Recent Advances in Vadose Zone Flow and Transport Modeling

M. TH. VAN GENUCHTEN
E. A. SUDICKY

The fate and transport of a variety of chemicals migrating from industrial and municipal waste disposal sites, or applied to agricultural lands, is increasingly becoming a concern. Once released into the subsurface, these chemicals are subject to a large number of simultaneous physical, chemical, and biological processes, including sorption–desorption, volatilization, and degradation. Depending upon the type of organic chemical involved, transport may also be subject to multiphase flow that involves partitioning of the chemical between different fluid phases. Many models of varying degree of complexity and dimensionality have been developed during the past several decades to quantify the basic physicochemical processes affecting transport in the unsaturated zone. Models for variably saturated water flow, solute transport, aqueous chemistry, and cation exchange were initially developed mostly independently of each other, and only recently has there been a significant effort to couple the different processes involved. Also, most solute transport models in the past considered only one solute. For example, the processes of adsorption–desorption and cation exchange were often accounted for by using relatively simple linear or nonlinear Freundlich isotherms such that all reactions between the solid and liquid phases were forced to be lumped into a single distribution coefficient, and possibly a nonlinear exponent. Other processes such as precipitation-dissolution, biodegradation, volatilization, or radioactive decay were generally simulated by means of simple first- and/or zero-order rate processes. These simplifying approaches were needed to keep the mathematics relatively simple in view of the limitations of previously available computers. The problem of coupling models for water flow and solute transport with multicomponent chemical equilibrium and nonequilibrium models is now increasingly being addressed, facilitated by the introduction of

more powerful computers, development of more advanced numerical techniques, and improved understanding of the underlying transport processes.

One major frustrating issue facing soil scientists and hydrologists in dealing with the unsaturated zone, both in terms of modeling and experimentation, is the overwhelming heterogeneity of the subsurface environment. Heterogeneity occurs at a hierarchy of spatial and time scales (Wheater and Cushman, 1991), ranging from microscopic scales that involve time-dependent chemical sorption and precipitation-dissolution reactions, to intermediate scales that involve the preferential movement of water and chemicals through macropores or fractures, and to much larger scales that involve the spatial variability of soils versus depth or across the landscape. Several lines of research are being followed to deal with the different types of soil heterogeneity. On the one hand, subsurface heterogeneity can be addressed in terms of process-based deterministic descriptions which attempt to consider the effects of heterogeneity at one or several scales (kinetic sorption, preferential flow, field-scale spatial variability). On the other hand, subsurface heterogeneity is often also addressed using stochastic approaches which incorporate certain assumptions about the transport process in the heterogeneous system (e.g., Jury and Roth, 1990; Degan, 1993; Russo, 1993). Much can be learned from both approaches.

In this chapter, we will focus on alternative conceptual approaches for deterministic modeling of solute transport in variably saturated media. Among the topics discussed are single-ion equilibrium and nonequilibrium transport, sorption, degradation, volatilization, and multicomponent transport. Transport in variably saturated structured systems is treated in somewhat more detail to illustrate the potential value of numerical models as useful tools for improving our understanding of the underlying transport processes at the field scale. We also briefly review recent developments in numerical techniques used for solving the governing flow and transport equations, including methods for solving large sparse matrices resulting from spatial and temporal numerical discretization.

Water Flow and Single-Species Solute Transport

Governing Flow and Transport Equations

Predictions of flow and transport in the vadose zone are traditionally based on the Richards equation that describes variably saturated water flow and the advection-dispersion equation that describes solute transport. For one-dimensional systems, these equations are given by

\[
\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left[ K(h) \frac{\partial h}{\partial z} - K(h) \right] - S \tag{6.1}
\]

\[
\frac{\partial (\theta c)}{\partial t} + \frac{\partial (\theta c)}{\partial z} = \frac{\partial}{\partial z} \left( \theta D \frac{\partial c}{\partial z} - q e \right) + \phi \tag{6.2}
\]

respectively, where \( \theta \) is the volumetric water content, \( h \) is the soil-water pressure head (negative for unsaturated conditions), \( t \) is time, \( z \) is distance from the soil surface.
downward, $K$ is the hydraulic conductivity as a function of $h$ or $\theta$, $s$ is the solute concentration associated with the solid phase of the soil, $c$ is the solute concentration of the liquid phase, $\rho$ is the soil bulk density, $D$ is the solute dispersion coefficient, $S$ and $\phi$ are sinks or sources for water and solutes, respectively, and $q$ is the volumetric fluid flux density given by Darcy’s law as

$$ q = -K(h) \frac{\partial h}{\partial z} + K(h) $$

(6.3)

Similar equations may be formulated for multidimensional flow and transport.

Assuming linear sorption such that the adsorbed concentration ($s$) is linearly related to the solution concentration ($c$) through a distribution coefficient, $k_D$ (i.e., $s = k_D c$), equation (6.2) reduces to the simpler form

$$ \frac{\partial (\theta Rc)}{\partial t} = \frac{\partial}{\partial z} \left( \theta D \frac{\partial c}{\partial z} - q c \right) + \phi $$

(6.4)

where $R = 1 + \frac{\rho k_D}{\theta}$ is the solute retardation factor. For conditions of steady-state water flow in homogeneous soils and in the absence of source or sink terms ($S$ and $\phi$), equation (6.4) further reduces to the standard advection–dispersion equation (ADE):

$$ R \frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial z^2} - \nu \frac{\partial c}{\partial z} $$

(6.5)

where $\nu = q/\theta$ is the average pore-water velocity.

While models based on equations (6.1) and (6.2) have proved to be important tools in research and management, they are subject to a large number of simplifying assumptions and complications that compromise or limit their applicability (Nielsen et al., 1986). It may be instructive to list here some of these assumptions and complications. For example, the equations assume that (1) the air phase plays a relatively minor role during unsaturated flow, and hence that a single equation can be used to describe variably saturated flow; (2) Darcy’s equation is valid at both very low and very high flow velocities (including those occurring in structured soils); (3) osmotic and electrochemical components of the soil water potential are negligible; (4) the fluid density is independent of the solute concentration; and (5) matrix and fluid compressibilities are relatively small. The equations are further complicated by (6) the hysteretic nature of especially the soil water-retention function, $\theta(h)$; (7) the extreme nonlinearity of the hydraulic conductivity function, $K(h)$; (8) the lack of accurate and inexpensive methods for measuring the unsaturated hydraulic properties; (9) the extreme heterogeneity of the subsurface environment; and (10) inconsistencies between the scale at which the hydraulic and solute transport parameters in equations (6.1) and (6.2) are usually measured, and the scale at which the predictive models are being applied. In addition, equations (6.1) and (6.2) are formulated for isothermal soil conditions. In reality, most physical, chemical, and microbial processes in the subsurface are strongly influenced by soil temperature. This also applies to water flow itself, including the effects of temperature (Constantz, 1982; Hopmans and Dane, 1986) and the concentration and ionic composition of the soil liquid phase (Dane and Klute, 1977; Suarez and Simůnek, 1996) on the unsaturated soil hydraulic
properties. Hence, a complete description of vadose zone transfer processes also requires consideration of heat flow and its nonlinear effects on most processes taking place in the soil–plant system.

Root Water Uptake

An important term in the variably saturated flow equation (6.1) is the source/sink term $S$ used to account for water uptake by plant roots. Widely different approaches exist for simulating water uptake (Molz, 1981). Many of the early studies of root water uptake (e.g., Whisler et al., 1968; Bresler et al., 1982) used uptake functions of the general form

$$S(z, t) = b_1(z)K[\theta(h_r - h(z, t))]$$  \hspace{1cm} (6.6)

where $h_r$ is an effective root-water pressure head at the root surface and $b_1$ is a depth-dependent proportionality constant often referred to as the root effectiveness function. Equation (6.6) may be viewed as a finite difference approximation of Darcy’s law across the soil–root interface. Another class of models for root water uptake is given by (Feddes et al., 1978; Vanclooster et al., 1994)

$$S(z, t) = -b_2(z)\alpha_1(h(z, t))T_p$$  \hspace{1cm} (6.7)

where $b_2$ is the potential root water uptake distribution function which integrates to unity over the soil root zone, $\alpha_1$ is a dimensionless water stress response function between 0 and 1, and $T_p$ is the potential transpiration rate.

The effects of soil salinity on water uptake may be accounted for by linearly adding the osmotic head, $\pi$, to the pressure head, $h(z, t)$ in equations (6.6) or (6.7) (Bresler and Hoffman, 1986; Cardon and Letey, 1992a, 1992b), or by incorporating into equation (6.7) a separate salinity response function, $\alpha_2(\pi)$, somewhat similar to $\alpha_1(h)$, to obtain (van Genuchten, 1987; Šimůnek and Suárez, 1994)

$$S(z, t) = -b_2(z)\alpha_1(h(z, t))\alpha_2(\pi(z, t))T_p$$  \hspace{1cm} (6.8)

Cardon and Letey (1992a, 1992b) showed that approaches based on equation (6.7) may be more appropriate than equation (6.6), particularly if suitably modified and used for saline conditions. Still, as pointed out by Nielsen et al. (1986), the above two classes of root water uptake models are essentially empirical by containing parameters that depend on specific crop, soil, and environmental conditions. Much research remains needed in the development of realistic process-based models of root growth and root water uptake as a function of various stresses (water, salinity, temperature, nutrients, and others) in the root zone, and to couple these descriptions with suitable crop growth models.

Linear Equilibrium Solute Transport

The term $\partial(k \varphi) / \partial t$ in equation (6.2) may be used to account for the effects of sorption or exchange on solute transport. Most often, a linear equilibrium isotherm, $s = k_p c$, is used to describe solute interactions between the liquid and solid phases of the soil, leading to a constant retardation factor $R$ in equation (6.4). The resulting advection-
dispersion equation given by equation (6.5) has been reasonably successful in describing solute transport data for relatively uniform laboratory or field soils. As an example, figure 6.1 shows solute breakthrough curves typical of the transport of an excluded anion, Cl⁻, an essentially nonreactive solute (tritiated water, \( ^3\text{H}_2\text{O} \)), and an adsorbed tracer, Cr\(^{6+} \), through homogeneous soil columns. The first two tracers pertain to transport through 30-cm-long columns that contain disturbed Glendale clay loam (P. J. Wierenga, 1972, unpublished data; van Genuchten and Cleary, 1979), while the Cr\(^{6+} \) data are for transport through a 5-cm-long column of sand (P. J. Wierenga, 1972, unpublished data). The data in figure 6.1 are plotted versus number of pore volumes \( T = vt/L \) of tracer solution leached through the columns. Analysis of the breakthrough curves in terms of the ADE by using inverse procedures (van Genuchten, 1981) yielded \( R \)-values of 0.681, 1.027, and 1.248, respectively, for the three tracers. Hence, Cl⁻ was strongly affected by anion exclusion \( (R < 1; k_p < 0) \) caused by the repulsion of chloride anions from negatively charged surfaces of clays and ionizable organic matter. Because water flow velocities are zero along pore walls, and maximum in the center of pores, anions such as Cl⁻ cannot travel much faster than water, especially in fine-textured soils. By comparison, \( ^3\text{H}_2\text{O} \) did travel with nearly the same velocity as water \( (R = 1.027) \), while Cr\(^{6+} \) was about 25% slower \( (R = 1.248) \).

### Nonlinear Adsorption

The assumption of a linear isotherm can greatly simplify the mathematics of a transport problem; unfortunately, sorption and exchange reactions are generally nonlinear and often depend also on the presence of competing species in the soil solution. The solute retardation factor for nonlinear adsorption is then not constant anymore, as was the case for linear adsorption, but will change as a function of the slope \( ds/dc \) of the adsorption isotherm \( s(c) \) as follows:

\[
R = 1 + \frac{\rho}{\theta} \frac{ds(c)}{dc} \quad (6.9)
\]

Figure 6.1 Observed and calculated breakthrough curves for Cl⁻, \( ^3\text{H}_2\text{O} \), and Cr\(^{6+} \).
A variety of models are available to describe nonlinear adsorption–desorption. Table 6.1 lists some of the most commonly used sorption isotherms that relate the adsorbed concentration, \(s\), to the solution concentration, \(c\). Although several of the equations in table 6.1 (e.g., the Langmuir and Freundlich equations) can be derived rigorously, such as for the adsorption of gases onto solids, the expression are generally used only in an empirical fashion. Of the equations listed in table 6.1, the most popular sorption models are the Langmuir, Freundlich, and Temkin equations.

A general classification of adsorption as reflected by different features of the adsorption isotherm, such as the initial slope, the presence or absence of a plateau, or the presence of a maximum, was proposed by Giles et al. (1960). They divided possible adsorption processes into four main classes: S, L (Langmuir), H (high-affinity), and C (constant partitioning) isotherms, and discussed mechanisms that explain the different types of isotherms. Increasing solution concentrations led to increasing or decreasing adsorption rates for the convex S and concave L isotherms, respectively (figure 6.2). An H isotherm is characterized by extremely high affinities

| Table 6.1 Equilibrium Adsorption Equations (van Genuchten and Cleary, 1979; Barry, 1992) |
|---------------------------------------------|-----------------|------------------|
| Equation                                      | Model            | Reference(s)     |
| \( s = k_1c + k_2 \)                           | Linear           | Lapidus and Amundson (1952); Lindstrom et al. (1967) |
| \( s = k_1c^{k_2} \)                           | Freundlich       | Freundlich (1909) |
| \( s = \frac{k_1c}{1 + k_2c} \)                | Langmuir         | Langmuir (1918)  |
| \( s = \frac{k_1c^{k_2}}{1 + k_2c^{k_3}} \)    | Freundlich-Langmuir | Sips (1950)  |
| \( s = \frac{k_1c + k_2c^{k_3}}{1 + k_2c} \)  | Double Langmuir  | Shapiro and Fried (1959) |
| \( s = k_1c^{k_2} \)                           | Extended Freundlich | Sibbesen (1981) |
| \( s = \frac{k_1c}{1 + k_2c + k_3\sqrt{c}} \) | Gunary           | Gunary (1970)    |
| \( s = k_1c^{k_2} - k_3 \)                    | Fitter-Sutton    | Fitter and Sutton (1975) |
| \( s = k_1(1 - [1 + k_2c^{k_3}]^{k_4}) \)    | Barry            | Barry (1992)     |
| \( s = \frac{RT}{k_1} \ln (k_2c) \)          | Temkin           | Bache and Williams (1971) |
| \( s = k_1c \exp(-2k_2c) \)                   | Modified Kielland | Lindstrom et al. (1971); van Genuchten et al. (1974) |

\( s/T = c[e^{c} + k_1(c_T - c)\exp(k_2(c_T - 2c))]^{-1} \)  Modified Kielland  

\( k_1, k_2, k_3, k_4, \) Empirical constants; \( R, \) universal gas constant; \( T, \) absolute temperature; \( c_T, \) maximum solute concentration; \( r_T, \) maximum adsorbed concentration.
of the exchanger for exchangeable ions, whereas a C isotherm reflects constant partitioning of the solute between the solution and adsorbed phases. We note that the shape of an isotherm can have significant impacts on the transport predictions. For example, S and L isotherms lead to unfavorable and favorable exchange situations, respectively, with the latter condition (e.g., for a Freundlich isotherm with \( k_2 \ll 1 \)) producing sharp concentration fronts during transport in a soil profile. The effects of isotherm nonlinearity on solute front sharpening and front broadening have been discussed at length in the literature (e.g., Bolt, 1979; Schweich and Sardin, 1981; van der Zee and van Riemsdijk, 1994).

**Nonequilibrium Transport**

Application of the above equilibrium models to single-ion transport through repacked laboratory or relatively uniform field soils has been fairly successful. The equilibrium approach, however, has not worked well in several situations, most notably for many strongly adsorbed solutes, many organic chemicals, and when used for simulating transport in structured (aggregated) media. A number of chemical-kinetic and diffusion-controlled "physical" models have been proposed to describe nonequilibrium transport.

Early models for nonequilibrium transport generally assumed relatively simple first-order type (one-site) kinetic rate equations. More refined nonequilibrium models introduced later invoked the assumptions of two-site or multisite sorption, and/or two-region (dual-porosity) transport that involves solute exchange between mobile and relatively immobile liquid regions. Models of this type generally resulted in better descriptions of observed laboratory and field transport data, mostly because of additional degrees of freedom in fitting observed concentration distributions. The different nonequilibrium approaches are briefly reviewed below.

**One-Site Sorption Models**

The simplest nonequilibrium formulation arises when a first-order linear kinetic rate process is assumed. Ignoring any solute production or decay in the adsorbed phase, equation (6.2) is then augmented with the equation

\[
\frac{\partial \alpha}{\partial t} = \alpha (k_\beta c - s) \tag{6.10}
\]
where \( \alpha \) is a first-order kinetic rate coefficient. Transport models using the above one-site kinetic sorption equation have generally resulted in only modest improvements in terms of their ability to match observed displacement data (e.g., Davidson and McDougal, 1973; van Genuchten et al., 1974). Success was usually limited only to experiments conducted at relatively low flow velocities; that is, for conditions where the equilibrium model already performed reasonably well. Moreover, one or both of the sorption parameters \((k_D, \alpha)\), when adjusted to get better transport predictions, were often found to vary as a function of the pore-water velocity. Similar limitations hold for most or all of the other nonequilibrium rate expressions listed in table 6.2.

Two-Site Chemical Nonequilibrium Transport

The one-site first-order kinetic model may be expanded into a two-site sorption concept by assuming that sorption sites can be divided into two fractions (Selim et al., 1976): sorption on one fraction (type 1 sites) is assumed to be instantaneous while sorption on the remaining (type 2) sites is considered to be time-dependent. Assuming a linear sorption process, the complete two-site transport model is given by (van Genuchten and Wagenet, 1989)

\[
\frac{\partial}{\partial t}(\theta + f \rho k_D c) = \frac{\partial}{\partial z}\left(\theta D \frac{\partial c}{\partial z} - qe\right) - \alpha \rho s_b(1 - f)k_D c - s_2 - \theta \rho s_a c - f \rho k_D s_2 c
\]

(6.11)

\[
\frac{\partial s_2}{\partial t} = \alpha((1 - f)k_D c - s_2) - \mu_2 s_2
\]

(6.12)

where \( \mu_2 \) and \( \mu_s \) are first-order decay constants for degradation in the liquid and sorbed phases, respectively, \( f \) is the fraction of exchange sites assumed to be at

<table>
<thead>
<tr>
<th>Equation</th>
<th>Model</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{dx}{dt} = \alpha(k_1 c + k_2 - s) )</td>
<td>Linear</td>
<td>Lapidus and Amundson (1952); Oddson et al. (1970)</td>
</tr>
<tr>
<td>( \frac{dx}{dt} = \alpha(k_1 c^2 - s) )</td>
<td>Freundlich</td>
<td>Horrsey and Davidson (1973); van Genuchten et al. (1974)</td>
</tr>
<tr>
<td>( \frac{dx}{dt} = \alpha k_1 c \frac{k_1 c}{1 + k_1 c} )</td>
<td>Langmuir</td>
<td>Hendricks (1972)</td>
</tr>
<tr>
<td>( \frac{dx}{dt} = \alpha k_1 c \frac{k_1 c^2}{1 + k_1 c^2} )</td>
<td>Freundlich-Langmuir</td>
<td>Šimůnek and van Genuchten (1994)</td>
</tr>
<tr>
<td>( \frac{dx}{dt} = \alpha(s_f - s) \sinh\frac{k_f}{s_f - s} )</td>
<td></td>
<td>Fava and Eyring (1956)</td>
</tr>
<tr>
<td>( \frac{dx}{dt} = \alpha \exp(k_2 s)(k_1 c \exp(-2k_2 s) - s) )</td>
<td></td>
<td>Lindstrom et al. (1971)</td>
</tr>
<tr>
<td>( \frac{dx}{dt} = \alpha \lambda s f )</td>
<td></td>
<td>Leenheer and Ahlrichs (1971); Enfield et al. (1976)</td>
</tr>
</tbody>
</table>
equilibrium, and the subscripts 1 and 2 refer to equilibrium (type 1) and kinetic (type 2) sorption sites, respectively. Note that if \( f = 0 \), the two-site sorption model reduces to the one-site fully kinetic sorption model, i.e., only type 2 kinetic sites are present. On the other hand, if \( f = 1 \), the two-site sorption model reduces to the equilibrium sorption model. The two-site sorption model has been quite successful in describing a large number of mostly laboratory-type miscible displacement experiments that involve a variety of organic and inorganic chemicals undergoing adsorption (Rao et al., 1979; Hoffman and Rolston, 1980; Parker and Jardine, 1986; Gonzalez and Ukrainczyk, 1999, among many others).

Two-Region Physical Nonequilibrium Transport

The two-region physical nonequilibrium transport model assumes that the liquid phase can be partitioned into distinct mobile (flowing) and immobile (stagnant) liquid pore regions, and that solute exchange between the two liquid regions can be modeled as an apparent first-order exchange process. Using the same notation as before, the two-region transport model is given by (van Genuchten and Wagenet, 1989)

\[
\frac{\partial}{\partial t}(\theta_m + f \rho \kappa_D) c_m = \frac{\partial}{\partial x} \left( \theta_m D_m \frac{\partial c_m}{\partial x} - q c_m \right) - \alpha (c_m - c_{im}) - (\theta_m \mu_{l,m} + f \rho \kappa_D \mu_{s,m}) c_m
\]

(6.13)

\[
[\theta_m + (1 - f) \rho \kappa_D] \frac{\partial c_{im}}{\partial t} = \alpha (c_m - c_{im}) - [\theta_m \mu_{l,im} + (1 - f) \rho \kappa_D \mu_{s,im}] c_{im}
\]

(6.14)

where the subscripts \( m \) and \( im \) refer to the mobile and immobile liquid regions, respectively, the subscripts \( l \) and \( s \) refer to the liquid and sorbed phases, respectively, \( f \) represents the fraction of sorption sites that equilibrates with the mobile liquid phase, and \( \alpha \) is a first-order mass transfer coefficient that governs the rate of solute exchange between the mobile and immobile liquid regions. The two-region physical nonequilibrium model has been successfully applied to laboratory-scale transport experiments that involve a large number of tracers (tritiated water, chloride, different organic chemicals, heavy metals) as shown in studies by Gaudet et al. (1977) van Genuchten et al. (1977), and Gaber et al. (1995), among others. As an example, figure 6.3 shows breakthrough curves for the pesticide 2,4,5-T (2,4,5-trichlorophenoxyacetic acid) obtained from a 30-cm-long soil column that contained aggregated (< 6 mm in diameter) Glendale clay loam (van Genuchten et al., 1987). Notice that the two-region model (TRM) provides an excellent description of the data, whereas the advection–dispersion equation (ADE) could not be made to fit the data.

A close comparison of the two-site and two-region nonequilibrium models shows that both have the same mathematical structure. As shown previously by Nkedikizza et al. (1984) and Toride et al. (1993), among others, the two models can be put into the same dimensionless form by using appropriately selected dimensionless parameters. Because the same dimensionless transport equations apply to conceptually different transport situations, it also follows that breakthrough curves such as
Figure 6.3 Observed and calculated effluent curves for 2,4,5-T movement through Glendale clay loam. The fitted curves were based on (a) the classical ADE and (b) two-region TRM transport models. (After van Genuchten et al., 1987.)

those shown in figure 6.3 generally contain insufficient information to differentiate between specific physical (mobile-immobile type) and chemical (kinetic type) processes that lead to nonequilibrium, unless nonadsorbing tracers are considered. Hence, independent parameter estimates are generally needed to effectively differentiate between presumed two-site and two-region nonequilibrium phenomena. On the other hand, the mathematical similarity of the two-site and two-region models also suggests that the two formulations may be used to macroscopically describe transport without having to delineate the exact physical and chemical processes at the microscopic level.

Vapor-Phase Transport and Volatilization

Vapor-phase transport and volatilization from the soil surface are increasingly recognized as being important processes that affect the field behavior of many organic
chemicals, including pesticides, gasoline, and such industrial solvents as TCE and CCl₄ (Glotfelty and Schomburg, 1989; Mercer and Cohen, 1990; Yates et al., 1996). While many organic pollutants dissipate by means of chemical and microbiological degradation, volatilization may be equally important for volatile substances. A thorough understanding of vapor-phase transport is important for the proper design of in situ remediation techniques [such as air sparging and soil venting (DiGiulio, 1992)] for cleaning up hazardous waste sites contaminated with nonaqueous-phase liquids (NAPLs). The process of volatilization has gained additional interest recently because of concerns of the effects of a variety of gases, such as methyl bromide, on stratospheric ozone. Methyl bromide has been used for many decades as an effective soil fumigant for the control of nematodes, weeds, and fungi, but is now suspected to cause significant damage to the ozone layer (Gan et al., 1997).

The volatility of a chemical is influenced by many factors, most important being the physicochemical properties of the chemical, soil texture and water content, and several environmental parameters, such as temperature and solar energy (Taylor and Spencer, 1990). Even though only a small fraction of a volatile chemical may exist in the gas phase, air-phase diffusion rates can be much larger than those in the liquid phase since gas-phase diffusion coefficients are about 10⁴ times greater than those in the liquid phase. The solute transport equation for volatile solutes may be written in the following general form (e.g., Wang et al., 1998):

\[
\frac{\partial (v \rho)}{\partial t} + \frac{\partial (v g)}{\partial t} + \frac{\partial (v g)}{\partial z} = \frac{\partial}{\partial z} \left( \theta D_w \frac{\partial c}{\partial z} + a D_a \frac{\partial \tilde{c}}{\partial z} - q_w c - q_a \tilde{c} \right) + \phi \tag{6.15}
\]

where \(v\) is the volumetric air content, \(g\) is the solute concentration associated with the gas phase, \(D_w\) and \(D_a\) are the solute dispersion coefficients in the liquid and gaseous phases, respectively, and \(q_w\) and \(q_a\) are the volumetric fluid flux densities of the liquid and gaseous phases, respectively.

Assuming linear equilibrium sorption and volatilization such that the adsorbed (s) and gaseous (g) concentrations are linearly related to the solution concentration (c) through distribution coefficients, \(k_D\) (i.e., \(s = k_D c\)) and \(k_H\) (i.e., \(g = k_H c\)), respectively, equation (6.13) reduces to the simpler form of equation (6.4), where \(q = q_w + q_a k_H\) is the effective fluid flux density, \(R = 1 + (\rho k_D + a k_H) / \theta\) is the solute retardation factor, and \(D = D_w + a D_a k_H / \theta\) is the effective dispersion coefficient. For conditions of steady-state water and gas flow in homogeneous soils, and neglecting the source/sink term \(\phi\), equation (6.13) reduces to the standard advection-dispersion equation (6.5), where \(v = (q_w + q_a k_H) / \theta\) is the average pore velocity.

Degradation

The source/sink term \(\phi\) in equation (6.2) may be used to account for nutrient uptake and/or a variety of chemical and biological reactions and transformations insofar as these processes are not already included in the sorption/exchange term \(\partial \rho c / \partial t\). Solute reactions and transformations can be highly dynamic and nonlinear in time and space, especially for nitrogen and pesticide products. For example, among the nitrogen transformation processes that may need to be considered are nitrification, deni-
trification, mineralization, and nitrogen uptake by plants (Stevenson, 1982). For microbially induced organic and inorganic transformations, the degradation process should also consider the growth and maintenance of soil microbes.

Alexander and Scow (1989) gave a review of some of the equations used to represent the kinetics of biodegradation. These equations include zero-order, half-order, first-order, three-half-order, mixed-order, logistic, logarithmic, Michaelis-Menten, and Monod type (with or without growth) expressions. Possible biological degradation equations are listed in Table 6.3. While most of these expressions have a theoretical basis, they are commonly used only in an empirical fashion by fitting the equations to observed data. Zero- and first-order kinetic equations remain the most popular models for describing the biodegradation of organic compounds, mostly because of the simplicity and ease in which these equations can be incorporated in solute transport models. Conditions for the application of zero- and first-order biodegradation equations are given by Alexander and Scow (1989).

One special group of degradation reactions involves decay chains in which solutes are subject to sequential (or consecutive) decay reactions. Problems of solute transport that involves sequential first-order decay reactions frequently occur in soil and groundwater systems. Examples are the migration of various radionuclides (Rogers, 1978), the simultaneous movement of interacting nitrogen species (Cho, 1971), organic phosphate transport, and the transport of certain pesticides and their metabolites (Wagenet and Hutson, 1987; Šimůnek and van Genuchten, 1994; Šimůnek et al., 1998).

### Multicomponent Solute Transport

Except for the above decay chains, thus far we have considered the transport of only one chemical species, and hence assumed that the behavior of a solute is independent

<table>
<thead>
<tr>
<th>Equation</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi = -k_1 c )</td>
<td>First-order kinetics</td>
</tr>
<tr>
<td>( \phi = -k_t )</td>
<td>Zero-order kinetics</td>
</tr>
<tr>
<td>( \phi = -k_t c^h )</td>
<td>Power rate kinetics</td>
</tr>
<tr>
<td>( \phi = -\mu_{\text{max}} \frac{c}{K_s + c} )</td>
<td>Monod, Michaelis-Menten kinetics</td>
</tr>
<tr>
<td>( \phi = -\mu_{\text{max}} \frac{c}{K_s + c} (c_0 + X_0 - c) )</td>
<td>Monod with growth kinetics</td>
</tr>
<tr>
<td>( \phi = -\frac{\mu_{\text{max}} c}{K_i + c^2/K_f} )</td>
<td>Haldane modification of Monod kinetics</td>
</tr>
<tr>
<td>( \phi = -\mu_{\text{max}} (c_0 + X_0 - c) )</td>
<td>Logarithmic kinetics</td>
</tr>
<tr>
<td>( \phi = -k_t c (c_0 + X_0 - c) )</td>
<td>Logistic kinetics</td>
</tr>
</tbody>
</table>

\( k_1, k_t \), empirical constants; \( \mu_{\text{max}} \), maximum specific degradation; \( K_s \), substrate concentration when the rate of decay is half the maximum rate; \( c_0 \), initial substrate concentration; \( X_0 \), amount of substrate required to produce the initial population; \( K_i \), inhibition constant that reflects the suppression of the growth rate by a toxic substrate rate.
of other species present in the soil solution. In reality, the soil liquid phase always contains a mixture of many ions which mutually may interact, create complex species, precipitate, dissolve, and/or compete with each other for sorption sites on the solid phase. In this section, we give a very brief review of such more complex situations that involve multicomponent transport. More comprehensive reviews are given by Yeh and Tripathi (1989), Mangold and Tsang (1991), and Suarez and Šimůnek (1996).

Attempts to model multicomponent transport initially focused primarily on the saturated zone, where changes in the fluid flux, temperature, and pH are relatively gradual and hence less important than in the unsaturated zone. Consequently, most multicomponent transport models assume one- or two-dimensional steady-state saturated water flow. Typical examples are given by Valocchi et al. (1981), Bryant et al. (1986), and Walter et al. (1994). Only recently have multicomponent transport models become more popular for application also to variably saturated flow problems (Liu and Narasimhan, 1989; Yeh and Tripathi, 1991; Šimůnek and Suarez, 1994).

In a recent review, Yeh and Tripathi (1989) identified three different approaches for mathematically solving multicomponent transport problems: (1) a mixed differential and algebraic approach, (2) a direct substitution approach, and (3) a sequential iteration approach. In the first approach, the sets of differential and algebraic equations that describe the transport processes and chemical reactions, respectively, are treated simultaneously (Miller and Benson, 1983; Lichtner, 1985). In the second approach, the algebraic equations that represent the nonlinear chemical reactions are substituted directly into the differential mass balance transport equations (Rubin and James, 1973; Jennings et al., 1982). The third approach considers two coupled sets of linear partial differential and algebraic equations, which are solved sequentially and iteratively (Walsh et al., 1984; Yeh and Tripathi, 1991; Šimůnek and Suarez, 1994; Walter et al., 1994). Based on a study of computer resource requirements, Yeh and Tripathi (1989) suggested that only the third method (sequential iteration) can be applied to realistic multidimensional problems.

As an example, the partial differential equations that govern one-dimensional multicomponent advective–dispersive chemical transport during transient variably saturated flow may be written as (Šimůnek and Suarez, 1994)

\[
\frac{\partial (\rho c_k)}{\partial t} + \rho \frac{\partial \bar{c}_k}{\partial t} + \rho \frac{\partial \bar{c}_k}{\partial z} = \frac{\partial}{\partial z} \left( \theta D \frac{\partial c_k}{\partial z} - q \bar{c}_k \right), \quad k = 1, 2, \ldots, N_c \tag{6.16}
\]

where \( c_k \) is the total dissolved concentration of aqueous component \( k \) (i.e., the sum of the component plus all complex species that contain component \( k \)), \( \bar{c}_k \) is the total sorbed concentration of the aqueous component \( k \), \( \bar{c}_k \) is the total precipitated concentration of aqueous component \( k \) (i.e., the sum of all precipitated species that contain the component \( k \)), and \( N_c \) is the number of aqueous components. The second and third terms in the left-hand side of equation (6.16) are zero for components that do not undergo ion exchange or precipitation–dissolution reactions. The total concentration of a component \( k \), defined as the sum of the dissolved, sorbed, and mineral concentrations, is influenced only by transport processes which act on the solution concentration \( c_k \), but not by chemical reactions (Zysset et al., 1994). However, the
relative fraction of a component in each of the three phases (solution, sorbed, mineral) depends strongly on the specific chemical processes in the system. Therefore, equation (6.16) must be augmented with a set of equations that describe the different equilibrium or nonequilibrium chemical reactions, such as complexation, cation exchange, adsorption–desorption, and precipitation–dissolution (Yeh and Tripathi, 1989). The above coupled approach that involves transport and chemical submodules was used also by Walsh et al. (1984), Cederberg et al. (1985), and Yeh and Tripathi (1991).

Transport in Structured Media

Field soils generally exhibit a variety of structural features, such as interaggregate pores, earthworm or gopher holes, decayed root channels, or drying cracks in fine-textured soils. Water and dissolved chemicals can move along preferred pathways in such structured media at rates much faster than what normally can be predicted with models based on the classical Richards and ADE equations. The resulting preferential-flow process has been shown to occur not only in aggregated field soils (Beven and Germann, 1982) and unsaturated fractured rock (Wang, 1991), but also in seemingly homogeneous soils because of fingering or some other unstable flow process (Parlange and Hill, 1976; Hillel, 1993). An important implication of preferential flow is the accelerated movement of surface-applied fertilizers, pesticides, or other pollutants into and through the unsaturated zone.

Deterministic descriptions of preferential flow in structured media are often based on dual-porosity, two-region, or bicontinuum models. Models of this type assume that the medium consists of two interacting pore regions, one associated with the macropore or fracture network, and one associated with the micropores inside soil aggregates or rock matrix blocks. Different formulations arise depending upon how water and solute movement in the micropore region are modeled, and how the micropore and macropore regions are coupled.

Geometry-Based Models

A rigorous analysis of transport in structured soils can be made when the medium is assumed to contain geometrically well-defined cylindrical, rectangular, or other types of macropores or fractures. Models may be formulated by assuming that the chemical is transported by advection, and possibly by diffusion and dispersion, through the macropores, while diffusion-type equations are used to describe the transfer of solutes from the larger pores into the micropores of the soil matrix. As an example, the governing equations for transport through media that contain parallel rectangular voids (figure 6.4) are given by (e.g., van Genuchten and Dalton, 1986).

\[
\frac{\partial f}{\partial t} + \frac{\partial y}{\partial x} \frac{\partial y}{\partial z} = \frac{\partial y}{\partial z} \left( \frac{\partial y}{\partial x} \right) - \frac{\partial y}{\partial x} \frac{\partial y}{\partial z}
\]  

(6.17)
Figure 6.4 Schematic of rectangular porous matrix blocks of width $2a$. The blocks, arranged as parallel slabs, are separated by a fracture pore system of width $2b$.

$$
\frac{3c_x}{\partial t} = D_a \frac{\partial^2 c_x}{\partial x^2}, \quad (-a \leq x \leq a) \tag{6.18}
$$

$$
c_m(z, t) = \frac{1}{a} \int_{-a}^{a} c_x(z, x, t) \, dx \tag{6.19}
$$

where the subscripts $f$ and $m$ refer to the interaggregate (fracture $f$) and intraaggregate (matrix $m$) pore regions, respectively, $c_x(z, x, t)$ is the local concentration in the aggregate, $x$ is the horizontal coordinate, and $D_a$ is the effective soil or rock matrix diffusion coefficient. Equation (6.17) describes vertical advective–dispersive transport through the fractures, while equation (6.18) accounts for linear diffusion in slab of width $2a$ in the horizontal ($x$) direction. Equation (6.19) represents the average concentration of the immobile soil matrix liquid phase. Equations (6.18) and (6.19) are coupled using the assumption of concentration continuity across the fracture–matrix interface:

$$
c_f(z, a, t) = c_f(z, t) \tag{6.20}
$$

The water contents $\theta_f$ and $\theta_m$ in equation (6.17) are given in terms of the bulk soil volume; that is,

$$
\theta_f = w_f \theta_f, \quad \theta_m = (1 - w_f) \theta_m \tag{6.21}
$$

where $w_f$ is the volume of the fracture pore system relative to that of the total soil pore system. The total water content ($\theta$) of the fracture–matrix system is given by the sum of $\theta_f$ and $\theta_m$.

Similar models to that above may be formulated for other aggregate or soil matrix geometries. Geometry-based transport models have been successfully applied to laboratory-scale experiments as well as to selected field studies that involve mostly saturated conditions. As an example, figure 6.5 shows calculated and observed Cl effluent curves from a 76-cm-long undisturbed column of fractured clayey till. The extremely skewed (nonsigmoidal) shape of the effluent curve is a direct result of water and dissolved chemical moving mostly through the fractures and bypassing the soil matrix, but with diffusion taking place between the fractures and the fine-textured matrix. Sudicky et al. (1985) also demonstrated the skewing effect of matrix
diffusion on the shape of an effluent curve by means of two-dimensional, saturated sandbox studies in which a Cl tracer migrated through a thin sand layer sandwiched between the two silt layers.

Several studies exist in which the above geometry-based approach has been extended to transient flow conditions. The approach assumes that the flow and transport equations of the macropore or fracture network of prescribed geometry can be solved simultaneously and in a fully coupled fashion with the corresponding equations for the porous matrix. Discrete-fracture numerical models of this type include those by Sudicky and McLaren (1992) for application to two-dimensional saturated flow and aqueous-phase transport problems, and those by Shikaze et al. (1994) for two-dimensional gas-phase flow and transport through a network of vadose zone fractures embedded in a variably saturated porous matrix. The discrete-fracture flow and transport model of Sudicky and McLaren (1992) was recently extended by Therrien and Sudicky (1996) to three dimensions and variably saturated conditions by solving the Richards equation both along the network of interconnected fracture planes and in the adjoining porous matrix. They superimposed a network of two-dimensional finite elements that represent the interconnecting fractures onto the mesh of three-dimensional elements that represent the matrix. Their fully coupled approach assumes continuity in pressure head and concentration at the fracture–matrix interface, thus permitting a simultaneous solution of the Richards and transport equations for both the fracture network and the porous matrix without a need to explicitly calculate fluxes between the two regions.

By solving the Richards equation also for the soil matrix region, the model of Therrien and Sudicky (1996) accounts for water flow into and through the matrix domain.
While geometry-based models are conceptually attractive, they may be too complicated for routine applications since structured field soils usually contain a mixture of aggregates of various sizes and shapes. More important, the problem of macropore and fluid flow continuity is not easily addressed with geometry-based flow models. The issue of fluid flow continuity may be especially critical in the vadose zone because of possible preferential flow and channeling within the fracture domain itself during unsaturated conditions. Also, preferential flow paths may well change with the degree of saturation during unsaturated flow. Some of these processes are more easily considered by using first-order models as discussed below.

Equivalent First-Order Exchange Models

Rather than using geometry-based transport models, many of the preferential flow features can also be accounted for by using models that assume simple first-order exchange of solutes by diffusion between the macropore (mobile) and micropore (immobile) liquid regions. The governing equations then become identical to those used previously for physical nonequilibrium transport; that is, equations (6.13) and (6.14). Ignoring the degradation terms in equations (6.13) and (6.14) and assuming steady-state water flow, the dual-porosity model becomes

$$\partial_t R_f \frac{\partial c_f}{\partial t} + \partial_t R_m \frac{\partial c_m}{\partial t} = \partial_j D_f \frac{\partial^2 c_f}{\partial z^2} - \partial_f \nu_f \frac{\partial c_f}{\partial z}$$

(6.22)

$$\partial_t R_m \frac{\partial c_m}{\partial t} = \alpha(c_f - c_m)$$

(6.23)

where \(\alpha\), as before, is a first-order solute mass transfer coefficient that characterizes diffusional exchange of solutes between the mobile and immobile liquid phases. Notice that equation (6.22) is identical to equation (6.17) for the rectangular geometry-based model. The mass transfer coefficient is of the general form

$$\alpha = \frac{\beta \rho_m D_a}{a^2}$$

(6.24a)

$$\alpha = \frac{\beta \rho_m D_a}{(\xi_a - 1)^2}$$

(6.24b)

where \(\beta\) is a geometry-dependent shape factor and \(a\) is the characteristic length of the aggregate (e.g., the radius of a spherical or solid cylindrical aggregate, or half the width of a rectangular aggregate). Equation (6.24b) holds for a hollow cylindrical macropore for which \(\xi_a \approx b/a\), where \(a\) now represents the radius of the macropore and \(b\) the outer radius of the cylindrical soil mantle that surrounds the macropore. The value of \(\beta\) ranges from 3 for rectangular slabs to 15 for spherical aggregates (Bolt, 1979; van Genuchten and Dalton, 1986; Sudicky, 1990).

Extension to Variably Saturated Flow

Different types of models have been proposed to extend the above first-order dual-porosity approach to variably saturated structured media (Wang, 1991; Zimmerman
et al., 1993). Here, we summarize the dual-porosity model developed by Gerke and van Genuchten (1993, 1996). This model assumes that the Richards equation for transient water flow and the advection–dispersion equation for solute transport can be applied to each of the two pore systems. Similar to the first-order mobile-immobile approach, water and solute mass transfer between the two pore systems is described with first-order rate equations. The flow equations for the fracture (subscript \( f \)) and matrix (subscript \( m \)) pore systems are, respectively,

\[
C_f \frac{\partial h_f}{\partial t} = \frac{\partial}{\partial z} \left( K_f \frac{\partial h_f}{\partial z} - K_f \right) - \frac{\Gamma_w}{w_f}
\]

(6.25)

\[
C_m \frac{\partial h_m}{\partial t} = \frac{\partial}{\partial z} \left( K_m \frac{\partial h_m}{\partial z} - K_m \right) + \frac{\Gamma_w}{1 - w_f}
\]

(6.26)

where \( \Gamma_w \) describes the rate of exchange of water between the fracture and matrix regions:

\[
\Gamma_w = \alpha_w (h_f - h_m)
\]

(6.27)

in which \( \alpha_w \) is a first-order mass transfer coefficient for water:

\[
\alpha_w = \frac{\beta}{\alpha} K_d y_w
\]

(6.28)

where \( \beta \) and \( \alpha \) are the same as before, \( K_d \) is the hydraulic conductivity of the fracture–matrix interface, and \( y_w \) (\( = 0.4 \)) is a dimensionless scaling factor. The solute transport equations for the fractures and matrix are given by, respectively,

\[
\frac{\partial}{\partial t} \left( \theta_f R_f c_f \right) = \frac{\partial}{\partial z} \left( \theta_f D_f \frac{\partial c_f}{\partial z} - q_f c_f \right) - \frac{\Gamma_s}{w_f}
\]

(6.29)

\[
\frac{\partial}{\partial t} \left( \theta_m R_m c_m \right) = \frac{\partial}{\partial z} \left( \theta_m D_m \frac{\partial c_m}{\partial z} - q_m c_m \right) + \frac{\Gamma_s}{1 - w_f}
\]

(6.30)

where \( \Gamma_s \) is the solute mass transfer term given by

\[
\Gamma_s = \alpha (c_f - c_m) + \begin{cases} 
\Gamma_w \frac{\partial c_f}{\theta}, & \Gamma_w \geq 0 \\
\Gamma_w \frac{\partial c_m}{\theta}, & \Gamma_w < 0 
\end{cases}
\]

(6.31)

in which \( \alpha \) is the same as used in the first-order mobile-immobile model. The first term on the right-hand side of equation (6.31) specifies the diffusion contribution to \( \Gamma_s \), while the second term gives the advective contribution. The above dual-porosity transport model reduces to the first-order model for conditions of steady-state flow in the fracture (macropore) region and no flow in the matrix pore system \( (q_m = \Gamma_w = 0) \).

The dual porosity model given by equations (6.25) through (6.31) contains two water-retention functions, one for the matrix and one for the fracture pore system, but three hydraulic conductivities functions: \( K_f(h_f) \) for the fracture network, \( K_m(h_m) \) for the matrix, and \( K_d(h) \) for the fracture–matrix interface. The \( K_f(h_f) \) function is determined by the structure of the fracture pore system; that is, the size, shape, continuity, and wall roughness of the fractures, and possibly the presence of fracture fillings. Similarly, \( K_m(h_m) \) is determined by the hydraulic properties of single matrix
blocks, and the degree of hydraulic contact between adjoining matrix blocks during unsaturated flow. Finally, \( K_d(h) \) is the effective hydraulic conductivity function to be used in equation (6.28) for describing the exchange of water between the two pore systems.

Several important features of preferential flow are illustrated here by using the above variably saturated dual-porosity model to calculate the infiltration of water at a constant rate of 50 cm/day into a 40-cm-deep structured soil profile that has an initially uniform pressure head of \(-1000\) cm. Water is allowed to infiltrate exclusively into the fracture pore system, thus assuming that the matrix pore system at the soil surface is sealed. The hydraulic properties of the fracture and matrix pore systems (figure 6.6) are indicative of relatively coarse- and fine-textured soils, respectively. The simulations assume a macroporosity of 5% \((\psi_f = 0.05)\), and rectangular aggregates \((\beta = 3)\) that have a width of 2 cm \((a = 1\) cm). The hydraulic parameters for \( K_d(h) \) were assumed to be the same as those for \( K_m(h_m) \), except for the saturated hydraulic conductivity, which was decreased by a factor of 100. Figure 6.7 shows simulated pressure head and water content distributions during infiltration. The results indicate a rapid increase in the pressure head of the fracture pore system, but a relatively slow response of the matrix (figure 6.7a). The resulting pressure head gradient between the two pore systems causes water to flow from the fracture into the matrix pore system (figure 6.7b), thus increasing the water content of the matrix (figure 6.7c). Significant pressure head differences between the two pore systems are still present when the infiltration front in the fracture system reaches the bottom of the soil profile after about 0.08 days (figure 6.7a). Notice that the water transfer rate, \( \Gamma_w \), is highest close to the infiltration front, and gradually decreases toward the soil surface (figure 6.7b). The shapes of the \( \Gamma_w \)-curves reflect the combined effects on \( \Gamma_w \) of the pressure head difference between the two pore regions (which decreases in time) and the value of \( K_d \) (which increases in time) at any point behind the wetting front.

Figure 6.8 shows the simulated concentration distributions. Results are for the infiltration of solute-free water into a structured medium that has a relative initial concentration of 1. As expected, the solute concentration in the fracture pore system initially decreases rapidly as solute-free water infiltrates (figure 6.8a). Water with a relatively low concentration subsequently flows from the fracture into the matrix pore system. At the same time, however, solutes begin to diffuse back from the matrix into the fracture pore system because of the large concentration gradients that develop between the two pore systems (figure 6.8a). The net solute transfer rate, \( \Gamma_s \), eventually becomes negative, indicating a net transfer from the matrix into the fracture pore system (figure 6.8b). The solute mass in the matrix pore system \((\theta_m c_m)\) initially decreases only slightly \((t = 0.01\) days in figure 6.8c), but starts to decrease more rapidly at later times \((t \geq 0.04\) days). The results in figure 6.8 illustrate the extremely transient and complicated nature of transport in a structured medium that involves vertical advective transport and dispersion, and horizontal mass transfer by advection and diffusion. Simulations such as those shown in figures 6.7 and 6.8 may be used to explain previously observed effects of several parameters on solute leaching during transient flow, including soil surface boundary condition (Bond and
Figure 6.6 (a) Water retention and (b) hydraulic conductivity functions of a dual-porosity medium that involves (1) the fracture network, (2) the matrix pore system, (3) the composite medium, and (4) the conductivity of the fracture-matrix interface. (After Gerke and van Genuchten, 1993.)
Figure 6.7 Simulated distributions versus depth of (a) the pressure head, \( h \), (b) the water transfer rate, \( \Gamma_w \), and (c) the volumetric water content, \( \theta \), for the matrix (dashed lines) and fracture (solid lines) pore systems at \( t = 0.01, 0.04, \) and \( 0.08 \) days. (After Gerke and van Genuchten, 1993.)
Figure 6.8 Simulated distributions versus depth of (a) the solute concentration, \( c \), (b) the solute transfer rates, \( \Gamma_r \), and (c) the solute mass, \( \vartheta c \), for the matrix (dashed lines) and fracture (solid lines) pore systems at \( t = 0.01, 0.04, \) and 0.08 days. (After Gerke and van Genuchten, 1993.)
Wierenga, 1990), water application rate (White et al., 1986; McLay et al., 1991), and initial condition (Kluitenberg and Horton, 1990).

The potential value of process-based preferential flow simulations is further illustrated in figure 6.9, which shows the sensitivity of the infiltration process to changes in the hydraulic conductivity $K_e$ of the fracture-matrix interface. Results obtained with a relatively large saturated conductivity, $K_{s,m}$, of 1 cm/day (equal to the matrix conductivity) closely approximate the limiting case of pressure head equilibrium (figure 6.9a) with little or no preferential flow. The moisture front in this case reached a depth of only 5 cm after 0.02 days. The water transfer rates (figure 6.9b) were so high that the two pore systems quickly approached equilibrium ($K_{s,m} = 1$ cm/day).

By comparison, for the smallest $K_{s,m}$ (0.001 cm/day), water percolated rapidly downward through the fracture pore system to a depth of 35 cm during the same time period ($t = 0.02$ days or 20 min). This last situation represents an extreme case of preferential flow with significant pressure head differences between the two pore systems (figure 6.9a).

The results in figure 6.9 indicate that equilibrium between the fracture and matrix pore systems should be expected when the hydraulic conductivity, $K_{s,m}$, of the matrix-fracture interface is roughly equal to the conductivity of the soil matrix (assuming a fracture spacing of 2 cm). For preferential flow to initiate in the present example, $K_{s,m}$ must be much less than $K_{s,m}$ of the matrix. This conclusion is consistent with experimental studies that suggest that a soil aggregate can have a higher local bulk density (and hence lower conductivity) near its surface than in the aggregate center, in part because of the deposition of organic matter, fine-texture mineral particles, or various oxides and hydroxides on the aggregate exteriors or macropore walls. For example, Wilding and Hallmark (1984) noted that ped argillans can markedly reduce rates of diffusion and mass flow from ped surfaces into the soil matrix. Cutans, which consist of coatings with modified physical, chemical, or biological properties, often have also preferred orientations parallel to soil aggregate surfaces. Unsaturated fractured rock formations may exhibit similar features—that is, fracture skins (Moench, 1984), or other types of coatings (Pruess and Wang, 1987) made up of fine clay particles, calcite, zeolites or silicates—which may reduce the hydraulic conductivity. Finally, preferential flow within the macropores or fractures themselves can also contribute to a lower effective $K_e(h)$. Situations like this can restrict water and solute exchange between the two pore systems (notably imbibition into the matrix) to only a small portion of the total interface area (Hoogmoed and Bouma, 1980), even in capillary-size pores (Omoti and Wild, 1979).

Hydrophobic fracture surfaces can similarly limit fluid exchange between the two pore systems.

Application of the variably saturated dual-porosity model requires several hydraulic and other parameters that are not easily measured. Estimates for the $K_f$- and $K_m$-functions (figure 6.6) may be obtained by assuming that $K_f$ is primarily the conductivity function in the wet range, while $K_m$ is the conductivity in the dry range (Peters and Klavetter, 1988; Othmer et al., 1991; Durner, 1994). Obtaining accurate estimates for the hydraulic properties of the fracture pore system from the composite curves requires that the hydraulic functions be very well defined in the wet range. This problem is indirectly demonstrated by figure 6.6a, which was obtained by assuming that the fracture pore system comprises 5% of the porous medium.
Figure 6.9 Simulated distributions versus depth of (a) the pressure head, $h$, (b) the water transfer rate, $T_w$, and (c) the total volumetric water content, $\theta$, for different values of the fracture-matrix interface hydraulic conductivity, $K_{f,m}$ ($t = 0.02$ days, $a = 1$ cm).
Notice that the retention function of the matrix differs only minimally from that of the composite medium. Hence, it may be very difficult, in practice, to estimate separate soil water-retention curves of the fracture and matrix pore systems by using bulk soil measurements that generally contain some noise. By contrast, it appears more promising to assess the contributions of macropores from carefully measured bulk hydraulic conductivity functions near saturation (e.g., Smettem and Kirby, 1990; Mohanty et al., 1997).

Finally, we note that the dual-porosity model discussed here assumes applicability of the Richards equation, and hence of Darcy's law. This assumption may not be strictly correct for the fracture pore system. However, given the uncertainties in all of the physical and chemical processes related to preferential flow, the real issue may not necessarily be the validity of Darcy's law as such, but whether Darcy's law—even if formally invalid—can still provide a useful qualitative description of the preferential-flow process. Alternative descriptions of the flow regime in fractures, such as Manning's equation for turbulent overland flow, kinematic wave theory, or simple gravity-flow models, may be too elaborate for routine use. Moreover, some of these approaches do not have provisions for flow to occur from the micropores back into the fractures—for example, at or near the bottom boundary of a coarse-textured soil horizon overlaying a fine-textured horizon.

Numerical Methods

A large number of analytical solutions have been published for one- and multidimensional transport problems (e.g., Javandel et al., 1984; Leij et al. 1993; Toride et al., 1993). While useful for simplified analyses, analytical solutions are generally not available for more complex situations, such as for transient variably saturated flow or situations that involve nonlinear sorption or degradation, in which case numerical models must be employed. In this section, we give a brief review of recent advances in numerical methods for solving subsurface flow and transport problems. We will not address issues that pertain to the discretization of multiphase "black oil" or compositional simulators. A detailed discussion of discretization issues for multiphase compositional problems can be found in Unger et al. (1996).

Numerical Solution of the Richards Equation

A variety of numerical methods may be used to solve the variably saturated flow and transport equations (e.g., Huyakorn and Pinder, 1983; Sudicky and Huyakorn, 1991). Early numerical variably saturated flow models generally used classical finite difference methods. Integrated finite differences (Narasimhan and Witherspoon, 1976), control-volume finite element techniques (Forsyth 1991; Therrien and Sudicky, 1996), and Galerkin finite element methods (Huyakorn et al., 1986; Šimůnek et al., 1994) became increasingly popular since the mid-1970s. Time and space discretization of the Richards equation using any of these methods leads to a nonlinear system of algebraic equations. These equations are most often linearized and solved using the Newton-Raphson or Picard iteration methods. Picard iteration
is widely used because of its ease of implementation, and because this method preserves symmetry of the final system of matrix equations. The Newton-Raphson iteration procedure is more complex and results in nonsymmetric matrices, but often achieves a faster rate of convergence and may be more robust than Picard iteration for highly nonlinear problems (Paniconi and Putti, 1994; Forsyth et al., 1995). In principle, the Picard scheme is linearly convergent, and therefore should converge more slowly than the quadratically convergent Newton-Raphson scheme.

The basic approach for discretizing and solving the Richards equation depends upon the flow formulation being used—that is, the $h$-based, the $\theta$-based, or the mixed formulation. Celia et al. (1990) suggested that numerical solutions based on the standard $h$-based formulation of the Richards equation often yield poor results, characterized by large mass balance errors and incorrect estimates of the pressure head distributions in the soil profile. They solved the mixed formulation of the Richards equation using a modified Picard iteration scheme which possesses mass-conserving properties for both finite element and finite difference spatial approximations. Therrien and Sudicky (1996) also solved the mass-conservative mixed form of the Richards equations, but implemented the more robust Newton-Raphson linearization method and a highly efficient algorithm to construct the Jacobian matrix (Forsyth and Simpson, 1991). Millly (1985) earlier presented two mass-conservative schemes for computing nodal values of the water capacity in the $h$-based formulation to force global mass balance. Several numerical schemes based on different types of pressure head transformations were recently also presented (Hills et al., 1989; Ross, 1990; Pan and Wierenga, 1995). Hills et al. (1989) showed that the $\theta$-based form of the Richards equation can yield fast and accurate solutions for infiltration into very dry heterogeneous soil profiles. However, the $\theta$-based numerical scheme cannot be used for soils that have saturated regions. Kirkland et al. (1992) expanded the work of Hills by combining the $\theta$-based and $h$-based models to yield a transformation applicable also to variably saturated systems. They defined a new variable that is a linear function of the pressure head and water content in the saturated and unsaturated zone, respectively. More recently, Forsyth et al. (1995) proposed a robust and highly efficient algorithm in which variable substitution is used to switch between $\theta$ or $h$ as the primary variables when constructing the Jacobian matrix for Newton-Raphson iteration. The primary variable switch is made after each Newton iteration in different parts of the computational domain as a function of the state of the degree of saturation in those parts. Using this approach, and also by employing a monotone discretization (i.e., upstream weighting of relative permeabilities) that guarantees that saturations always remain in the physical range, they demonstrated that an order-of-magnitude execution speedup can be achieved for difficult problems that involve infiltration into dry, heterogeneous soils. They also pointed out that the method of Kirkland et al. (1992) is not necessarily monotone because of its partially explicit nature, and that mass balance errors can occur at the transition between the saturated and unsaturated zones. Upstream weighting of relative permeabilities is very much recommended over central weighting since the latter method can introduce oscillations (i.e., negative saturation values) for difficult problems, and thus can cause complete failure of the nonlinear iteration process (Forsyth, 1991; Therrien and Sudicky, 1996). Because of the self-sharpening properties of soil moisture fronts,
the use of upstream weighting in the numerical solutions of the Richards equation generally does not introduce excessive smearing of saturation fronts, unlike its use in solutions of the advection-dispersion equation.

Numerical Solution of the Transport Equation

Numerical methods for solving the advection-dispersion equation may be classified into three groups: (1) Eulerian, (2) Lagrangian, and (3) mixed Lagrangian-Eulerian methods. In the Eulerian approach, the transport equation is discretized by means of a usual finite difference or finite element method that uses a fixed grid system. For the Lagrangian approach, the mesh either deforms and moves along with the flow path, or the mesh is assumed to be stable in a deforming coordinate system. A two-step procedure is followed for a mixed Lagrangian-Eulerian approach. First, advective transport is considered, using a Lagrangian approach in which Lagrangian concentrations are estimated from particle trajectories. Subsequently, all other processes, including sinks and sources, are modeled using the standard Eulerian approach that involves any finite element or finite differences method, thus leading to the final concentrations.

Standard finite difference and Galerkin or control-volume type finite element methods belong to the first group of Eulerian methods. Finite differences and finite elements methods provided the early tools for solving solute transport problems and, in spite of some limitations as discussed below, are still the most popular methods being used at present. Numerical studies have shown that both methods give good results for transport problems where dispersion is relatively dominant as compared with advective transport (e.g., as indicated by the grid Péclet number). However, both methods can lead to significant numerical oscillations and/or dispersion for advection-dominated transport problems. The Eulerian methods have been found to be very reliable and accurate when applied to quasi-symmetric problems when diffusion dominates the transport process. The advection term brings non-symmetry into the governing solute transport equation and, as a result, compromises the success of Eulerian methods when applied to advection-dominated transport problems. By selecting an appropriate combination of relatively small space and time steps, it is still possible to virtually eliminate most or all oscillations. Alternatively, the spatial grid system may be refined using a “zoomable hidden fine-mesh” approach (Yeh, 1990), or by implementing local adaptive grid refinement (Wolfsberg and Freyberg, 1994). However, there is an additional computational cost with this approach, and the handling of natural grid irregularities due to material heterogeneity or other domain features can be problematic. Criteria for minimizing or eliminating oscillations and reducing numerical dispersion when solving the linear advection-dispersion equations are well known; that is, the product of the local Péclet ($\nu\Delta x/D$) and Courant ($\nu\Delta t/\Delta x$) numbers should be less than 2 ($\Delta t$ is the time step and $\Delta x$ is the nodal spacing). When small oscillations in the solution can be tolerated, this criterion can be increased to about 5 or 10. Monotonicity conditions and numerical smearing are also influenced by the type of temporal discretization being used. For example, while fully implicit time-weighting schemes are monotone (i.e., concentrations always fall within the physical range), they are more prone to
numerical dispersion than central-in-time (Crank-Nicolson type) weighting which is second-order correct. By comparison, monotonicity cannot be guaranteed with central weighting unless the grid Peclet and Courant criteria are appropriately satisfied. We refer to Unger et al. (1996) for a more thorough discussion of discretization strategies that involve both the linear and nonlinear forms of the transport equation in the context of multiphase compositional modeling. In particular, they discuss the use of flux limiters and total variation diminishing (TVD) schemes for reducing numerical dispersion in nonlinear multiphase compositional transport problems.

Upstream weighting methods virtually eliminate numerical oscillations, even for purely advective transport, but a disadvantage is that they may create unacceptable numerical dispersion. Huyakorn and Nilkuha (1979) and Yeh (1986) used weighting functions that are different for the spatial derivatives than for other terms in the finite element solution of the transport equation. Their approach places greater weight on the upstream nodes within a particular element. Huyakorn and Nilkuha (1979) suggested, for this purpose, nonorthogonal basis functions, whereas Yeh (1986) used orthogonal functions. Petrov-Galerkin methods require the use of higher order weighting functions, which makes their implementation more difficult and more costly than classical Galerkin finite element methods. Another alternative for overcoming numerical dispersion is the use of higher order temporal and spatial approximation (e.g., van Genuchten and Gray, 1978). Such higher order approximations, however, are computationally more expensive and often produce numerical oscillations.

While Lagrangian methods (or methods of characteristics) may substantially reduce or even eliminate problems with numerical oscillations (e.g., Neuman and Sorek, 1982), they can also introduce other problems, such as nonconservation of mass. Lagrangian methods are also relatively difficult to implement in two and three dimensions when an unstructured (nonrectangular) spatial discretization scheme is used. Instabilities that result from inappropriate spatial discretization sometimes occur during longer simulations as a result of a deformation of the stream function. Furthermore, unrealistic distortions of the results may occur when modeling the transport of solutes that are undergoing certain sorption/exchange or precipitation reactions.

Mixed Eulerian-Lagrangian approaches have been reported by several authors (e.g., Molz et al., 1986; Sorek, 1988; Yeh, 1990, among others). In view of the different mathematical character of the diffusive (parabolic) and advective (hyperbolic) terms in the advection-dispersion equation, the transport equation can be decomposed into a mixed problem that consists of a purely advection problem, followed by a diffusion-only problem. Advective transport then is solved with the Lagrangian approach, while all other terms of the transport equation are solved using Eulerian methods. The trajectories of the flowing particles may be obtained in a variety of ways. For example, Molz et al. (1986) used single-step reverse particle tracking in which the initial position of particles arriving at the end of a time step at fixed nodal points is calculated for each time step. The use of continuous forward particle tracking has similar disadvantages as the Lagrangian approach since complex geometric regions are, again, difficult to handle. To obtain good results, it may
be necessary to follow a large number of particles, thereby quickly leading to excessive computer time and memory.

Still other solution methods exist, such as the use of a combination of analytical and numerical techniques. For example, Sudicky (1989) and Sudicky and McLaren (1992) modeled solute transport by using Laplace transforms with respect to time and Galerkin finite elements for the spatial domain. The use of Laplace transforms avoids the need for intermediate simulations (time-stepping) between the initial condition and the points in time for which solutions are needed, while also needing less stringent requirements for the spatial discretization. These features lead to computational efficiency, especially for large-time simulations. Recently, several methods have also been suggested that make use of local analytical solutions of the advection–dispersion equation in combination with finite differences (Li et al., 1992). We emphasize that combinations of analytical and numerical techniques have one important limitation. Because, for example, the Laplace transform eliminates time as an independent variable in the governing solute transport equation, all coefficients, such as water content, flow velocity, and retardation factors, must be independent of time. This limitation precludes the use of combination methods for solving coupled, transient variably saturated flow and transport problems typical of many field situations.

Matrix Equation Solvers

Discretization of the governing partial differential equations for water flow and solute transport generally leads to a system of linear matrix equations:

\[ [A][x] = [b] \] (6.32)

in which \([x]\) is an unknown solution vector, \([b]\) is the known right-hand side vector of the matrix equation, and \([A]\) is a sparse banded matrix that is symmetric for water flow if the modified Picard procedure is used, but asymmetric for water flow if the Newton-Raphson method is used. Matrix \([A]\) is generally asymmetric for solute transport, unless advection is not considered in the formulation. Technological breakthroughs in computer hardware and increased incentives to simulate complex coupled flow and transport problems in large three-dimensional systems has spurred the development and use of highly efficient and robust iterative matrix solvers. Robustness of the solver is essential to handle stiff matrices that result from extreme contrasts in material properties and, in the situation of variably saturated flow, severe nonlinearity.

Traditionally, matrix equations have been solved by means of such direct methods as Gaussian elimination and LU decomposition. Although these methods usually take advantage of the banded nature of the coefficient matrices, they have several disadvantages as compared with iterative methods. For example, for two-dimensional problems, the operation count for a direct solver increases approximately by the square of the number of nodes, whereas for an iterative solver the operation count is typically 1.5 or less (Mendoza et al., 1991; VanderKwaak et al., 1995, among others). A similar reduction also holds for the
memory requirement since iterative methods do not require storage of nonzero matrix elements. Memory requirements, therefore, increase at a much smaller rate with the size and dimensionality of a problem when iterative solvers are used (Mendoza et al., 1991; VanderKwaak et al., 1995). Round-off errors also represent less of a problem for iterative methods as compared with direct methods. This is because round-off errors in iterative methods are self-correcting (Letniowski, 1989). Finally, for time-dependent problems, a reasonable approximation of the solution (i.e., the solution at the previous time step) exists for iterative methods, but not for direct methods (Letniowski, 1989). In general, direct methods are more appropriate for relatively small problems that involve a couple of thousand nodes, while iterative methods are more suitable for the larger problems. This issue is of critical importance given that many research problems, as well as certain practical vadose zone field applications, may require discretizations that involve tens to hundreds of thousands of nodes.

Many iterative methods have been used in the past for handling large sparse matrix equations. These methods include Jacobi, Gauss-Seidel, alternating direction implicit (ADI), successive overrelaxation (SOR), block successive overrelaxation (BSSOR), successive line overrelaxation (SLOR), and strongly implicit procedures (SIP), among others (Letniowski, 1989). More powerful preconditioned accelerated iterative methods, such as the preconditioned conjugate gradient method (PCG) (Meijerink and van der Vorst, 1977; Kershaw, 1978; Behie and Forsyth, 1983), were introduced more recently. Sudicky and Huyakorn (1991) gave three advantages of the PCG procedure as compared with other iterative methods: (1) PCG can be readily modified for finite element methods with irregular grids, (2) the method does not require arbitrary iteration parameters, and (3) PCG usually outperforms its iterative counterparts for situations that involve relatively stiff matrix conditions. The PCG methods can be used only for symmetric matrices. Since the system of linear equations that results from discretization of the solute transport equation is nonsymmetrical (the same is true when linearizing numerical solutions of the Richards equation using Newton-Raphson iteration), it is necessary to either formulate the transport problem such that a symmetric matrix results (Leismann and Frind, 1989), or use an extension of PCG for nonsymmetrical matrices. Examples for such an extension are the ORTHOMIN (generalized conjugate residual method, Behie and Forsyth, 1984), GMRES (generalized minimal residual method, Saad and Schultz, 1986), CGSTAB (conjugate gradient stabilized method, van der Vorst, 1992), or the conjugate gradient squared (Letniowski, 1989; Paniconi and Putti, 1994) procedures.

Competitive iterative methods generally involve two operations: (1) initial preconditioning, and (2) iterative solution with a particular acceleration method—such as CGSTAB. Incomplete lower-upper (ILU) factorization (among other methods, such as incomplete Cholesky for symmetric matrices) can be used to precondition matrix [4]. This factorization leads to more easily inverted lower and upper triangular matrices by partial Gaussian elimination. The preconditioned matrix is subsequently repeatedly inverted using updated solution estimates, thus leading to a new approximation of the solution. More details about the CGSTAB and ORTHOMIN methods are given in the user’s guides of the WATSOLV (VanderKwaak et al., 1995)
and ORTHOFEM (Mendoza et al., 1991). Letniowski (1989) also gave a comprehensive review of accelerated iterative methods, as well as preconditional techniques. The ORTHOMIN procedure is becoming increasingly popular in variably saturated flow and contaminant transport simulations (Gambolati et al., 1986; Kirkland et al., 1992; Šimánek et al., 1994; Thierien and Sudicky, 1996). Recent numerical experimentation for problems that involve difficult-to-solve variably saturated flow problems indicate that CGSTAB outperforms ORTHOMIN. This is because ORTHOMIN sometimes tends to stagnate and hence fails to satisfy a specified convergence criterion after many iterations.

Concluding Remarks

The past few decades have produced tremendous advances in our ability to mathematically describe and simulate vadose zone flow and chemical transport processes. Careful laboratory and field experimentation has yielded much new information, not only on the form of fundamental constitutive relations, but also on the controlling effects of heterogeneities, fractures, and macropores on flow and transport at the field scale. Process-oriented deterministic and stochastic theories have achieved reasonable success in providing new qualitative and quantitative information on the hierarchical nature of heterogeneities and the scaling-up of relevant parameters and constitutive relations for use in large-scale simulators. The development and implementation of modern numerical algorithms for solving the nonlinear Richards equation in an efficient and robust fashion now makes it possible to routinely simulate large-grid, three-dimensional vadose flow problems on modern workstations. Similar strides in algorithm development have been made with regard to solutions of the solute transport equation and, for example, its coupling to multicomponent geochemical speciation models. One important key enabling us to handle increasingly larger scale three-dimensional flow and transport problems has been the implementation of highly efficient iterative sparse-matrix equation solvers, such as those based on ORTHOMIN and CGSTAB acceleration.

In spite of the modeling advances, much remains to be done. For example, field testing of recently developed scale-up theories for application to heterogeneous field settings is generally lacking, and numerical models designed to predict vadose zone flow and chemical transport processes often yield only qualitative similarities to real-world field observations. Attempts to improve the predictive capabilities of recent models have typically involved the introduction of additional fitting parameters that are elusive and perhaps impossible to measure independently. This aspect is disconcerting given that practitioners, planners, and regulators are increasingly relying upon model predictions to establish far-reaching policies. It is our belief that a harmonious blend between laboratory research, field-scale experimentation, and modeling-based research is key to at least maintaining, and perhaps accelerating, the pace of advancement that we have seen over the last couple of decades.
References


Freundlich, H., 1909, Kapillardchemie; eine Darstellung der Chemie der Kolloide und verwandter Gebiete, Akademische Verlagsgesellschaft, Leipzig, Germany.


