EVALUATION OF POLLUTANT TRANSPORT IN THE UNSATURATED ZONE

M. Th. VAN GENUCHTEN and J. ŠIMŮNEK
U.S. Salinity Laboratory, USDA, ARS
450 West Big Springs Road, Riverside, CA, 92507-4617, USA

Abstract

The past several decades has seen considerable progress in the conceptual understanding and mathematical description of water flow and solute transport processes in the unsaturated or vadose zone. A variety of analytical and numerical models are now available to predict water and/or solute transfer processes between the soil surface and the groundwater table. The most popular models remain those based on the Richards equation for variably-saturated flow, and the Fickian-based convection-dispersion equation for solute transport. Deterministic solutions of these classical equations likely will continue to be used in the near future for predicting water and solute movement in the unsaturated zone. In this paper we review recent developments in variably-saturated flow and transport research, especially from the point of view of process-based modeling. Among the topics being discussed are single-ion equilibrium transport, physical and chemical nonequilibrium transport, volatilization and degradation, multiple-species solute transport, parameter estimation, and recent advances in numerical modeling. Also briefly discussed are alternative stochastic approaches for evaluating contaminant transport in spatially variable field soils.

1. Introduction

The fate and transport of a variety of chemicals migrating from industrial and municipal waste disposal sites, or intentionally applied to agricultural lands, is increasingly becoming a concern. Once released into the subsurface environment, these chemicals are subject to a large number of simultaneous physical, chemical, and biological processes, including sorption-desorption, volatilization, photolysis, and biodegradation. For example, the persistence of many organic chemicals, including pesticides, in soils is largely determined by the extent of degradation, sorption and volatilization. Several processes such as liquid and gaseous molecular diffusion, and convective-dispersive transport, act only on solutes that are not adsorbed. Degradation of organic compounds likely occurs mainly, or even exclusively, in the liquid phase. On the other hand, radioactive decay takes place equally in the solution and adsorbed phases, while other reactions may occur...
only in the sorbed phase. Depending upon the type of organic chemical, transport may also be subject to multiphase flow involving 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 physical and chemical processes affecting pollutant 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 models. 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 lumped into a single distribution coefficient, $k_d$, and possibly a nonlinear exponent. Other processes such as precipitation-dissolution, biodgradation, volatilization or radioactive decay were generally simulated by means of simple first- and/or zero-order rate processes. These simplifying approaches were needed so as 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 equilibria and nonequilibrium models is now increasingly being addressed, facilitated in part by the introduction of more powerful computers, and the concomitant development of more advanced numerical techniques.

In the first part of this paper we will describe several conceptual approaches for modeling solute transport in variably-saturated media, and discuss some of their limitations. Next, we will 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 discretizations. We also briefly discuss the problem of subsurface heterogeneity and alternative stochastic approaches for quantifying solute transport in such systems.

2. Water Flow and Single Species Solute Transport

2.1. GOVERNING EQUATIONS

Deterministic descriptions of water and solute movement in the vadose zone are generally based on the classical Richards equation for variably-saturated water flow and the convection-dispersion equation for solute transport. For one-dimensional vertical transfer, these equations are respectively

$$C(h) \frac{dh}{dt} = \frac{\partial}{\partial z} \left( K(h) \frac{dh}{dz} K(h) \right) + S$$

(1)
\[ \frac{\partial (\rho z)}{\partial t} + \frac{\partial (\theta c)}{\partial t} - \frac{\partial (\theta D \frac{\partial c}{\partial z} - q c)}{\partial z} + \Phi = 0 \] (2)

where \( \theta \) is the soil water capacity, \( z \) is the slope of the soil water retention curve, \( \theta (0) \); \( 0 \) is the volumetric water content, \( k \) is the soil water pressure head (negative for unsaturated conditions), \( z \) is time, \( z \) is distance from the soil surface downward, \( K \) is the hydraulic conductivity as a function of \( h \) or \( \theta \), \( z \) 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 sources and sinks for water and solutes, respectively, and \( q \) is the volumetric fluid flux density given by Darcy’s law as
\[ q = -K(\theta) \frac{\partial h}{\partial \theta} \]

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., \( s = K_d c \)), Eq. (2) reduces to the simpler form
\[ \frac{\partial (\rho z c)}{\partial t} + \frac{\partial (\theta D \frac{\partial c}{\partial z} - q c)}{\partial z} + \Phi = 0 \] (4)

where \( R = 1 + \rho K_d \theta \) is the solute retardation factor. For conditions of steady-state water flow in homogeneous soils, and neglecting the source/sink terms \( S \) and \( \Phi \), Eq. (4) further reduces to the standard convection-dispersion equation:
\[ \frac{\partial (\rho z c)}{\partial t} + \frac{\partial (\rho c \rho z c)}{\partial z} - \frac{\partial (\rho D \frac{\partial c}{\partial z} - q c)}{\partial z} = 0 \] (5)

where \( \rho = \rho(\theta) \) is the average pore water velocity.

While models based on Eqs. (1) and (2) have proved to be important tools in research and management, they are subject to a large number of complications and simplifying assumptions which compromise or limit their applicability [115]. It may be instructive in list here some of those assumptions and limitations. For example, the equations assume that (a) 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, (b) Darcy’s equation is valid at both very low and very high flow velocities (including those occurring in structured soils), (c) the osmotic and electro-chemical gradients in the soil water potential are negligible, (d) the fluid density is independent of the solute concentration, and (e) matrix and fluid compressibilities are relatively small. The equations are further complicated by (f) the hysteretic nature of the retention function, \( \theta(\theta) \), (g) the extreme nonlinearity of the hydraulic conductivity function, \( K(\theta) \), (h) the lack of accurate and cheap methods for measuring the unsaturated hydraulic properties,
(i) the extreme heterogeneity of the subsurface environment, and (ii) inconsistencies between the scale at which the hydraulic and solute transport parameters in Eqs. (1) and (2) are usually measured, and the scale at which the predictive models are being applied. In addition, Eqs. (1) and (2) are formulated for inothermal 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 [32, 67] and the concentration and ionic composition of the soil liquid phase [36, 149] on the unsaturated soil hydraulic properties. Hence, a complete description of unsaturated zone transfer processes requires also consideration of heat flow and its nonlinear effect on most processes taking place in the soil-plant system.

2.2. ROOT WATER UPTAKE

An important term in the variably-saturated flow equation (Eq. 1) is the source/sink term \( S \) which may be used to account for water uptake by plant roots. Widely different approaches exist for simulating water uptake [106]. Many of the early studies of root water uptake [183, 145] have used uptake functions of the form

\[
S(x, t) = -b_h(x) K(0) \left[ h_h - h(x, t) \right]
\]

(6a)

where \( h_h \) is an effective root-water pressure head at the root surface, and \( b_h \) is a depth-dependent proportionality constant often referred to as the root effectiveness function. Equation (6a) 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 [144, 158]:

\[
S(x, t) = -b_h(x) h(x, t) \cdot T_p
\]

(6b)

where \( b_h \) is the potential root water uptake distribution function which integrates to unity over the soil root zone, \( \alpha_h \) is a dimensionless water stress response function between 0 and 1, and \( T_p \) is the potential transpiration rate. The effects of salinity on water uptake have been accounted for by linearly adding the osmotic head, \( \pi \), to the pressure head, \( h(x, t) \) in Eq. (6a) or (6b) [14, 23, 24, 25], or by incorporating into Eq. (6b) a separate salinity stress response function, \( \alpha_s(h) \), somewhat similar to \( \alpha_h(h) \) [105, 149]. Cardon and Letey [23] showed that approaches based on Eq. (6b) may be more appropriate than (6a), particularly if suitably modified and used for saline conditions. Still, as pointed out by Nieber et al. [115], 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 to develop realistic process-based descriptions of root growth and root water uptake as a function of various stresses (water, salinity, temperature; others) in the root zone, and to couple these descriptions with suitable crop growth models.
2.3. LINEAR EQUILIBRIUM SOLUTE TRANSPORT

Several simplifying assumptions are usually invoked when defining the sorption/exchange term \( \delta \) in Eq. (2). Most often, a linear equilibrium isotherm, \( \delta = k_{eq}c \), is used to describe solute interactions between the liquid and solid phases of the soil, leading to a constant retardation factor \( R \) in Eq. (4). The resulting convection-dispersion equation (CDE) given by Eq. (5) has been reasonably successful in describing displacement data for relatively uniform laboratory or field soils. Figure 1 shows solute breakthrough curves typical of the transport of an excluded anion, \( \text{Cl}^- \), a nonreactive solute (irradiated water, \( \text{H}_2\text{O} \)), and an adsorbed tracer, \( \text{Cr}^{6+} \), through homogeneous media. The first two tracers pertain to transport through 30-cm long columns containing disturbed Glendale clay loam (Wierenga, 1972, unpublished data; 168), while the \( \text{Cr}^{6+} \) data are for transport through a 5-cm long column of sand (Wierenga, 1972, unpublished data). The data in Figure 1 are plotted versus number of pore volumes \( (T = v(t)/v) \) of tracer solution leached through the columns. Analysis of the breakthrough curves in terms of the CDE using inverse procedures [163] yielded \( R \)-values of 0.681, 1.027 and 1.248, respectively, for the three tracers. Hence, \( \text{Cl}^- \) was strongly affected by anion exclusion \((P < 1; k_p < 0)\) resulting from the repulsion of chloride ions away from negatively charged surfaces of clays and ionicizable organic matter. Because water flow velocities are zero along pore walls, and maximum in the center of pores, anions such as \( \text{Cl}^- \) can travel much faster than water, especially in fine-textured soils.
2.4. NONLINEAR ADSORPTION

While the use of a linear isotherm can greatly simplify the mathematics of a transport problem, sorption and exchange reactions are generally nonlinear and usually depend also on the presence of competing species in the soil solution. The solute retardation factor for nonlinear adsorption is not constant, as in the case for linear adsorption, but changes as a function of the slope of the adsorption isotherm:

\[ R = 1 + \frac{D \frac{dc}{dc}}{0} \]  

(7)

A variety of models have been used in the past to describe the nonlinear nature of adsorption desorption processes. Table 1 lists some of the most commonly used sorption models relating the adsorbed concentration, \( t_s \), to the total concentration, \( c \), of an ion in solution. Although several of the equations in Table 1 (e.g., the Langmuir and Freundlich models) can be derived rigorously such as for the adsorption of gases onto solids, the expressions are generally used only in a purely empirical fashion. Of the equations listed in Table 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. [53]. They divided possible adsorption processes into four main classes: S, L (Langmuir), H (high affinity), and C (constant partitioning) isotherms, and discussed mechanisms explaining the different types of isotherms. Increasing solution concentrations result in increasing or decreasing adsorption rates for the S and L isotherms, respectively. An H isotherm is characterized by extremely high affinities of the exchanger for exchangeable ions, whereas a C isotherm results from constant partitioning of the solute between the solution and adsorbed phases. The shape of an isotherm can have significant effects on 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 \( k_L < 1 \)) producing relatively sharp concentration fronts during transport in a soil profile. The effects of isotherm nonlinearity on solute front sharpening are discussed at length by Hilt [9], van Genuchten and Cleary [108] and, more recently, by van der Zee and van Rimsdijk [110].

2.5. 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.
<table>
<thead>
<tr>
<th>Equation</th>
<th>Model</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q = k_1 c + k_2$</td>
<td>Linear</td>
<td>Lapidot and Amielzamk [87] Lindstrom et al. [98]</td>
</tr>
<tr>
<td>$q = k_1 c^b$</td>
<td>Freundlich</td>
<td>Freundlich [48]</td>
</tr>
<tr>
<td>$q = q_0 + q_e (1 - e^{-k_1 c})$</td>
<td>Langmuir</td>
<td>Langmuir [96]</td>
</tr>
<tr>
<td>$q = k_1 c (1 + k_q c)^{-n}$</td>
<td>Freundlich-Langmuir</td>
<td>Sips [1144]</td>
</tr>
<tr>
<td>$q = k_1 c (1 + k_q c)^{-1} + k_2 c$</td>
<td>Double Langmuir</td>
<td>Shapiro and Fried [138]</td>
</tr>
<tr>
<td>$q = k_1 c^{n_m}$</td>
<td>Extended Freundlich</td>
<td>Sibbenton [159]</td>
</tr>
<tr>
<td>$q = \frac{k_1 c}{1 + k_q c + k_q^{m_p} c^{m_p}}$</td>
<td>Gouary</td>
<td>Gouary [61]</td>
</tr>
<tr>
<td>$q = k_1 c^{n_m} + k_2 c$</td>
<td>Finet-Sumas</td>
<td>Finet and Setton [46]</td>
</tr>
<tr>
<td>$q = k_1 (1 - 1 + k_q c^{n_q})$</td>
<td>Hart</td>
<td>Hart [5]</td>
</tr>
<tr>
<td>$q = \frac{RT}{k_b} \ln (1/k_1 c)$</td>
<td>Temkin</td>
<td>Bache and Williams [4]</td>
</tr>
<tr>
<td>$q = k_1 c \exp (-2k_1 c)$</td>
<td>Lindstrom et al. [97] van Genuchten et al. [130]</td>
<td></td>
</tr>
<tr>
<td>$q = q_e (c + k_2 (c_{max} - c))$</td>
<td>modified Kieland</td>
<td>Lai and Jurinak [85]</td>
</tr>
</tbody>
</table>

$k_1, k_2, k_3, k_4$: empirical constants
$R$: universal gas constant
$T$: absolute temperature
$c_{max}$: maximum solute concentration
$c$: maximum solute concentration

Most of the early attempts to model nonequilibrium transport invoked relatively simple first-order (one-site) kinetic rate equations. More refined nonequilibrium models subsequently used the assumptions of two-site or multi-site sorption, and/or two-region
(dual-porosity) transport involving solute exchange between mobile and relatively immobile liquid regions. Models of this type have resulted in better descriptions of observed laboratory and field transport data, in part by providing additional degrees of freedom (fitting parameters) for describing observed concentration distributions. The different modeling approaches are briefly summarized below.

2.5.1. One-Site Sorption Models

The simplest nonequilibrium formulation arises when a chemically controlled first-order linear kinetic rate process is assumed. Ignoring any solute production or decay processes, Eq. (2) in that case is augmented with the equation

\[
\frac{\partial s}{\partial t} = \alpha \left( k_p c - s \right)
\]

(8)

where \(\alpha\) is a first-order kinetic rate coefficient. Transport models using the above one-site kinetic sorption equation have resulted in modest improvements in terms of their ability to match observed displacement data [37, 170]. Success in general was limited only to experiments conducted at relatively low pore water velocities where the equilibrium model already did perform reasonably well. Moreover, several of the sorption parameters (\(k_p\) and \(\alpha\)), when adjusted to get better transport predictions, were found to frequently vary as a function of the pore water velocity. Similar conclusions hold for most or all of the other nonequilibrium rate expressions listed in Table 2.

2.5.2. Two-Site Chemical Nonequilibrium Transport

The one-site first-order kinetic model may be expanded into a two-site sorption concept by assuming that the sorption sites can be divided into two fractions [137]: 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 [173]

\[
\frac{\partial}{\partial t} \left[ (0.7p_0k_{pc} - c) \right] + \frac{\partial}{\partial z} \left[ (0.3p_0k_{pc} - c) \right] - \alpha P \left[ (1 - f)k_{pc} - c \right] = - \frac{\partial}{\partial z} \left[ k_p c - s \right] + \alpha P \left[ (1 - f)k_{pc} - s \right]
\]

(9)

\[
\frac{\partial s_1}{\partial t} = \alpha \left[ (1 - f)k_{pc} - s_1 \right] - \alpha s_1
\]

(10)

where \(p_0\) and \(s_1\) 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 equilibrium, and the subscripts \(e\) and \(k\) 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 involving a variety of organic and inorganic chemicals [31, 129, 65, 122].
<table>
<thead>
<tr>
<th>Equation</th>
<th>Model</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{dx}{dt} = \alpha \left( k_c c + k_e^{-1} - x \right)$</td>
<td>Linear</td>
<td>Lapidus and Arrott [67]</td>
</tr>
<tr>
<td>$\frac{dx}{dt} = \alpha \left( k_e c^{-1} - x \right)$</td>
<td>Fickian</td>
<td>Homoky and Davidson [68]</td>
</tr>
<tr>
<td>$\frac{dx}{dt} = \alpha \left( \frac{k_c c}{1 + k_c c} \right)$</td>
<td>Langmuir</td>
<td>Hancké [63]</td>
</tr>
<tr>
<td>$\frac{dx}{dt} = \alpha \left( \frac{k_e c^{-1}}{1 + k_e c^{-1}} \right)$</td>
<td>Fickian-Langmuir</td>
<td>Snijkerk and van Genuchten [642]</td>
</tr>
<tr>
<td>$\frac{ds}{dt} = \alpha \left( s_x - s \right) \left( \frac{s_x - s}{s_y - s_y} \right)$</td>
<td>Fickian</td>
<td>Fick and Frenkel [62]</td>
</tr>
<tr>
<td>$\frac{dx}{dt} = \alpha \exp(k_c c) \left( \frac{k_c c}{1 + k_c c} \right)$</td>
<td>Lineat</td>
<td>Lineat</td>
</tr>
<tr>
<td>$\frac{dx}{dt} = \alpha \left( \frac{k_e c^{-1}}{1 + k_e c^{-1}} \right)$</td>
<td>Fickian-Langmuir</td>
<td>Fickian-Langmuir</td>
</tr>
<tr>
<td>$\frac{ds}{dt} = \alpha \left( s_x - s \right) \left( \frac{s_x - s}{s_y - s_y} \right)$</td>
<td>Fickian</td>
<td>Fickian</td>
</tr>
</tbody>
</table>

2.5.3. 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 a first-order exchange process [175, 181]. Using the same notation as before, the two-region transport model is given by [173]

$$\frac{\partial}{\partial t} \left[ (1-f) \rho_k \mu_{m} \right] - \frac{\partial}{\partial x} \left[ (1-f) \rho_k \mu_{m} \right] = -\alpha (c_m - c_w) - (0 \mu_m + (1-f) \rho_k \mu_{m} c_w) \alpha (11)$$

$$\frac{\partial}{\partial t} \left[ f \rho_k \mu_{m} \right] - \frac{\partial}{\partial x} \left[ f \rho_k \mu_{m} \right] = -\alpha (c_m - c_w) - (0 \mu_m + (1-f) \rho_k \mu_{m} c_w) \alpha (12)$$

where the subscripts $m$ and $m$ refer to the mobile and immobile liquid regions, respectively; the subscripts $f$ and $m$ refer to the liquid and sorbed phases, respectively; $f$ represents the fraction of sorption sites that equilibrate with the mobile liquid phase, and $\alpha$ is a first-order mass transfer coefficient governing 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 involving a large number of tracers (irradiated water, chloride,
different organic chemicals, heavy metals) as shown in studies by [51, 176, 114, 49].

As an example, Figure 2 shows breakthrough curves for the pesticide 2,4,5-T (2,4,5-
Trichlorophenoxy-acetic acid) obtained from a 30-cm long soil column containing
aggregated (<6 mm in diameter) Glendale clay loam [172]. Notice that the two-region
model (TRM) provides an excellent description of the data, whereas the CDE model
performed relatively poorly.

![Figure 2](image)

Figure 2. Observed and fitted effluent curves for 2,4,5-T movement through Glendale clay loam. The fitted
curves were based on the classical CDE (A) and two-region TRM transport models.

A close comparison of the two-site and two-region nonequilibrium models shows that
both have the same mathematical structure. As indicated previously by Nkedi-Kizza et
al. [113] and Toride et al. [155], among others, the two models can be put into the same
dimensionless form 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 those shown in Fig. 2
generally contain insufficient information to differentiate between specific physical
(mobile-immobile type) and chemical (kinetic type) processes leading 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 describe macroscopic transport without having to delineate the exact physical and chemical processes at the microscopic level.

2.5.4. Nonequilibrium Transport in Structured Media
The mass transfer coefficient $\alpha$ in the two-region model, Eqs. (11) and (12), is a quasi-empirical parameter which accounts for the overall effects of intra-aggregate diffusion as determined by soil aggregate size, aggregate geometry and the diffusion coefficient $D$.

Expression for $\alpha$ may be derived theoretically for transport through geometrically well-defined structured porous media, i.e., media containing uniformly-sized cylindrical, rectangular, spherical or other types of macropores or aggregates or fractures [169]. Still, one important limitation of mobile-immobile two-region type transport models is the assumption that water flow is limited to the macropore region. This assumption is inconsistent with experimental observations which indicate that water in the soil matrix is generally also mobile.

Different types of models have been used to simulate transport in variably-saturated structured media [180, 193]. As an example, we will summarize here the dual-porosity model developed by Gerke and van Genuchten [54, 55]. This model assumes that the Richards equation for transient water flow and the convection-dispersion equation for solute transport can be applied to each of the two pore systems (Fig. 3), i.e.,

$$
C \frac{\partial h}{\partial t} = \frac{\partial}{\partial z} \left( \frac{\partial h}{\partial z} \left( \frac{\partial h}{\partial z} - K_a \right) \right) - \frac{\Gamma_s}{w_f},
$$

$$
C \frac{\partial K_r}{\partial t} = \frac{\partial}{\partial z} \left( K_r \frac{\partial h}{\partial z} - K_a \right) + \frac{\Gamma_r}{1-w_f},
$$

and

$$
\frac{\partial}{\partial t} \left( \frac{\partial R}{\partial z} \right) = \frac{\partial}{\partial z} \left( \frac{\partial R}{\partial z} \left( \frac{\partial R}{\partial z} - D_a \right) \right) - \frac{\Gamma_s}{w_f},
$$

$$
\frac{\partial}{\partial t} \left( \frac{\partial R_s}{\partial z} \right) = \frac{\partial}{\partial z} \left( \frac{\partial R_s}{\partial z} \left( \frac{\partial R_s}{\partial z} - q_s a_s \right) \right) + \frac{\Gamma_r}{1-w_f}
$$

where the subscripts $f$ and $m$ refer to the fracture and matrix pore systems, respectively; $\Gamma_s$ and $\Gamma_r$ describe the rate of exchange of water and solute between the fracture and matrix regions, respectively; $w_f$ is the volume of the fracture pore system relative to that of the total soil pore system. Similarly as for the first-order mobile-immobile transport models, water and solute mass transfer between the two pore systems is described with first-order rate equations.
\[ \Gamma_x = \alpha_w (b_y - b_w) \]  
\[ \Gamma_z = \alpha_c (c_y - c_w) \begin{cases} c_w (1 - w) \theta_x c / \theta y & \Gamma_x > 0 \\ c_w (1 - w) \theta_x c / \theta y & \Gamma_x < 0 \end{cases} \]  

in which \( \alpha_w \) and \( \alpha_c \) are first-order mass transfer coefficients for water and solute, respectively. The first term on the right-hand side of (18) specifies the diffusion contribution to \( \Gamma_w \), while the second term gives the convective contribution. The above variably-saturated 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 (\( q_w = \Gamma_w = 0 \)). Simulations using the above dual-porosity model are given by Gerke and van Genuchten [54] and van Genuchten [166].

2.6. VOLATILIZATION

While in the past most chemical pollutants were regarded as involatile, volatilization is now increasingly recognized as being an important process affecting the fate of many organic chemicals, including pesticides, in field soils [57, 146]. While many organic
pollutants dissipate by means of chemical and microbiological degradation, volatilization may be equally important for volatile substances, such as certain pesticides. The volatility of pesticides is influenced by many factors, most important being the physicochemical properties of the chemical and several environmental parameters such as temperature and solar energy. Even though only a small fraction of pesticide may exist in the gas phase, air-phase diffusion rates can be comparable to liquid-phase diffusion since gas-phase diffusion is about 10^4 times greater than liquid-phase diffusion.

The solute transport equation for volatile solutes has three additional terms as compared to Eq. (2) as follows

$$\frac{\partial (\rho \tau)}{\partial t} + \frac{\partial (\rho c)}{\partial t} + \frac{\partial (\rho g)}{\partial t} = \nabla \cdot (\rho D_{v} \nabla c) - \rho \tau \frac{\partial g}{\partial t} - q_{a} c - q_{v} \tau \Phi \frac{\partial g}{\partial \Phi}$$ (19)

where $\alpha$ is volumetric air content, $c$ is the solute concentration associated with the gas phase, $D_{v}$ and $D_{g}$ are the solute dispersion coefficients in the liquid and gaseous phases, respectively; and $q_{a}$ and $q_{v}$ are the volumetric fluid flux densities of the liquid and gaseous phases, respectively.

Assuming linear 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 $k_{g}$ (i.e., $g = k_{g}c$), respectively, Eq. (19) reduces to the simpler form (8), where $Q = q_{a} + q_{v}k_{d}g$ is the effective fluid flux density, $R = 1 + (q_{a} + q_{v}k_{d})/\rho \tau$ is the solute retardation factor, and $D_{e} = D_{v} + q_{v}k_{d}/\rho \tau$ is the effective dispersion coefficient. For conditions of steady-state water and gas flow in homogeneous soils, and neglecting the source/sink term $\Phi$, Eq. (19) reduces to the standard convection-dispersion equation (3), where $v = (q_{a} + q_{v}k_{d})/\rho \tau$ is the average pore velocity.

2.6. DEGRADATION

The source/sink term $\Phi$ in Eq. (2) may be used to account for nutrient uptake and/or a variety of chemical and biological reactions and transformations in soil; as these processes are not already included in the sorption/exchange term $\partial (\rho \tau \Phi)/\partial t$. 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 must be considered are nitrification, denitrification, mineralization, and nitrogen uptake by plants [448]. For microbiologically induced organic and inorganic transformations, the degradation processes should also consider the growth and maintenance of soil microorganisms.

Alexander and Scow [2] 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 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-
First-order kinetic equations remain the most popular for describing the biodegradation of organic compounds, mainly because of the simplicity and ease in which these equations can be incorporated into solute transport models. Conditions for the application of zero- and first-order biodegradation equations are described by Alexander and Scow [2].

Table 3. Biological Degradation Equations

<table>
<thead>
<tr>
<th>Equation</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi = -k_1c$</td>
<td>First-order kinetics</td>
</tr>
<tr>
<td>$\phi = -k_2$</td>
<td>Zero-order kinetics</td>
</tr>
<tr>
<td>$\phi = -k_3c^{\lambda_1}$</td>
<td>Power rate kinetics</td>
</tr>
<tr>
<td>$\phi = \frac{\rho_{\text{max}}c}{k_4 + c}$</td>
<td>Monod, Michaelis-Menten kinetics</td>
</tr>
<tr>
<td>$\phi = \frac{\rho_{\text{max}}c(c_0 + X_0 - c)}{k_5 + c}$</td>
<td>Monod with growth kinetics</td>
</tr>
<tr>
<td>$\phi = \frac{\rho_{\text{max}}c}{k_5 + c + c^2/k_4}$</td>
<td>Haldane modification of the Monod kinetics</td>
</tr>
<tr>
<td>$\phi = \rho_{\text{max}}(c_0 + X_0 - c)$</td>
<td>Logarithmic kinetics</td>
</tr>
<tr>
<td>$\phi = -k_6c(c_0 + X_0 - c)$</td>
<td>Logistic kinetics</td>
</tr>
</tbody>
</table>

$k_1, k_2$: empirical constant
$\rho_{\text{max}}$: maximum specific degradation
$k_4$: 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
$k_6$: inhibition constant that effects the suppression of the growth rate by toxic substrate

One special group of degradation reactions involves decay chains in which solutes are subject to sequential (or consecutive) decay reactions. Problems of solute transport involving sequential first-order decay reactions frequently occur in soil and groundwater systems. Examples are the migration of various radionuclides [131, 164], the
3. Multicomponent Solute Transport

Thus far we have considered the transport of only one chemical species and assumed that the behavior of this solute is independent of that of other species which may be present in the soil solution. In reality, the soil solution is always a mixture of many ions which mutually may interact, create complex species, precipitate, dissolve, and/or could compete with each other for sorption sites. This section considers such more complex situations leading to multicomponent transport.

3.1. THE AQUEOUS EQUILIBRIUM CHEMISTRY MODELS

Aqueous equilibrium chemistry models [156, 181] which solve for solution equilibrium without consideration of transport were originally developed mostly independent of the models for water flow and solute transport. Only during the last decade has there been a significant effort to couple the different types of models.

Chemical processes may be broadly divided into kinetic reactions and thermodynamic equilibrium reactions (i.e., reactions which proceed sufficiently fast to be considered instantaneous). While, mathematically, kinetic processes lead to a system of ordinary partial differential equations, thermodynamic equilibrium processes normally result in systems of algebraic equations. The term 'component' is generally used when dealing with chemical equilibrium systems. Following Westall et al. [181], components may be defined as a set of linearly independent chemical entities such that every species in the system can be uniquely represented as the product of a reaction involving only these components, and no component can be represented as the product of a reaction involving only the other components. As a typical example, the chemical species CaCO₃ consists of the two components Ca²⁺ and CO₃²⁻. However, from the definition of a component it follows that any other combination of two species selected from the above system as basis components, such as CaCO₃⁻ and Ca²⁺; or CaCO₃⁻ and CO₃²⁻, may be used also since the third species can always be represented by a linear combination of the two selected components.

Chemical equilibrium can be formulated and solved mathematically by means of two different but thermodynamically equivalent approaches, i.e., (1) by minimizing the Gibbs free energy of the system subject to the constraints of mass balance, or (2) by using equilibrium constants. The solution procedure for the second approach begins with an initial guess for a set of components from which the minimum Gibbs free energy composition can be readily calculated using equilibrium constants, after which the mass balance equations are solved by iteration [181]. This approach is a popular method when dealing with chemical equilibrium system [156, 181].

Chemical reaction processes are generally represented mathematically using mass action laws which relate thermodynamic equilibrium constants to activities (the
thermodynamic effective concentration) of the reactants and products involved \([100]\). For example, the reaction

\[ bB + cC \rightarrow dD + eE \]  

(20)

where \( b, c, d, \) and \( e \) are the number of moles of substances \( B \) and \( C \) which react to yield products \( D \) and \( E \), is represented at equilibrium by the law of mass action

\[ K = \frac{a_D^d a_E^e}{a_B^b a_C^c} \]  

(21)

where \( K \) is a temperature-dependent thermodynamic equilibrium constant, and \( a_i \) is the ion activity, being defined as the product of the activity coefficient \( \gamma_i \) and the ion molality \( m_i \), i.e., \( a_i = \gamma_i m_i \). Single-ion activity coefficients may be calculated using either an extended version of the Debye-Hückel equation \([156]\) for dilute to moderately saline solutions, or by means of the more complicated Pitzer expressions \([127]\) which are considered more accurate for solutions having very high ionic strength down to solutions having infinite dilution. Equation (21) can be used to describe all of the major chemical processes, such as aqueous complexation, sorption, precipitation-dissolution, and acid-base and redox reactions. Acid-base and redox reactions are usually formulated by using the hydrogen ion \( \text{H}^+ \) and the electron \( e^- \) as components, respectively \([100]\). Acid-base reactions can be alternatively treated by using the electroneutrality condition. A Newton-Raphson method is usually employed to solve the final set of nonlinear algebraic equations.

3.2. MULTICOMPONENT TRANSPORT MODELS

Most modeling efforts involving multicomponent transport have thus far focused on the saturated zone where changes in water velocity, temperature and pH are relatively gradual and hence less important than in the unsaturated zone. Consequently, most multicomponent transport models assume one-dimensional steady-state saturated water flow with fixed water flow velocity, temperature and pH \([157, 179, 27, 19]\), among others. Only recently have several multicomponent transport models been published which also consider variably-saturated flow \([19, 191, 141]\).

In a recent review, Yeh and Tripathi \([190]\) identified three different approaches for mathematically solving multicomponent transport problems: (1) a mixed differential and algebraic approach, (2) a direct substitutional approach, and (3) a sequential iteration approach. In the first approach, the sets of differential and algebraic equations describing the transport processes and chemical reactions, respectively, are treated simultaneously \([103, 90]\). In the second approach, the algebraic reactions representing the nonlinear chemical reactions are substituted directly into the differential mass balance transport equations \([133, 157, 73]\). The third approach considers two coupled sets of linear partial differential and algebraic equations, which are solved sequentially and iteratively \([191, 179, 141]\). Based on a study of computer resource requirements, Yeh
and Tripathi [190] suggested that only the third method (sequential iteration) can be applied to realistic two-dimensional problems, and that this method "provides perhaps the only hope for realistic three-dimensional applications".

Rigorous modeling of multicomponent transport of charged aqueous species requires several refinements to Fick's laws. Actual diffusion rates are influenced by the effects of Coulomb interactions which maintain electroneutrality and which lead to a coupling between each solute flux and all concentration gradients [88]. Ion pairing represents another coupling mechanism since it couples the fluxes of the particular ions in a special way. These refinements to Fick's laws are beyond the scope of this paper and will not be considered here.

The partial differential equations governing one-dimensional multicomponent convective-dispersive chemical transport during transient water flow in variably saturated porous media may be taken as [141]

\[
\frac{\partial (\theta \epsilon_k)}{\partial t} + \frac{\partial \rho_k}{\partial t} = \frac{\partial \rho_k}{\partial x} \frac{\partial q_k}{\partial x} \epsilon_k \quad k=1,2,...,N_c
\]  

(22)

where \( \epsilon_k \) is the total dissolved concentration of the aqueous component \( k \) (i.e., the sum of the components plus all complex species containing component \( k \)), \( \rho_k \) is the total sorbed concentration of the aqueous component \( k \), \( \epsilon_k \) is the total precipitated concentration of aqueous component \( k \) (i.e., the sum of all precipitated species containing the component \( k \)), and \( N_c \) is the number of aqueous components. The second and third terms on the left side of Eq. (22) are zero for components that do not undergo ion exchange or precipitation/dissolution. The above coupled approach involving transport and chemical submodules was previously also followed by [179, 27, 19, 191].

4. Numerical Methods

A large number of analytical solutions have been published for simplified one- and multidimensional transport problems [167, 72, 58, 91, 92, 155]. Unfortunately, analytical solutions for more complex situations, such as for transient water flow or the nonequilibrium solute transport with nonlinear reactions, are not available, in which case numerical models must be employed.

4.1. NUMERICAL SOLUTION OF RICHARDS EQUATION

A variety of numerical methods may be used to solve the variably-saturated flow and transport equations. The popularity of numerical methods stems from the fact that the highly nonlinear Richards equation can be solved analytically only for a very limited number of cases involving homogeneous soils and relatively simple constitutive relationships describing the unsaturated soil-hydraulic properties. Even so, the nonlinear nature of the Richards equation also hampered the development of computationally efficient numerical methods which would be stable in all situations, particularly for
infiltration in very dry soils.

Early applications of numerical methods for solving variably-saturated flow problems generally involved classical finite difference methods [62, 47]. Integrated finite difference methods [109] and especially finite element methods [110, 71, 189, 142] became increasingly popular in the seventies and eighties. Time and space discretization of the Richards equation using any of these methods leads to a nonlinear system of 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 equations. The Newton-Raphson iteration procedure is more complex and results in nonsymmetric matrices, but often achieves a faster rate of convergence and can be more robust than Picard iteration for certain types of problems [120]. Paniconi et al. [119] observed that 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 differs as a function of the type of flow formulation being used, i.e., the h-based, the B-based, or the mixed formulation. Celia et al. [28] 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. Milly [115] 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 highly efficient numerical schemes based on different types of pressure head transformations were presented recently by [64, 132, 80, 118]. 10ths et al. [64] showed that the B-based form of the Richards equation can yield fast and accurate solutions for infiltration into very dry heterogeneous soil profiles. However, the h-based numerical scheme cannot be used for soils having saturated regions. Kirkland et al. [80] expanded the work of Hills by combining the h-based and h-based models to yield a transformation method applicable also to variably-saturated systems. They defined a new variable which is a linear function of the pressure head and water content in the saturated and unsaturated zone, respectively.

4.2. NUMERICAL SOLUTION OF THE TRANSPORT EQUATION

A large number of numerical methods may be used to solve the convection-dispersion solute transport equation. These methods may be classified into three groups [111]: (1) Eulerian, (2) Lagrangian, and (3) mixed Lagrangian-Eulerian methods. In the Eulerian approach the transport equation is discretized by a usual finite difference or finite element method using a fixed grid system. For the Lagrangian approach the mesh either deforms and moves along with the flow path, or the mesh is assumed stable in a deforming coordinate system. A two-step procedure is followed for a mixed Lagrangian-Eulerian approach. First, convective 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 Eulierian approach involving any finite element or finite differences methods, thus leading to the final concentrations.

Standard finite difference [13, 11] and Galerkin-type finite element [61, 59, 126] 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, still are the most popular methods being used at present. Numerical studies have shown that both methods give good results for transport where dispersion is a relatively dominant process (e.g., as indicated by the grid Peclet number). However, both methods can lead to significant numerical oscillations and/or dispersion for convection-dominated transport problems. The Eulerian methods are very successful when applied to symmetric problems such as diffusion-dominated solute transport. The convection term brings nonsymmetry into the governing solute transport equation and, as a result, the success of Eulerian methods for symmetric cases is lost when convection dominates the transport process [38]. Still, by selecting an appropriate combination of relatively small space and time steps, it is possible to virtually eliminate all oscillations. Alternatively, the spatial grid system may be refined using a "zoomable hidden fine-mesh" approach [188] or by implementing local adaptive grid refinement [185]. Perrochet and Berot [124] developed criteria for minimizing or eliminating numerical oscillations based on a "performance index". They conclude that all oscillations should be eliminated when the performance index, defined as the product of the local Peclet (\(\text{Peclet} / D\)) and Courant (\(\text{Courant} / \Delta t\)) numbers, is less than 2. When small oscillations in the solution can be tolerated, the performance index can be increased to about 5 or 10 [124]. Unfortunately, one may not always be able to decrease the spatial step size, while for pure convection it is impossible to obtain reasonable results. Therefore, improved numerical methods that limit or prevent oscillations are still actively being sought.

One alternative is the use of upwind finite difference methods [29, 174, 171]. This method virtually eliminates numerical oscillations, even for purely convective transport [174]. A disadvantage is that this method may create significant and often unacceptable numerical dispersion. Similarly, upstream weighting has been proposed for finite elements [170, 187]. The method uses weighting functions which are different for terms having spatial derivatives than for other terms in the transport equation. This approach places greater weight on the upstream nodes within a particular element. Methods based on upstream weighting are called Petrov-Galerkin methods. Huyakorn and Nithara [70] suggested for this purpose the nonorthogonal base function, whereas Yeh [187] used orthogonal functions. Petrov-Galerkin methods require the use of higher-order weighting functions, which makes their implementation more difficult than classical Galerkin finite element methods [38]. A second alternative for overcoming numerical dispersion is the use of higher order temporal and spatial approximation [66, 171, 89, 39], although, such higher order approximations sometimes may increase numerical oscillations.

While Lagrangian methods (or method of characteristics) virtually eliminate problems with numerical oscillations [177, 116, 112], they may introduce other problems, notably
artificial dispersion and nonconservative solution. Lagrangian methods are also relatively difficult to implement in two and three dimensions. Instabilities resulting from inappropriate spatial discretization may occur during longer simulations as a result of a deformation of the stream function. Furthermore, nonrealistic distortions of the results may occur when modeling transport of solutes which are subjected to different sorption/exchange or precipitation reactions.

Mixed Eulerian-Lagrangian approaches have been reported by several authors [82, 111, 107, 145, 188]. In view of the different mathematical character of the diffusive (parabolic) and convective (hyperbolic) terms in the convection-dispersion equation, the transport equation is best decomposed into a mixed problem consisting of a pure convection problem, followed by a pure diffusion-only problem. Methods based on this approach are called operator-splitting or splitting-up methods [38]. Convective 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 with three different methods. First, Konikow and Bredehoef [82] suggested continuous forward particle tracking, i.e., to follow a set of particles as they move through the flow domain. Second, Mose et al. [107] 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. Third, a combination of both approaches as suggested by Neuman [111]. Continuous forward particle tracking method has similar disadvantages as the Lagrangian approach since complex geometrical regions are again difficult to handle. To obtain good results it is necessary to follow large number of particles which quickly lead to excessive computer time and memory. The single-step reverse particle tracking method [107] is very straightforward, especially if problem with numerical dispersion can be resolved [69].

A method proposed by Ashtrom et al. [1] and later used by Pockett et al. [128], among others, represents a different way of following solute particles during convective-dispersive transport. Dispersion in this case is modeled by a so-called "random walk" process as used for each particle individually. The method is conceptually very simple and can be programmed relatively easy. Displacement of each particle during each time step is given by a certain distance, being the sum of two velocity contributions - a deterministic and stochastic contribution. Studies with this method indicate that it may be necessary to use many thousands of particles in order to obtain relatively precise results.

Still other solution methods exist, such as the use of a combination of analytical and numerical techniques. For example, Sudicky [150] modeled solute transport 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 less stringent requirements are needed for the spatial discretization. Park and Ligget [121] developed a more general method for solving similar problems, i.e., the Taylor least-squares technique. Recently several methods were suggested which make use of local analytical solutions of the convection-dispersion equation in combination with finite differences [95]. This combination of analytical and
Numerical techniques have one important limitation. Since Laplace transforms eliminate 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 means that combination methods can only solve solute transport problems during steady-state water flow, and hence are inappropriate for transient variably-saturated flow typical of most field situations.

4.3. MATRIX EQUATION SOLVERS

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

$$[A] \{x\} = \{b\} \quad (23)$$

in which $\{x\}$ is an unknown solution vector, $\{b\}$ is the known right-hand side vector of the matrix equation, and where $[A]$ is a sparse banded matrix which 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 convection is not considered in the formulation. Since the water flow equation is most often solved with the modified-Picard procedure and solute transport usually considers convection, we will assume here that matrix $[A]$ is always symmetric for water flow, and asymmetric for solute transport.

Traditionally, matrix equations have been solved by means of such direct methods as Gaussian elimination or LU decomposition. These methods usually take advantage of the banded nature of the coefficient matrices and, in the case of water flow, of the symmetric properties of the matrix. Direct solution methods have several disadvantages as compared to iterative methods. For example, they require a fixed number of operations (depending upon the size of the matrix) which increases approximately by the square of the number of nodes [102]. Iterative methods, on the other hand, require a variable number of repeated steps, with the number increasing at a much smaller rate (about 1.5) with the size of a problem [102]. A similar reduction also holds for the memory requirement since iterative methods do not require one to store non-zero matrix elements. Memory requirements, therefore, increase at a much smaller rate with the size of the problem when iterative solvers are used [102]. Round-off errors also represent less of a problem for iterative methods as compared to direct methods. This is because round-off errors in iterative methods are self-correcting [94]. 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 [94]. In general, direct methods are more appropriate for relatively small problems, while iterative methods are more suitable for larger problems.

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 over-relaxation (SOR), block successive over-relaxation (BSSOR), successive line over-relaxation (SLOR), and strongly implicit procedures (SIP), among
others [94]. More powerful preconditioned accelerated iterative methods, such as the preconditioned conjugate gradient method (PCG) [79, 8], were introduced more recently. Sildick and Huyakorn [151] gave three advantages of the PCG procedure as compared to other iterative methods: PCG can be readily modified for finite element methods with irregular grids, the method does not require iterative parameters, and PCG usually outperforms its iterative counterparts for situations involving relatively stiff matrix conditions. The PCG methods can be used only for symmetric matrices. Since the system of linear equations resulting from discretization of the solute transport equation is nonsymmetrical, it is necessary to either formulate the transport problem such that a symmetric matrix results [93], or use the extension of PCG for nonsymmetrical matrices, such as ORTHOMIN (generalized conjugate residual method) [6], GMRES (generalised minimal residual method), bi-conjugate gradients, TFQMR (transpose-free quasi-minimal residual algorithm), or conjugate gradient squared procedures [94, 120].

The preconditioned conjugate gradient and ORTHOMIN methods consist of two essential parts: initial preconditioning, and iterative solution with either conjugate gradient or ORTHOMIN acceleration [102]. Incomplete lower-upper (ILU) factorization (among other methods, such as polynomial, Newton-Raphson, diagonal scaling, UDR factorization, red-black ordering, block preconditioning, or Crout lower-upper decomposition) can be used as preconditioning of matrix \( A \), which is factorized into lower and upper triangular matrices by partial Gaussian elimination. The preconditioned matrix is subsequently repeatedly inverted using updated solution estimates to provide a new approximation of the solution. The orthogonalization-minimization acceleration technique is used to update the solution estimate. This technique insures that the search direction for each new solution is orthogonal to the previous approximate solution, and that either the norm of the residuals (for conjugate gradient acceleration (101)) or the sum of the squares of the residuals (for ORTHOMIN; [8]) is minimized. More details about the two methods is given in the user's guide of the ORTHOFEM software package [102, 94]. Letowski [94] also gave a comprehensive review of accelerated iterative methods, as well as preconditioning techniques. The ORTHOMIN procedure [7] is becoming increasingly popular in variably-saturated flow and contaminant transport simulations [50, 30, 143].

5. Parameter Estimation

Realistic application of computer models to field-scale flow and transport problems implies a commensurate effort in quantifying a large number of model parameters. As increasingly more complicated flow and transport models are being developed, the accuracy of future simulations may well depend upon the accuracy with which various model parameters can be estimated. This is especially true for the unsaturated hydraulic conductivity which is a key parameter determining the rate and direction at which water and dissolved chemicals move in the subsurface. Accurate measurement of the hydraulic properties is confounded by the extreme spatial heterogeneity of the subsurface environment. The hydraulic properties frequently also shows significant variations in
time because of cultivation or other agricultural activities, shrink-swell phenomena of fine-textured soil, the effects of particle dispersion and soil crusting, and changes in the concentration and ionic composition of the soil solution.

A variety of laboratory and field methods are available for measuring the hydraulic conductivity, \( K \), or the soil water diffusivity, \( D_t \), as a function of the pressure head, \( A \), and/or the water content, \( \theta \) [81, 60]. Most laboratory methods are steady-state procedures based on direct inversion of Darcy’s law. Transient methods generally involve some type of approximation or simplification of the Richards equation. Popular transient methods include the Bruce and Klute [17] horizontal infiltration method, and various modifications thereof such as the hot-air method, and the sorptivity method. Popular field methods include the instantaneous profile method, various unit-gradient type approaches, sorptivity methods following ponded infiltration, the use of tension infiltrometers, and the crust method based on steady water flow. While relatively simple in concept, these direct measurement methods have a number of limitations that restrict their use in practice. For example, most methods are very time-consuming to execute because of the need to adhere to relatively restrictive initial and boundary conditions. This is especially true for field gravity-drainage experiments involving medium- and fine-textured soils. Methods requiring repeated steady-state flow situations, or other equilibrium conditions are also tedious, while linearizations and other approximations or interpolations to allow analytic or semi-analytic inversions of the flow equation introduce additional errors. Finally, information about uncertainty in the estimated hydraulic parameters is not readily obtained using direct inversion methods.

A relatively more flexible approach for solving the inverse problem is the use of parameter optimization methods. Optimization procedures also make it possible to simultaneously estimate the retention and hydraulic conductivity functions from transient flow data [84]. Early parameter optimization studies focused primarily on solute transport [163, 123]. Starting with the studies of Zachmann et al. [192] and Due and Hruska [35], the method is now increasingly being used also for estimating the unsaturated soil hydraulic functions. Computer models applicable to one-step and multi-step laboratory outflow measurements are given by [83, 159] and [40], respectively. While initially applied primarily to laboratory type experiments, inverse methods are equally well applicable to field data, or some appropriate combination of field and laboratory data. An important advantage of inverse procedures, if formulated within the context of a parameter optimization problem, is that a detailed error analysis of the estimated parameter is more easily considered. Inverse procedures may prove to be very appropriate for estimating regional-scale effective soil hydraulic parameters, either by appropriately manipulating in-situ measurement of the hydraulic properties as shown by Feddes et al. [43], or by using remotely-sensed measurements of the soil surface water content [45]. While parameter optimization methods pose several advantages, a number of problems related to computational efficiency, convergence, and parameter uniqueness, remain to be solved, especially when many hydraulic parameters must be estimated simultaneously.

In contrast to direct methods, somewhat less attention has been paid to the development of indirect methods which predict the hydraulic properties from more easily
measured data, including water retention data, and pore- or particle-size distributions. This is unfortunate since indirect methods generally are more convenient and far less costly to implement. Moreover, indirect methods often give hydraulic conductivity estimates which may well be accurate enough, or are close to being accurate enough, for many applications. Among the indirect approaches are theoretical methods which estimate the unshrunk hydraulic properties from water measured soil retention data, and the use of pedotransfer functions which correlate the hydraulic functions with soil texture and other data routinely available from soil survey databases.

Theoretical methods are usually based on statistical pore-size distribution models which assume water flow through cylindrical soil pores, and incorporate the equations of Darcy and Poiseuille. Several models for predicting the unshrunk hydraulic conductivity from measured water retention data have been derived for this purpose [10, 20, 108]. Bourque’s model [20] was applied by Brooks and Corey [16] to derive their classical function for the unshrunk hydraulic conductivity. Numerous other approaches have been formulated [104, 162, 134]. Because of their simplicity and ease of use, predictive models for \( K(h) \) or \( K(0) \) have become very popular in numerical studies of unshrunk flow using Richards’ equation.

Several attempts have been made to derive pedotransfer functions [10] which correlate the soil hydraulic properties to soil texture and other soils data, including bulk density, organic matter content, cation exchange capacity, clay mineralogy, and/or soil structural information [184, 22, 130]. Two alternative approaches are generally followed: (i) parameters in specific hydraulic models are correlated directly with soil texture and related data, and (ii) water contents at selected pressure heads are estimated from the surrogate soils data and subsequently used in a curve-fitting exercise to estimate the complete retention function. As indicated by Thomsen and Carter [152] and others, the second approach has been the more popular one since retention measurements in the past were usually limited to only a few points (notably field capacity and the permanent wilting point), leading to accurate predictions for those points only, but relatively poor estimates of other parts of the retention curve. Also, there was generally limited need for complete descriptions of the soil hydraulic functions until the recent widespread use of computer models and Geographic Information Systems for a wide range of applications in research and management. Pedotransfer functions have been reasonably successful in some cases of giving approximate orders of magnitude of the hydraulic properties [186, 153].

6. Stochastic Transport

There is ample evidence to suggest that solutions of the classical Richards and CDE models, no matter how refined to include the most relevant chemical and biological processes and soil properties, fail to accurately describe transfer processes in most natural field soils. The one major factor responsible for this failure is the overwhelming heterogeneity of the subsurface environment. Heterogeneity occurs at a hierarchy of spatial and time scales [182], ranging from microscopic scales involving time-dependent chemical sorption and precipitation/dissolution reactions, to intermediate scales involving...
the preferential movement of water and chemicals through macropores or fractures, and to much large scales involving the three-dimensional spatial variability of soils across the landscape. Several approaches are possible to address the different types of soil heterogeneity. On the one hand, subsurface heterogeneity may be addressed in terms of process-based 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 increasingly being addressed using stochastic approaches which incorporate certain assumptions about the transport process in the heterogeneous system.

Among the stochastic approaches currently being pursued are Monte Carlo methods, stochastic-continuum models, and various stochastic-convective approaches. A common assumption of stochastic transport models [33, 34, 77] is that parameters are treated as random variables with values assigned according to a given probability distribution. In practice, the stochastic approach is generally used with several simplifying assumptions, including (1) the stationarity hypothesis which assumes that a random parameter has the same probability density function (pdf) at every point in the field, and (2) the ergodicity hypothesis which states that ensemble averages can be replaced by spatial averages, and that spatial replicates can be used to construct the appropriate pdf's for the transport parameters.

Monte Carlo simulations assume that the flow and transport parameters are random variables with values assigned from a joint pdf. The water flow or solute transport equations are repeatedly solved using coefficient values from the assumed pdf until a large number of possible outcomes has been generated. These outcomes are then used to calculate sample means and variances of the underlying stochastic transfer process. The method may be used to demonstrate, among other things, that macrodispersion in field soils is affected primarily by pore-water velocity variations, and much less by local dispersion phenomena [3, 125].

Stochastic-continuum models were initially used primarily in groundwater studies [33, 52, 78], but have recently also found application to vadose zone transport processes [135]. In these models all random variables are represented by the sum of their mean value plus random fluctuations which, when substituted into the CDE model, lead to a new mean transport model with additional terms. The modified model may be evaluated by deriving first-order approximations for the fluctuations and solved by means of Fourier transforms. The approach leads to, among other things, a macro-scale dispersion coefficient whose value is reached asymptotically as distance and/or time increase. Spatial correlations of solute velocity variations characterized by its autocorrelation function, have been shown to play important roles in the derivation of the asymptotic convection-dispersion equation [147].

A simpler stochastic approach arises when the field viewed as a series of independent vertical columns (Fig. 4), generally referred to as "stream tubes", through which chemicals move downward from the soil surface to the groundwater table [12]. Solute mixing between the stream tubes is usually assumed negligible, while transport in each
convection-dispersion equation. The mean solute concentration for an entire field is
given by the ensemble average of the local concentrations in all stream tubes. The one-
dimensional CDE (assuming perfect mixing in a plane perpendicular to the flow
direction) and the stream-tube model (assuming no mixing between individual tubes)
may be viewed as two limiting cases for solute transport at the field scale [76].

The stochastic-convective approach was used by Simmons [140] who neglected the
dispersion coefficient $D$ in Eq. (4), and used the pore-water velocity and travel time as
random variables. Jury [75] initially also neglected $D$ in his development of the transfer
function model (TFM) of solute transport. The TFM involves the use of a probability
density function $f(t)$, of travel times from the soil surface down to some reference depth
$L$. The travel time probability density function for many transport experiments is given
by a lognormal distribution. The flux concentration in the profile is represented with a
convolution integral of $f(t)$ and the imposed flux concentration at the soil surface.
Stream-tube and related transfer function models are expected to find increasingly wider
applications in subsurface solute transport as the underlying theory is being strengthened
by the incorporation of a variety of physical, chemical and biological phenomena [77,
136].

Several other statistical approaches exist, including the use of continuous Markov
processes, fractal-mathematics, random walk methods, and a variety of procedures based
on moment analysis. More work in these areas of research can be expected in the near
future.

7. References

transport model theory and numerical implementation, Report BNL-21277, Battelle Pacific Northwest Laboratory, Richland, WA.


363-368.