvented, so, with progressive desaturation, tortuosity increases. In coarse-textured soils, water may be confined almost entirely to the capillary wedges at the contact points of the particles, thus forming separate and discontinuous pockets of water (see Fig. 8.1). In aggregated soils, too, the large interaggregate spaces that confer high conductivity at saturation become (when emptied) barriers to liquid flow from one aggregate to another.

For all these reasons, the transition from saturation to unsaturation generally entails a steep drop in hydraulic conductivity, which may decrease by several orders of magnitude (sometimes down to one-millionth of its value at saturation) as suction increases from 0 to 1 MPa. At still higher suctions, or lower wetness values, the conductivity may be so low that very steep suction gradients, or very long times, are required for any appreciable flow to occur at all.

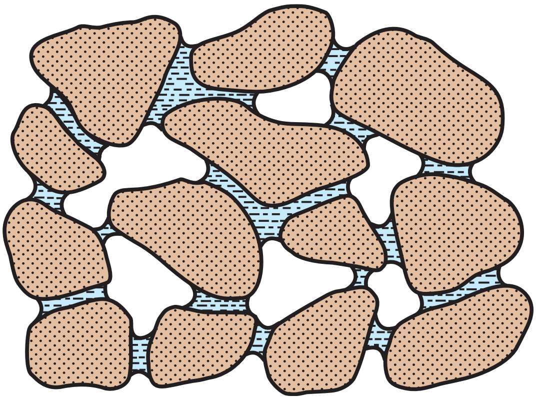


Fig. 8.1. Water in an unsaturated coarse-textured soil.

RELATION OF CONDUCTIVITY TO SUCTION AND WETNESS

Consider an unsaturated, horizontal body of soil in which water is flowing horizontally (i.e., in the absence of gravity gradients) in response to a suction gradient, as illustrated schematically in Fig. 8.2. In this model, the potential difference between the inflow and outflow ends is maintained not by different levels of positive hydrostatic pressure, but by different imposed suctions. In general, as the suction varies along the sample, so will the wetness and conductivity. If the suction head at each end of the sample is maintained constant, the flow rate will be steady but the suction gradient will vary along the sample's axis. Since the product of gradient and conductivity must be constant for steady flow, the hydraulic potential gradient must steepen as the conductivity diminishes with the increase in suction along the length of the sample. This phenomenon is illustrated in Fig. 8.3.

In view of the fact that the gradient along an unsaturated column is generally variable, we should not, strictly speaking, divide the flux by the overall ratio of the head drop to the distance ($\Delta H/\Delta x$) to obtain the conductivity. Rather, we should divide the flux by the local gradient at each point to evaluate the exact conductivity and its variation with suction. In the following treatment, however, we assume that the column of Fig. 8.2 is sufficiently short (and the hydraulic gradient is not too steep) to allow us to evaluate at least an average conductivity for the sample as a whole (i.e., $K = q\ \Delta x/\Delta H$).

The average negative head, or suction, acting in the column is

$$-H_{\text{average}} = \psi_{\text{average}} = -(H_1 + H_2)/2$$

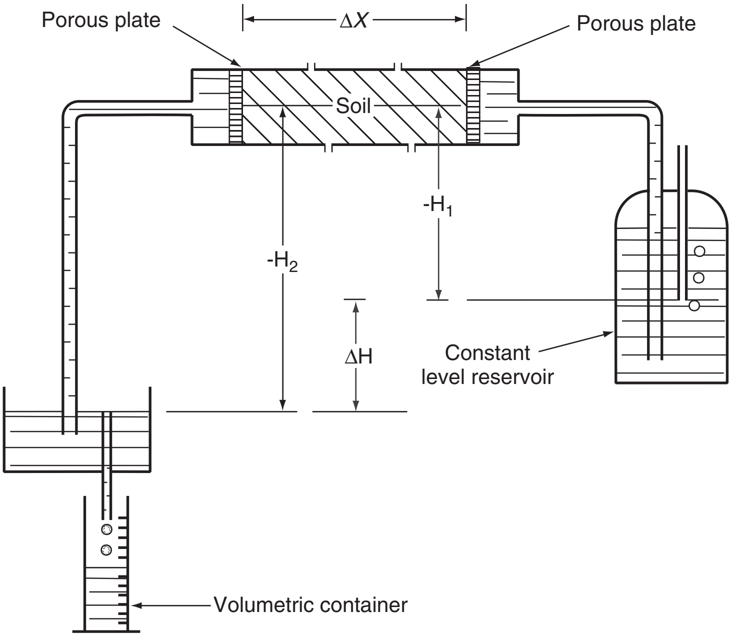


Fig. 8.2. A model illustraning unsaturated flow (under a suction gradient) in a horizontal column.

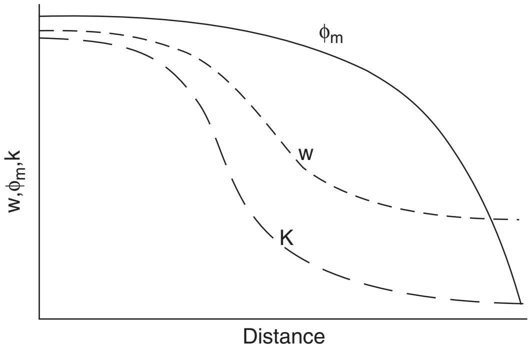


Fig. 8.3. The variation of wetness *w*, matric potential ϕ_m, and conductivity *K* along a hypothetical column of unsaturated soil conducting a steady flow of water.

We further assume that the suction everywhere exceeds the air-entry value so that the soil is unsaturated throughout.

Let us now make successive and systematic measurements of flux versus suction gradient for different values of average suction. The hypothetical results of such a series of measurements are shown schematically in Fig. 8.4. As in the case of saturated flow, we find that the flux is proportional to the gradient. However, the slope of the flux versus gradient line, being the hydraulic conductivity, varies with the average suction. In a saturated soil, by way of contrast, the hydraulic conductivity is usually independent of the water pressure.

Figure 8.5 shows the general trend of the dependence of conductivity on suction in soils of different texture. (Note that curves of *K* versus suction are usually drawn on a log–log scale, because both *K* and ψ vary over several orders of magnitude within the suction range of general interest, say, 0–100 m of suction head). It is seen that, although the saturated conductivity of the

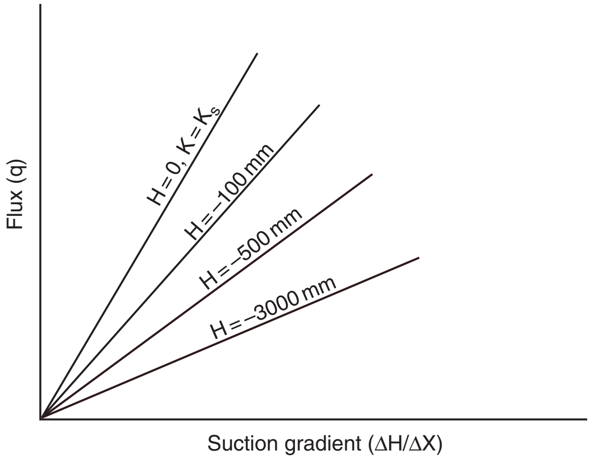


Fig. 8.4. Hydraulic conductivity, the slope of the flux versus gradient relation, depends on the average suction in an unsaturated soil.

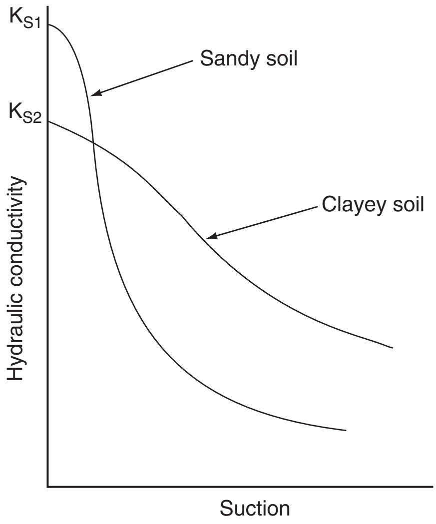


Fig. 8.5. Dependence of conductivity on suction in different soils (log–log scale).

sandy soil K_{s1} is typically greater than that of the clayey soil K_{s2}, the unsaturated conductivity of the former decreases more steeply with increasing suction and eventually becomes lower.

The steep decline of hydraulic conductivity with rising matric suction carries important implications regarding soil-water dynamics. It suggests that processes taking place in wet soil conditions are inherently faster than those occurring in drier soil conditions. Thus, the process of infiltration, during which water moves into the soil profile through a generally saturated surface zone, is much more rapid than evaporation, which typically involves the transfer of water from the interior of the soil to the atmosphere through a drying surface zone.

The fact that a sandy soil is more conductive than a clayey soil at low suction values but is less conductive in the higher suction range is the reason why a sandy soil absorbs water more rapidly during infiltration (while the surface is wetted to saturation or nearly so) but then does not sustain the evaporation process (during which the surface zone becomes desiccated) as can a clay soil.

Various empirical equations have been proposed for the relation of conductivity to suction or to wetness (e.g., Gardner, 1960b), including the following:

$$K(\psi) = a/\psi^n \tag{8.1a}$$

$$K(\psi) = a/(b + \psi^n) \tag{8.1b}$$

$$K(\psi) = K_s/[1 + (\psi/\psi_c)^n] \tag{8.1c}$$

$$K(\theta) = a\theta^n \tag{8.1d}$$

$$K(\theta) = K_s s^n = K_s(\theta/f)^n \tag{8.1e}$$

where K is the hydraulic conductivity at any degree of saturation (or unsaturation), K_s is the saturated conductivity of the same soil, a, b, and n are empirical constants (different in each equation), ψ is matric suction head, θ is volumetric wetness, s is the degree of saturation, and ψ_c is the suction head at which $K = K_s/2$. Note that $s = \theta/f$, where f is porosity.

In all of the equations, the most important parameter is the exponential constant, since it controls the steepness with which conductivity decreases with increasing suction or with decreasing water content. The n value of the first two equations is about 2 or less for clayey soils and may be 4 or more for sandy soils. For each soil, the equation of best fit and the values of the parameters must be determined experimentally.

Sample Problem

The hydraulic conductivity versus matric suction functions of two hypothetical soils (sandy and clayey loams) conform to the empirically based equation

$$K = a/[b + (\psi - \psi_a)^n \qquad \text{for} \qquad \psi \geq \psi_a$$

where K is hydraulic conductivity (cm/sec), ψ is suction head (cm H_2O) and ψ_a is air-entry suction, and a, b, n are constants, with a/b representing the saturated soil's hydraulic conductivity K_s. The exponential parameter n characterizes the steepness with which K decreases with increasing ψ. Assume that in the sandy soil $K_s = 10^{-3}$ cm/sec, $a = 1$, $b = 10^3$, $\psi_a = 10$ cm, and $n = 3$, whereas in the clayey soil $K_s = 2 \times 10^{-5}$, $a = 0.2$, $b = 10^4$, $\psi_a = 20$, and $n = 2$.

Plot the K versus ψ curves. Note that the curves cross, and the relative conductivities are reversed: The sandy soil with the higher K_s exhibits a steeper decrease of K and falls below the clayey soil beyond a certain suction value ψ_c. Calculate the values of ψ (designated $\psi_{1/2}$) at which each K equals $(1/2)K_s$, and estimate the common value of ψ_c at which the two curves intersect.

For the sandy soil:

$$K = [10^3 + (\psi - 10)^3]^{-1}$$

The suction $\psi_{1/2}$ at which $K = 0.5K_s = a/2b$ can be obtained by substituting $\tilde{\psi}$ for $(\psi - \psi_a)$ and setting $\tilde{\psi}^3 = 10^3$ (thus doubling the denominator). Therefore,

$$\tilde{\psi} = 10 \quad \text{and} \quad \psi_{1/2} = \tilde{\psi} + \psi_a = 10 + 10 = 20 \text{ cm}$$

For the clayey soil:

$$K = 0.2/[10^4 + (\psi - 20)^2]$$

Following the previous procedure, and setting $\tilde{\psi}^2 = 10^4$, we obtain

$$\tilde{\psi} = 10^2 \quad \text{and} \quad \psi_{1/2} = \tilde{\psi} + \psi_a = 100 + 20 = 120 \text{ cm}$$

Crossover suction value (ψ_c): We can attempt to obtain this value algebrically (by setting the two expressions for K equal to each other and solving for the common ψ value) or graphically (Fig. 8.6) by reading the ψ value at which the $K(\psi)$ curves intersect. The latter procedure is easier in this case, and it shows ψ_c to be about 48 cm.

(*Note*: The soils depicted are completely hypothetical and are not to be taken as typical of real sandy and clayey soils.)

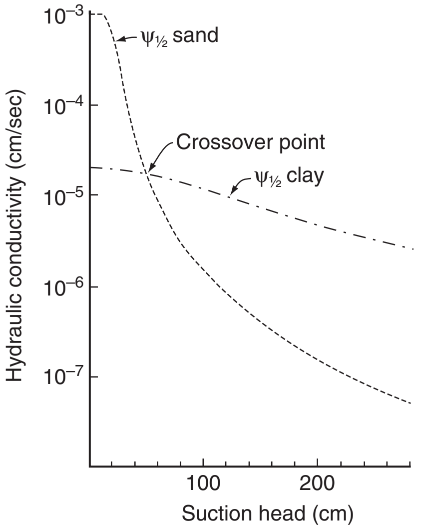


Fig. 8.6. Hydraulic conductivity of a sand and a clay versus suction.

EQUATIONS OF UNSATURATED FLOW

The Continuity Equation

Students may feel that the *equation of continuity*, based on the mass conservation law, is intuitively understandable or logically self-evident so that it requires no formal proof. It is a simple statement, in mathematical form, that for a conserved substance such as water (i.e., a substance that is neither created nor destroyed in the soil), the amount entering minus the amount exiting a specified volume of soil must be equal to the change in water content of the same body. For those who like formalism, however, we offer the following.

Consider a volume element of soil as a rectangular parallelepiped inside a space defined by the rectangular coordinates x, y, z (Fig. 8.7). Assume that the sides of the volume element are Δx, Δy, and Δz and that its volume is $\Delta x\ \Delta y\ \Delta z$. Now consider the net flow in the x direction. If the flux emerging from the right-hand face exceeds the flux entering the left-hand face by the amount $(\delta q/\delta x)\Delta x$, then the difference in discharge (volume per unit time) flowing through the two faces must be (discharge being equal to flux multiplied by area $\Delta y\ \Delta z$)

$$\text{Change in volume discharge} = q\ \Delta y\ \Delta z - [q + (\partial q/\partial x)\Delta x]\Delta y\ \Delta z \qquad (8.2)$$
$$\text{(net inflow rate)} = \text{(inflow rate minus outflow rate)}$$

The net inflow rate must equal the rate of gain of water by the volume element of soil per unit time:

$$\text{Net inflow rate} = -(\partial q/\partial x)\Delta x\ \Delta y\ \Delta z \qquad (8.3)$$

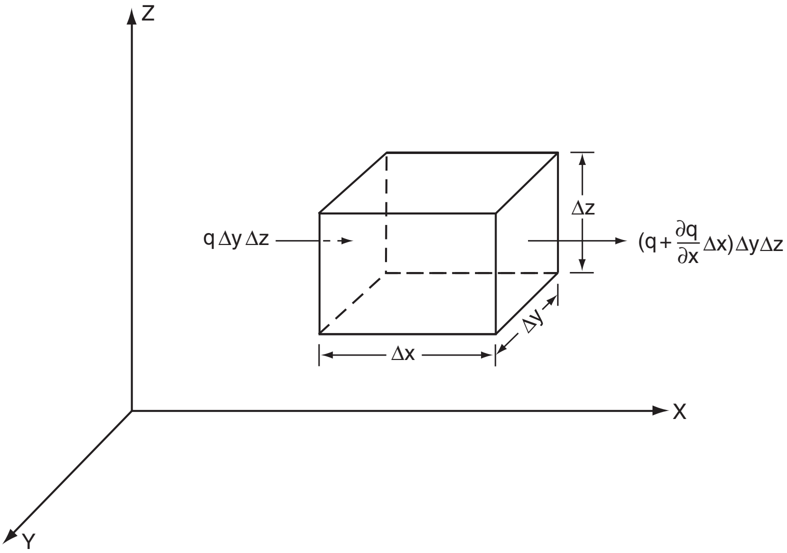


Fig. 8.7. The continuity principle: a volume element of soil gaining or losing water in accordance with the divergence of the flux.

The rate of gain of water by the volume element of soil can also be expressed in terms of the time rate of change of the volume concentration of water, θ, multiplied by the volume of the element:

$$\text{Rate of gain} = (\partial\theta/\partial t)\Delta x\ \Delta y\ \Delta z$$

Setting the two alternative expressions equal to each other, we get

$$(\partial\theta/\partial t)\Delta x\ \Delta y\ \Delta z = -(\partial q/\partial x)\Delta x\ \Delta y\ \Delta z$$

$$\text{or}\quad \partial\theta/\partial t = -\partial q/\partial x \tag{8.4}$$

If we also consider the fluxes in the y and z directions, we obtain the three-dimensional form of the continuity equation:

$$\frac{\partial\theta}{\partial t}\Delta x\Delta y\Delta z = -\left(\frac{\partial q_x}{\partial x}+\frac{\partial q_y}{\partial y}+\frac{\partial q_z}{\partial z}\right)\Delta x\Delta y\Delta z \tag{8.5}$$

$$\frac{\partial\theta}{\partial t} = -\left(\frac{\partial q_x}{\partial x}+\frac{\partial q_y}{\partial y}+\frac{\partial q_z}{\partial z}\right) \tag{8.6}$$

where q_x, q_y, q_z are the fluxes in the x, y, z directions, respectively.

In shorthand mathematical notation, the last equation is written as

$$\partial\theta/\partial t = -\nabla \cdot q \tag{8.7}$$

where the symbol ∇ (del) is the *vector differential operator*, representing the three-dimensional gradient in space (in our case it is the spatial gradient of the flux q). The scalar product of the del operator and a vector function is called the *divergence* and is designated *div*. An alternative expression of Eq. (9.7) is therefore

$$\partial\theta/\partial t = -\text{div}\ q \tag{8.8}$$

The Combined Flow Equation

Darcy's law, though originally conceived for flow in saturated porous media, has been extended to unsaturated flow, with the provision that the conductivity is now a function of the matric suction head [i.e., $K = K(\psi)$]:

$$q = -K(\psi)\ \nabla H \tag{8.9}$$

where ∇H is the hydraulic head gradient, which may include both suction and gravitational components. This equation, alons with its alternative formulations, is known as *Richards' equation*.

Equation (8.9) as written here fails to take into account the hysteresis of soil-water characteristics. In practice, the hysteresis problem can sometimes be evaded by limiting the use of this equation to cases in which the suction (or wetness) change is monotonic — either increasing or decreasing continuously. However, in processes involving successive wetting and drying phases, the $K(\psi)$ function may be highly hysteretic. However, the relation of conductivity to volumetric wetness $K(\theta)$ or to degree of saturation $K(S)$ is affected by hysteresis to a lesser extent than is the $K(\psi)$ function. Darcy's law for unsaturated soil can thus be written

$$q = -K(\theta)\ \nabla H \tag{8.10}$$

which, however, still leaves us the problem of dealing with the hysteresis between ψ and θ.

To account for transient flow processes, we introduce the continuity principle:

$$\partial\theta/\partial t = -\nabla \cdot q$$

Thus,

$$\partial\theta/\partial t = \nabla \cdot [K(\psi)\nabla H] \tag{8.11}$$

Remembering that the hydraulic head is, in general, the sum of the pressure head (or its negative, the suction head ψ) and the gravitational head (or elevation z), we can write

$$\partial\theta/\partial t = \nabla \cdot [K(\psi)\nabla(\psi - z) \tag{8.12}$$

Since ∇z is zero for horizontal flow and unity for vertical flow, we can rewrite Eg. (8.6) as follows:

$$\frac{\partial\theta}{\partial t} = -\nabla \cdot (K(\psi)\nabla\psi) + \frac{\partial K}{\partial z} \tag{8.12a}$$

or

$$\frac{\partial\theta}{\partial t} = -\frac{\partial}{\partial x}\left(K\frac{\partial\psi}{\partial x}\right) - \frac{\partial}{\partial y}\left(K\frac{\partial\psi}{\partial y}\right) - \frac{\partial}{\partial z}\left(K\frac{\partial\psi}{\partial z}\right) + \frac{\partial K}{\partial z} \tag{8.13}$$

BOX 8.1 Two-Phase Flow

Most theoretical treatments of water flow ignore the presence of the air phase in an unsaturated soil or at least imply that soil air does not hinder water movement. As mentioned, transient-state flow processes generally involve changes in the amount of water contained in a given volume element of soil. If the bulk density of the soil body does not change, then a net inflow of water implies expulsion of air, and an outflow of water implies entry of air. As long as the soil's air phase is continuous (i.e., not encapsulated in the form of isolated pockets or bubbles) and open to the external atmosphere, it presents no appreciable resistance to water flow. This is so because the fluidity of air (inversely related to the viscosity) is some 50 times greater than that of water.

However, when the free entry or escape of air is obstructed by some barrier (e.g., a saturated or tightly compacted layer with constricted pores), the air phase is confined and necessarily affects water movement. Under such conditions, the simultaneous flows of the two fluids should be considered jointly (Kutilek and Nielsen, 1994).

Processes may occur in which ∇z (the gravity gradient) is negligible compared to the strong matric suction gradient $\nabla\psi$. In such cases,

$$\partial\theta/\partial t = \nabla \cdot [K(\psi)\nabla\psi] \tag{8.14}$$

or, in a one-dimensional horizontal system,

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x}\left[K(\psi)\frac{\partial\psi}{\partial x}\right] \tag{8.15}$$

HYDRAULIC DIFFUSIVITY

Efforts have been made to simplify the mathematical treatment of unsaturated flow processes by casting the flow equations into a form analogous to the equations of diffusion and heat conduction, for which ready solutions are available (e.g., Carslaw and Jaeger, 1959; Crank, 1975), in some cases involving boundary conditions applicable to soil-water flow processes. It is sometimes possible to relate the flux to the water content (wetness) gradient rather than to the potential (suction) gradient. We illustrate the approach using the one-dimensional form of the flow equation, disregarding gravity.

The matric suction gradient $\delta\psi/\delta x$ can be expanded by the chain rule:

$$\frac{\partial\psi}{\partial x} = \frac{d\psi}{d\theta}\frac{\partial\theta}{\partial x} \tag{8.16}$$

Here $\partial\theta/\partial x$ is the wetness gradient and $\partial\psi/\partial\theta$ is the reciprocal of *specific water capacity*, $C(\theta)$:

$$C(\theta) = d\theta/d\psi \tag{8.17}$$

which is the slope of the soil-moisture characteristic curve at a particular value of wetness θ.

We can now rewrite the Darcy equation as follows:

$$q = K(\theta)\frac{\partial\psi}{\partial x} = -\frac{K(\theta)}{c(\theta)}\frac{\partial\theta}{\partial x} \tag{8.18}$$

To cast this equation into a form analogous to Fick's law of diffusion, a function is introduced called the *diffusivity*, D:

$$D(\theta) = K(\theta)/C(\theta) = K(\theta)(d\psi/d\theta) \tag{8.19}$$

D is thus defined as the ratio of the hydraulic conductivity K to the specific water capacity C, and since both are functions of soil wetness, so D must be also. To avoid confusion between the classical concept of diffusivity pertaining to the diffusive transfer of components in the gaseous and liquid phases (see Chapters 9 and 11 on solute movement and gas exchange in the soil) and this borrowed application of the same term to describe convective flow, we propose to qualify it with the adjective *hydraulic*. Here, therefore, we shall employ the term *hydraulic diffusivity* when referring to D of Eq. (8.19). We can now rewrite Eq. (8.10):

$$q = -D(\theta)\nabla\theta \tag{8.20}$$

or, in one dimension,

$$q = -D(\theta)(\partial\theta/\partial x) \tag{8.21}$$

which is mathematically identical to Fick's first law of diffusion.

Hydraulic diffusivity can thus be viewed as the ratio of the flux (in the absence of gravity and of hysteresis effects) to the soil-water content (wetness) gradient. As such, D has dimensions of length squared per unit time (L^2T^{-1}), since K has the dimensions of volume per unit area per time (LT^{-1}) and the specific water capacity C has dimensions of volume of water per unit volume of soil per unit change in matric suction head (L^{-1}). In the use of Eq. (8.21), the gradient of wetness is taken to represent, implicitly, a gradient of matric potential, which is the true driving force.

Introducing the hydraulic diffusivity into Eq. (8.15), for one-dimensional flow in the absence of gravity, we obtain

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x}\left[D(\theta)\frac{\partial\theta}{\partial x}\right] \tag{8.22}$$

which has only one dependent variable (θ) rather than the two (θ and ψ) of Eq. (9.15).

In the special case that the hydraulic diffusivity remains constant (though it is not generally safe to assume this except for a very small range of wetness), Eq. (8.22) can be written in the form of Fick's second diffusion equation:

$$\partial\theta/\partial t = D(\partial^2\theta/\partial x^2) \tag{8.23}$$

A word of caution is now in order. In employing the diffusivity concept and all relationships derived from it, we must remember that the process of liquid water movement in the soil is not one of diffusion but of *mass flow*, also termed *convection*. As we have already suggested, the borrowed term *diffusivity*, if taken literally, can be misleading. Furthermore, the diffusivity equations become awkward whenever the hysteresis effect is appreciable or where the soil is layered or in the presence of thermal gradients. Under such conditions, flow may not occur down a water-content gradient or bear any simple or consistent relation to that gradient, and it may actually be in the opposite direction to it.

Whenever the use of hydraulic diffusivity is appropriate, an important advantage to doing so is that its range of variation is generally much smaller than that of hydraulic conductivity. The maximum value of D found in practice is of the order of 1 m^2/day (10^4 cm^2/day). D generally diminishes to about 10^{-3}–10^{-4} m^2/day (1–10 cm^2/day) at the lower limit of wetness normally encountered in the root zone. It thus varies about a thousandfold rather than about a millionfold, as does the hydraulic conductivity in the same wetness range. Hence small changes in θ are likely to affect the value of $D(\theta)$ to a much small degree than that of $K(\theta)$. Conversely, the use of an approximate or "average" value of D is less likely to entail large errors in calculation of flux than the use of an inaccurate value of K.

To take account of gravity, a diffusivity equation can be written in the form

$$\frac{\partial\theta}{\partial t} = \nabla \cdot [D(\theta)\nabla\theta] + \frac{\partial K(\theta)}{\partial z} = \nabla \cdot [D(\theta)\nabla\theta] + \frac{dK}{d\theta}\frac{\partial\theta}{\partial z} \tag{8.24}$$

The relation of hydraulic diffusivity to wetness is shown in Fig. 8.8. Note that the right-hand section of the curve, shows a rise in diffusivity with wetness. In the very dry range, however, the diffusivity often indicates an opposite trend — a rise with diminishing soil wetness, apparently due to the contribution of vapor movement. In the very wet range, as the soil approaches complete saturation, the diffusivity becomes indeterminate [since $C(\theta)$ tends to zero]. The latter effect is due to the shape of the soil moisture characteristic, θ versus ψ, in the near-saturation range.

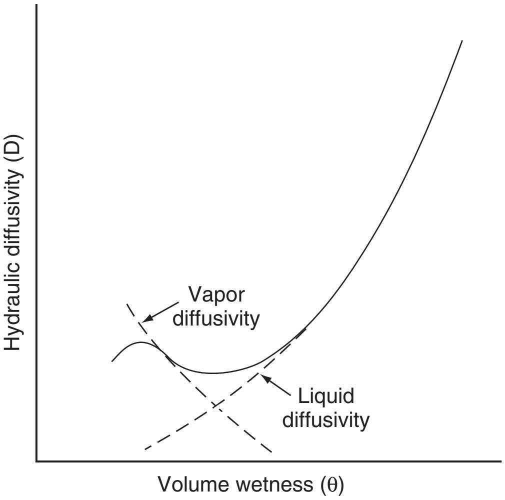


Fig. 8.8. Relation of diffusivity to soil wetness.

LABORATORY MEASUREMENT OF CONDUCTIVITY AND DIFFUSIVITY

Knowledge of the unsaturated hydraulic conductivity and diffusivity at different suction and wetness values is generally required before any of the mathematical theories of water flow can be applied in practice. Since there is as yet no universally accepted way to predict these values from more basic or easily obtainable soil properties, K and D must be measured experimentally. In principle, these parameters can be obtained from either *steady-state* or *transient-state* flow systems. In steady flow systems, the flux, gradient, and water content are constant in time, whereas in transient flow systems they vary. Hence measurements based on steady flow may seem to be more convenient and more accurate. The difficulty, however, lies in setting up a steady flow system, which may take a very long time to stabilize.

Techniques for measurement of conductivity and diffusivity of soil samples or models in the laboratory were described by Klute and Dirksen (1986). The conductivity is usually measured by applying a constant hydraulic head difference across the sample and measuring the resulting steady flux of water. Soil samples can be desaturated either by tension-plate devices or in a pressure chamber. The measurements are made at successive levels of suction and wetness so as to obtain the functions $K(\psi)$, $K(\theta)$, and $D(\theta)$. The $K(\psi)$ relation may be strongly hysteretic, so a complete determination requires measurements in desorption and in sorption (as well, perhaps, as in intermediate scanning). This is difficult, however, and requires specialized apparatus, so all too often only the desorption curve is measured (starting at saturation and proceeding to increase the suction in increments).

Such laboratory techniques can be applied to the measurement of undisturbed soil cores taken from the field. This is certainly preferable to measurements taken on fragmented and artificially packed samples, though it should be understood that no field sampling technique yet available provides truly undisturbed samples. Moreover, any attempt to represent a field soil by means of extracted samples incurs the problem of field soil heterogeneity as well as the associated problem of determining the appropriate scale (i.e., the representative volume) for realistic measurement of parameters.

A widely used transient flow method for measurement of conductivity and diffusivity in the laboratory is the outflow method. It is based on measuring the diminishing rate of outflow from a sample in a pressure cell after the pressure is raised by a certain increment. In the application of this method, the *hydraulic resistance* of the porous plate or membrane on which the sample is placed must be taken into account. Laboratory measurements of conductivity and diffusivity can also be made on long columns of soil, not only on small samples contained in cells. In such columns, steady-state flow can be induced by evaporation or by infiltration. If the column is long enough to allow the measurement of suction gradients (e.g., by a series of tensiometers) and of wetness gradients (by sectioning or, preferably, by a nondestructive technique such as gamma-ray scanning), the $K(\theta)$ and $K(\psi)$ relationships can be obtained for a range of θ values with a single column or with a series of similarly packed columns.

IN SITU MEASUREMENT OF UNSATURATED HYDRAULIC CONDUCTIVITY

Application of the theories of soil physics to the description or prediction of actual processes in the field (processes involved in infiltration, runoff control, irrigation, drainage, water conservation, groundwater recharge and pollution, etc.) depends on knowledge of soil hydraulic characteristics, including the functional relation of hydraulic conductivity and of matric suction to soil wetness, as well as the their spatial and temporal variation.

It seems fundamentally unrealistic to measure the unsaturated hydraulic conductivity of field soil by making laboratory determinations on discrete samples removed from their natural continuum. Such samples are generally dried, fragmented, and repacked into experimental containers, so the original structure is destroyed. Hence it is necessary to devise and test practical methods for measuring soil hydraulic conductivity on a realistic scale in situ. We now proceed to give a brief description of several of the methods available for this purpose. A comprehensive review was given by Green et al. (1986) and more recently by Clothier and Scotter (2002).

Sprinkling Infiltration

The principle of this method is that a continued supply of water to the soil at a constant rate lower than the saturated hydraulic conductivity of the soil eventually establishes a steady moisture distribution in the conducting profile. Once steady-state conditions are established, a constant flux exists. In a uniform soil the suction gradients will tend to zero, and with only a unit gravitational gradient in effect the hydraulic conductivity becomes effectively equal to the flux. This test can be performed on an initially dry soil to which successively increasing sprinkling intensities are applied incrementally. It then becomes possible to obtain different values of hydraulic conductivity corresponding to different values of soil wetness.

The difficulty of the steady sprinkling infiltration test in the field is that it requires rather elaborate equipment, which must be maintained in continuous operation for long periods of time. The requirement of maintaining continuous operation becomes increasingly important, and difficult, as one attempts to extend the test toward the greater suction range by reducing the application rate below 1 mm/hr. Another difficulty is to avoid the raindrop impact effect, which can cause the exposed surface soil to disperse and seal, thus reducing infiltrability. This problem can be avoided by mulching the soil surface with straw. A more serious problem is the possible presence at some depth of soil layers that might impede flow, thus preventing the attainment throughout the profile of a condition of a zero matric suction gradient. Ideally, therefore, this test applies to uniform (or nearly uniform) soil profiles rather than to distinctly layered ones.

Infiltration Through an Impeding Layer

This method was first suggested by Hillel and Gardner (1970a), who showed that an impeding layer at the surface of the soil can be used to achieve the desired boundary conditions for measuring the underlying soil's unsaturated hydraulic conductivity and diffusivity as functions of soil wetness. The effect of an impeding layer present over the top of the profile during infiltration is to decrease the hydraulic potential in the profile under the impeding layer. Thus, the soil wetness and, correspondingly, the conductivity and diffusivity values of the infiltrating profile are reduced.

When the surface is covered by an impeding layer ("crust") with a saturated conductivity smaller than that of the underlying soil, steady flow conditions are eventually established. At that stage, the hydraulic head gradient in the subcrust soil is unity (i.e., gravitational only), while the head gradient through the impeding layer (owing to its low conductivity) is necessarily greater than unity. If the ponding depth of water over the surface is negligible, such an impeding layer induces the development of suction in the sublayer, the magnitude of which increases with increasing hydraulic resistance of the crust. After steady infiltration is achieved, the conductivity of the subcrust soil becomes equal to the flux.

A variant of the crust method is the use of a so-called *tension infiltrometer* (Elrick and Reynolds, 1992; Warrick, 1992; Bootlink and Bouma, 2002). It consists of a reservoir of water designed to maintain a constant subatmospheric pressure by means of a Mariotte tube, at the bottom of which is a fixed a porous plate. That plate introduces hydraulic resistance (analogous to a surface crust of the soil itself), and the hydraulic gradient through the plate when it is placed on the soil surface creates a suction at its contact place with the soil. This instrument has been used in soils with macropores to assess their potential contribution to the process of infiltration under ponding (see Chapter 14).

A serious problem that may invalidate the crust test in some cases pertains to the so-called *unstable flow* phenomenon. It has been observed that in transition from a fine-textured zone to a coarse-textured zone during infiltration, the advance of the water may not be even (or planar, as it tends to be in a uniform profile), but that sudden breakaway flows may occur in specific locations where fingerlike intrusions take place. This phenomenon is discussed in Chapter 14.

Internal Drainage

This method is based on monitoring the transient-state internal drainage of a profile in the field (Fig. 8.9). The method can help to minimize the alteration of soil hydraulics that can result from methods that disrupt soil structure. The method requires frequent and simultaneous measurements of the soil wetness and matric suction profiles under conditions of drainage alone (evapotranspiration prevented). The use of strain gauge pressure transducers in tensiometry facilitate the rapid and automatic acquisition of suction data, while the required soil wetness data are obtainable with a neutron moisture gauge or a time-domain reflectometer. From these measurements it is possible to obtain instantaneous values of the potential gradients and fluxes operating within the profile and hence also of hydraulic conductivity values. Once the hydraulic conductivity at each elevation within the profile is known in relation to wetness, the data can be applied to the analysis of drainage and evaporation in a bare field or of evapotranspiration in a vegetated field.

To apply this method in the field, one must choose a characteristic fallow plot that is large enough (say, at least 5 × 5 m) so that processes at its center are

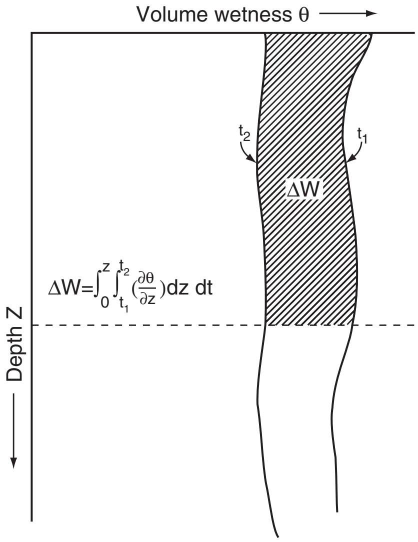


Fig. 8.9. Change in profile water content ΔW in the depth interval $Z = 0$ (soil surface) to depth Z in the time interval t_1 to t_2 during internal drainage, assuming no evaporation or lateral flow.

unaffected by the boundaries. Within this plot, at least one neutron access tube is installed through and below the root zone. The desirable depth will sometimes exceed 2 m. A series of tensiometers is installed at various depths near the access tube so as to represent profile horizons. (The depth interval between succeeding tensiometers should not exceed 30 cm.) Water is then ponded on the surface and the plot is irrigated long enough so that the entire profile becomes as wet as it can be. After this, the soil surface is covered by an opaque plastic sheet so as to minimize soil heating and to prevent evaporation from the soil surface. As internal drainage proceeds, periodic measurements are made of distribution and tension of soil moisture throughout the profile. The handling of the data and a sample calculation of the hydraulic conductivity were described by Hillel et al. (1972) and more recently by Vachaud and Dane (2002).

The method as described is not applicable where lateral movement of soil moisture is appreciable. This movement is not normally significant when the soil profile is unsaturated, but lateral movement can become significant wherever an impeding layer occurs on which saturated conditions might prevail for some time. In practice, the moisture range for which conductivity can be measured by the internal drainage method is generally limited to suctions not exceeding about 0.5 bar, since the drainage process often slows down within a few days or weeks to become practically imperceptible thereafter.

VAPOR MOVEMENT

As already stated, liquid water moves in the soil by mass flow, a process by which the entire fluid flows in response to differences in hydraulic potential. In certain circumstances, the movement of water vapor can also occur as mass flow, for instance, when wind gusts induce bulk movement of air (containing vapor) in the surface zone of the soil. In general, however, vapor movement within the soil profile occurs by diffusion, a process in which different components of a mixed fluid move independently, and at times in opposite directions, in response to differences in concentration (or partial pressure) from one location to another. Water vapor is always present in the gaseous phase of an unsaturated soil, and vapor diffusion occurs whenever differences in vapor pressure develop within the soil.

The diffusion equation for water vapor is

$$q_v = -D_v(\partial \rho_v / \partial x) \tag{8.25}$$

where ρv is the vapor density (or concentration) in the gaseous phase and *D*v is the diffusion coefficient for water vapor. *D*v in the soil is lower than in open air because of the restricted volume and the tortuosity of air-filled pores (Currie, 1961).

By considering that the liquid water in the soil serves as both a source and a sink for water vapor, and assuming that changes in liquid water content with time are much greater than the changes in vapor density with time, Jackson (1964a) described nonsteady vapor transfer in terms of the liquid water content (volumetric wetness) θ:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left[D_v \frac{\partial \rho_v}{\partial \theta}\frac{\partial \theta}{\partial x}\right] \tag{8.26}$$

For the simultaneous transfer of both liquid and vapor, the following equation applies:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left[\left(D_v \frac{\partial \rho_v}{\partial \theta} + D_\theta\right)\frac{\partial \theta}{\partial x}\right] \tag{8.27}$$

where D_θ is the hydraulic diffusivity for liquid water.

The foregoing equations consider water vapor diffusion as an isothermal process, assuming that both viscous flow in the liquid phase and diffusion of vapor are impelled by the forces of adsorption and of capillarity. No explicit account is thus taken of osmotic or solute effects on vapor pressure, though such effects can sometimes be significant. More importantly, this approach disregards the simultaneous and interactive transport of both water and heat in nonisothermal situations, to be described in a subsequent chapter.

At constant temperature, vapor-pressure differences in a nonsaline soil are likely to be very small. For example, at standard temperature a change in matric suction from 0 to 10 MPa (100 bar) is accompanied by a vapor pressure reduction from 17.54 to 16.34 torr, i.e. a difference of only 1.2 torr (1.6×10^{-4} MPa, or 1.6 mbar). (The torr is a unit of pressure equal to 1 mm of mercury under standard conditions.) Under normal field conditions (except at extreme dryness) soil air is nearly vapor saturated at almost all times.

When temperature differences do occur, they can cause considerable differences in vapor pressure. For example, a change of 1 degree in water temperature, from 19 to 20°C, results in an increase in vapor pressure of 1.1 torr.

Thus, a change in temperature of just 1°C has nearly the same effect on vapor pressure as a change in suction of 10 MPa (100 bar)!

In the range of temperatures prevailing in the field, the variation of saturated vapor pressure (that is, the vapor pressure in equilibrium with pure, free water) is as follows:

Temperature (°C)	0	20	30	40
Vapor pressure (torr)	4.6	17.5	38.0	55.8

Vapor movement tends to take place from warmer to cooler zones in the soil. Since the soil surface is warmer during the daytime and colder during the night than the deeper layers, vapor movement tends to be downward during the day and upward during the night. Temperature gradients can also induce liquid flow, because higher temperatures tend to reduce matric suction and thus can induce hydraulic potential gradients. As commonly observed, liquid flow is the dominant mode in moist, nearly isothermal soils; hence the contribution of vapor diffusion to overall water movement is probably negligible in the main part of the root zone, where diurnal temperature fluctuations are slight.

9. SOLUTE MOVEMENT AND SOIL SALINITY

THE SOIL SOLUTION

In the preceding chapters, we discussed soil water with only occasional reference to solutes. However, water present in the soil and constituting its liquid phase is never chemically pure. In the first place, the water entering the soil as rain or irrigation is itself a solution. Rainwater is of course distilled and essentially pure when it first condenses to form clouds, but as it descends through the atmosphere it generally dissolves such atmospheric gases as carbon dioxide and oxygen, often together with such products of our industrial civilization as oxides of sulfur and nitrogen, as well as — along the coast — appreciable quantities of salt that enter the air as sea spray. Solute concentrations (mostly electrolytic salts) typically found in rainwater are of the order of 5–20 mg/L (units roughly equivalent to mg/kg, also designated parts per million, or ppm).

During its residence in the soil, the infiltrated water tends to dissolve additional solutes, including soluble products of mineral and organic matter decomposition as well as fertilizers and pesticides. The concentration of salts in irrigation water (generally obtained from surface reservoirs or subterranean aquifers) may range between 100 and 1000 mg/L. Finally, the concentration may be as high as 10,000 mg/L in drainage from saline soil. (*Note*: The concentration of ocean water is of the order of 35 grams per liter.)

As it moves through the profile, soil water carries its solute load in its convective stream, leaving some of it behind to the extent that the component salts are adsorbed, taken up by plants, or precipitated whenever their concentration exceeds their solubility (mainly at the soil surface during evaporation). Solutes move not only with soil water, but also within it, in response to concentration gradients. At the same time, solutes react among themselves and interact with the solid matrix of the soil in a continuous cyclic succession of interrelated physical and chemical processes. These interactions involve and are strongly influenced by such variable factors as acidity, temperature, oxidation–reduction potential, composition, and concentration of the soil solution.

In former times, processes involving solutes were considered to belong in the exclusive realm of soil chemistry and outside the scope of a treatise on soil physics. Nowadays, however, we have come to recognize that arbitrary separations among "traditional" disciplines that are in fact complementary and overlapping should not be allowed to hinder our quest for an ever more comprehensive understanding of interactive phenomena in the environment.

A better understanding of the simultaneous movement and interactions of water and solutes in the soil is essential to the improvement of soil fertility through control of nutrients in the root zone as well as to the prevention of soil salinity, alkalinity, and other occurrences of toxicity. Such an understanding has also become crucial in the area of environmental management whenever soil-borne solutes migrate to, and affect the quality of, groundwater or surface water resources (Hendrickx et al., 2002).

CONVECTIVE TRANSPORT OF SOLUTES

The convection (or *mass flow*) of soil water, sometimes called the *Darcian flow*, carries with it a flux of solutes J_c proportional to their concentration c:

$$J_c = qc = -c[K\ (dH/dx)] \tag{9.1}$$

where $q = -K(dH/dx)$ is Darcy's law, discussed in the preceding two chapters. Since q is usually expressed as volume of liquid flowing through a unit area (perpendicular to the flow direction) per unit time and c as mass of solute per unit volume of solution, J_c is given in terms of mass of solute passing through a unit cross-sectional area of soil per unit time.

To estimate the distance of travel of a solute per unit time, we consider the *average apparent velocity* $\bar{v}$ of the flowing solution:

$$\bar{v} = q/\theta \tag{9.2}$$

where θ is volumetric wetness and $\bar{v}$ is taken as the straight-line length of path traversed in the soil in unit time. In this formulation we disregard the roundabout path caused by the geometric tortuosity of the water-filled soil pores. Since actual velocities vary over several orders of magnitude within pores and between pores, the concept of *average velocity* is obviously a gross approximation (much like lumping insects and elephants in computing an "average" size animal). However, assuming $\bar{v}$ to be a working approximation, we get

$$J_c = \bar{v}\theta c \tag{9.3}$$

It is sometimes useful to have an estimate of the average *residence time* t_r of a solute within a layer of soil of thickness L (especially if we are concerned with some time-dependent interactive process involving the solute under consideration). Accordingly,

$$t_r = L/\bar{v} \tag{9.4}$$

If the flow is impelled by gravity alone, with no pressure gradients, then the downward flux of liquid is equal to the hydraulic conductivity K of the medium, a function of wetness θ:

$$q = K(\theta)$$

Using Eq. (9.2), we thus obtain for Eq. (9.4)

$$t_r = L\ \theta/K(\theta) \tag{9.5}$$

The foregoing equations allow us to estimate, for instance, the distance of travel of a soluble pollutant from the bottom of, say, a septic tank or sanitary landfill to the water table, through the so-called *unsaturated zone*, as follows:

$$L_t = tK(\theta)/\theta \tag{9.6}$$

where L_t is the average distance of convective transport in time t. If the values of wetness and hydraulic conductivity vary within the soil profile, as often happens, the foregoing calculations must be carried out layer by layer to determine the solution's space-variable flux, average distance of travel, and per-layer residence time.

The serious shortcoming of this approach is that one cannot be sure that the transport of solutes occurs by convection alone. In fact, we know that solutes do not merely move with the water as sedentary passengers in a train, but also move within the flowing water in response to concentration gradients in the twin processes of *diffusion* and *hydrodynamic dispersion*. Moreover, solutes are not always as inert as we have thus far assumed, because they tend to interact with the biological system within the soil (e.g., to be taken up or released by microbes and the roots of plants) and with the physicochemical system (e.g., to be adsorbed or exchanged within or by the soil). Additionally, solutes may undergo chemical reactions and may also be removed from solution by precipitation or volatilization.

Sample Problem

A soluble pollutant was inadvertently spilled on the ground. Suppose that it is nondegradable, nonvolatile, not taken up by plants, not adsorbed by the soil, and not immobilized by any other mechanism. If the annual rainfall is 1500 mm, the annual evapotranspiration is 1250 mm, the water table is 20 m deep, and the so-called unsaturated zone underlying the soil has a constant volumetric wetness of 25%, estimate the residence time in the unsaturated zone and the time required for the pollutant to reach the groundwater.

As a rough approximation, we assume that the solute is transported only by the convective stream of water draining out of the soil, vertically downward through the unsaturated zone toward the water table. The possible effects of diffusion and dispersion are thus disregarded. We further assume that this drainage occurs under steady-state conditions, that is, that temporal perturbations at the soil surface resulting from intermittent periods of rainfall and evapotranspiration are damped out in the upper layer of the soil.

To estimate the distance of travel of a solute per unit time, we take the average velocity $\bar{v}$ of the flowing solution to be [Eq. (9.2)]

$$\bar{v} = q/\theta$$

where $\bar{v}$ is the straight-line length of path traversed through the soil or subsoil by the solution in unit time, θ is the volumetric wetness, and q is the flux (volume flowing through unit area per unit time). Substituting the given values we obtain

$$\bar{v} = (1500 - 1250)/0.25 = 1000 \text{ mm/yr}$$

To estimate the residence time t_r of the solute within the unsaturated zone, we refer to Eq. (9.4):

$$t_r = L/\bar{v}$$

where L is the thickness of the zone considered. Thus,

$$t_r = 20 \text{ m}/1 \text{ m/yr} = 20 \text{ y}$$

Thus, the bulk of the solute can be expected to reach the water table and enter the groundwater in about 20 years. Since in fact diffusion and dispersion phenomena do operate, we can expect some of the solute to move faster than the bulk of the solution and some to lag behind. However, the calculation based on convective transport alone can serve as a useful rough estimation of the movement rate.

DIFFUSION OF SOLUTES

Diffusion processes commonly occur within multicomponent gaseous or liquid phases, in consequence of the random thermal motion (often called *Brownian motion*) and repeated collisions and deflections of molecules in the fluid. The net effect is a tendency toward equalizing the spatial distribution of the components in a nonhomogeneous fluid.

Diffusion processes are very important in the soil. As we show in Chapter 11 on soil aeration, diffusion in the air phase of such gases as oxygen, carbon dioxide, nitrogen, and water vapor can have a decisive influence on the soil's chemical and biological processes. Equally important are diffusion processes involving solutes in the soil's liquid phase, including nutrients as well as salts and toxic compounds that affect plant growth.

Wherever solutes are not distributed uniformly throughout a solution, concentration gradients exist. Consequently, solutes tend to diffuse from zones where their concentration is higher to where it is lower. In bulk water at rest, the rate of diffusion J_d is related by *Fick's first law* to the gradient of the concentration c:

$$J_d = -D_0(dc/dx) \tag{9.7}$$

in which D_0 is the diffusion coefficient for a particular solute diffusing in bulk water and dc/dx is that solute's effective concentration gradient.

For diffusion in the soil's liquid phase, the effective diffusion coefficient is generally less than the diffusion coefficient in bulk water, D_0, for several reasons. In the first place, the liquid phase occupies only a fraction of the soil volume; at most, in a state of saturation, its volume fraction equals the soil's porosity. Secondly, the soil's pore passages are tortuous, so the actual path length of diffusion is significantly greater than the apparent straight-line distance. In an unsaturated soil, as soil wetness diminishes, the fractional volume available for diffusion in the liquid phase decreases still further and the tortuous length of path increases.

If the sole factors affecting the diffusion coefficient in the soil (D_s) are fractional water volume θ and *tortuosity* ξ, we can write

$$D_s = D_0\theta\xi \tag{9.8}$$

where ξ, the tortuosity factor, is an empirical parameter smaller than unity, expressing the ratio of the straight-line distance to the average roundabout-path length through the water-filled pores for a diffusing substance. This parameter has been found to depend on both the fractional volume and the geometric configuration of the water phase and, hence, to decrease with decreasing θ. Thus, D_s is strongly dependent on θ, both directly and through its dependence on ξ, which itself is a function of θ. To show this dependence, we write $D_s(\theta)$.

Other factors in addition to the geometric ones considered in Eq. (9.8) also tend to decrease the effective diffusion coefficient, particularly in unsaturated soils with an appreciable content of clay. As soil wetness is reduced and the water films coating the particles contract, the increasing density of the exchangeable cations adsorbed to the clay surfaces and the corresponding exclusion of anions, as well as the possible increase in viscosity of the adsorbed liquid phase, might combine to further retard diffusion. Because these and other complicating factors are not mutually independent, it has seemed impossible thus far to formulate them separately. Hence, it is tempting to lump them all together into a single *complexity factor*, which we can designate α. Thus,

$$D_s = D_0\theta\alpha \tag{9.9}$$

We can now rewrite Eq. (9.7) for diffusion in the liquid phase of an unsaturated soil:

$$J_d = -D_s(\theta)(dc/dx) \tag{9.10}$$

By itself, this equation can only describe *steady-state diffusion processes*. In order to proceed to a more generalized diffusion equation capable of describing *transient-state processes* (in which the rate and concentration vary in time), we must (as in the case of water flow alone) invoke the *mass conservation law*, expressed in the *continuity equation*.

Let us assume that there are no sources or sinks for the diffusing solute in the soil body in which diffusion is occurring. Now let us consider a rectangular volume element of soil that contains a liquid phase and is bounded by two parallel square planes, of area A, separated by a distance Δx. If diffusion takes place from left to right, the amount of solute diffusing through the left-hand plane into the volume element per unit time is AJ_d, and the amount diffusing out through the right-hand plane is $A[J_d + (\delta J_d/\delta x)\Delta x]$. The rate of accumulation of the solute in the volume element is $A(\delta c/\delta t)\Delta x$, where $\delta c/\delta t$ is the time rate of change of concentration. Thus,

$$A(\partial c/\partial t)\Delta x = A[J_d + (\partial J_d/\partial x)\Delta x] - AJ_d$$

which reduces to

$$\partial c/\partial t = -\partial J_d/\partial x \tag{9.11}$$

The negative sign indicates that any increase of diffusive flux along the direction of diffusion from the entry face to the outlet face of the volume element necessarily depletes the concentration in the volume element; and vice versa (the concentration increases if the rate of out-diffusion is less than the in-diffusion).

Combining this with Eq. (9.10), we obtain a second-order equation as follows:

$$\partial c/\partial t = \partial[D_s(\theta)(\partial c/\partial x)]/\partial x \tag{9.12}$$

In the special case when the diffusion coefficient D_s is constant, Eq. (9.12) assumes the classical form of *Fick's second law of diffusion*:

$$\partial c/\partial t = D_s(\partial^2 c/\partial x^2) \tag{9.13}$$

However, D_s is generally not constant and in fact might be dependent not only on wetness but also on concentration, i.e., be a function of both θ and c.

Attempts to apply the foregoing equations to describe the diffusion of solutes in the soil solution, particularly in unsaturated soil conditions, have difficulties. The soil varies both in space and time. Solutes interact with, and modify, the solid matrix — and hence may affect the pore space. Different species of solutes in the liquid phase interact with one another as well as with the adsorbed phase. Evaporation and condensation of water in pores further modify the concentration gradients and the pattern of diffusion. Finally, the convective flow of the solution also affects the diffusion process by changing the distribution of solutes and by inducing a process called *hydrodynamic dispersion* (Skaggs and Leij, 2002).

HYDRODYNAMIC DISPERSION

The motion of any inhomogeneous solution in a porous body brings about another process that differs from diffusion in its mechanism but that tends to produce an analogous or synergetic tendency to mix and eventually to even-out the differences in concentration or composition among different portions of the flowing solution. This process, which at times predominates over diffusion, is called *hydrodynamic dispersion*. It results from the microscopic nonuniformity of flow velocity in the soil's conducting pores. Since water moves faster in wide pores than in narrow ones and faster in the center of each pore than along its walls, some portions of the flowing solution move ahead while other portions lag behind.

Consider, for example, the laminar flow of a solution through a single capillary pore hypothetically shaped like a cylindrical tube. From our earlier derivation of Poiseuille's law (Chapter 7) we know that flow velocity v in such a tube is a decreasing function of radial distance r from the center:

$$v = 2\bar{v}(1 - r^2/R^2) \tag{9.14}$$

Here $\bar{v}$ is average velocity and R is the radius of the tube. Thus the velocity of a solute molecule carried by the convective stream depends on its position within the pore passage. At $r = R$, which is at the wall of the tube, the velocity is zero, whereas at $r = 0$, in the center of the tube, the velocity is maximal and equal to twice the average velocity.

Added to the nonuniformity of velocity within each pore is the fact that pores vary widely in radius over several orders of magnitude, say from 1 μm to 1 mm. We recall Poiseuille's law [Eq. (7.3)],

$$Q = 4R^4(\Delta p/8\eta L)$$

which states that the discharge Q (volume flow per unit time) is proportional to the fourth power of the radius R. Hence a pore with an effective radius of 1 mm will conduct a volume of water $(10^3)^4 = 10^{12}$ (a million million) times greater than a pore having an effective radius of 1 μm. Great indeed is the microscopic-scale variation of pore water velocity in the soil!

The fact that some portions of a flowing solution move faster than other portions causes an incoming solution to mix or disperse within the antecedent solution. The degree of mixing depends on such factors as average flow velocity, pore size distribution, degree of saturation, and concentration gradients. When the convective velocity is sufficiently high, the relative effect of hydrodynamic dispersion can greatly exceed that of molecular diffusion and the latter can be neglected in the analysis of solute movement. On the other hand, when the soil solution is at rest, hydrodynamic dispersion does not come into play at all.

Mathematically, hydrodynamic dispersion can be formulated in a manner analogous to diffusion, per Eqs. (9.10) and (9.12), except that a *dispersion coefficient* D_h is used instead of a diffusion coefficient. D_h has been found (Bresler, 1972b) to depend more or less linearly on the average flow velocity $\bar{v}$. Thus

$$D_h = a\bar{v} \tag{9.15}$$

with a an empirical parameter.

Because of the similarity in effect (though not in mechanism) between diffusion and dispersion, it is tempting to assume the two effects to be additive. Accordingly, the diffusion and dispersion coefficients are often combined into a single term, called the *diffusion–dispersion coefficient* (D_{sh}, which is a function of both the fractional water volume θ and the average velocity $\bar{v}$:

$$D_{sh}(\theta, \bar{v}) = D_s(\theta) + D_h(\bar{v}) \tag{9.16}$$

MISCIBLE DISPLACEMENT AND BREAKTHROUGH CURVES

When a liquid different in composition or concentration from the preexisting pore liquid is introduced into a column of soil, and the outflow from the end of the column is collected and analyzed, its composition is seen to change in time as the old liquid is displaced and replaced by the new one. If the two liquids are not mutually soluble (as is the case, for example, with oil and water), then the process is called *immiscible displacement*. If, on the other hand, the two liquids mix readily — as do many aqueous solutions — the process just described is referred to as *miscible displacement*. Plots of the outflowing solution's composition versus time or versus cumulative discharge are called *breakthrough curves*.

If the soil were saturated throughout the experiment and if neither diffusion nor hydrodynamic dispersion were to take place at the boundary between the displacing and the displaced solutions, then that boundary would remain a sharp front moving along the length of the column at a velocity equal to the flux. If we were to monitor the composition at the column's outlet, we would notice an abrupt change in composition at the moment the last portion of the old solution was completely driven out and the new solution arrived. Such a pattern of displacement without mixing is commonly called *piston flow*. It is seldom if ever encountered in practice. What normally happens at the boundary or front between the two solutions is a gradual mixing resulting from both diffusion and hydrodynamic dispersion, so that the boundary becomes increasingly diffuse about the mean position of the advancing front, as shown in Figs. 9.1 and 9.2.

Ideally, breakthrough curves should be sigmoidal in shape and symmetrical about the front of the advancing solution, with the inflection representing 50% displacement at a cumulative flow of one "pore volume" if the soil is saturated. A typical curve of this sort is shown in Fig. 9.2 for a sand. Curves obtained with finer-textured soils differ from the symmetrical ideal, particularly in the case of structured (aggregated) soils, owing to various interactions between solutes and the soil matrix (e.g., positive or negative adsorption) and to the possible existence of pore spaces where water movement is so sluggish that portions of the antecedent solution lag behind and are displaced only very slowly.

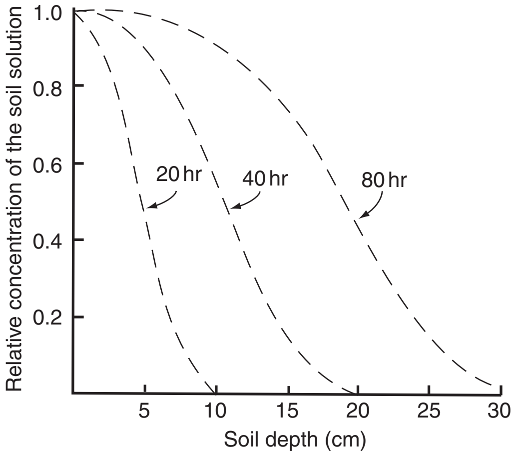


Fig. 9.1. Solute concentration versus soil column depth at various times during the infiltration of a saline solution into a nonsaline soil. Note that the concentration "front" becomes increasingly diffuse about its mean position as it advances in the soil.

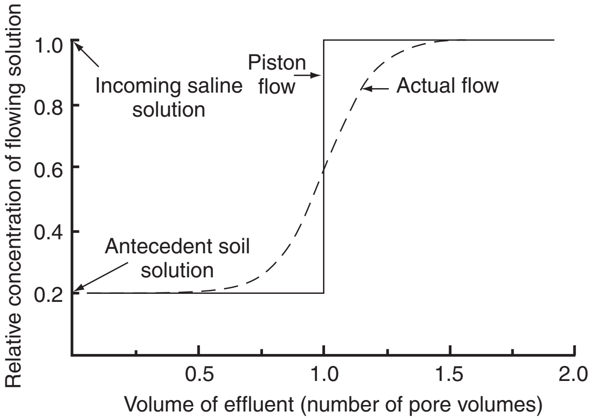


Fig. 9.2. Schematic illustration of a sigmodial breakthrough curve in a saturated sandy soil during the replacement of a dilute antecedent solution by an entering saline solution of fivefold greater concentration.

Miscible displacement phenomena and breakthrough curves are not merely of theoretical interest but are indeed relevant to the solution of many real-world problems, such as the leaching of excess salts from saline soils, the distribution of nutrient solutions, and the pollution of groundwater by the migration of soil-borne solutes of various types, including radioactive wastes, toxic chemicals, and agrochemical (pesticide and fertilizer) residues.

COMBINED TRANSPORT OF SOLUTES

To take into account the three mechanisms of solute movement described thus far [mass flow (convection), molecular diffusion, and hydrodynamic dispersion], we combine Eqs. (9.3), (9.10), and (9.16):

$$J = \bar{v}\theta c - D_s(\theta)(dc/dx) + D_h(\bar{v})(dc/dx)$$

which can be stated in words:

[combined solute flux] = [flux due to convection] + [flux due to diffusion]
+ [flux due to hydrodynamic dispersion]

Since in practice the diffusion and dispersion phenomena cannot be separated, the foregoing equation is usually written in the form

$$J = v\theta c - D_{sh}(\theta, \bar{v})(dc/dx) \tag{9.17}$$

Here J is the total mass of a solute transported across a unit cross-sectional area of soil per unit time, D_{sh} is the lumped diffusion–dispersion coefficient (a function of volumetric wetness θ and average pore-water velocity $\bar{v}$), c is the solute concentration, and dc/dx is the solute gradient.

Equation (9.17) by itself can describe only steady-state (time-invariant) processes. Moreover, it is limited to "noninteracting" (inert) solutes, by which term we designate solutes not subject to adsorption by soil solids or subject to chemical or biological reactions. Strictly speaking, truly noninteracting solutes scarcely exist. Moreover, parameters D_s, D_h, $\bar{v}$, θ, and c can only be defined in macroscopic terms as gross spatial averages. Hence Eq. (9.17) is only an approximation of the process it is meant to depict.

Turning now to transient-state processes, in which fluxes and concentrations can vary both in time and in space, we once again invoke the conservation ("continuity") principle, which for combined convective–diffusive–dispersive transport can be written

$$\partial(c\theta)/\partial t = -\partial J/\partial x \tag{9.18}$$

The rate of change of the solute mass in a volume element of soil should equal the difference between the incoming and outgoing fluxes of the solute, provided there are no gains or losses of the solute by any mechanisms operating within the volume element itself (i.e., there are no sources or sinks).

Combining Eqs. (9.8) and (9.17), we get

$$\frac{\partial(c\theta)}{\partial t} = -\frac{\partial(v\theta c)}{\partial x} + \frac{\partial}{\partial x}\left(D_{sh}\frac{\partial c}{\partial x}\right) \tag{9.19}$$

For steady water flow (which, however, does not necessarily imply steady solute movement) θ, $\bar{v}$, and D_{sh} can be taken as constant, and Eq. (9.19) simplifies to

$$\frac{\partial c}{\partial t} = -\bar{v}\frac{\partial c}{\partial x} + \frac{D_{sh}}{\theta}\frac{\partial^2 c}{\partial x^2} \tag{9.20}$$

Certain solutes are generated within the soil. An example is the nitrate ion, which evolves under appropriate circumstances from nitrification of organic matter. In contrast, some solutes may disappear from the soil volume (as in the case of nitrates being taken up by plants or undergoing denitrification). To account for such "sources" and "sinks," we must modify Eq. (9.18) to include a composite source–sink term, S, expressing the rate of production or disappearance of the particular solute:

$$\partial(c\theta)/\partial t = -\partial J/\partial x + S \tag{9.21}$$

The term S represents the net sum of all n possible sources $\Sigma\, s_i$ and all m possible sinks $\Sigma\, s_j$. Accordingly,

$$S = \sum_{i=1}^{n} s_i - \sum_{j=1}^{m} s_j \tag{9.22}$$

The terms of this equation must be specified as quantities of the solute generated or dissipated per unit volume of soil and per unit length of time.

An additional possibility to consider is the existence of a dynamic storage for the solute inside the soil body but outside the mobile liquid phase, as, for instance, in precipitated form or in the soil's exchange complex. In this case, the left-hand side of Eq. (9.21) can be expanded to include the quantity of solute in storage (σ_s). Accordingly, the left-hand side of Eq. (9.21) becomes $\partial(\theta c + \sigma_s)/\partial t$. The time derivative of σ_s (namely, $\partial\sigma_s/\partial t$)

expresses the rate of increase of storage outside the solution phase for the solute under consideration.

We can now write a comprehensive equation of transient-state solute dynamics to include convective–dispersive–diffusive movement as well as sources, sinks, and storage changes. For a vertical soil profile, we substitute the axis z (representing depth below the soil surface) for x, to obtain

$$\frac{\partial(\theta c + \sigma_s)}{\partial t} = \frac{\partial}{\partial z}\left(D_{sh}\frac{\partial c}{\partial z}\right) - \frac{\partial(qc)}{\partial z} + S \qquad (9.23)$$

wherein, as previously defined, the solution's convective flux q is equal to the product of its average velocity and volume fraction in the soil; that is, $q = \bar{v}\theta$.

The adsorption of ions in the soil can be positive, as is the attachment of cations to clay surfaces, or it can be negative, as is the repulsion or partial exclusion of anions from the electrostatic double layer of the same clay. Since, in consequence, the anions are relegated to the centers of the pores, they tend to travel somewhat faster than cations when an electrolytic solution is passed through a soil, even as the cations undergo adsorption and exchange phenomena within the soil's cation exchange complex and their motion is thereby retarded. If the attainment of equilibrium between the ions in the solution phase and the soil's exchange complex is rapid enough to be considered instantaneous, then the amount adsorbed A can be taken as dependent only on the concentration c of the ion species in the soil solution. Assuming that storage outside the solution phase is due entirely to adsorption (i.e., there is no other storage mechanisms, such as precipitation of a component of limited solubility), then the amount in storage σs is equal to A. Using these assumptions, Eq. (9.23) can be rewritten

$$[\theta + A(c)]\frac{\partial c}{\partial t} = \frac{\partial}{\partial z}\left(D_{sh}\frac{\partial c}{\partial z}\right) - \frac{\partial(qc)}{\partial z} + S \qquad (9.24)$$

Here $A(c)$ is the slope of the *adsorption isotherm*, a function of concentration; θ, c, s, and q are functions of depth and time; and D_{sh} is a function of θ and average velocity $\bar{v} = q/\theta$.

EFFECTS OF SOLUTES ON WATER MOVEMENT

The first phenomenon to consider in this regard is the hydration of clay particles and consequent swelling of the soil. Recall that negatively charged clay particles attract cations, with which they form an *electrostatic double layer* (see Chapter 4). This process of imbibition, causing swelling, is especially pronounced when the ambient solution (i.e., the soil solution away from the particle surfaces) is the more dilute. Imbibition between clay platelets is constrained by interparticle bonds and usually ceases at ambient solution concentrations above 200–400 meq/L (McNeal, 1974). With more dilute solutions, continued swelling — in a tendency to relieve the osmotic pressure differential between the clay domains and the ambient solution — drives the particles apart and weakens the interparticle bonds. A combination of this osmotic swelling with mechanical disturbance of the soil system ("puddling") can lead to a rupturing of interparticle bonds, so adjacent particles separate and the clay fraction undergoes dispersion, which in turn alters the geometry of soil pores and results in a decrease of intrinsic permeability.

Combinations of low salt concentration and high *exchangeable sodium percentage* (ESP) are the conditions most likely to cause swelling, dispersion, and reduction of permeability. The collapse of aggregates resulting from dispersion of clay tends to plug the large interaggregate pores, particularly in the top layer, so an "open" soil surface can become sealed. Moreover, dispersed particles can move with percolating water and migrate into the soil profile. Evidence of such migration can be seen in the occurrence of clay skins over aggregates deeper in the profile and in the natural deposition of clay in the formation of distinct layers called *clay pans*.

Loss of permeability resulting from the adverse effect of salt concentration and exchangeable sodium percentage can become a major problem in irrigated agriculture. It can cause problems, as well, in the operation of municipal and industrial waste disposal systems, such as septic tank drainage fields, where decreases in the hydraulic conductivity by one or more orders of magnitude have been reported.

A widely accepted index for characterizing the soil solution with respect to its likely influence on the exchangeable sodium percentage is the *sodium adsorption ratio* (SAR). It has also been used to assess the quality of irrigation water. SAR is defined as follows:

$$\mathrm{SAR} = [\mathrm{Na}^+]/\{([\mathrm{Ca}^{2+}] + [\mathrm{Mg}^{2+}])/2\}^{0.5} \tag{9.25}$$

In words, it is the ratio of the sodium ion concentration to the square root of the average concentration of the divalent calcium and magnesium ions. In this context, all concentrations are expressed in milliequivalents per liter. SAR is thus an approximate expression for the relative activity of Na^+ ions in exchange reactions in soils. A high SAR, particularly at low concentrations of the soil solution, causes high ESP and a decrease of soil hydraulic conductivity. The relation of hydraulic conductivity to composition and concentration of the soil solution and to the composition and capacity of the soil's exchange complex is influenced by the quantity and nature of the predominant clay (e.g., whether kaolinitic or montmorillonitic) as well as by the presence and content of cementing agents (e.g., humus and sesquioxides).

SOIL SALINITY AND ALKALINITY

An excessive accumulation of salts in the soil causes a decline in productivity. *Soil salinity* is the term used to designate a condition in which the soluble salt content of the soil reaches a level harmful to crops. Soil salinity affects plants directly through the reduced osmotic potential of the soil solution and the toxicity of specific ions, (such as boron, chloride, and sodium. If the salts are mainly sodic salts, their accumulation increases the concentration of sodium ions in the soil's exchange complex, which in turn affects soil properties and behavior. Thus, salinity can also have indirect effects on plant growth through deleterious modification of such soil properties as swelling, porosity, water retention, and permeability.

Soluble salts may be present in the original soil material or be brought to the soil by invading surface water or rising groundwater. The ocean is a source of salt for low-lying soils along the coast, where it may also contribute salt to rainfall by sea spray. Except along the seacoast, however, saline soils seldom occur in humid regions, thanks to the typically downward net seepage of water through the soil profile brought about by the excess of rainfall over evapotranspiration. In arid regions, however, there may be no net downward percolation of water and no effective leaching, so salts can accumulate in the root zone, especially in its top layer. Here the combined effects of meager rainfall, a high evaporation rate, the occurrence of salt-bearing sediments, and the presence of shallow, brackish groundwater give rise to a group of soils known in classical pedology as *solonchaks*.

Less obvious than the appearance of naturally saline soils, but perhaps more ominous, is the *induced salination* of originally productive soils by the injudicious practice of irrigation in certain arid regions. Typically located in the river valleys of the dry zone, irrigated soils are particularly vulnerable to the insidious, and for a time invisible, rise of the water table caused by failure to provide adequate groundwater drainage. *Proper irrigation* maintains a supply of good-quality moisture needed by plants to answer the climatically imposed evaporational demand while at the same time ensuring a favorable salt balance and nutrient supply together with adequate aeration and temperature regimes throughout the root zone. *Efficient irrigation*, furthermore, avoids wasting water through runoff or excessive drainage, except insofar as some drainage is necessary to flush out potentially harmful salts that would otherwise accumulate in the root zone. Crop plants extract water from the soil while leaving most of the salt behind. Unless leached away (preferably continuously, but at least periodically) such salts will sooner or later begin to hinder crop growth.

The classical and still pertinent publication on soil salinity is the U.S. Department of Agriculture *Handbook 60* (L. A. Richards, editor, 1954). Although newer and much revised publications have since been issued (e.g., Bresler et al., 1982; Tanji, 1990; Rhoades et al., 1999; and Hillel, 2000), some of the definitions and concepts of the original publication are still used widely. The accepted classification of soil salinity (as well as of irrigation water quality) is based on total salt concentration, as measured by its electrical conductivity, and on the relative concentration of sodium. *Soil alkalinity* (or *sodicity*) is characterized by the ESP, the content of exchangeable sodium ions (expressed in milliequivalents per 100 grams of soil) as a fraction (percent) of the total cation exchange capacity.

According to these criteria, a *saline soil* is one whose saturation extract indicates an electrical conductivity exceeding 0.4 S/m (4 mmho/cm) at 25°C but whose ESP is less than 15. (A saturation extract is obtained after adding an amount of water just sufficient to saturate a soil sample, which is then stirred to form a slurry or “paste” in a standard manner. The water is later sampled by filtration for determination of salt concentration.) Such soils usually have pH values less than 8.5 and are ordinarily well flocculated.

Reclamation of saline soils requires removal of excess salts by leaching, with due care to prevent a rise in the exchangeable sodium percentage. *Saline-alkali soils* have been defined as having a saturation-extract conductivity greater than 0.4 S/m and an ESP greater than 15. Such soils, when leached, become highly dispersed and exhibit higher pH values. To be reclaimed, these soils require, in addition to leaching, treatment with soluble calcium salts (e.g., gypsum) to replace the excess of exchangeable sodium. Finally, *nonsaline alkali soils* exhibit an ESP greater than 15 but a salinity less than the 0.4-S/m level. Such soils, which often exhibit very high pH values (8.5–10) have traditionally been called *solonetz* by pedologists and sometimes *black alkali* (in reference to the black color resulting from the highly dispersed state of organic matter at the surface).

SALT BALANCE OF THE SOIL PROFILE

The *salt balance* is a quantified summary of all salt inputs and outputs for a defined volume or depth of soil during a specified period of time. If salts are conserved (that is to say, if they are neither generated nor decomposed chemically in the soil), then the difference between the total input and output must equal the change in salt content of the soil zone monitored. Accordingly, if input exceeds output, then salt must be accumulating, and vice versa. The salt balance has been used as an indicator of salinity trends and the need for salinity control measures in large-scale irrigation projects as well as in single irrigated fields.

The following equation applies to the amount of salt in the liquid phase of the root zone per unit area of land:

$$[\rho_w(V_r c_r + V_i c_i + V_g c_g) + M_s + M_a] - (M_p + M_c + \rho_w V_d c_d) = \Delta M_{sw} \qquad (9.26)$$

Here V_r is the volume of rainwater entering the soil with a salt concentration c_r; V_i is the volume of irrigation with a concentration c_i; V_g is the volume of groundwater with a concentration c_g entering the root zone by capillary rise; V_d is the volume of water drained from the soil with a concentration c_d; M_s and M_a are masses of salt dissolved from the soil and from agricultural inputs (fertilizers and soil amendments), respectively; M_p and M_c are the mass of salt precipitated (or adsorbed) in situ and the mass removed by the crop, respectively; ρ_w is the density of water; and, ΔM_{sw} is the change in mass of salt in the soil's liquid phase during the period considered.

A simplified form of the salt balance equation was given by Buras (1974) for a complete annual period. Assuming that the net change in total soil water content from one year to the next (as monitored either at the end of the wet season or at the end of the dry season) is close to zero; disregarding surface runoff, the *water balance* per unit area is:

$$V_i + V_r + V_g = V_{et} + V_d \qquad (9.27)$$

Here V_i, V_r, V_g, V_{et}, and V_d are total annual volumes of irrigation, atmospheric precipitation, capillary rise of groundwater, evapotranspiration, and drainage, respectively. As evapotranspiration removes no salt and crops generally remove only a negligible amount, and if we disregard agricultural inputs and in situ precipitation and dissolution of salts, the salt balance corresponding to Eq. (9.29), assuming no accumulation, is

$$V_i c_i + V_r c_r = (V_d - V_g) c_d \qquad (9.28)$$

where c_i, c_r, and c_d are the average concentrations of salt in the irrigation, rainfall, and drainage waters, respectively. If the water table is kept deep enough so that no substantial capillary rise into the root zone occurs, and since the salt content of atmospheric precipitation is usually negligible, Eq. (9.30) further simplifies to

$$V_i c_i = V_d c_d \tag{9.29}$$

Such an overall "black box" approach disregards the mechanisms and rates of salt and water interactions in the root zone as well as the changing pattern of salt distribution throughout the soil profile. A more detailed treatment of the dynamic balance of salts in the root zone under irrigation was given by Bresler (1972b), based on the following equation:

$$\frac{\partial}{\partial t}\int_0^z [c(z,t)\,\theta(z,t)]\,dz = q(o,t)c_0(t) - q(z,t) \tag{9.30}$$

where the flux q, soil wetness θ, and salt concentration c are considered as functions of depth z and time t, and c_0 is the salt concentration of the applied water. This equation was solved explicitly for $c(z, t)$, via numerical methods, for different initial and boundary conditions.

Figure 9.3 demonstrates the results obtained from these computations for a fallow (unvegetated) loam soil during cycles of infiltration, redistribution of soil moisture, and evaporation. The computation of salt and water dynamics in the presence of an active root system, though more complicated, is amenable to the same type of quantitative approach (e.g., Hillel et al., 1975).

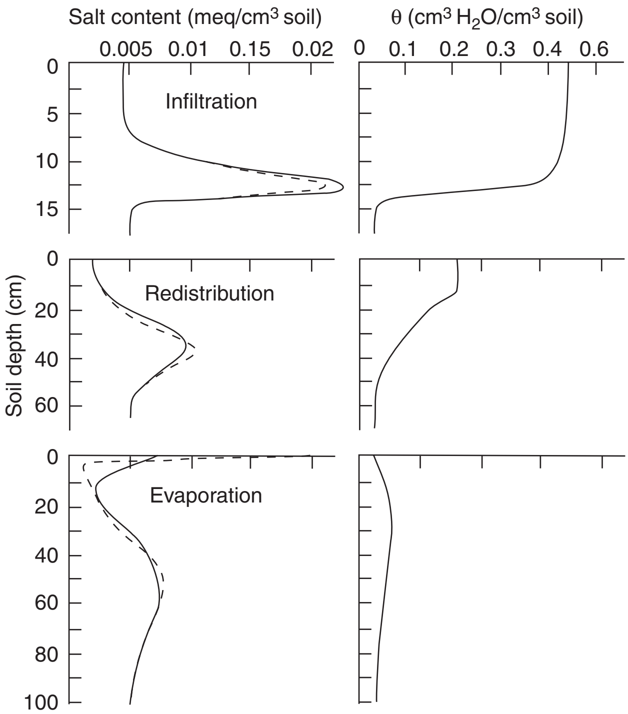


Fig. 9.3. Computed salt and water content distributions following successive stages of the field water cycle. The dispersion term was considered (solid lines) and omitted (dashed lines). (From Bresler, 1972b.)

Sample Problem

The following data were obtained in a field in an arid zone:

(a) Rainfall occurred only in winter and amounted to 300 mm, with a total salt concentration of 40 ppm.
(b) Capillary rise from shallow, saline groundwater during spring and autumn totaled 100 mm at a concentration of 1000 ppm.
(c) Irrigation was applied during the summer season and amounted to 900 mm (400 ppm salts).
(d) Drainage occurred only during the irrigation season and amounted to 200 mm, with a soluble salt concentration of 800 ppm.
(e) An additional increment 0.12 kg/m^2 of soluble salts was added in the form of fertilizers and soil amendments, whereas an amount of 0.1 kg/m^2 was removed by the harvested crops.

Disregarding dissolution and/or precipitation of salts within the soil, compute the annual salt balance. Is there a net accumulation or release of salts by the soil?

We begin with a slightly modified version of Eq. (9.28) to calculate the salt balance of the root zone per unit of land area:

$$\Delta M_s = \rho_w(V_r c_r + V_i c_i + V_g c_g - V_d c_d) + M_a - M_c$$

where ΔM_s is the change in mass of salt in the root zone (in the liquid phase as well as in the solid phase, i.e., the sum of all dissolved, adsorbed, and precipitated salt); ρ_w is the density of water; V_r, V_i, V_g, and V_d are the volumes of rain, irrigation, groundwater rise, and drainage, respectively, with corresponding salt concentrations of c_r, c_i, c_g, c_d; and M_a, M_c are the masses of salts added agriculturally and removed by the crops, respectively.

Using 1 m^2 as the unit "field" area, calculating water volumes in terms of cubic meters and masses in terms of kilograms (water density being 1000 kg/m^3), we can substitute the given quantities into the previous equation to obtain the change in mass content of salt in the soil:

$$\begin{aligned}\Delta M_s = {} & [10^3\ \text{kg/m}^3(0.3\ \text{m} \times 40 \times 10^{-6} + 0.1\ \text{m} \times 1000 \times 10^{-6} \\ & + 0.9\ \text{m} \times 400 \times 10^{-6} - 0.2\ \text{m} \times 800 \times 10^{-6}) \\ & + 0.12\ \text{kg/m}^2 - 0.1\ \text{kg/m}^2] \\ = {} & 0.322\ \text{kg/m}^2\ \text{yr}\end{aligned}$$

Thus, the soil's root zone is accumulating salt at a rate equivalent to 3220 kg (3.22 metric tons) per hectare per year.

BOX 9.1 Silt and Salt in Ancient Mesopotamia

Ancient Mesopotamia owed its prominence to its agricultural productivity, based on its soil and water resources and its favorable climate. The soils of this extensive alluvial valley are deep and fertile. The topography is level and the climate is warm and dry, with abundant sunshine year-round. The scant rainfall poses no appreciable problems of water erosion or of nutrient leaching, such as occur on sloping lands in more humid regions. The main constraint to crop production in this arid valley is, of course, water. Fortunately, water is abundant in the twin rivers, the Euphrates and the Tigris.

However, the diversion of river water onto valley land led to a series of interrelated problems. The first problem was sedimentation. Early in history, the upland watersheds of the twin rivers were deforested and overgrazed. Erosion resulting from seasonal rains proceeded to strip off the soil of sloping uplands and pour it into the streams. As the floodwaters wound their way toward the lower reaches of the valley, the sediment settled along the bottoms and sides of the rivers, raising their beds and banks above the adjacent plain. During periodic floods the rivers overtopped their banks and inundated large tracts of land.

The second insidious problem was salt. It resulted from the inexorable rise in the water table, which, in the absence of adequate drainage, naturally follows the flood irrigation of low-lying lands. Elevated rivers continually seep into the groundwater. So do diversion canals and distribution ditches. Finally, irrigation itself causes seepage from the entire surface of the land. Since all irrigation waters contain dissolved salts, and since crop roots normally exclude salts while extracting soil moisture, the salts tend to accumulate in the soil. Unless leached out, the salts eventually poison the root zone.

In arid regions, natural rainfall is generally insufficient for annual leaching; irrigation must hence be applied in excess of crop water requirements so as to remove harmful salts by downward percolation beyond the root zone. For some years or even generations, the processes of groundwater salination and water table rise are invisible and go unnoticed. Then, when the water table comes close to the ground surface, a secondary process of capillary rise comes into play. The rising groundwater evaporates at the surface, thus infusing the topsoil with salt. This was the process that eventually doomed the civilizations of Mesopotamia. Those civilizations — Sumer, Akkad, Babylonia, and Assyria — each in turn rose and then declined, as the center of population and culture shifted over the centuries from the lower to the central to the upper parts of the Tigris–Euphrates valley (Adams, 1981; Jacobsen, 1982; Artzy and Hillel, 1988).

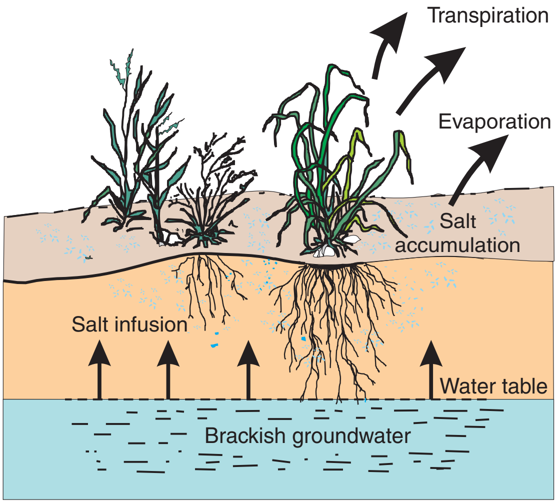


Fig. B9.1. The process of water table rise and salination.

LEACHING OF EXCESS SALTS

It is a startling fact that 1-m depth of even reasonably good-quality irrigation water (an amount normally applied in a single irrigation season) contains sufficient salt to salinize an initially salt-free soil (i.e., about 3000–5000 kg, or 3–5 tons, per hectare).

To prevent salts from accumulating during repeated irrigation–evapotranspiration cycles, the obvious remedy is to apply water in an amount greater than evapotranspiration so as to cause a portion of the applied water to flow through the root zone and wash away the excess salts. However, unless the water table is very deep or lateral groundwater drainage is sufficiently rapid, the excess irrigation often causes a progressive rise of the water table. Once the water table comes within a meter or so of the soil surface, groundwater tends to seep upward into the root zone by capillary action and to reinfuse the soil with salt.

From the foregoing it should be obvious that any attempt to leach without provision of adequate drainage is not merely doomed to fail but can indeed exacerbate the problem. In many areas where natural drainage is slow and artificial drainage is not provided, it becomes impossible to sustain irrigation and the land must sooner or later be abandoned owing to progressive salination.

Much attention has been devoted to assessing the optimal quantity of water that must be applied to cause leaching. The application of too much water can be as harmful as the application of too little. Exaggerated leaching not only wastes water but also tends to remove essential nutrients and to impede aeration by waterlogging the soil. The *leaching requirement* concept was developed by the U.S. Salinity Laboratory. It has been defined as the fraction of the irrigation water that must be leached out of the bottom of the root zone so as to prevent average soil salinity from rising above some specifiable limit.

According to the standards of the Handbook of the U.S. Salinity Laboratory, the maximum concentration of the soil solution, expressed in terms of electrical conductivity, should be kept below 0.4 S/m (4 mmhos/cm) for sensitive crops. Tolerant crops such as beets, alfalfa, and cotton may give good yields at values up to 0.8 S/m, while very tolerant crops such as barley may give good yields at salinity values as high as 1.2 S/m.

The leaching requirement depends on the salt concentration of the irrigation water, on the amount of water extracted from the soil by the crop (the evapotranspiration), and on the salt tolerance of the crop, which determines the maximum allowable concentration of the soil solution in the root zone. Assuming steady-state conditions of through-flow (thus disregarding short-term changes in soil moisture content, flux, and salinity) and assuming no appreciable dissolution or precipitation of salts in the soil and no removal of salts by the crop or capillary rise of salt-bearing water from below, we obtain from Eq. (9.29),

$$V_{\mathrm{d}}/V_{\mathrm{i}} = c_{\mathrm{i}}/c_{\mathrm{d}} \tag{9.31}$$

in which V_{d} and V_{i} are the volumes of drainage and of irrigation, respectively, and c_{d} and c_{i} are the corresponding concentrations of salt. Water volumes are normally expressed per unit area of land as equivalent depths of water, and salt concentrations are generally measured and often reported in terms of electrical conductivity. Since the volume of water drained is the difference between the volumes of irrigation and evapotranspiration (i.e., $V_d = V_i - V_{et}$), we can transform the last equation as follows:

$$V_i = [c_d/(c_d - c_i)]V_{et} \tag{9.32}$$

which is equivalent to the formulation by Richards (1954):

$$d_i = [E_d/(E_d - E_i)]d_{et} \tag{9.33}$$

Here d_i is depth of irrigation, d_{et} is the equivalent depth of the crop's "consumptive use," and E_d, E_i are the electrical conductivities of the drainage and irrigation waters, respectively.

The leaching requirement equation implies that by varying the fraction of applied water percolated through the root zone, it is possible to control the concentration of salts in the drainage water and hence to maintain the concentration of the soil solution in the main part of the root zone at some intermediate (optimal) level between c_i and c_d.

Leaching soils at a water content below saturation (e.g., under low-intensity sprinkling irrigation or under intermittent irrigation) can produce more efficient leaching and thereby reduce the amount of water required as well as mitigate drainage problems in areas of high water tables. If a soil with large vertical cracks is ponded with water, much of the water moves through the cracks and is ineffective in leaching. Under rainfall or slow sprinkling, in contrast, a greater portion of the applied water moves through the soil blocks and micropores, thus producing more efficient leaching of the soil matrix per unit volume of water infiltrated.

Sample Problem

Estimate the "leaching requirements" of a field subject to a seasonal evapotranspiration of 1000 mm, if the electric conductivity of the irrigation water is 0.1 S/m (equivalent to a salt concentration of about 650 ppm) and that of the drainage water is allowed to reach 0.4 S/m (about 2600 ppm salts). What would be the leaching requirement if the irrigation water were half as concentrated? And what if the drainage water were allowed to become twice as concentrated? Finally, what would be the electrical conductivity of the drainage water if the amount of irrigation were 1500 mm?

Using Eq. (9.33):

$$d_i = [E_d/(E_d - E_i)]d_{et}$$

where d_i and d_{et} are the volumes of water per unit land area (in depth units) of irrigation and of evapotranspiration, respectively, and E_d and E_i are the electrical conductivities of the drainage and irrigation waters, respectively. Substituting the appropriate values, we get

$$d_i = [0.4/(0.4 - 0.1)]\ 1000\ \text{mm} = 1333\ \text{mm}$$

The "leaching depth" is

$$d_e = d_i - d_{et} = 1333 - 1000 = 333\ \text{mm}$$

If the irrigation water were half as concentrated ($E_i = 0.05$ S/m),

$$d_i = [0.4/(0.4 - 0.05)]\ 1000\ \text{mm} = 1143\ \text{mm}$$

Thus, the leaching depth would be only 143 mm, which is less than half the previously required leaching volume. If the drainage water were permitted to be twice as concentrated, that is, if E_d were 0.8 S/m instead of 0.4, then

$$d_i = [0.8/(0.8 - 0.1)]\ 1000\ \text{mm} = 1143\ \text{mm}$$

In words: Doubling the allowable concentration of the drainage water is equivalent to halving the concentration of the applied irrigation water, in terms of its effect on reducing leaching requirements. If the depth of irrigation water applied were 1500 mm and its electric conductivity were 0.1 S/m, the electric conductivity of the drainage water would be

$$E_d = E_i/[1 - (d_{et}/d_i)]$$

which is obtained by simple transformation of Eq. (9.33). Thus,

$$E_d = 0.1\ \text{S/m}/[1 - (1000/1500)] = 0.3\ \text{S/m}$$

BOX 9.2 How Ancient Egypt Escaped the Scourge of Salinity

In sharp contrast to Mesopotamia, the civilization of Egypt survived and continued in the same location for several millennia. What explains the persistence of irrigated farming in Egypt in the face of its demise in Mesopotamia?

The answer lies in the different soil and water regimes of the two river valleys. Neither clogging by silt nor poisoning by salt was as severe along the Nile as in the Tigris–Euphrates valley. The silt of Egypt is delivered by the Blue Nile from the volcanic highlands of Ethiopia, and it is mixed with the organic matter delivered by the White Nile from its swampy sources. It was not so excessive as to choke the irrigation canals, but it was fertile enough to add nutrients to the fields. Whereas in Mesopotamia the inundation usually comes in the spring and summer evaporation tends to make the soil saline, the Nile crests at a much more favorable time: after the summer heat has killed the weeds and aerated the soil, and in time to wet the soil for the prewinter planting.

The narrow floodplain of the Nile (except in the Delta) precluded the widespread rise of the water table. The water table was controlled by the stage of the river, which, over most of its length, normally lies below the level of the adjacent land. When the river inundated the land, the seepage naturally raised the water table. As the river's level dropped, it pulled the water table down after it. This annual pulsation of the river and the associated fluctuation of the water table under a free-draining floodplain created an automatically self-flushing cycle by which the salts were leached from the soil and carried away by the Nile itself (Hillel, 1994).

Unfortunately, the soil of Egypt is now threatened with degradation. The Aswan High Dam has blocked the fertile silt that had formerly been delivered by the Nile. Hence Egyptian farmers must rely increasingly on chemical fertilizers. The river, now running clear of silt, has increased its erosivity and has been scouring its own banks. Along the estuaries of the Delta there is no more deposition, so coastal erosion has begun, along with seawater intrusion. The artificial maintenance of a nearly constant water level in the river, needed to allow easy pumping of irrigation water throughout the year, has resulted in raising the water table and making drainage more difficult. So Egypt is now experiencing the maladies of waterlogging and salination from which it had for so long seemed immune.