Transport processes from soil surfaces to groundwaters

Donald R. Nielsen
Department of Land, Air and Water Resources, University of California, Davis, California 95616, USA

M. Th. van Genuchten
U.S. Salinity Laboratory, Riverside, California 92510, USA

William A. Jury
Department of Soil and Environmental Sciences, University of California, Riverside, California 92521, USA

ABSTRACT The management of water and its dissolved constituents from the soil surface through the unsaturated zone to groundwater supplies requires an accurate conceptual understanding and a quantitative description of all relevant physical, chemical and biological processes and properties in the unsaturated zone. Several conceptual problems and opportunities for modeling vadose zone water flow and solute transport processes are reviewed. Alternatives to classical modeling approaches are also outlined.

INTRODUCTION

Groundwater is contaminated as a result of anthropogenically induced perturbations of physical, chemical and biological processes occurring from soil surfaces through the unsaturated vadose zone to subterranean resources. Motivation for developing a quantitative understanding of those processes in relation to the movement and retention of water and its dissolved solutes as well as other liquids into and within groundwater resources has never been higher than it is today. The public is increasingly focusing its attention on the intentional and unintentional release of surface-applied or soil-incorporated contaminants, their immediate effects on soil and water pollution, and their indirect harm by entering the food chain for all living organisms. Fertilizers and pesticides intentionally applied to agricultural lands inevitably move below the root zone, percolate through the vadose zone and eventually contaminate ground waters. Chemicals migrating from municipal and industrial disposal sites as well as radionuclides from nuclear energy and waste storage facilities also present environmental hazards. Although accidental releases of inorganic and organic chemicals and wastes onto the land and into surface water supplies enjoy a great deal of notoriety in the public media, of equal importance to the quality of the global environment are the subtle, yet enormous releases of organic constituents from field soils caused by changing land management practices. That is, the normally cyclic variations in the flux of constituents moving through the vadose zone related to crop rotation and annual weather patterns are occasionally confounded by irregular perturbations caused by major changes in land use. Because intentional and unintentional releases of contaminants at the soil surface will prevail under a variety of local and regional conditions, our
abilities to understand, model and manage contaminant transport must be enhanced and integrated into a more unified technology.

Our incomplete ability to understand transport and transformation of contaminants stems from the fact that they have not been the consistent focus of attention of any scientific discipline. Biological research for agriculture and silviculture considers transport and transformation of only plant-essential nutrients limited to the root zone of commercially important cultivars. Crop yield is foremost with the contamination of the underlying vadose environment of secondary importance. Until very recently, groundwater hydrologists, not vested in biology or microbiology, including related biochemical processes, focused on water transport in saturated soils and aquifers. Their attention to solute transport was oftentimes limited to the behavior of "ideal, non-reactive" tracers in nonisotropic media. Geochemists have tended to concentrate on chemical equilibria and the relative isotopic abundances of naturally occurring or human-released elements for age dating or estimating transfer processes within the earth. Owing to the dearth of readily available carbon sources usually encountered in the unsaturated horizons below the root zone of plants, microbiologists have not been inclined to extensively investigate the nature of its microbial communities. Hence, segmented, disciplinary research contributes to our inability to fully comprehend the intricacies of transport phenomena. Transport models remain more or less unpredictable to predict when, where and how much of a contaminant released at the soil surface will reach the groundwater.

On the other hand, researchers, albeit dispersed and sparsely populated throughout several disciplinary sciences, have provided considerable information about contaminant transport during the past decade. Much has been learned from elaborate laboratory, plot and field experiments. A number of challenges remain for both the theoretician and the practitioner. They include: (a) how to best deal with preferential flow and transport, (b) how to best model the effects of local and regional spatial and temporal variabilities of soil hydraulic properties on solute transport, (c) how to efficiently couple multicomponent geochemical subroutines with available unsaturated-saturated water flow models, (d) how to improve field methods for measuring vadose zone transport parameters, and (e) how to best estimate the long-term consequences of short-term land management decisions regarding the quality of ground waters. We briefly review here some of conceptual opportunities for modeling transport, and delineate promising areas for further research and management.

TRANSPORT MODELS

It is convenient to conceptually divide the large number of transport models into two groups, process models and stochastic models. Process models develop a description of transport based on mass conservation and flux laws, leading to differential equations to predict values of the water and solute variables as functions of position and time. In contrast, stochastic models describe the variables as random functions, which depend on the distribution of values of the soil properties which determine their movement. Rather than predict values of concentration as a function of position and time, stochastic models predict concentration averages
Transport processes

and variances, and are used to calculate the probability of having a given value appear at a given depth or time.

Process models

Water flow in soil  The well-known Darcy law, commonly assumed to describe the flux of water through homogeneous isotropic unsaturated soil in one dimension is

\[ J_w = -K(h) \frac{d\theta(h)}{dx} - \frac{d\theta(h)}{dx} \]  

(1)

where \( J_w \) is water flux, \( K \) is the unsaturated hydraulic conductivity, \( \theta \) is soil water content, \( h \) is matric potential head and \( z \) is the vertical coordinate measured positively downward. For many field soils to date, it is also assumed that \( \theta(h) \), the soil water characteristic, is spatially and temporarily invariant.

Combining (1) with the equation of continuity

\[ \frac{d\theta}{dt} = -\frac{d\theta}{dz} + \Phi \]  

(2)

leads to Richards' equation (Richards, 1931) for liquid flow in unsaturated soils. The unsaturated flow equation becomes

\[ \frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left( \frac{\partial \theta}{\partial z} - \frac{\partial K(h)}{\partial z} \right) - \frac{\partial K(h)}{\partial z} + \Phi \]  

(3)

where \( C(h) = d\theta/dh \) is the water capacity or the slope of the soil water characteristic curve, \( \theta(h) \); \( t \) is time; while \( \Phi \) represents sources and sinks of water in the system, notably those resulting from plant water extraction in the soil root zone. A close scrutiny of the theoretical basis of (3) reveals several assumptions that may be invalid for various field situations. For example, the equation ignores soil matrix and fluid compressibilities and assumes that the fluid density \( \rho_w \) is independent of concentration. Spatial variations in \( \rho_w \) are also taken to be insignificant. More importantly, (3) assumes that the air phase plays a negligible role in unsaturated flow processes and hence that a single equation can be used to describe unsaturated-saturated flow. A number of alternative two-phase flow models exist that explicitly consider simultaneous flow of water and air (or other immiscible fluids) in soils. Equation (3) also assumes that Darcy's law originally derives for saturated flow can be extended to unsaturated conditions. Unfortunately, the functions \( K(h) \) and \( \theta(h) \), especially those of fine-textured soils, are greatly affected by the total concentration and ionic composition of the soil solution. Because of the significant effects of concentration and ionic composition on these functions, attempts must be made to include them in existing transport models.

Solute in the soil  A complete consideration of a solute mass balance for a solute species in soil is extremely complicated. The solute might be present as an aqueously dissolved species or a gas; it may absorb to stationary mineral or organic surfaces; it may precipitate out as a salt or dissolve into soil solution; it may undergo chemical or biological transformation to a new species or be created from a transformation of another soil constituent; it may be removed from the soil solution by plant uptake or volatilized into the atmosphere. In the case of an organic
contaminant or solute, it may exist in the presence of a complete mixture involving several different organic chemicals with different immiscibilities in both aqueous and non-aqueous phases. Its molecular size, structure and polar characteristics lead to a wide range in its solubility in both aqueous and organic phases. Corresponding to (3) the equation assumed to describe an aqueous solute species is

\[
\frac{\partial (\Phi_c)}{\partial t} + \frac{\partial (\rho_{a_s})}{\partial t} + \phi_s \frac{\partial \rho_s}{\partial z} = \frac{\partial J_s}{\partial z}
\]  

(4)

where \( c \) and \( z \) are the dissolved and adsorbed concentrations, \( \rho_s \) is soil bulk density, \( \phi_s \) is a solute reaction term, \( J_s \) is the solute mass flux and \( t \) is time. The term \( \phi_s \) is used to describe adsorption-desorption, plant root uptake, microbial transformations, chemical reactions, etc. The correct expression to use for the solute mass flux \( J_s \) is currently a matter of debate for two reasons. First, the convection of solute may involve substantial contributions from only that part of the wetted pore space through which water is flowing significantly. However, since all parts of the wetted pore space are accessible to solute through diffusion, the solute storage term will include contributions from the entire wetted pore space. The second area of uncertainty in quantifying the solute flux is in the representation of the hydrodynamic dispersion term. Depending on the scale over which (4) is written, different models for dispersion may be appropriate.

Convective-dispersive flow: The convection-diffusion equation, the cornerstone during the past 20 years for modeling transport is obtained when

\[
J_s = -\phi_s \frac{\partial C}{\partial z} + J_s \rho_s
\]  

(5)

is substituted into (4) with \( D \) being the apparent diffusion coefficient. In practice, the \( D \) used is an empirical parameter that includes all of the solute spreading mechanisms that are not directly included in (5) through the definitions of \( s \) and \( \phi_s \), such as nonlinear adsorption or exchange, nonlinear decay, or various physical and chemical nonequilibrium conditions. In concept, however, the coefficient is commonly assumed to reflect two additional phenomena: ionic or molecular diffusion arising from the natural thermal motion of dissolved constituents, and mechanical dispersion resulting from the fact that local fluid velocities inside individual pores and between pores of different shapes, sizes and directions, deviate from the average fluid flux. Because the inter- and intra-aggregate concentration and pore water velocity distributions depend on geometric configurations of the pores as well as on the concentration and ionic composition of the soil solution, previously established relationships between \( D \) and macroscopically observable parameters must be considered to be strictly empirical. Also, since diffusion is an active process in response to concentration gradients irrespective of flow while dispersion is a passive process in response to fluid flow, the additivity of the two processes must be questioned.

Structural models: The convection-dispersion equation assumes that at a given location water is flowing at a constant average
velocity and that its associated solute concentration has an average value. This assumption is reasonably accurate in repacked soil columns or in relatively structureless sandy soils, but may be erroneous in aggregated media or soils with macropores or other preferential flow pathways. In such structured media, watered regions exist where water is relatively stagnant, but which are accessible to dissolved solutes by diffusion. As a consequence, the leaching process is less efficient when solutes are initially present in the soil, and solutes which are added with water at the surface move more rapidly through soil than predicted by (4). Hence, (4) has not performed well in several situations, especially for strongly absorbed chemicals. Another reason for this poor performance is the failure to accurately describe the kinetic nature of sorption and exchange processes (Wagener, 1983).

A transport model which accounts for stagnant regions of soil water has been developed by van Genuchten & Wiersma (1976). In this approach, the wetted pore space is divided into a mobile water content within which solute moves by convection and dispersion, and an immobile water content which exchange solute with the mobile region by a rate-limited diffusion process. Selim et al. (1976) presented the notion that two kinds of sites exist simultaneously—one described by a first-order reaction and another by an instantaneous, equilibrium adsorption. An alternative conceptualization is a physical nonequilibrium model that partitions the liquid phase into mobile and immobile regions. For this model, the sorption rate is limited by the diffusion rate to reaction sites. Diffusion into and out of immobile regions is generally modeled as an apparent first-order exchange process.

Hendriks et al. (1984) demonstrated that the two-site and two-region models have the same dimensionless form and showed that effluent curves from laboratory soil columns alone cannot be used to differentiate between chemical and physical phenomena that cause an apparent nonequilibrium condition. Recent reviews (MacKay et al., 1985; Wagener & Rao, 1985; and Abrilés, 1987) indicate these same problems of apparent nonequilibria also pertain to the transport of organic solutes and are further complicated by volatilization and the potential presence of coevolvents.

A large number of analytical two-region or "bicontinuum" models are presently available for such structured (or fractured) systems. A commonality of these models is the assumption that the chemical is transported through a single, well-defined pore or crack of known geometry, or through the unit between well-defined uniformly sized aggregates. In addition, diffusion-type equations are used to describe the transfer of solute from the larger pores into the micropores of the soil matrix. Application of these solutions to laboratory-scale experiments involving disturbed and undisturbed soil columns has been encouraging.

Still another approach for water and solutes moving together along well-defined macropores is the concept of a kinematic wave (Germann & Beven, 1985). It is too early to evaluate its success.

Stochastic models

Spatial variability of water and solute transport parameters in natural field soils has made it difficult to apply detailed process models of the type discussed above to field scale solute transport. For example, Jury (1985) reviewed field experiments of solute
transport and reported coefficients of variation of the order of 60-90% for apparent solute velocity. Recognition of soil variability is not new (Pendleton, 1979). Research during the past decades has led to the view that transport phenomena are intrinsically stochastic processes susceptible to quantitative characterization by stochastic models.

Monte Carlo simulation. Monte Carlo simulations of the solution of a deterministic equation such as (4) allow transport and storage coefficients to be random variables of the nature expected within a heterogeneous or nonhomogeneous field soil. The variable may be independent, correlated or manifest a variance structure. Based upon a preliminary sampling, parameters selected for the assumed probability density function (pdf) permit repeated solutions of the deterministic equation [e.g., \( c_s(s,t) \)]. These solutions \( c_s(s,t) \) are then used to calculate sample moments (mean, variance, etc.) which are assumed to represent the statistical properties of the underlying stochastic transport process. Amoregar-Fard et al. (1982) used Monte Carlo simulations of the solutions of (4). Two kinds of \( c_s(s,t) \) of interest. The first is that expected at any location within a field, and the second is that obtained by averaging \( c_s(s,t) \) from many locations within the field. The former is important because it is commensurate with that solute profile associated with a single crop plant or a small neighborhood of plants. The latter is important because it is the expectation of solute retention and mass emission of solutes across the entire field. Future research may allow the determination of autocorrelation and/or cross correlation lengths of the dependent variables in (4). With those lengths known, Monte Carlo simulations of \( c_s(s,t) \) could be associated with a small site or known domain in the vicinity of a particular field location. Without those lengths, the extent of the field characterized by a single simulation is not known. The opportunities afforded by using Monte Carlo simulations of the solutions of (4) hinge upon the development of methods to measure and ascertain the pdf of the transport coefficients within prescribed limits of vadose zone depth and time.

Stochastic continuum approach. This approach was initiated by Gelhar et al. (1979) studying transport processes in water-saturated aquifers and has been extended by Yeh et al. (1985) to describe steady flow in the unsaturated zone. The interest in the development of a stochastic approach was based upon recognizing the paucity of solute concentration data usually available to permit calibration of the solution of (5) for the value of the apparent diffusion coefficient. The stochastic continuum approach considers the hydraulic conductivity as a random field with spatial persistence characterized by its covariance function. However, when random variables such as \( K \) or \( D \) in (1) or (5), respectively, are represented by the sum of their mean value plus random fluctuations, a mean transport model with additional terms is obtained. By solving these stochastic equations of the local-scale water and solute transport, the functional form of the D for macro-regions is related to the statistics describing the variability. The effective parameters so obtained can then be used in process models provided that the transport has developed to the point where asymptotic analysis can be used (Sudicky, 1983; Freyberg, 1986).
Stochastic convective approach Although several stochastic convective models have been suggested, the transfer function model (TFM) of solute transport initiated by Jury (1982) provided the impetus for the more generalized concept to describe movement of a solute that may undergo physical, chemical or biological transformations as it moves through the vadose zone. The general three-dimensional TFM was derived by Jury et al. (1986) from the principle of superposition and solute mass balance. In this approach, the volume of soil between the soil surface and the outflow surface at depth \( L \) where the solute is monitored is characterized with respect to its solute transport properties by the solute travel time probability density function (pdf) \( f_s(t) \). The travel time pdf describes how long it takes solute to travel from the surface to the outflow end (which may be a depth in the soil monitored by solution samplers, a soil column outflow boundary, or a tile drain outlet). It is measured by adding a pulse of solute mass at \( t = 0 \) to the inlet end, and measuring the relative mass arriving as a function of time at the outflow end. When normalized to enclose a unit area, this relative mass versus time curve represents the pdf \( f_s(t) \).

Once the soil has been calibrated by the travel time pdf, transport of solute added as a function of time to the surface at a concentration \( C_s(t) \) emerges at the outlet end \( x = L \) at a concentration

\[
C_L(t) = \frac{1}{G} \int_0^t f_s(t', C_L(t - t')) dt'
\]

provided that the water flux remains the same as it was during the calibration of \( f_s(t) \). A modified form of this model with much greater applicability relates the transport to the outflow end of the cumulative drainage flux \( J \) (Jury, 1982).

The advantages of this approach for solute transport in the field are at least twofold. First, the formulation of transport described in (6) is general enough to encompass all linear solute transport models. Second, the transfer function may be calibrated without measuring the soil water by transport properties directly. This calibration is most easily accomplished by adding a solute pulse to the soil surface and recovering it by soil coring after a known amount of water has been added to the soil.

Scaling Scaling theory applied to field-scale transport remains immature and undeveloped. It stems from the pioneering work of Miller & Miller (1956) in which different regions of a heterogeneous field are conceptualized as being scale transformations for a hypothetical reference location. Warren et al. (1977), Tillotso & Nielsen (1984), Abouja et al. (1984), and Sposito & Jury (1985) have pursued the scaling concept to identify a scaling factor \( \lambda \) which characterizes the hydraulic properties of a particular region in a field. Bresler & Dagen (1983), assuming that \( \lambda \) is a random variable represented by a probability density function, has relied on scaling to describe water and solute transport in field soils. More recently, Vachaud et al. (1988) have shown that heterogeneity from location to location within a field or at the scale of a watershed may be approximated with scaling techniques.

Other approaches Chromatographic considerations of solute transport (Dutt & Tanji, 1982) offer possibilities that remain
relatively unexplored. Under these considerations the vadose zone can be considered as a finite number of layers with different chemical, physical and biological processes occurring simultaneously within each or any layer. Recently, Knighton & Wagner (1986) have simulated such solute transport using a continuous time Markov process and even more recently Knighton has extended the analysis to allow for short-circuiting relative proportions of the water flow and its dissolved constituents from one layer to a distribution of downstream layers in a fashion consistent with the concept of macrosopic flow.

Fractal distributions of soil heterogeneity and their influence on water and solute transport are emerging opportunities. Fractals were originally introduced by Mandelbrot to relate the length of a coastline $F$ to the length of measuring length $A$ by $F = A^{-D}$, where $D$ is the fractal dimension. If a fractal relation is found, a scale invariant phenomenon is implied. Hewitt (1986) has discussed the geometric properties and spatial correlation structure of fractal distributions and has reviewed methods for measuring the fractal character of field data and synthesizing fields with a similar correlation structure.

There are, no doubt, other intellectual frameworks available and being developed to understand and manage the quality of water moving through the vadose zone. Basic research on physical and chemical mechanisms and processes affecting solute transport remains important. The relationships between molecular and macrosopic transport requires continual investigation (Sposito, 1986; and Cushman, 1988). The development of a data base on multiscale flow parameters is no less important (Finder & Abriola 1986).

THE FUTURE

New methods and technologies of measurement are needed to keep pace with our ability to simulate field systems. A number of potentially powerful methods based on parameter estimation of various unsaturated flow (Dawson & Brussa, 1982; Noil et al., 1987) and solute transport (Parker & van Genuchten, 1984; Jury & Sposito, 1985; Wagner & Gesclicki, 1985) parameters have recently been introduced. Other papers have contributed significantly to a better technology of measurement (Paetkold et al., 1985; Topp & Davis, 1985; Dabezé & Dalton, 1985). It is imperative that research in these and related areas continue in order to couple specific sampling design with various parameter estimation methods. The key to the future is accelerated efforts in field experimentation, with the development of management tools at the farm or basin scale providing the greatest challenge.

REFERENCES


