Theory for Source-Responsive and Free-Surface Film Modeling of Unsaturated Flow

A new approach to preferential flow emphasizes rapid response to hydraulic inputs at the soil surface—an effect that can be conceptualized as laminar flow in free-surface films along the walls of pores. A mathematical formulation based on these concepts is useful to represent traditionally difficult cases of unsaturated-zone flow.

Dual-domain models are often used to quantify fluxes of water and other substances through unsaturated zone materials with complex structure. Typically, the separate domains correspond to preferential and nonpreferential (diffuse) flow. Most commonly the nonpreferential domain is formulated according to the traditional Darcy–Buckingham law for fluxes and Richards’ (1931) equation for transient water contents. The preferential domain may use Richards’ equation also (e.g., Gerke and van Genuchten, 1993) or kinematic waves (e.g., Larsbo and Jarvis, 2003) or other approaches. All options neglect or simplify some of the processes and geometries of preferential flow. There is the additional major problem that many processes critical to preferential flow are significantly different from those of nonpreferential flow and are poorly understood. Capillarity, for example, is a dominant influence in diffuse unsaturated flow but probably not in many types of preferential flow. Other needs for preferential flow models are to maximize the use of commonly available information (e.g., visually observable characteristics of rocks, soils, and weather) and reduce the need for data that are rarely known (e.g., unsaturated hydraulic conductivity as a spatially and temporally variable function of water content). Thus there is much possible benefit in exploring models that differ from what has been tried before. This study developed such an alternative and a two-domain combination of it with a more standard representation of unsaturated flow.

This work builds on source-responsive concepts and generalizations proposed in the Nimmo (2007) model of maximum speed of transport and further detailed by Ebel and Nimmo (2009). Source-responsive refers to the thesis that preferential flow may be triggered and modulated in response to the source of water to the unsaturated medium (e.g., infiltration), which may be some distance away from the subsurface point under consideration. Dependence on local potential gradients is much less than in traditional unsaturated flow models. For computation of minimum travel times, this model was formulated empirically using a wide variety of observed cases of preferential flow. Some generalizations in the source-responsive model based on these observed cases are (i) little dependence on...
The objectives of this study were: (i) to formulate a one-dimensional source-responsive flux model for unsaturated flow; (ii) to develop a plausible visualization and empirical parameterization of the source-responsive model in a free-surface film framework; (iii) to show that the source-responsive model can be combined with the Darcy–Buckingham law and with Richards’ equation for deterministic two-domain models of unsaturated-zone flow; and (iv) to illustrate the source-responsive model’s application to unsaturated-zone flow situations that are not well treated with Darcy–Buckingham or Richards formulations alone.

Unavoidably, in a multidisciplinary field with diverse applications, some key terms have multiple definitions. There is not complete agreement on the precise definitions of Darcy–Buckingham law, macropore, and others. In many cases, the differences have no significant consequence, but sometimes they can affect understanding or application of the technical subject matter. Here, certain concepts or relations make sense only with a particular definition. The appendix gives the intended meanings of terms as used here. This is not to imply that other usages are invalid but to permit more precise distinctions in terminology as necessary.
where the subscript \( D \) refers to the diffuse-flow domain and \( S \) to the source-responsive domain. No fixed relation between \( \theta_D \) and \( \theta_S \) is specified; various proportions of source-responsive flow may occur for various values of \( \theta_D \). For example, \( \theta_S \) may be nonzero even for infiltration into dry soil (low \( \theta_D \)) because the diffuse-flow and source-responsive domains are not necessarily in equilibrium. At a given time, there is division of the pore space, but the spatial domains of the two flow modes are not necessarily fixed; parts of the pore space may sometimes be of one type and sometimes the other.

From the Darcy–Buckingham law, the diffuse-flow portion of flux density is

\[
q_D(z,t) = -K(\theta_D) \frac{\partial \Phi}{\partial z}
\]

where \( z \) is depth, \( t \) is time, \( K \) is the unsaturated hydraulic conductivity, and \( \Phi \) is the total potential relevant to diffuse flow. Typically, \( \Phi = \psi + \rho g z \) (in SI units), \( \psi \) being the matric potential, \( \rho \) the density of water, and \( g \) the acceleration of gravity.

The source-responsive portion of flux density \( q_S \) can vary in magnitude from 0 to some maximum value \( q_{S_{\text{max}}}(z) \). This value represents the maximum volumetric flux density sustainable within the source-responsive domain at \( z \). Interpretable as the maximum amount of preferential flow, \( q_{S_{\text{max}}}(z) \) is a constitutive property of the medium. In general, \( q_{S_{\text{max}}} \) varies with \( z \) as the size or number of preferential flow paths varies with \( z \).

The maximum possible source-responsive flow is taken to be modulated at any given \( t \) and \( z \) by a dimensionless factor \( f(z,t) \), referred to as the active area fraction, in reference to the concept of film flow along some of the internal surface area of pores, as discussed below. The value of \( f \) ranges from 0 to 1, depending on some yet-unspecified conditions of the flow system. Then the source-responsive flux density is

\[
q_S(z,t) = f(z,t) q_{S_{\text{max}}}(z)
\]

and the total flux density is

\[
q(z,t) = -K(\theta_D) \frac{\partial \Phi}{\partial z} + f(z,t) q_{S_{\text{max}}}(z)
\]

Model with Film-Flow Concepts

To elaborate and constrain the model, assume that flow within unsaturated macropores occurs in free-surface films, probably <100 \( \mu \)m thick, lining pore walls (Fig. 1). These films are assumed to obey laminar flow principles. Film flow concepts are easily adaptable into the mathematical framework presented so far, and are likely to realistically represent significant aspects of unsaturated flow (Tokunaga and Wan, 1997; Tokunaga et al., 2000). This treatment does not rule out alternatives for representing the source-responsive domain, however. It may additionally represent other modes of flow as long as they are cast in the film-flow framework.

Within a given medium, the area potentially available for free-surface film flow might be the internal wall area of all macropores. For a pore that approximates a parallel-plane fracture, this area could be the total area of both facing walls. For one that approximates a cylinder, it could be the area of the cylinder wall, \( 2\pi \) times the product of the effective radius and length. A useful quantification as a property of the medium is a facial area density \( M(z) \) \([L^{-1}]\). For a given three-dimensional parcel of intact porous medium, \( M \) within that parcel would equal the total internal facial area of macropores within the parcel, divided by the volume of the parcel. Geometrically this quantity is equivalent to the contact length per unit cross-sectional area (Germann and Hensel, 2006) and to common definitions of fracture density. It is that portion of the total specific surface area of the medium that may allow free-surface film flow in a source-responsive manner.

From the geometry of flow domains described so far, and with film thickness \( L \), the local source-responsive water content is

\[
\theta_S = f M L
\]

In the most general case, \( f, M, \) and \( L \) could be functions of both space and time. This represents the volume of the source-responsive domain filled with actively flowing free-surface film. An assumption of laminar flow on a vertical plane leads to relations between flow speed and film thickness. The laminar-flow maximum and average vertical flow speeds of liquid in the film are

\[
V_{\text{max}} = \frac{1}{2} \frac{g L^2}{\nu}
\]

and

\[
V_{\text{avg}} = \frac{1}{3} \frac{L^2}{\nu} = \frac{2}{3} V_{\text{max}}
\]

where \( \nu \) is the kinematic viscosity of the liquid (Bird et al., 2002, p. 45). The situation of maximum source-responsive flow corresponds to all macropore facial area having a maximum supportable free-surface film thickness, \( L_{\text{max}}(z) \). The source-responsive formula for flux density in films is

\[
q_{S_{\text{max}}} = M L_{\text{max}} V_{\text{avg}} = \frac{1}{3} \frac{g L_{\text{max}}^3}{\nu} M
\]
For constraining the values given to parameters, again assuming laminar flow, film thickness $L_u$ relates to the film flux density $V_u$, the film-profile averaged flow velocity, as

$$V_u = \frac{1}{3} \frac{g}{\nu} L_u f^2$$  

Substituting into Eq. [11], the combined-domain flux density is

$$q(z,t) = -K(\theta_D) \frac{\partial \Phi}{\partial z} + V_u L_u M(z) f(z,t)$$  

With the laminar flow assumption of Eq. [12], the functional relation between $V_u$ and $L_u$ is fixed, so the coefficient $V_u L_u$ acts as a single parameter. Writing it as $V_u L_u$ recalls its geometric interpretation and its relation to flux density through multiplication by the facial area density.

It is useful to relate $V_u$ to the nominal standard maximum transport speed $V_0$ for the continuous-supply conditions of Nimmo (2007). The speed $V_0$ takes a value of $1.5 \times 10^{-4} \text{ m s}^{-1}$, empirically determined as the geometric mean of the maximum transport speeds for a set of 23 case studies of preferential flow maintained by ponded or high-surface-flux conditions. Though computed for high-input conditions, the value $V_0$ can be also be used in applying a source-responsive model to smaller fluxes by use of an on-off intermittent flow hypothesis (Nimmo, 2007). Assuming the film thickness is such that the maximum of the parabolic (laminar) distribution of velocity equals $V_0$, then the average flow velocity $V_u$ is $2/3$ of $V_0$ (Eq. [8]). Consequently, the suggested value is $V_u = 1.0 \times 10^{-4} \text{ m s}^{-1}$ and $L_u = 6 \mu\text{m}$. By taking this $V_u$ or some other value (ideally based on more data) to be a standard for problems of water flow in the near surface of the earth, independent of the type of porous medium, then $V_u L_u$, being dependent only on $g$, $\nu$, and $L_u$ is a constant characteristic of the earth and the designated liquid, water. With $\nu = 1.0 \times 10^{-6} \text{ m}^2 \text{s}^{-1}$, $V_u L_u$ is $5.5 \times 10^{-10} \text{ m}^3 \text{s}^{-1}$.

**Water Content Dynamics during Transient Flow**

For the general treatment of transient flow, changing water contents are important as well as fluxes. Explicitly considering the transfer of water between domains and defining $r(z,t)$ [$\text{T}^{-1}$] as the interdomain transfer term (taken positive in the direction from the source-responsive domain to the diffuse flow domain, analogous to absorption of water from macropores into the matrix), the continuity equations are

$$\frac{\partial \theta_D}{\partial t} = -\frac{\partial q_D}{\partial z} + r(z,t)$$  

and

$$\frac{\partial \theta_S}{\partial t} = -\frac{\partial q_S}{\partial z} - r(z,t)$$  

**Simplifications for Practical Use**

The two-domain flux Eq. [5] and [10] have too many degrees of freedom for most practical applications in the unsaturated zone, where subsurface characterizations are typically sparse and inexact.

Few, if any, direct measurements are normally available to assign values to $\theta_\psi(z), K(\theta,z), f(z,t), M(z)$, and $L_{max}(z)$, even ignoring basic complications such as hysteresis and temperature dependence. Further simplification is possible with constraints based on the known or probable character of film flow in earth materials. In particular, the values given to $M, f$, and $L_{max}$ can be constrained using the quantitative generalizations of Nimmo (2007).

A possible assumption for constraining film dimensions is a uniform thickness $L_u$ for all $z$ during free-surface film flow. The value of $L_u$ would be a characteristic of the general nature of the flow system. To account for variations in $\theta_\psi$ and $q_S$, the film can be construed to cover a fraction $f(z,t)$ of the available surface area. In effect, the medium’s macro pore facial area would be coated by a film thickness either $L_u$ or 0.

An alternative option for constraining $q_S$ between 0 and $q_{Smax}(z)$ would be to let $L$ vary with $z$ but consider film to cover all the available surface area. The previous option, however, is more likely to correspond to commonly observed features of preferential flow, such as flow speed and film thickness, based on observational evidence, for example, (i) from Nimmo (2007), that variation in $V_{max}$ is modest, (ii) from Tokunaga and Wan (1997), that film flow velocities can remain little changed across a wide range of matric potentials, and (iii) from Germann et al. (2007), that $L$ varies only within a small range.

With film thickness uniform and $f(z,t)$ representing the film-covered fraction of the macro pore facial area $M(z)$, this model could be considered an active area model, analogous to the active fracture model of Liu et al. (1998). A value $f = 0$ would mean that no liquid water is undergoing source-responsive flow in free-surface films, and $f = 1$ would mean that films of thickness $L_u$ cover all the potentially active macro pore facial area. This concept also relates to several earlier propositions, including the “bands” of Bouma and Dekker (1978), “rivulets” of Germann et al. (2007), films of Dragila and Wheatcraft (2001), and pulsed flow used for the intermittent-case source-responsive travel-time model of Nimmo (2007). With this simplifying constraint of uniform thickness, the combined-domain flux density is

$$q(z,t) = -K(\theta_D) \frac{\partial \Phi}{\partial z} + \frac{1}{3} \frac{\partial}{\partial z} L_u^3 M(z) f(z,t)$$  

For constraining the values given to parameters, again assuming laminar flow, film thickness $L_u$ relates to the film flux density $V_u$, the film-profile averaged flow velocity, as
Because the direction of transfer would, in effect, be perpendicular to the flow direction, and the development here is for one-dimensional flow, the interdomain transfer does not need to explicitly appear in equations for combined-domain flow. In other words, \( r \) represents neither diffuse flow nor film flow. The definition of \( r \) here is analogous to other models and concepts, for example the domain exchange of Beven and Germann (1981), Germann and al Hagrey (2008) and Hincapié and Germann (2009b) used the term water abstraction, although this implies water loss in the preferential-to-nonpreferential direction driven by a pressure difference between domains, whereas \( r \) here represents a general domain transfer that can go in either direction and may have drivers supplemental to pressure differences.

Putting Eq. [3] into Eq. [14], the diffuse flow transient equation is

\[
\frac{\partial \theta_{D}}{\partial t} = \frac{\partial K(\theta_{D})}{\partial z} \frac{\partial \Phi}{\partial z} + K(\theta_{D}) \frac{\partial^{2} \Phi}{\partial z^{2}} + r(z,t) \tag{16}
\]

which is Richards’ (1931) equation with a source–sink term. Similarly, for the source-responsive domain, putting the flux density Eq. [4] into the continuity Eq. [15] gives

\[
\frac{\partial \theta_{S}}{\partial t} = -\frac{\partial f q_{S_{\text{max}}}}{\partial z} - r(z,t) \tag{17}
\]

Adding these two equations gives the combined-domain equation for transient conditions:

\[
\frac{\partial \theta}{\partial t} = \frac{\partial K(\theta_{D})}{\partial z} \frac{\partial \Phi}{\partial z} + K(\theta_{D}) \frac{\partial^{2} \Phi}{\partial z^{2}} - \frac{\partial f q_{S_{\text{max}}}}{\partial z} \tag{18}
\]

equivalent to Richards’ equation with an additional term for the source-responsive domain. Putting the source-responsive maximum film flux density of Eq. [9] into Eq. [18], the general source-responsive film equation for transient flow in the combined domains is

\[
\frac{\partial \theta}{\partial t} = \frac{\partial K}{\partial z} \frac{\partial \Phi}{\partial z} + K \frac{\partial^{2} \Phi}{\partial z^{2}} - \frac{1}{3} \frac{g}{\nu} \frac{\partial f M L_{\text{max}}}{\partial z} \tag{19}
\]

With the assumption of uniform film thickness, this simplified combined transient equation becomes

\[
\frac{\partial \theta}{\partial t} = \frac{\partial K}{\partial z} \frac{\partial \Phi}{\partial z} + K \frac{\partial^{2} \Phi}{\partial z^{2}} - \frac{V_{u}L_{u}}{\nu} \frac{\partial f M}{\partial z} \tag{20}
\]

Strictly, in Eq. [18–20], \( K \) is a function of \( \theta_{D} \) rather than \( \theta \). The use of \( K(\theta) \) instead of \( K(\theta_{D}) \) may sometimes be an adequate approximation. Otherwise, Eq. [6] may be applied and the function taken as \( K(\theta - f M L) \).

## Results

Tests against direct observations are essential to demonstrate the value of a theory or model. Application of the theory presented here is dependent on the valuation of the \( M, \) and \( f \) parameters. Because these are not standardized properties for which measurement techniques have been established, a means of determining \( M \) and \( f \) values needs to be designated before the model is applied in a predictive sense. Thus, a test against measured results is a test of both the model and the method for \( M \) and \( f \) determination. The development of a general method for determining \( M \) and \( f \) from measured data is highly desirable but beyond the scope of this study. Instead, case-specific means of estimating the values of \( M \) and \( f \) from data for two examples here illustrate possible approaches to such determination. In these examples, the parameter-estimation problem is kept to a minimum by choosing situations for which a very simple form of the \( M(z) \) and \( f(z,t) \) functions is intuitively suitable. In the absence of direct property measurement techniques, some applications might use a similar approach, or more generally, inverse methods to estimate \( M \) and \( f \) on the basis of observations corresponding to model output.

The main purpose of these case studies was to assess whether the effort involved in substantial and detailed testing is worthwhile. More specifically, the case studies served to demonstrate that the source-responsive component of the model carries a strong possibility of significant improvement in prediction of preferential flow effects in situations where the Darcy–Buckingham or Richards’ formulations alone do not produce plausible results.

## Case of Rapid Water Table Response

In evaluating a fluctuating water table, fluxes (Eq. [13]) are typically of greater interest than variations in the unsaturated zone water content (Eq. [20]). An important application of water-table fluctuations is for estimation of aquifer recharge (Healy and Cook, 2002). At certain sites such fluctuations occur rapidly, deviating substantially from expectations based on diffuse flow (Gleeson et al., 2009). The case explored here is from the Masser site in Pennsylvania, where responses often show a strong and rapid hydraulic effect across distances of many meters within the unsaturated zone. For example, on 20 Oct. 1995, 110 mm of rainfall caused a 9.7-m rise, delayed by only a few hours (Gburek and Folmar, 1999; Heppner et al., 2007).

The basic formula relating recharge flux to water table depth, \( z_{w} \), is

\[
q(z_{w},t) = -Y \frac{dz_{w}}{dt} \tag{21}
\]

where \( Y \) is the specific yield of the medium at the water table. In this example, the unsaturated zone is assumed thick enough that whatever diffuse-flow recharging flux occurs is steady by the time it reaches the water table. In other words, fluctuations in the diffuse
flux have been completely damped and it represents a long-term average recharging flux. Let \( z_{w0} \) be the level the water table would attain if no preferential-flow recharge occurred for a long time. Because the continuous diffuse flux raises the water table, \( z_{w0} \) is shallower, by a constant amount, than the level the water table would attain if there were no recharge whatsoever. Assume also that the rate of decline of the water table is directly proportional to its height above \( z_{w0} \), such that the master recession curve is linear (Rorabaugh, 1960). For convenience, the proportionality constant is considered to be the reciprocal of a value \( \tau \). Also, \( H = z_{w0} - z_w \) is the height of the water table above its steady-state level, as shown in Fig. 2. Then the recession rate is

\[
\frac{dH}{dt} = -\frac{H}{\tau}
\]  

and the accretion rate due to recharge is

\[
\frac{dH}{dt} = \frac{q(z_w,t)}{Y}
\]

In the flux Eq. [13], with a steady diffuse component of recharge implicit in the value of \( z_{w0} \), the diffuse term is not needed, so that

\[ q_S(z,t) = V_u L_u M(z) f(z,t) \]

Assuming rainfall goes to source-responsive flow and contributes immediately to recharge, the rate of recharge would be limited to

\[ q_{\text{lim}} = V_u L_u M_{\text{lim}} \]

where the limiting value \( M_{\text{lim}} \) may be considered the smallest value of \( M \) for \( z \) between 0 and \( z_w \). The effect of \( M_{\text{lim}} \) is to establish \( q_{\text{lim}} \) as a critical rainfall rate analogous to the infiltration capacity, rain in excess of which does not become recharge by preferential flow. The function \( f \) is taken to be uniform in the unsaturated zone, with values

\[ f = \frac{q_{\text{lim}}}{q_{\text{lim}}} \quad \text{for} \quad q_{\text{in}} \leq q_{\text{lim}} \]

\[ f = 1 \quad \text{for} \quad q_{\text{in}} > q_{\text{lim}} \]

where \( q_{\text{in}} \) is the rainfall rate. The net rate of water table fluctuation is the sum of Eq. [22] and Eq. [23], with the recharging \( q \) taken as the source-responsive flux density Eq. [24]:

\[
\frac{dH}{dt} = -\frac{H}{\tau} + \frac{q_{\text{lim}} f}{Y}
\]

With constant \( f \), this equation has an analytical solution. In general, \( f \) will vary with the water input rate, but the analytical solution can be applied across discrete time intervals within each of which \( f \) is approximated as being constant. Then the solution in recursion form is

\[
H_{i+1} = H_i \exp\left(\frac{t_i - t_{i+1}}{\tau}\right) + \frac{q_{\text{lim}} f_{i+1} \tau}{Y} \left(1 - \exp\left(\frac{t_i - t_{i+1}}{\tau}\right)\right)
\]

This formula can predict \( H \) for a sequence of brief intervals, given one value of \( H_i \). The first term on the right-hand side represents recession by Darcian processes, and the second term represents accretion by source-responsive processes.

For this case study, precipitation and water levels were taken for measurements during a high-intensity, medium-sized (28-mm) storm at the Masser site on 10 July 1995 and a more prolonged, large (70-mm) storm on 15 Sept. 1999. The initial parameter values are: \( V_u L_u = 2.1 \times 10^{-6} \text{ m}^2 \text{ h}^{-1} \), \( Y = 0.01 \), \( M_{\text{lim}} = 4000 \text{ m}^{-1} \），\( \tau = 72 \text{ h} \)，and \( z_{w0} = 16 \text{ m} \). The value of \( V_u L_u = 6 \mu \text{m} \), as before; \( Y \) is based on direct estimates for the site (Gburek and Folmar, 1999); the other parameter values were estimated on the basis of Masser site data and interpretations (Gburek and Folmar, 1999; Heppner et al., 2007).

For \( H \) pegged at its measured value at a time shortly before the start of precipitation, Fig. 3 shows the model predictions. The simulation produced strong and rapid rises of the water table, as occur at the Masser site. In comparing the predicted to measured \( H \) values, some features are not strongly pertinent to evaluation of the source-responsive model, for example, the recession rate (behavior after peak \( H \)), which is not a source-responsive feature and is improvable with a more realistic treatment of recession than the linearity embodied in Eq. [28]. Features that are pertinent to this comparison include the rate of rise before the peak and the magnitude of \( H \) increase. Both of these compare well with the...
measured data for the more prolonged storm (15 Sept. 1999). For the brief, intense storm (10 July 1995), the magnitude of the $H$ peak is reasonably well approximated by the model, but the predicted rate of rise is somewhat faster than observed. This result suggests that additional, unaccounted-for influences may be significant in this case, but overall the predictions indicate that the source-responsive model has value for this fast-response situation that traditional unsaturated flow theory predicts poorly.

The assumption of instantaneous accretion of the water table by any amount of rainfall neglects several obvious effects. Correction for these by additional modification of Eq. [27] and [28] could improve the model predictions. One is that before source-responsive flow can begin for a given storm, a threshold amount of rainfall is probably needed to prime the subsurface medium, directly analogous to the threshold storm-response behavior described by Tromp-van Meerveld and McDonnell (2006). Another is the finite time required for water introduced at the land surface to travel to the water table.

Examining the sensitivity to $M$ by reducing the value of $M_{\text{lim}}$ by a factor of 10, with other parameters the same as before, gives the Fig. 3 curves labeled “lesser macropore density.” The reduced $M_{\text{lim}}$ limits how much rain, when it occurs at high intensity, becomes recharge. The result is a much-reduced amplitude of fluctuations, as would occur at a site where there is not such an extreme response to individual storms. With $M_{\text{lim}} = 4000 \text{ m}^{-1}$, the critical rate $q_{\text{S,lim}}$ is 0.008 $\text{ m} \text{ h}^{-1}$, just slightly below the largest rainfall rates given in this example, and has little effect, whereas $M_{\text{lim}} = 400 \text{ m}^{-1}$ creates a severe limitation. Much greater values of $M_{\text{lim}}$ would have little effect on the water table rise unless the storm were much more intense. In practice, a site that does not exhibit fast responses to rainfall could be represented using a small $M_{\text{lim}}$, implying that at least one layer of material in the unsaturated zone conveys so little preferential flow that it damps the rapidity of the response.

Considering sensitivity to the other source-responsive parameter $V_{\text{u,L,}}$, if $V_{\text{u}}$ increased by a factor of 2, entailing an increase in $L_{\text{u}}$ by a factor of 1.414 according to Eq. [12], $V_{\text{u,L,}}$ would increase by a factor of 2.83. As shown by the simulated curves in Fig. 3 labeled “greater film thickness,” this causes a slight increase in both the speed and amplitude of the rise. The simulated response appears to have small sensitivity to $V_{\text{u}}$, which is helpful in practice because there is substantial uncertainty in the value of $V_{\text{u}}$.

**Case of Transient Conditions Affected by Preferential Flow**

During transient conditions affected by preferential flow, variations in the water content as expressed by Eq. [20] are important in addition to fluxes. Infiltration into a relatively dry macroporous soil can cause rapid changes at depth during a short time. Sometimes the wetting of lower layers precedes the wetting of upper layers, contrary to Richards’ equation (Lin and Zhou, 2008). Nimmo and Perkins (2008) observed effects of this sort in a study of disturbed and undisturbed soil on the Snake River plain, Idaho. In the undisturbed experiments, in a virgin silt loam soil with highly developed macroporous structure, water was ponded for 24 h, during which time $\theta(z,t)$ was measured at small intervals of $z$ and $t$ with a neutron soil-moisture probe. Water content increased rapidly even in the lower portions of the root zone (for example at about 2-m depth <6 h after infiltration started). Especially at one of two neutron-access holes in the ponded area of undisturbed soil, there was no evidence of typical wetting-front behavior, but rather a nearly immediate flow of significant water throughout the depth range of the greatest root and animal activity. In contrast, the same experiments in disturbed, less macroporous soil showed a dominance of sharp, more slowly moving wetting fronts. Comparing the results with predictions of Richards’ equation, even using unsaturated hydraulic properties carefully measured in extensive field and laboratory experiments, Nimmo and Perkins (2008) found that for about the first 3 h of infiltration, the predictions could not successfully represent the $\theta(z,t)$ behavior of the soil. Richards’ equation could, however, reasonably predict the...
succeeding 21 h of infiltration and 76 d of redistribution. For the identical field experiments in the less macroporous disturbed soil, Richards’ equation was found consistently useful for the entire time of infiltration and redistribution.

Other investigators have observed similar water-content dynamics in the early stages of infiltration. Cey and Rudolph (2009) noted rapid subsurface spreading of a blue-stained infiltrated water in a silt loam soil with near-cessation immediately after infiltration ceased. Mahmood-ul-Hassan and Gregory (2002, p. 40) observed that most water movement “takes place within a few hours of the start of rain.” Hodnett and Bell (1990) found increases in matric potential at 0.40-m depth to cease almost immediately on the cessation of rainfall, and at 0.80 m in a few hours.

For application of the source-responsive model, root-zone conditions during ponded infiltration suggest that the active-area fraction \( f \) might be a declining function of time, nearly uniform with depth. A reasonable representation could be

\[
f = \exp \left( \frac{t_0 - t}{\tau} \right)
\]  

which equals unity (indicative of maximal preferential flow) at the starting time \( t_0 \) of the ponded interval, and then declines exponentially with a time constant \( \tau \), which would probably have a value of a few hours. The decline in \( f \) represents the increasing relative importance of diffuse flow after the early stages of infiltration; the processes involved might include a transition from source-responsive film flow to Darcian flow within macropores that fill completely with water, or a simple diminution of the rate of possible increase of water content with the increasing occupation of fillable pore space by water.

During the early infiltration period, as with the previous case study, assume that the diffuse component of flow is negligible. Then the Richards terms of the transient water content Eq. [20] can be ignored, leaving

\[
\frac{\partial \theta}{\partial t} = -V_u L_u \frac{\partial f M}{\partial z} 
\]

Incorporating the expression for declining \( f \) yields

\[
\frac{\partial \theta}{\partial t} = -V_u L_u \frac{\partial M}{\partial z} \exp \left( \frac{t_0 - t}{\tau} \right)
\]

Time-domain integration gives the change in water content during the interval \( t_0 \) to \( t \), and

\[
\theta(z,t) = \theta(z,t_0) + \tau V_u L_u \frac{\partial M}{\partial z} \left[ \exp \left( \frac{t_0 - t}{\tau} \right) - 1 \right]
\]

If there is a depth \( z_c \) below which \( \theta \) does not change during times of source-responsive flow, \( M = 0 \) at \( z_c \). Then spatial integration from \( z_c \) to \( z \) yields

\[
M(z) = \frac{\int_{z_c}^z [\theta(t_0 + \Delta t) - \theta(t_0)] \, dz}{\tau V_u L_u \left[ \exp \left( \frac{\Delta t}{\tau} \right) - 1 \right]}
\]

where \( \Delta t \) is a time interval long enough to show a significant source-responsive change in \( \theta \) at depths shallower than \( z_c \).

For the case study of \( \theta(z,t) \) measured at location NAH3 (Nimmo and Perkins, 2008, electronic supplementary material), parameter values are taken to be: \( \tau = 3 \) h (previously noted by Nimmo and Perkins as a roughly optimized characteristic time for substantial preferential flow), \( \Delta t = 3.05 \) h, \( z_c = 4.52 \) m, and again \( V_u L_u = 2.1 \times 10^{-6} \) m² h⁻¹. Figure 4 shows the calculated values of \( M(z) \) and \( dM/dz \) and Fig. 5 shows a selection of measured and modeled \( \theta(z,t) \) for comparison. The \( M \) values decline with depth, consistent with a greater intensity of macropore-creating activity near the surface. The modeled curves correspond to the general character of the observed \( \theta(z,t) \) behavior that Nimmo and Perkins (2008) found could not be adequately simulated using Richards’ equation. Water content increases fastest at the shallower depths,
where the slope of $M(z)$ has the greatest magnitude. The variation in slope of $M(z)$ between the 0.6- and 1.0-m depths corresponds to a pronounced effect on the rate of increase near $z = 1$ m, demonstrating how the source-responsive model might represent some effects of natural soil horizionation. More pronounced non-monotonicity in $dM/dz$ could represent the sometimes-observed phenomenon of water contents increasing at some depths before any evident change at one or more intervals above those depths.

**Discussion**

**Principal Features of the Source-Responsive Model**

The model presented here relates closely to other concepts and previous models, for example the film-flow domain of Beven and Germann (1981), the film-flow concepts of Dragila (1999) and Dragila and Wheatcroft (2001), the film-flow portion of the model of Tuller and Or (2001), and the film–capillary model of Peters and Durner (2008). There are also similarities to kinematic waves (Germann and Beven, 1985) and to the “water content wave” of Hincapié and Germann (2009b), conceived as an assemblage of films. These models, although similar in many respects, also have important differences, for example the water content of the medium plays a smaller role in the source-responsive model than is commonly assumed in the kinematic wave and other models.

A key concept in this source-responsive flow model is the incorporation of nonlocal influences, for example in that preferential flow behavior may depend weakly on local $\theta$ and strongly on rainstorm conditions at the land surface. This nonlocal emphasis facilitates the treatment of concurrent transport phenomena that differ drastically in their characteristic time scales. It allows the possible greater physical importance of nonlocal influences under some conditions. There is a further advantage in that the relevant nonlocal influences (e.g., rainfall rate) are usually known with far less uncertainty.

Another striking feature is the thinness of films and the smallness of pores that may contribute significantly to preferential flow. A film thickness of $6 \mu$m is much less than the size of conduit usually considered for preferential flow, but it is consistent with other thickness measurements and estimates of flowing films in unsaturated media. In the measurements of Tokunaga and Wan (1997), film thickness ranged from 2 to 70 $\mu$m, with average flux densities of $0.2 \times 10^{-6}$ to $5 \times 10^{-6}$ m s$^{-1}$. The film thickness estimates of Germann et al. (2007) are mostly in the range of 2 to 8 $\mu$m. The 6-$\mu$m value falls within the lower part of the observed range of Hincapié and Germann (2009a) and Germann and Hensel (2006). Although hypothesized as filled capillaries rather than films, the pore sizes active in preferential flow estimated by Kung et al. (2006) were predominantly about 10 $\mu$m or less. If films this thin convey a major portion of preferential flow, then the high volumes of preferential flow observed in some cases may imply that a large amount of surface area (large $M$) is involved. The lower aperture size limit of pores contributing to preferential flow, which might be little more than 10 $\mu$m, suggests a large range of pore sizes to consider with respect to preferential flow, and therefore the possibility that such pores are more abundant than is usually thought.

The hypothesized thinness of the films and corresponding high abundance of effective macropores could mean that the two domains are tightly interwoven. An important implication is that to transfer domains, a water molecule does not have to travel far. This effect may be related to observations by Cey and Rudolph (2009) and others that preferential flow within an unsaturated profile may behave as if abruptly switching on and off. It also may be related to processes of preferential flow in isolated fractures (Su et al., 2003). The effectiveness of the travel-time model of Nimmo (2007) for representing intermittently supplied preferential flow as sharply pulsed behavior may also be related to the ease of transfer between domains.

The characteristics of film flow relate well to the idea of a maximum transport speed with little sensitivity to water content or medium (Nimmo, 2007). In the source-responsive film-flow model, if the flow rate doubles, it can be explained in terms of a doubling of the film surface area (doubling of $f$) without changing the film thickness or the speed of transport. This can help explain why preferential flow may be less sensitive to properties of the medium than diffuse flow is.

Many dye-tracer studies in soil have found that within some depth intervals, only a few narrow preferential flow paths are evident (e.g., Weiler and Naef, 2003; Kasteel et al., 2005). Sometimes, at depths below such an interval, there may be a zone where preferential flow again appears to affect much internal surface area.
(e.g., Flury et al., 1994). The resulting dye patterns often show a near-surface zone (e.g., the mixing layer noted by Steenhuis et al., 1994) with much dye coverage and a similarly dyed zone a few decimeters deeper, with an in-between zone of few, narrow, widely spaced dyed paths. In the source-responsive framework, the zones of much dye coverage could be interpreted as intervals of high $M$. The middle zone would have much lower $M$, and might be dominated by filled macropores of high flux carrying capacity rather than unsaturated macropores that are film lined. Such filled macropores may be representable as part of the diffuse-flow domain. This middle layer might then be represented with low $M$ but a very large hydraulic conductivity, at least at high water content. Although large macropores do probably play a large and sometimes dominant role, smaller pores may be more important to preferential flow than is generally thought.

Relevant Properties and Parameters

The material properties in this model have a straightforward physical interpretation. The $M(z)$ property has a geometric basis and depends on the particular medium being considered, much as, for diffuse flow, $K(0)$ is a property of the medium. The fraction $f(z,t)$ depends on the medium and also on the hydraulic conditions at time $t$. At given $z$ and $t$, $f$ can be thought of in terms of the area of internal pore surface currently participating in preferential flow.

Sometimes the surface area on which preferential flow takes place may represent a substantial fraction of the total particulate surface area. The value of $M$ can be considered in terms of the fraction of the total specific surface area that it represents. For direct comparisons, the soil specific surface area $[L^2 M^{-1}]$ may be converted to volumetric form $[L^{-1}]$ by multiplying by the bulk density. For example, the silt loam soil in the infiltration case study above, with a bulk density of about 1.5 Mg m$^{-3}$ (Nimmo and Perkins, 2008), might have a specific surface of about 40 m$^2$ g$^{-1}$, or $6 \times 10^7$ m$^{-1}$. Given $M(z)$ values of $6 \times 10^4$ m$^{-1}$ and less (Fig. 4), for this soil perhaps about one-thousandth of the total particle surface area participates in source-responsive flow.

Of the three quantities representing the source-responsive domain in equations such as Eq. [10] and [11], the $V_u L_u$ parameter is most nearly constant and uniform. Its value depends on laminar flow theory and fixed properties of earth and water, and on $V_u$ or $V_0$ (also dependent primarily on the nature of typical earth materials), and so can be approximated as a constant for water flow in the earth’s unsaturated zone.

Further Development

Various additional tests, extensions, and applications may be fruitful. The character and valuation of $f(z,t)$ needs to be explored. Three-dimensional adaptation of the source-responsive model may be very useful and should be feasible, although nontrivial because of the inherent directionality of the gravitational force, which is enfolded into the formulation. The domain transfer is analogous to root water uptake with a sign change, so that the two-domain model calculations may be easier using existing solutions of Richards’ equation with root-uptake terms.

It remains to be determined how the case of a totally saturated medium best fits into the two-domain, source-responsive, diffuse-flow model. Field measurements of gravity-driven flow in such cases may suggest that source-responsive processes are active. Because forces of buoyancy oppose gravity in saturated media, however, and because Darcy’s law works well for a wide range of saturated flow situations, the best assumption may be that saturated flow occurs entirely as diffuse-domain flow. An implication would be that some fraction of the pore volume is in the source-responsive domain when the medium is unsaturated but in the diffuse domain when saturated.

The free-surface film concept can be useful in a general sense for a wide variety of unsaturated-flow situations. It may be able to represent certain equivalent geometries and flow modes better than capillarity does. Examples of other possible flow modes include fingered flow, “snapping rivulet flow” (which bridges opposing fracture walls and is sometimes connected throughout the depth of the medium), the “pulsating blob flow” of Su et al. (1999), and fracture surface-zone flow as described by Tokunaga and Wan (2001b). Any of these could be represented by an equivalent film-flow geometry, analogously to the representation of unsaturated flow with other geometric forms. Other types of flow, not considered directly here, may be represented by films if key phenomenological features are not significantly different.

The number of domains need not be limited to two. One could hypothesize three domains characterized as immobile–diffuse–preferential flow. Or, for cases of low $M$ and high flux, as mentioned above, diffuse–film–conduit flow domains. There may also be value in defining a separate near-saturation domain (e.g., Hincapié and Germann, 2009a). Another possibility, already demonstrated in the case studies here, is a one-domain model. This could apply in situations where matrix transport is negligible across the time scale of concern, in other words, where the diffuse flow domain is effectively an immobile domain. Where plausible, this approach would have the major advantage that without the diffuse flow terms, the source-responsive flux and transient water content equations are much easier to solve than traditional formulations of unsaturated flow.

Conclusions

Concepts from the Nimmo (2007) source-responsive speed-of-transport model are useful in a more comprehensive model for predicting preferential fluxes and dynamic conditions. Important among these are the constant characteristic travel velocity and the on–off behavior of preferential flow. The treatment of unsaturated flow can be improved by an approach that is not based entirely on
The issue of what distinguishes a macropore from the other pores of a porous medium may be made in terms of different types of criteria (Luxmoore, 1981), such as size, genesis, or hydraulic function. Among scientists who agree on the appropriate type of criteria, there are further differences in what the particular criteria should be. In terms of genesis, for example, investigators who agree that wormholes should be included may disagree concerning interaggregate pores or shrink–swell cracks. The pores capable of supporting source-responsive free-surface film flow could be thought of as macropores on the basis of hydraulic function, although in fact they might be smaller than the typical lower limits considered for that category. For example, a pore with an aperture of 30 μm or less might have flowing 10-μm free-surface films on one or both walls. The pores of interest to the source-responsive model include both large and intermediate sizes, and do not fit any particular pore classification in widespread use. This category would probably include most pores that would typically be considered macropores (including fractures in rocks as well as soil macropores) in addition to smaller pores that can support source-responsive flow.

The commonly used formula expressing the flux of water in unsaturated soil as proportional to a gradient of the relevant potentials is sometimes called Darcy’s law, as a generalization of the principle that Darcy (1856) presented for the case of flow through finite lengths of saturated sand. Because for unsaturated flow this generalization requires the concepts of matric potential and water-content-dependent conductivity as introduced by Buckingham (1907), this formula is sometimes called the Darcy–Buckingham law when applied to unsaturated media. This convention is adopted here, although not intending to imply that this is the only or superior name for it. Richards (1931) combined the Darcy–Buckingham law with the continuity equation to produce a formula for the rate of change of water content, and it is this formula, for this scalar quantity, that is here referred to as Richards’ equation. Thus the Darcy–Buckingham law relates to Richards’ equation in the same way Fick’s first law of diffusion relates to Fick’s second law.

The term diffuse-flow domain, as used here, refers to the idealized portion of the pore space of a medium within which flow is driven by local gradients of potential or concentration. In this domain, flow-affecting influences are transmitted across macroscopic distances only by pore-by-pore transmission through a sequence of numerous microscopic pores. In many respects, it is equivalent to what often is called a matrix domain. Microscopic here is taken to mean that the size is much smaller than a statistically significant representative elementary volume, whereas macroscopic is taken to mean the size of a representative elementary volume or greater. Here it is assumed that water flux in the diffuse-flow domain is correctly described by the Darcy–Buckingham law and transient water content by Richards’ equation.

The preferential domain is here considered as the portion of the pore space within which the water flux moves across macroscopic distances through conduits of macroscopic length, conduit being defined as a linearly extended volume through which water can flow faster than can be explained in terms of diffuse flow. In general, such a conduit could be a macropore filled with water, or a fraction of the pore volume that is water filled, or a narrow path composed of many microscopic pores of markedly higher water content than the average water content outside that path.

The source-responsive domain is the preferential domain or that portion of it within which flow rates are dominated by influences transmitted across macroscopic distances faster than they would be transmitted by diffuse flow. Within this study, all of the preferential domain is considered to be in the source-responsive domain, although this in general would not have to be the case.

Appendix: Definitions

The issue of what distinguishes a macropore from the other pores of a porous medium may be made in terms of different types of criteria (Luxmoore, 1981), such as size, genesis, or hydraulic function. Among scientists who agree on the appropriate type of criteria, there are further differences in what the particular criteria should be. In terms of genesis, for example, investigators who agree that wormholes should be included may disagree concerning interaggregate pores or shrink–swell cracks. The pores capable of supporting source-responsive free-surface film flow could be thought of as macropores on the basis of hydraulic function, although in fact they might be smaller than the typical lower limits considered for that category. For example, a pore with an aperture of 30 μm or less might have flowing 10-μm free-surface films on one or both walls. The pores of interest to the source-responsive model include both large and intermediate sizes, and do not fit any particular pore classification in widespread use. This category would probably include most pores that would typically be considered macropores (including fractures in rocks as well as soil macropores) in addition to smaller pores that can support source-responsive flow.

The commonly used formula expressing the flux of water in unsaturated soil as proportional to a gradient of the relevant potentials is sometimes called Darcy’s law, as a generalization of the principle that Darcy (1856) presented for the case of flow through finite lengths of saturated sand. Because for unsaturated flow this generalization requires the concepts of matric potential and water-content-dependent conductivity as introduced by Buckingham (1907), this formula is sometimes called the Darcy–Buckingham law when applied to unsaturated media. This convention is adopted here, although not intending to imply that this is the only or superior name for it. Richards (1931) combined the Darcy–Buckingham law with the continuity equation to produce a formula for the rate of change of water content, and it is this formula, for this scalar quantity, that is here referred to as Richards’ equation. Thus the Darcy–Buckingham law relates to Richards’ equation in the same way Fick’s first law of diffusion relates to Fick’s second law.

The term diffuse-flow domain, as used here, refers to the idealized portion of the pore space of a medium within which flow is driven by local gradients of potential or concentration. In this domain, flow-affecting influences are transmitted across macroscopic distances only by pore-by-pore transmission through a sequence of numerous microscopic pores. In many respects, it is equivalent to what often is called a matrix domain. Microscopic here is taken to mean that the size is much smaller than a statistically significant representative elementary volume, whereas macroscopic is taken to mean the size of a representative elementary volume or greater. Here it is assumed that water flux in the diffuse-flow domain is correctly described by the Darcy–Buckingham law and transient water content by Richards’ equation.

The preferential domain is here considered as the portion of the pore space within which the water flux moves across macroscopic distances through conduits of macroscopic length, conduit being defined as a linearly extended volume through which water can flow faster than can be explained in terms of diffuse flow. In general, such a conduit could be a macropore filled with water, or a fraction of the pore volume that is water filled, or a narrow path composed of many microscopic pores of markedly higher water content than the average water content outside that path.

The source-responsive domain is the preferential domain or that portion of it within which flow rates are dominated by influences transmitted across macroscopic distances faster than they would be transmitted by diffuse flow. Within this study, all of the preferential domain is considered to be in the source-responsive domain, although this in general would not have to be the case.

Acknowledgments

For helpful discussion of the ideas presented here, I am indebted to many, including Grace Su, Maria Dragila, Brian Ebel, Kim Perkins, Bill Herkelrath, Peter Germain, and Dani Or.