Macropores and water flow in soils revisited

Keith Beven¹,²,³ and Peter Germann⁴

Received 8 February 2012; revised 16 February 2013; accepted 18 February 2013; published 4 June 2013.

¹ The original review of macropores and water flow in soils by Beven and Germann is now 30 years old and has become one of the most highly cited papers in hydrology. This paper attempts to review the progress in observations and theoretical reasoning about preferential soil water flows over the intervening period. It is suggested that the topic has still not received the attention that it deserves, in part because of the ready availability of software packages rooted firmly in the Richards domain, albeit that there is still not received the attention that its importance deserves, in part because of the ready availability of software packages rooted firmly in the Richards domain, albeit that there is convincing evidence that this may be predicated on the wrong experimental method for natural conditions. There is still not an adequate physical theory linking all types of flow, and there are still not adequate observational techniques to support the scale dependent parameterizations that will be required at practical field and hillslope scales of application. Some thoughts on future needs to develop a more comprehensive representation of soil water flows are offered.


1. Introduction

[2] The general topic of macropore flow in soils and similar permeable media, and related topics like preferential flow, nonequilibrium flow, and dual-porosity flow continues to increase in popularity among researchers in various fields. Gerke et al. [2010], for instance, reported a decadal increase of more than 50% of annually published relevant papers on these topics. The review of macropores and water flow in soils of Beven and Germann [1982] (BG82 from hereon) continues to be cited and is now one of the most frequently referenced papers in hydrology journals [Koutsoyiannis and Kundzewicz, 2007]. The review paper stemmed from work that was started at the Institute of Hydrology, Wallingford, UK, during a study visit by PG supported by the Swiss National Science Foundation [Figure 1; Germann and Beven, 1981a, 1981b; Beven and Germann, 1981] and continued when we were both based at the University of Virginia in Charlottesville. We shared the experience from the UK and Switzerland of faster infiltration and downslope macropore flows than the Richards [1931] equation would normally predict.

[3] At that time, there had already been other reviews about the influence of macropores on water flows in the soil profile [Thomas and Phillips, 1979; Bouma, 1981a], discussions of microporosity, mesoporosity and macroporosity [Luxmoore, 1981; Bouma, 1981b; Skopp, 1981; Beven1981a], and recognition of the much earlier work of Schumacher [1864] and Lawes et al. [1882]. We also now know that Robert Horton was aware of the work of Lawes et al. and, in an unpublished monograph on infiltration, recognized the importance of macropores in both water and air flows in the upper layers of the soil [Beven, 2004a]. He also rejected the idea of profile (capillary gradient) controls on infiltration rates in favor of surface controls. He saw his infiltration equation as an “extinction” equation, with surface processes restricting inputs of water into larger flow pathways [Beven, 2004a].

[4] In 1982, however, the vast majority of soil hydrologists and soil physicists still adhered strongly to the Richards approach to water movement and our experience was that papers suggesting a need for alternative concepts tended to be refereed quite harshly. Only 10 years later, 400 people turned up at the ASAE special meeting on Preferential Flow in Soils in Chicago [Gish and Shirzamadi, 1991]. A major reason for this change was the need to understand why pesticides and other pollutants that should sorb strongly onto soil particles in the near surface were being found widely in routine water quality observations of groundwaters and field drains. Even if the pollutants had sorbed onto fine colloidal materials, it was difficult to understand how there could be continuous flow pathways that would allow even fine particles to reach significant depths without being filtered out by the soil matrix. It seemed then that there would be a real impetus towards a more realistic approach to representing preferential water movement in field soils and exploring the implications at larger hillslope and catchment scales. Thus, 30 years after the BG82 review, it seems worthwhile to assess the innovations in the study of soil water processes at different scales since that time. So what progress has been made?

[5] Well, there has certainly been lots of activity in the field but the dominant concept of soil physics in recent hydrological textbooks remains the Darcy-Richards equation [e.g., Brutsaert, 2005; Shaw et al., 2010]. The dominant concept underlying “physically based” hydrological
models remains the Darcy-Richards equation [e.g., Loague et al., 2006; Qu and Duffy, 2007; Ivanov et al., 2008]. The widespread use of pedotransfer functions (mostly derived from experiments on small soil samples) presumes that the Darcy-Richards equation holds at larger scales of application, and that the parameter values are constant in time and space for a given soil horizon [see, for example, Wösten, 1999; Acuitis and Donatelli, 2003; Schaap et al. 2001; McBratney et al., 2002]. The dominant concept in solute transport remains the advection-dispersion equation, normally implemented as a symmetric Gaussian velocity distribution around a local mean pore water velocity derived from Darcy or Darcy-Richards theory. If preferential flows are important, this will not properly represent their effect on transport, even if the mean velocities are of the correct magnitude. A reason for continuing to use such models has been expressed in terms of understanding why they fail in comparison with field observations, so that they can be improved in future [Ebel and Loague, 2006; Loague et al., 2006; James et al., 2010] even though there should already have been an expectation of failure based on the field tracer evidence and from a consideration of the physics of Darcy-Richards flow in unsaturated heterogeneous soils [Beven, 1989, 2001].

So why has the Darcy-Richards representation of flow in partially saturated soils remained so popular, despite the evidence? One reason surely is that computing power has increased to such an extent in the last thirty years that the use of software packages to solve the Richards equation has become comfortable for a wide range of users. In addition, modifications of the Richards approach evolved to take some account of preferential flows. Generally accessible codes, like HYDRUS now provide the capability for applying different dual porosity and dual permeability representations [Simůnek and van Genuchten, 2008; Radcliffe and Simůnek, 2010]. Other codes have been developed to represent preferential flows as an additional nonequilibrium flow component but often without a secure physical foundation (see later under Theoretical Studies).

[7] Arguably, the availability of such tools has diverted attention from more fundamental research on macropore and preferential flow, but there has been a wide range of experimental and modeling studies published on these topics since BG82. These have predominantly been at the profile to lysimeter or small plot scale, where vertical flows dominate. In moving from the profile scale to hillslope scale subsurface stormflows it is necessary to consider the integration of vertical and lateral or downslope flows, including the effects of the capillary fringe on celerities and displacement of stored water mentioned in BG82. It is important to differentiate between pore water velocities that control transport processes, and celerities that control the propagation of perturbations through the system and therefore control hydrograph responses [e.g., Beven, 2012, chap. 5]. At local scales, a plot or field is often treated as a collection of vertical columns for both flow and transport calculations [e.g., Beven et al., 2006], but in shallow soils, or where the water table is close to the surface, spatial interactions become important. In particular, the importance of macropores and preferential flows in facilitating rapid subsurface stormflow responses has been incorporated into developing perceptual models of hillslope hydrology (See section 5 below). There is now a better appreciation of the different roles of wave celerities and water velocities in these interactions (though note that the discussion of “translatory waves” goes back at least to the 1930s, see Beven [2004b]).

[8] As noted earlier one of the primary drivers for interest in preferential flows and macropores in soils was the problem of explaining how pesticides and other sorbing pollutants were being transported to field drains, groundwater and rivers [Flury, 1996]. This interest has not abated, and there are many recent papers that address this problem [see, for example, Kladivko et al., 2001; Zehe and Flühler, 2001a; Reichenberger et al., 2002; Ciglasch et al., 2005; Beven et al., 2006], including the transport of viruses and bacteria [Darnault et al., 2003; Germann et al. 1987]; the role of colloids in facilitating transport of sorbing compounds through large soil pores [Villholth et al., 2000; Germann et al., 2002a]; and factors such as water repellency in inducing preferential flow in certain locations [e.g., Bauters et al., 1998; Blackwell, 2000; Cerda et al. 1998]. The importance of macro pore connectivity has also been demonstrated [Andreini and Steenhuis, 1990; Allaire-Leung et al., 2000; Rosenbom et al., 2008].

[9] This has also become a regulatory issue. The licensing of potentially harmful products depends on some assessment of how easily they are transported in the environment. That assessment, however, must be done across a wide range of soil characteristics, land management, and soil water conditions. Classification or indexing [e.g., Rao et al., 1985; Quisenberry et al., 1993; Schlather and Huwe, 2005; Sinkevich et al., 2005; Stenemo et al., 2007; McGrath et al., 2009] and modeling [Steenhuis et al., 1994; Jarvis et al., 1994, 1997; Stewart and Loague, 1999] approaches have been used to assess the relative risk of transport of different types of products that might be harmful to the environment in ways that reflect the possibility of preferential flow.
[10] The questions raised in BG82 about “When does water flow through macropores in the soil? How does water flow through macropores in the soil? How does water in a macropore interact with water in the surrounding soil? How important are macropores in terms of volumes of flow at the hill slope or catchment scale? What are the implications of macropores for movement of solutes and chemical interactions in the soil?” are still relevant today [see also Jury, 1999] but relatively few testable concepts have evolved since that time. Although all of these questions have been worked on during the last 30 years, revisiting BG82 may offer an opportunity to reassess the achievements and the needs in this area of soil physics.

2. The Aims of This Paper

[11] Preferential flows of different types have been the subject of significant interest over this period and there have been a number of review articles and special issues of journals published in the last decade or so [see Bryan and Jones, 1997; Blackwell, 2000; Sidle et al., 2000; Uchida et al., 2001; Simůnek et al., 2003; Gerke, 2006; Jarvis, 2007; Coppola et al., 2009; Alaire et al., 2009; Köhne et al., 2009a, 2009b; Clothier et al., 2008; Morales et al., 2010; Beven, 2010; Chappell, 2010; Jones, 2010; Hencher, 2010; Bachmair and Weiler, 2011]. Our intention here is not to repeat the information presented in these papers, but rather to focus on the issues that remain and make some suggestions about ways in which they might be addressed. Throughout, given the large body of relevant literature, the choice of citations is intended to be illustrative of the issues rather than complete. We will suggest more radical alternatives to the “safe” Darcy-Richards based concepts incorporated into dual-continuum and dual-permeability flow models. We will follow the broad structure of the original paper in moving from local scale to catchment scale issues, and differentiating experimental evidence from theoretical studies, with the aim of encouraging a more realistic physical basis for future studies on water flows in soils.

3. The Occurrence of Preferential Flow in Soils

[12] BG82 cautiously staked out some common ground by stating that: “There has long been speculation that large continuous openings in field soils (which we will call macropores) may be very important in the movement of water—at least under certain conditions. Such voids are readily visible, and it is known that they may be continuous for distances of at least several meters in both vertical and lateral directions.” The implication then was that water flow in such readily visible pores would be subject to rather different process controls than water flow in the soil matrix. There is basic agreement that soil macropores originate from the processes of desiccation, growth and decay of roots and mycelia, and burrowing animals [BG82, and more recently Coppola et al., 2009 and Bachmair and Weiler, 2011]. Extension to consideration of sediments and rocks may include fissures and karst formations that have been the subject of interest in hydrogeology, radioactive waste disposal and petroleum engineering. Soil features, including cracks, root channels, worm holes and other biologically induced macropores, are typically restricted to the depth of the soil profile, but may link to openings in the regolith and bedrock which may extend over tens or hundreds of meters in fractured rock systems. Dubois [1991], for instance, injected dye into the granite formation approximately 2000 vertical meters above the Mt. Blanc tunnel between France and Italy. About a hundred days later he found the tracer in water samples taken from the tunnel’s drainage system. Hillslope tracer measurements also suggest connected flow pathways extending over at least tens of meters [e.g., Nyberg et al., 1999; Weiler and McDonnell, 2007; McGuire et al., 2007; Graham and McDonnell, 2010].

[13] There is an important point to be made here about the Darcy and Richards equations. Both are based on experimentation under particular conditions. Both are consistent with those conditions by back-calculation of a coefficient or function of proportionality that we now call the hydraulic conductivity. In the case of Darcy [1856], who looked at saturated conditions in steel cylinders filled with sorted sand, the linearity between hydraulic gradient and flux rate should hold in the field, provided that the flow remains in the laminar and stable flow regime (although Darcy’s data were only very nearly linear, Davis et al. [1992]). In the case of Richards [1931], who looked at unsaturated conditions imposed by sequentially decreasing capillary pressure in a confined sample using a hanging water column, the experimental conditions preclude preferential flow in larger soil pores, which are a priori drained at each step of the decrease in capillary pressure. By then imposing a pressure gradient, Richards experimentally created a steady flux at different points on the retention curve.

[14] Richards [1931, p. 322] explained: “When the conditions for equilibrium under gravity … are fulfilled, the velocity and acceleration of the capillary liquid are everywhere zero . . . which means that the force arising from the pressure gradient just balances gravity.” He used pressure interchangeably with capillary potential in unsaturated porous media. Because neither velocity nor acceleration occurs, hydrostatic conditions are assumed according to the material-specific retention curve. Further “If this condition does not obtain there will be a resultant water moving force and in general there will be capillary flow.” (p. 322). Thus, in his view (and resulting equations), only gravity and capillarity drive capillary flow, while the flow-restraining forces are summarized in the bulk “. . . factor of proportionality . . .” (p. 323) as Richards called the hydraulic conductivity.

[15] While Richards’ equation may be valid within that experimental framework, it should not be a surprise that the concept may not carry over to field conditions when the air pressure within well-aerated soils is atmospheric, the soil is heterogeneous in its characteristics and the fluxes are subject to dynamic effects (including preferential flows). Already in the capillary range of pore sizes, where it is commonly assumed that a Richards type of description is appropriate, there is a problem in applying the theory at scales of interest in heterogeneous, unsaturated soils. The problem is that capillary potentials are then rarely in equilibrium such that in there is no consistent hydraulic gradient, even in the absence of preferential flows. Since the constant of proportionality is a nonlinear function of
capillary potential or water content, it will also vary with the heterogeneity. Thus, at the scale of a useful landscape element, the heterogeneity means that the Richards equation will not hold the physics tells us that some other form of equation should be used even if the Richards equation holds at the small scale. Some attempts have been made to define such an equation given assumptions about the heterogeneity [e.g., Yeh et al., 1985a, 1985b; Russo, 1995, 2010] or to simulate the effect of randomly generated fields [Mantoglou and Gelhar, 1987; Binley et al., 1989; Binley and Beven, 1992; Russo et al., 2001; Fiori and Russo, 2007, 2008]. Basically, however, knowledge about the heterogeneity will never be available so that the impact of heterogeneity has been consistently ignored. Richards equation is applied as if the soil were homogeneous at the scale of the calculation element. Similar considerations apply when equilibrium, dual porosity or dual permeability representations of preferential flow are applied within a Richards domain (see below). This is not a new insight, it has been known for decades [e.g., Dagan and Bresler, 1983], but with very little effect on modeling practice [see discussion in Beven, 1989, 2001, 2006, 2012].

Given these limitations of the Richards approach when used in practice (that relationships derived under steady equilibrium assumptions should not be expected to apply in dynamic cases and that local equilibration will not apply in heterogeneous unsaturated soils) we hope to encourage a radical rethink of how preferential flows are represented at the plot, field, hillslope and catchment scales of interest. We suggest that the Richards approach to representing fluxes in field soils should not be considered to be physically based, but a convenient conceptual approximation.

[16] Flow in the structural voids of the soil will result in nonequilibrium conditions when water cannot move fast enough into the smaller pores of the surrounding matrix to spontaneously and continuously achieve equilibrium according to the retention curve. Germann et al. [1984], for instance, reported bromide concentrations to decrease with increasing horizontal distance from stained macropores. Jarvis [2007] describes nonequilibrium conditions as follows: “As water starts to flow into large structural macropores, the sharp contrast in pore size and tortuosity with the surrounding textural pores leads to an abrupt increase in water flow rate for only a small increase in soil water pressure. The resulting non-uniform flow (physical non-equilibrium) can be illustrated by imagining a soil block that contains macropores wetting up towards saturation during infiltration.” This implies that macropores carry water quite independently from antecedent soil moisture and capillary flow such that (at least) a dual-pore structure containing the matrix and the preferential flow supporting macropores is needed.

4. Experimental Evidence

4.1. Profile and Plot Scale

[17] Some of the most convincing evidence of preferential flow in soils has come from the use of tracers and dye staining. Most of these experiments have been carried out at the profile and plot scales [e.g., Figure 1; Flury et al., 1994; Abdulkadir et al., 1996; Henderson et al., 1996; Vilholth et al., 1998; McIntosh et al., 1999; Kung et al., 2000, 2001b; Zehe and Flühler, 2001a, 2001b; Stamm et al., 2002; Weiler and Naef, 2003b; Weiler and Flühler, 2004; Bachmair et al., 2009; Blume et al., 2009; van Schaik, 2009; and many others]. Preferential flow has to be inferred from tracer data; Kung et al. [2000] for example conclude from the similar early breakthroughs of both sorbing and conservative tracers that flow must be through preferential pathways. Dye staining may not reveal all preferential pathways; only those that have been connected to a source of dyed input water. It is in fact quite common in such studies to see readily visible macropores that are not stained, or have been only partly stained. This may be because of a lack of connectivity to the source, or because they were not flowing at capacity, or that they contained water displaced from the matrix by the input water.

[18] Watson and Luxmoore [1986] were probably the first to apply tension infiltrometers to relate flow rates to pore diameters in situ (according to Clothier and White [1981]). They applied water to the soil surface at preset capillary potentials of 0, −3, −6, and −15 hPa, thus implementing a Dirichlet-boundary condition. An analysis of their results suggested that 95% of flow occurred in pores that were wider than 250 μm, which occupied just 0.32% of the pore volume. These field data support the model results of Beven and Germann [1981] who demonstrated the paramount contribution to flow by only a very small portion of a soil volume containing wide continuous voids. The resulting parameters are frequently related with the corresponding water contents in order to construct an apparently physical relationship. These sequential steady state experiments are consistent with the equilibrium conditions of Richards. However, the imposed Dirichlet boundary condition is far from realistic. Water typically arrives at the ground surface either in discrete drops, as in rain, in streams as in concentrated stem flow, or ponded in surface hollows, such that pressure in the arriving water remains positive or at least atmospheric. Thus, preference should be given to the more realistic flux-controlled infiltration observing a Neumann boundary condition, albeit that the spatial pattern of fluxes might be difficult to assess under other than controlled experimental conditions as a result of throughfall, stemflow, surface irregularities, collection area to a surface connected macropore, other surface controls, etc. [e.g., Weiler and Naef, 2003a]. This raises the question when and where in a permeable medium significant capillary forces start to act on the infiltrating water. Wide enough macropores will preclude any significant effect of capillarity. Turbulent pipe flows [see, for instance, Jones, 2010] are the most impressive representation of such flows. Capillary sorption of water has still to be considered regardless of the width of the preferred flow path but may be restricted by the local microstructure of the soil (such as the cutans of translocated clay particles, earthworm mucus at the edges of larger pores, or hydrophobic excretions by plants, e.g., Cerda et al. [1998]).

[19] Other configurations may also lead to flows that occur close to atmospheric pressure in nonsaturated permeable media. Germann and al Hagrey [2008], for example, reported from studies of data collected in the 2 m deep sand tank in Kiel that the capillary potential collapsed to close to atmospheric pressure behind the progressing wetting front. The tank was uniformly filled with sand in the
were between...saturating that water contents were always far from saturation. But maximum capillary potentials during infiltration...constant average velocity followed from the slope of the linear relationship...volumetric water contents at nine depths. Depth averages of the initial and maximum water contents were 0.08 and 0.26 m$^3$ m$^{-3}$, respectively, while the maximum degree of saturation was 0.54 of the porosity of 0.47 m$^3$ m$^{-3}$, demonstrating that water contents were always far from saturation. But maximum capillary potentials during infiltration were between $-17$ and $-25$ hPa, indicating that infiltration behind the wetting front occurred close to atmospheric pressure. The data from tensiometers and TDR waveguides showed that the wetting front moved with a constant average velocity of $3.2 \times 10^{-3}$ m s$^{-1}$ from top to bottom during the 16 h of sprinkler infiltration. The average wetting front velocity followed from the slope of the linear relationship between the depth of the wetting front and its respective arrival time with a coefficient of determination of 0.994 (Figure 2). Richards equation will predict a linear propagation of the wetting front only if the effect of the capillary gradient is negligible relative to gravity but this should not be the case, of course, for wetting of a dry soil.

[20] A variety of other researchers have found constant velocities of wetting fronts in sand boxes, including in the absence of natural macro pores (see Table 1). The boxes of Sellker et al. [1992] and Hincapié and Germann [2010] were 0.9 m and 0.4 m long, respectively. At larger scales, Rimon et al. [2007] and Ochoa et al. [2009] found in situ constant wetting front velocities due to percolation in undisturbed sediments to depths of 21 and 3 m. Constant wetting front velocities that persist over prolonged time periods and considerable depths are a strong indication of non-Richards behavior during infiltration and percolation demanding a fundamentally different approach to the representation of the flow. Figure 3 shows the results of an analysis of 215 such sets of observations in the form of a distribution of derived wetting front velocities.

4.2. Field Scale

[21] Studies at the profile and plot scale are already difficult. At the field scale, additional complexity is introduced. Transport to field drains, groundwaters, and rivers will depend on preferential flows induced by the heterogeneity of soil properties, including in some cases natural pipes or agricultural drains (e.g., Figure 1). Such heterogeneities are not necessarily obvious from the soil surface and may be the result of the history of soil development and land use, ranging from deep tertiary weathering and Holocene periglacial structures to recent land management and drainage histories. Some larger scale staining experiments have been carried out [e.g., Kung, 1990; Noguchi et al., 1999; Wiennhöfer et al., 2009; Anderson et al., 2009a, 2009b] but the evidence for the importance of preferential flows at this and larger scales tends to be indirect, inferred from the bulk responses of natural or artificial tracer concentrations at some measurement point [e.g., Hornberger et al., 1990; Nyberg et al., 1999; Rodhe et al., 1996, Kienzler and Naef, 2008; McGuire and McDonnell, 2010]. Kung et al. [2005, 2006], however, show how tracer experiments to field drains under different imposed flux rates can be interpreted, under laminar flow assumptions, in terms of consistent distributions of effective pore radii.

[22] One particular area of interest in respect of preferential flow at the field scale is for understanding and predicting groundwater recharge. There have been many recharge studies where observations of rapid water table responses have been reported, or where tracers or pollutants have been reported at depths much greater than would be expected without inferring preferential flow. A recent example is the study of Cuttbert et al. [2010], where rapid groundwater responses in a sandstone aquifer overlain by a till soil, even during summer periods, were interpreted as recharge that occurred when a dry soil matrix was bypassed by preferential flow through fractures in the clay.

[23] Macro pores are considered to be the soil structures most vulnerable to mechanical compaction due to heavy machinery used in agriculture, forestry, and on construction sites [e.g., Chappell, 2010]. In addition to the earlier papers cited in BG82, Alaoui and Helbling [2006] investigated with sprinkler infiltration and dye-staining tests the effects of cattle trampling, driving with a six-row sugar beet harvester, and soil reconstruction procedures on macropore flow. They concluded that trafficking sealed the soil surface almost completely. Cattle trampling, on the other hand, negatively affected mostly fine pores, while increasing macroporosity in some instances. Infiltration into reconstructed soils was mainly through the soil matrix because no macro pores had formed since reconstruction.

4.3. Hillslope Scale

[24] The transition from preferential vertical flow in soil profiles to preferential lateral flow i.e., subsurface
stormflow in hillslopes is typically related to morphological features like low permeability horizons in soil profiles or regolith, compact glacial till or bed rock. Common to these features are their decreased hydraulic conductivities in comparison with the strata closer to the soil surface, and they are related to the perching of water tables at least under conditions of high degrees of saturation. Greiner [1984] installed more than 700 tensiometers in a hillslope segment of about 38 m × 7 m at seven layers above the low permeability sandstone bedrock at the 1.6-m depth. Average slope angle was about 28%. He demonstrated that capillary potentials were insignificant relative to gravity potentials under conditions of pronounced lateral flow. Ochoa et al. [2007] demonstrated that low permeability sandstone bedrock at the 1.6-m depth.

Table 1. Constant Wetting Front Velocities From Sand Boxes, a Sand Tank, an Alluvial Deposit a Sand Dune, and an EutricCambisol

<table>
<thead>
<tr>
<th>Author</th>
<th>System, Dimensions (m)</th>
<th>Recording</th>
<th>Infiltration Rates (m s⁻¹)</th>
<th>Wetting Front Velocities (m s⁻¹)</th>
<th>Samples</th>
<th>Coefficient of determination, r², of ΔWO(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selker et al. [1992]</td>
<td>Sand box 0.98 m high</td>
<td>Video taping slanted F-TDR probes</td>
<td>1.7 × 10⁻⁴</td>
<td>3.1–4.4 × 10⁻³</td>
<td>1 run 5 fingers</td>
<td>0.99</td>
</tr>
<tr>
<td>Selim et al. [2007]</td>
<td>Sand dune &gt; 21 m deep</td>
<td></td>
<td></td>
<td></td>
<td>1 run 34 depths</td>
<td>0.94</td>
</tr>
<tr>
<td>Germann and al Hagrey [2008]</td>
<td>Sand tank 2 m deep</td>
<td>Horizontal TDR probes</td>
<td>4.3 × 10⁻⁶</td>
<td>3.3 × 10⁻⁵</td>
<td>1 run 9 depths</td>
<td>0.99</td>
</tr>
<tr>
<td>Ochoa et al. [2009] (Plot 5)</td>
<td>Alluvium 1.6 m deep</td>
<td>Horizontal TDR probes</td>
<td>3.7 × 10⁻⁵</td>
<td>2.7–5.4 × 10⁻⁴</td>
<td>8 runs 4 depths</td>
<td>0.89–0.98</td>
</tr>
<tr>
<td>Hincapié and Germann [2010]</td>
<td>Fingerflow in sand box 0.05 × 0.4 m, 0.4 m high</td>
<td>Neutron radiography</td>
<td>3.3 × 10⁻⁵</td>
<td>7.1 × 10⁻⁴</td>
<td>1 run 1 finger 15 depths</td>
<td>0.99</td>
</tr>
<tr>
<td>Schütz [2002] in Germann [2013]</td>
<td>Eutric Cambisol at 0.26, 0.37, 0.47 m</td>
<td>TDR probes</td>
<td>2.1 × 10⁻⁵</td>
<td>1.7 × 10⁻⁴</td>
<td>1 run, 3 depths</td>
<td>0.99</td>
</tr>
</tbody>
</table>

25 [BG82] noted the possibility that macropores and pipes might be important in controlling the hydrological response and transport processes at the hillslope scale and summarized the published evidence available at that time (Table 2 in the original paper). Recently, Jones [2010] and Chappell [2010] have reviewed the occurrence of preferential flows and piping of different types on hillslopes in numerous environments from the Amazon to Alaska. There are now a wide variety of papers suggesting that preferential flows might often be important in controlling hillslope hydrology, and by implication patterns of solute transport, but that these subsurface pathways might, in some environments, only be revealed by digging into the soil. The networks revealed by such destructive sampling can, however, be extensive, even if not all obvious macropores are connected to a flow network [e.g., Noguchi et al., 1999; Terajima et al., 2000; Weiler and Flühler, 2004; Anderson et al., 2009a; Graham and McDonnell, 2010]. The networks of obvious macropores are often disconnected, such that continuous downslope flows require some mechanism for transmitting flow from one part of the network to another. Where the soil is saturated, this can be achieved by the propagation of pressure waves through the system, such that the build up of pressure in an upslope macropore link will cause the displacement of stored water into a pathway further downslope. This can also be a mechanism for the displacement of water from the matrix into an empty

Figure 3. Distribution of 215 wetting front velocities determined in situ for 25 soils covering Udalf, Ochrept, Umbrept, Aquod, Ferrod, Humod, and Udult suborders from Germann and Hensel [2006]; and wetting front velocities from (1) al Hagrey et al. [1999], (2) Dubois [1991], and (3) Selker et al. [1992].
macropore. There is also evidence that networks of connected lateral flow pathways may develop, including at the base of the soil profile [e.g., Freer et al., 2002; Battle and MacDonald, 2002; Anderson et al., 2009b; Graham and McDonnell, 2010], even when there is percolation into the bedrock [e.g., Tromp-van Meerveld et al., 2007].

4.4. Catchment Scale

[26] If preferential flows are important at the hillslope scale then it implies that they are also important in catchment scale responses [see Uhlenbrook, 2006; Graham and Lin, 2011]. Determining if preferential flows are significant at catchment scale, however, remains a difficult issue when, very often, the only information available is the sequence of rainfall (or other) inputs and observed discharges, unless there are directly observable pipe sources of runoff [e.g., Uchida et al., 1999]. Even where local measurements of water table responses, or solute or isotopic concentrations in discharges are available, it can be difficult to determine the significance of preferential flows because of the variability in space and time of flow pathways and processes [e.g., Uchida and Asano, 2010; Jensco and McGlynn, 2011; Bachmair et al., 2012]. It might be difficult to extend knowledge at the hillslope scale to larger catchment scales because of the variability of responses in space and time and the potential for deeper flow pathways to become more significant at the catchment scale [Uchida et al., 2005b; Tromp-van Meerveld et al., 2007; Bachmair and Weiler, 2013]. Direct tracing of flow pathways remains difficult, even at the zero-order catchment scale: the logistics of providing sufficient quantities of a sufficiently conservative artificial tracer with a relatively homogeneous application remain demanding.

[27] One of the most successful attempts to do this was carried out at the Gårdstång catchment in Sweden, where the aim was to understand the flow and geochemical processes involved in the response of catchment to acid rain inputs [Rodhe et al., 1996]. In this study, a headwater catchment of 6.3 ha was roofed. All the rainfall inputs on the roof were collected and then an equivalent amount of water was added to the catchment using a network of sprinklers. The added water was taken from a nearby less acid lake with different isotopic characteristics. The change in isotopic signal was compared to a similar nearby catchment that was subjected only to natural rainfalls.

[28] Rodhe et al. [1996] analyzed the results from this experiment in a number of ways, one of which was to find a transfer function that best fit the observations. Their results showed that transfer function is highly skewed towards the origin, suggesting that there was a significant fast transfer of new water during rainfall application periods (although the mean travel time was 65 days). The soil at Gårdstång exhibits a strong exponential decline in saturated hydraulic conductivity with depth [Nyberg, 1995] so that these rapid transfers are most likely to occur in the upper, high conductivity, soil layers.

[29] There are suggestions that flow from pipes in the soil, with very rapid responses to rainfall, can, at least under some circumstances, be dominated by “old” or pre-event water [e.g., Sklash et al., 1996]. This led Kirchner [2003] to suggest that there was a double paradox of hillslope hydrology: that the subsurface response could be so fast and that the chemical concentrations could change significantly if the response was dominated by pre-event water. This need not be such a surprise, however, if the type of displacement process referred to in the previous section is invoked. We should not expect preferential flow lines to be continuous links between rainfall inputs and stream channels or rills. Very often there will be discontinuities in any preferential flow network, but the build up of saturation at a discontinuity can be effective in displacing stored water into downstream fast flow pathways. That water might be pre-event, or “old” water, but with different chemical characteristics to that contributing to pre-event baseflows [see, for example, the increasing proportion of soil water in the wetting up sequence in Iorgulescu et al., 2005, 2007]. This provides one explanation of the old water paradox of Kirchner [2003] [see also Rasmussen et al., 2000; Williams et al., 2002; Bishop et al., 2004; Weiler and McDonnell, 2007; Kienzler and Naef, 2008]. At the hillslope scale, in the simplest kinematic wave analogy, this can be expressed as the difference between the wave celerity that depends on storage deficits and the pore water velocities that will depend on storage [see Beven, 2012, p.145]. Resolving the paradox might also depend on the distribution of preferential flow velocities and the hydraulic conductivity profile as the upper soil layers contribute to downslope flows and the mix of waters being displaced in a “transmissivity feedback” mechanism [Bishop et al., 2004], although, as noted earlier, zones of higher preferential flow are not always found close to the surface.

5. Can We Now Clarify the Perceptual Model of Preferential Flows?

[30] A perceptual model of a system need only be qualitative, and can encompass all the complexity that is perceived as being important in the nature of the processes involved [Beven, 1987]. A variety of perceptual models of the role of macropores and preferential flows in profile, hillslope and catchment responses have been presented [e.g., Beven, 2004a, 2012, chap.1; McDonnell, 1990; Brammer and McDonnell, 1996; Noguchi et al., 1999; Battle and MacDonald, 2002; McGlynn et al., 2002; Williams et al., 2002; Scherrer et al., 2007; Weiler and McDonnell, 2007; van Schaik et al., 2008; Graham et al., 2010; Bachmair and Weiler, 2011]. Our own perceptual model recognizes that there may be a wide range of flow velocities in a wide range of pore sizes in the soil. Preferential flow is not just associated with “macropores” in the soil. Typical film thicknesses for preferential flows appear to be in the approximate range from 3 to 100 μm (see discussion above, and Hincapié and Germann [2009a]). Under such conditions we expect the viscous shear force to be weak enough for gravity to dominate flow in accord with laminar flow principles of momentum dissipation at water-solid interfaces. In smaller pores, capillarity and capillary gradients can play a role in controlling water movement, but we expect the range of such effects to be small scale and local. In some larger pores, there may be transitional or turbulent flows at higher flux rates.

[31] The supply of water to preferential flow pathways does not require the pathways to be directly connected to the surface supply of water (as suggested by Horton and
others since, see Beven [2004a]). This might occur, of course, in which case the generation of preferential flows might be dependent on the complex redistribution of incoming rainfalls as throughfall and stemflow by a vegetation cover [e.g., Liang et al., 2009]. However, the supply might be the result of local saturation of the matrix resulting in displacement into larger pore spaces, leading to the suggestion that water in preferential flow pathways during wetting may be “pre-event” water displaced into faster flow pathways (e.g., the responses of natural ephemeral pipes studied by Sklash et al. [1996]). Such preferential flows are likely to be in the form of film flows, with transitions to turbulent flows in larger macropores in the soil.

[32] Preferential flow can take place regardless of the antecedent moisture content of the bulk soil. The propagation of preferential flow will be practically limited by the duration and intensity of inputs. When the input ceases, a drying front starts. It eventually catches up with the wetting front of preferential flow after which the front will dissipate (see Appendix A). An important criterion is whether the distance at which this takes place is long enough to induce a response at a less permeable layer, for instance, at the base of the profile or hillslope. This is therefore a critical link between profile, hillslope and catchment scale responses that may require a reinterpretation of macropore volume filling, “field capacity” or “fill-and-spill” hypotheses of hillslope scale responses in terms of a multiple process mechanism under wetting and drying events.

[33] In moving from profile to hillslope to catchment scales, as with any representation of hydrological processes, it is also necessary to invoke greater heterogeneities of input conditions, soil characteristics, regolith structure and geology [Freer et al., 2002; Kienzler and Naef, 2008; Graham et al., 2010; Graham and McDonnell, 2010]. Kosugi et al. [2008], Tromp-van Meerveld and Weiler [2008], and others criticize the common assumption that the bedrock can be considered impermeable, while Tromp-van Meerveld et al. [2007] demonstrate that losses into granite bedrock are important at the Panola site.

[34] We expect that at the hillslope scale, connectivity of larger pores will be an issue in bulk water movement when film thicknesses start to reach the flow capacities of smaller pores, but it is not necessary to have direct connectivity of larger pathways to have longer range preferential flow effects; the buildup of local saturation between larger, discontinuous pathways can act to link pathways by displacement and replenishment of water storage between larger pores [see, for example, Stauffer and Dracos, 1982; Weiler and McDonnell, 2007; Nieber and Sidle, 2010; or Klaus and Zehe, 2010], although some of these investigations treat preferential flow as discontinuous lines of higher hydraulic conductivity in a Darcian framework.

[35] This perceptual model implies certain requirements for a new strategy for representing the effects of preferential flow in soils when moving from profile to hillslope and catchment scales. Based on the discussion above, a functional specification for a new approach can be outlined. As with any functional specification it is probable that not all the requirements can be met at the current time. The functional specification does, however, set out a goal for the future that can act as a driver for technological and scientific advances. Any such functional elements would need (at least) the following components for the case of the response of preferential flow to an input:

[36] 1. A representation of the interaction between input variability, (nonstationary) surface characteristics, antecedent conditions, and the generation of preferential flow in ways that reflect the (non-Richards) dynamics of capillary driven flow in a heterogeneous matrix, viscosity dominated laminar preferential flows, and transition to turbulence at higher flux rates in larger macropores.

[37] 2. A representation of the propagation and dissipation of preferential flow wetting and drying fronts that distinguishes velocities and celerities in a way that reflects the scale of a calculation element.

[38] 3. A way of determining the parameters required by these representations for particular applications.

6. Theoretical Studies: Meeting the Functional Requirements

6.1. Profile and Plot Scales

[39] In 1982, we observed that there had been no significant advances in modeling hillslope flows that involved macropores (although reference was made to the potential for using kinematic wave solutions for saturated downslope flows for cases where the effects of macropores and preferential flows could be included within a storage-flux relationship as in Beven [1981b]). There have now been a variety of attempts to incorporate preferential flows into the Darcy-Richards framework, particularly at the profile and plot scales. Gerke [2006] provides an earlier review of modeling approaches [see also Simůnek et al., 2003; Köhne et al., 2009a, 2009b].

[40] There are four main approaches to modeling preferential flows in current use. These are:

[41] 1. the single continuum approach in which preferential flows are treated simply as a modified relative conductivity curve close to saturation;

[42] 2. a dual-continuum approaches in which one domain is treated as immobile and one is a Darcy-Richards domain, with some exchange term between the two;

[43] 3. a dual–permeability approach in which preferential flows are represented by high permeability inclusions into a Darcy-Richards domain (as represented by DUAL of Gerke and van Genuchten [1993]) or in fractured permeable rocks where the matrix is represented as blocks exchanging water with Darcy-Richards flow in the fractures [e.g., Doughty, 1999; van der Hoven et al., 2002; Ireson and Butler, 2011]; and

[44] 4. a dual-porosity approach in which simple volume filling or a kinematic equation is used for the preferential flows and exchanges with an immobile or Darcy-Richards matrix are treated separately (as in the kinematic model of Beven and Germann [1981]; the PREFLO model of Workman and Skaggs [1990]; the MACRO model for vertical preferential flow of Jarvis et al. [1991, 1994, 1997]; Larsson and Jarvis [1999]; Larsson et al. [1999]; Acutis et al. [2001]; and the SWAP model of van Dam et al. [2008]).

[45] Models (1) to (3) are all options in the latest versions of the HYDRUS code of Simůnek and van Genuchten [2008]. A somewhat different dual porosity model (IN’M) has been introduced by Weiler [2005, see also Bachmair et
al., 2010] in which Green-Ampt wetting models are used for both domains and for the infiltration of water from cylindrical macropores into the matrix, but where a concept of surface or subsurface contributing area for individual macropores is introduced, building on the conceptual framework of Beven and Clarke [1986] and the RZWQM model of Ahuja et al. [2000]. The SWAP model of van Dam et al. [2008] allows for distributions of the occurrence and depths of vertical cylindrical macropores in a soil volume. Beven and Clarke [1986] also allowed for spatial and depth distributions within an analytical solution framework.

While all these approaches aim to satisfy some components of the functional specification defined above, they are all deficient in important respects, in particular the common dependence on Richards equation and its associated assumptions of equilibrium gradients and homogeneous soil properties (see earlier arguments). It is also the case that they cannot always reproduce observations of infiltration into macroporous soils. In a comparison of IN’M and MACRO, Bachmair et al. [2010] showed some success in reproducing plot scale wetting in grassland sites but not for tilled, untilled or forest sites. Gerke et al. [2007] showed some success in reproducing 2-D flow to drains but suggested that there were still considerable challenges in applying dual-permeability models to predict preferential flows.

So if Richards equation is not an adequate basis for representing the effects of preferential flows when wetting heterogeneous soils, is there another framework that might be useful? It is worth noting that Richards [1931] made no claim that the equation he presented was a complete description of what might happen under other nonequilibrium conditions (although he did discuss the issue of hysteresis in soil moisture characteristics at the end of his paper implying a recognition of dynamic effects on fluxes). The issues in developing an adequate theory to representing water flow in soils with preferential flow are those of the three functional requirements set out above, in this case for flows dominated by gravity and viscosity rather than capillarity effects. If, as the evidence suggests, preferential flows involve fluxes that are not strongly related to capillarity, then viscosity should have an important role to play.

There are alternatives to continuum approaches based on Richards equation. Pore-based numerical approaches have also included particular solutions of the Navier-Stokes equations, to 2- and 3-D images of pore arrangements [e.g., Heijs and Lee, 1995]; the application of Lattice-Gas simulations [Di Pietro et al., 1994; Di Pietro, 1996] to simulated structures of assumed pore dimensions; and the use of particle tracking of water parcels at the profile scale has been presented by Ewen [1996] and Beven et al. [1989]. The former of these methods are limited by computational constraints to small scales; the latter seem promising but have been subject to only limited testing.

Germann and Di Pietro [1999] have proposed a viscosity-based Stokes law approach to describing laminar preferential flow as films of a certain thickness (see the summary in Appendix A). The theory allows wetting front velocity to be described as a function of input rate. The theory is consistent with the linear progression of wetting fronts that has been observed in a variety of soils (e.g., Figures 2 and 3, Table 1). It is suggested that it applies, at least over a certain range of pore sizes, during fast preferential flows. Note that, unlike the Richards equation, there is no need to invoke a “representative elementary volume” for the averaging of potentials. In Stokes flow, the only two local requirements are that the gravity potential is balanced by the dissipation of momentum due to viscosity at the scale of the film thickness and that the definitions of mobile water content in the films, contact length, film thickness, and volume flux density apply to the same cross-sectional area. However, it also matches a variety of experimental evidence during infiltration at the profile scale up to the limits of laminar flows [Hincapié and Germann, 2009b]. Film thickness will be sensitive to spatial and temporal changes in input rates, but since flux rates are proportional to the cube of film thickness such adjustments should be rapid regardless of soil heterogeneity.

One interesting feature that has come out the analysis of such wetting front experiments is that the wetted length of films, \( L \), does not change dramatically with flow rate (see equation (A9) of Appendix A). Hincapié and Germann [2009b] determined \( v \) at the depth of 0.28 m in a column of an undisturbed forest soil as response to four input rates of \( q = 5, 10, 20, 40 \text{ mm h}^{-1} \) (i.e., 1.4, 2.8, 5.6, 11.1 \( \times 10^{-6} \text{ m s}^{-1} \)). The observed wetting front velocities (averages from 7 experimental runs each) amounted to 7.3, 9.7, 12.0, and 25 \( \times 10^{-3} \text{ m s}^{-1} \). Figure 4 illustrates the results and shows that assumption of a constant value of \( L \) is reasonable in this case. The calculation of \( L \) for the four data points produced an average value of 181 m m\(^{-2}\). The velocities from these experiments lie in the upper 10% of the frequency distribution of \( v \) of Hincapié and Germann [2009a] while \( L \) is the shortest found so far. Both values of \( L \) and \( v \) hint at flow along well established preferred flow paths. However, neither \( L \) nor the variation of the corresponding film thicknesses of free-surface flow in the

![Figure 4. Wetting front velocity, \( v \), as function of the input rate, \( q \), according to equation (A9). (Data from Hincapié and Germann [2009b]; figure from Germann [2013], with permission.)](image-url)
approximate range of about 15 ≤ F ≤ 30 μm from equation (A1), correspond with the typical perception of preferential flow along well-visible macropores. Many more cases have been studied across a wide range of soil types and suggest that the constant L assumption is worth testing experimentally as a hypothesis within the range of viscosity dominated flows [see for example Germann and Bürgi, 1996; Mdaghri-Alaloui et al., 1997; Germann and Niggli, 1998; Germann et al., 2002b, 2007; see also Table 1]. It would, of course, be useful for such tests to be carried out with conservative tracers so that predictions of both celerities and velocities could be checked, at least at local scales. Even locally, however, it is likely that the heterogeneity of local flux rates and displacement of stored water might make hypothesis testing difficult, but future experiments should hopefully give a guide to the process representations necessary to be consistent with such joint observations.

Stokes law will not of course apply where film thicknesses are greater than can be supported by laminar flow at high-flux rates in the macropore size range. Lin and Wang [1986] suggest that the limit for a laminar film flow occurs at a Reynolds number of Re = F v/ν = 3, implying a maximum film thickness of 100 < F max/150 μm and an associated range of netting front velocities of 0.03 < v max < 0.08 m s⁻¹ in vertical flow [Germann et al., 2007]. Flow with Re > 3 requires faster velocities and, therefore, thicker films that Ghezzehei and Or [2005] suggest will break up into drops and bridges. Macropore flows with much higher velocities and Reynolds numbers have been observed experimentally [e.g., Logsdon, 1995; Anderson et al., 2009b], while flow in larger pipes might be expected to be fully turbulent [Chappell, 2010; Jones, 2010]. Additional flux relationships will then be required, but it is clear that a Richards equation based approach will not be appropriate. The Manning equation has been used to provide such a flux relationship [e.g., Logsdon, 1995; Sidle et al., 1995] but might also not be appropriate (Manning himself favored another form, and that was for the case of fully developed open channel flows at steady discharges and near hydrostatic pressures). There are also relationships derived from experiments on engineered pipes that might be a better approximation with the use of an effective roughness coefficient.

6.2. Field, Hillslope and Catchment Scales

Again, as at the plot and profile scales, there are models that are based on the Richards equation that have been modified to incorporate representations of preferential flow at the field and hillslope scales. These include the equilibrium, dual-component and dual-permeability versions of HYDRUS2D/3D of Šimůnek and van Genuchten [2008], the CATFLOW model of Klaus and Zehe [2010, 2011]; the 2-D model of Nieber and Warner [1991], and Nieber and Sidle [2010]; the 2-D QSOIL model of Faeh et al. [1997]; the DFSDM model for lateral preferential flows of Mutungu et al. [2005]; and the recent model of Dusek et al. [2012] that combines a dual-permeability approach for vertical flow with a diffusion equation for downslope flow. Beckers and Allia [2004] added a Green-Ampt preferential flow model to the DVSHM catchment model. These have similar limitations to the 1-D profile and plot scale models, with the additional difficulty of parameterizing the characteristics of the preferential flow domain at the hillslope scale. There have also been models developed specifically for pipeflows including the model of Jones and Connelly [2002], and the HillVi model of Weiler and McDonnell [2007] and Tromp-van Meerveld and Weiler [2008].

It is also interesting to speculate the way in which the description of vertical fluxes based on Stokes flow could be modified to represent lateral flow of water in a structured soil (see equations (A9)–(A12) of Appendix A). Note that this description does not require that the soil saturation rate from its base to provide lateral flow. This is a common assumption in models of hillslope hydrology, but this Stokes flow description also admits the possibility of fingering in the downslope direction to maintain the flux q. Certainly, Anderson et al. [2010] suggest that saturation, as observed by piezometric responses is not a sufficient guide to the occurrence of downslope preferential flows, although Jenson and McGlynn [2011] have had some success in using piezometers at the bottom of a hillslope as a basis for connectivity of subsurface responses to the stream. For a Stokes flow, the only assumption is essentially that the effective film thickness adjusts to satisfy the downslope flux, given momentum loss over a surface L per unit volume of soil. In fact, although hillslope kinematic wave models [e.g., Beven, 1981b, 1982a, 1982b; Davies and Beven, 2012] are often represented as if they involve saturation from the base of the profile, such an assumption is not necessary. They require only a functional relationship between storage of mobile water and flux (which can be provided by a Stokes flow representation, but which might also be extended to transitional and turbulent flows at higher flux rates).

Consider then the case of a steady vertical recharge rate per unit area, qlat, over a contributing area A. upstream of the point. Then the downslope, viscosity controlled, volume flux density at the point will be:

\[ q_{lat} = q_{v}A_c. \]

For laminar Stokes flow, values of film thickness, velocity and celerity and an effective value of the momentum dissipation surface per unit volume of soil (L) can then be derived from equations (A9)–(A12) of Appendix A. Figure 5 shows the resulting relationship between the required film thickness Flat and mobile water content for the case of a number of different downslope flux densities (reflecting changes in either upslope contributing area or vertical recharge rate. These relationships can also be used to show how the required film thickness would change under different slope angles (Figure 6). What is interesting about these plots is that, for reasonable slopes > 0.05 or so, the film thicknesses suggested by these calculations are such as to allow downslope preferential flows, induced by the structure of the flow pathways, without complete saturation for a range of vertical flux rates and upslope contributing areas. In contrast, it is worth noting that Battle and MacDonald [2002] argue that macropores are only important in supplying vertical flux to the base of the profile. At their site in Ontario downslope fluxes take place in a high-conductivity layer at the base of the profile and as a saturated matrix flow in the overlying soil.

That argument does not imply that saturated flow in soil macropores and rock fissures is not important in hillslope scale water fluxes, only that saturation is not
necessarily required to support significant downslope flow as films in a structured soil. Saturated flow might be very important in many situations, even if only locally at the base of the soil and in rock fractures. In particular, saturated flow in fractures resulting in excess pressure was thought to be important in the Coos Bay slope failure [Montgomery et al., 2002] and might be a common feature of many landslides and debris flows [Uchida et al., 2001; Hencher, 2010] including peat failures [Dykes and Warburton, 2007]. Because of very low effective storage coefficients, the celerities controlling subsurface output responses can be fast, and far in excess of pore water flow velocities controlling transport. Event responses will then be predominantly made up of pre-event water.

7. Defining Model Parameters for Application Scales

Regardless of the representation of flow processes used, one of the difficulties in specifying any conceptualization of preferential and matrix flows is in knowing what the appropriate soil characteristics are at any site of interest; the third functional requirement listed above. That is why, in applications of models based on the Richards equation, pedotransfer functions have been so popular in defining soil hydraulic characteristics given more readily available texture and other information. But pedotransfer functions have been based on measurements on small samples of soil, and so cannot reflect the effects of larger scale heterogeneities, structure and macroporosity (both measurements and pedotransfer function estimates are also associated with large uncertainty that is often neglected in applications, e.g., Sherlock et al. [2000]). Identification of parameters of dual-permeability models from fitting field or column observations is challenging [e.g., Arora et al., 2011, 2012].

Thus if new, more realistic, element scale, constitutive relationships are to be defined, there will still be the question of how the parameters of those relationships might be determined either from measurements, or a priori based on experience at sites where they have been determined from measurements. This might be a matter of trying to use local scale observations to inform larger scale model parameters. In the study of van Schaik et al. [2010], local destructive tracer measurements have been used to inform the parameterization of the SWAP model applied at larger scales. Weiler and McDonnell [2007] also based their parameterization of the Hill-Vi model on destructive sampling of pipe characteristics at the Maimai catchment. Some reinterpretation of existing data might be required, such as in the inference of Stokes flow from TDR measurements of water content waves described in section A3 of Appendix A.

However, in general, this is not at all a simple question. In fact, Beven [2006, 2012, chap. 9] suggests that it is the fundamental limitation on the development of hydrological science. This is because in order to fully characterize the response of a plot, a field, a hillslope or a catchment, both data types, hydrograph and tracer concentrations are required to allow the differentiation of celerity effects that control the hydrograph response and the distributions of pore water velocities that control the tracer response. At small plot scales it is easier to control tracer applications, but more difficult to collect the discharge without significant disturbance. At small catchment scales it can be somewhat easier to measure discharge in a channel (with significant uncertainty about whether the catchment is really watertight), but more difficult to use artificial tracers, while inferences from environmental tracers have their own difficulties of interpretation and measurement.

Those difficulties are being reduced by a new generation of isotope measurement devices that allow higher frequency data to be collected [see for example Berman et al. 2009], but the interpretation of the data will remain subject to the vagaries of uncontrolled inputs with concentrations that vary in space and time. However, the joint use of such data will, even allowing for the uncertainties associated with such measurements, exert constraints on the potential model parameterizations that might be consistent with those data within a hypothesis testing framework [e.g.,

Figure 5. Effect of upslope flux rate per unit width (m² s⁻¹) and upslope contributing area $A_c$ on film thickness given a slope angle of 0.05 and steady vertical recharge rate of 1 mm h⁻¹.

Figure 6. Effect of slope angle on film thickness and upslope mobile water content given a steady vertical recharge rate of 1 mm h⁻¹ and upslope flux rate per unit width of 0.001 m² s⁻¹ (equivalent to the same steady recharge rate applied over an upslope contributing area $A_c = 3600$ m²).
8. Self-Organization and Constructal Networks

[64] A number of authors have suggested that water flow pathways in a soil system within a hillslope might be self-organizing into a network structure analogous to the evolution of a dendritic drainage network of surface channels. The surface channel network has been the subject of research on self-organization for many years [e.g., Rodríguez-Iturbe and Rinaldo, 1997; Bejan, 2000, 2007; Bejan and Lorente, 2010]; the soil system much more recently [Sidle et al., 2001; Lorente and Bejan, 2006; Hunt, 2005, 2009]. The concept behind such a theory is that flows will develop a drainage system that is both space filling and efficient. Where there is evolutionary growth of a system (as in surface channel networks, root systems, lungs, blood circulation, etc.) then the connectivity and size distributions of flow pathways within the system will organize such that there will be a form of fractal space filling. The fractal dimensions of pore networks in permeable media have been reported, but only for rather small samples [e.g., Brakensiek et al., 1992; Hatano and Booltink, 1992; Preston et al., 1993; Baveye et al., 1998; Perret et al., 1999, 2003]. This type of development may, however, be the case for the development of piping systems by erosion in dispersive soils found in some parts of the world, with an interaction between fluxes, soil properties and the development of pipe networks [e.g., Bryan and Jones, 1997; Gutierrez et al., 1997; Faulkner et al., 2004; Wilson, 2011]. In these cases, the flow is implicated in the development of the network; in the terminology of Bejan [2000] the pattern is constructal. While fractal descriptions are a way of describing scaling phenomena, there are two problems in applying them usefully. One is that large amounts of data are required to estimate a fractal dimension precisely, but small changes in the fractal dimension can have significant effects on the resulting flow properties. The second is that the scaling may not hold over all scales.

[65] There is clearly an attraction, in considering macropores and preferential flow pathways in soils, that such a scaling concept might provide a way of theorizing about the nature of such pathways because they are so difficult to observe in the field [Sidle et al., 2001; McDonnell et al., 2007]. The concept is that the patterns of preferential flow will have coevolved with the soil, vegetation cover and biota of a slope under the constraint of the history of meteorological conditions above and the geology below [Zehe et al., 2010; Reinhardt et al., 2010]. However, while connectivity is increasingly recognized as an important control on hillslope responses [e.g., Uchida et al., 2005a, 2005b; Bracken and Croke, 2007; Lehmann et al., 2007; Anderson et al., 2009b; Michaelides and Chappell, 2009], there does appear to be a problem with this type of theorizing. This is that many of the causes of macroporosity in the soil are quite independent of any preferential flow process. Causes such as cracking due to drought, tree death from disease slowly creating root channel macropores, burrowing by soil fauna, are related to but do not depend strongly on water fluxes. For example, many worm species and plant roots actively avoid saturated soils. In the same way, the causes of loss of connectivity in the system by blocking and collapse of macropore flow pathways may also be independent of the water flows. Thus the potential for self-organization will be constrained by boundary conditions exogenous to the flow itself. It can therefore be argued that the patterns are, to a large extent, constructed rather than constructal.

[66] This does not preclude the possibility that within all the discontinuous macropore pathways in a volume of soil, as created by a whole variety of processes, efficient preferential flow pathways may develop into a particular space filling and efficient flow network, so that some flow pathways will consistently take more flow than others. Certainly, tracing experiments in which dye is used to follow the actual flow pathways of water within soil show that there are macropore pathways that have been used by the flowing water and others that have not, at both profile scale and on the scale of pipe networks developed on hillslopes. Whether this is a consistent pattern (as it would need to be for a constructal theory to be considered useful) or whether these are rather random, path of least resistance, choices for the organization of water films that depend on the particular boundary conditions of an experiment or natural event remains unclear. Certainly, it seems to be too simplistic to suggest that these networks are purely self-organizing in any sense other than following paths of least resistance (more akin to the percolation theory approach of Hunt [2005, 2009], see also Nieber et al. [2006]; Nieber and Sidle [2010]).

9. Advances in Measurement Techniques

[66] One way of learning about macropore and preferential flows in soils is to find some method of visualizing the flow. In the past, the main methods of visualization have been through the use of tracers, dyes and moulds (see Figure 1 and the review of Allaire et al. [2009]). These all have their limitations [e.g., Wienhöfer et al., 2009]. Conservative tracers give some direct information about Lagrangian velocities (the integral velocity of a tracer particle from input to output) but not directly about the pathways or local pore water velocities involved. Environmental tracers are subject to spatial and temporal variations that might not be detected by other than high frequency sampling [Berman et al., 2009; Herbstreit et al., 2012] and which will lead to difficulties in interpretation. Dyes require destructive sampling and may not reveal all pathways of the moving water. Moulds require destructive sampling and, because of viscosity effects, may not sample all the potential water pathways.

[67] Thus, the advent of 2-D and 3-D computer aided tomography of different types has been a real advance in allowing the structure of the flow within a sample to be studied. Various techniques of tomographic imaging have been applied to the study of preferential flow in soil including X-ray (CT), electrical resistance (ERT), radar and ultrasound. X-ray imaging allows very fine structure to be determined [e.g., Pierret et al., 2002], but on generally rather small samples. Shichota et al. [2010] have investigated preferential flows in ‘‘undisturbed’’ samples demonstrating using CT.
The other techniques have been applied to larger cores and in the field. An early ERT study by Binley et al. [1996] on an undisturbed soil core showed how localized dye patterns on sections taken after infiltration were broadly matched by nondestructive tomographic images. Later, Koestel et al. [2009] showed that Brilliant Blue dye could be detected directly and nondestructively using 3-D ERT. Both electrical resistance and radar techniques have been used in the field with either sampling from the surface or between boreholes demonstrating rapid and localized water movements [e.g., French and Binley, 2004; Amidu and Dunbar, 2007; Cuthbert et al., 2009; Garré et al., 2010]. Slater et al. [1997] demonstrated rapid recharge to unconfined groundwater using cross-borehole ERT.

There is, however, a trade-off between the resolution of the images and the scale of application that results from the time needed to record all the signals required for the tomographic inversion, particularly under dynamic conditions. Furthermore, the data acquisition for tomographic methods, particularly in 3-D, can add further constraints to the spatial resolution when studying dynamic processes. The inverse methods for these techniques often require regularization in order to overcome the problems of instability and nonuniqueness in the solution; however, these can often limit the sensitivity to highly localized changes in geophysical properties (e.g., due to over-smoothing). This implies a degree of uncertainty associated with the resulting images and few attempts have provided supporting evidence of inferred flow pathways. Added to this, one must consider the impact of changes in the hydrological state (water content, solute content) on the imaged geophysical property (resistivity, permittivity, etc.). Under preferential flow the net change in hydrological state over the imaged pixel/voxel volume may be extremely small (e.g., 100% change in solute content but in only 1% of the volume) and perhaps beyond the level of detectability according to the sensitivity of the petrophysical relationship. Since the inverse methods applied can be highly susceptible to overfitting without adequate noise assessment then one can risk misinterpretation of preferential flow when the true cause is inappropriate assessment of signal to noise. Binley et al. [1995] illustrate how the tomographic inversions can be sensitive to data noise.

Tomographic techniques have been primarily applied at the core, profile and plot scale for good logistical reasons. At the hillslope scale, there has not been much novelty in instrumentation with the spatial and temporal resolutions required to observe preferential flows. Techniques of trenching to detect subsurface stormflows and the installation of recording piezometers have been used much more intensively in some experimental sites. There have been some directly measured macropore fluxes in trenched slopes (e.g., the root channels at Panola) but tracing and excavation experiments have revealed that preferential flows can also be more diffuse (as has been argued above). Thus, at these larger scales it is necessary to rely on high frequency flow, soil moisture and artificial or environmental tracer data to infer the importance of preferential flows (as in the hypothesis testing of Davies et al. [2011]). But inference of process from bulk responses, of course, is uncertain and future experiments will need to consider how best to constrain that uncertainty.

## 10. Prospects

In looking forward to future research needs in 1982 we suggested that more experimental data were required to support the integration of preferential flow concepts into hydrological models. We also suggested that the difficulty of doing so would not be one of theoretical development but of obtaining the necessary experimental data at the field scales of interest in real applications for predicting infiltration rates, subsurface stormflows and transport of nonpoint source pollutants. To some extent these suggestions have been borne out by events but what we did not anticipate was the persistent dominance of the Richards equation approach as a result of easy-to-use software packages being made widely available. The Richards equation has even dominated representation of soil processes in land-surface parameterizations at the grid scale of global circulation models. It is really difficult to see how such a misuse of “physics” is justified.

While it is true that many models based on soil physics now allow for some representation of preferential flow as a dual continuum, dual porosity or dual permeability options, these have generally been “bolted on” to the Richards solution without any rigorous physical underpinning and despite the fact that the physics suggests that the Richards approach itself is not an adequate representation of flow processes in a heterogeneous unsaturated soil matrix. It is perhaps still reasonable to suggest that this is because of the limited observational techniques available for studying flow departures from Richards equilibrium theory. It is still, however, the case that nonequilibrium capillary and preferential flows are important in heterogeneous field soils: an extreme, but to us rather attractive, view would be to suggest that much of soil physics (at least during significant infiltration) is predicated on the wrong experimental technique as used by Richards in 1931.

A reconsideration of the representation of preferential flows as a Stokes flow provides a new impetus to addressing the problem (albeit only for preferential flows within which the flow remains laminar – there has been even less study of transitional, droplet and turbulent flows in the largest pathways). Past experimental work has shown how the nonequilibrium propagation of wetting fronts can be represented in this way, at least at the soil profile scale. At the hillslope scale, there have been many studies of preferential flows in visually obvious macropores, but the possibility of downslope Stokes film flows induced by sloping soil structures remains speculative. At both profile and hillslope scales, there is also a need to combine flow measurements with tracer experiments to test whether assumptions about flow velocities can equally reproduce the celerities controlling the hydrograph. One such modeling study, using a particle tracking model for both flow and transport, led to some interesting hypothesis testing at the Gårdsjön site without the need to invoke Richards capillary gradients for this highly heterogeneous domain [see Davies et al., 2011]. This Multiple Interacting Pathways (MIPs) model might be one way of introducing different process representations into profile, hillslope and catchment applications in future.

So we can conclude that there has been some progress in understanding and methodologies over the last
30 years. However, despite the increasing number of papers published on the topic, there has not been as much attention paid to macroscopic and preferential flows as we might have expected, given its significance in all areas of soil and catchment hydrology, water quality, slope stability, and agricultural management. A fully convincing integrated physical theory still has not yet been achieved at practical scales of interest (given the practical impossibility of integrating or directly simulating the Navier-Stokes equations for other than very restricted hypothetical networks of pores). Such a theory would need to deal with the transition between capillary and viscous flow regimes, between viscous and nonlaminar flow regimes, and with the expected effects of scale on process representations and effective parameter values (see the discussion of Beven [2006]). Again, however, we suggest that this really should not be a surprise. Soil physics remains measurement technique limited, particularly in the field. Even the dramatic impact of computer aided tomography on the visualization of patterns of water movement has been limited in either scale or resolution in contributing to the development of an integrated field scale theory, while all tracer experiments remain somewhat difficult to interpret, even at core or profile scales.

[73] Three decades on, the fundamental questions posed by BG82 have not been resolved, which then poses the question as to how they might be resolved. We have suggested that a multiprocess concurrent representation combining both capillary effects in heterogeneous soil, a Stokes flow representation of preferential flows and some extension to higher nonlaminar flux rates is worth investigating. The assumptions that underlie the Stokes flow representation of preferential flows, and its consequences in terms of constant wetting front velocities and linked drying front celerities, require only micrometer-scale equilibrium assumptions and provide a number of testable hypotheses. The necessary assumptions will not hold under all circumstances as outlined earlier but can be tested for consistency with experimental data for suitable ranges of flux conditions. There is more hypothesis testing to be done in looking at the range of conditions over which a Stokes flow representation of preferential flow might be useful.

Appendix A: Preferential Flow as a Viscosity Dominated Stokes Flow and Kinematic Wave


[74] Stokes flow provides a framework for representing the effects of preferential flow within which gravity provides the driving force that is just balanced by the effects of viscosity in dissipating momentum within water films moving between a solid-water interface with a no-slip boundary condition and an air-water interface with no friction boundary condition, with capillarity being assumed to play a negligible role. Germann and Di Pietro [1999] present the following basic expressions derived from the force balance and Newton’s shear hypothesis, for the mobile water content \( w \) (m³ m⁻³), the wetting shock-front velocity \( v \) (m s⁻¹), the volume flux density of mobile water \( q \) (m s⁻¹), and the celerity \( c \) (m s⁻¹) of mobile water as:

\[
w(F, L) = F \cdot L, \quad (A1)
\]

\[
v(F) = \frac{zW(t)}{t - T_B} = \frac{q(F, L)}{w(F, L)} = \frac{g}{3\eta} \cdot F^2. \quad (A2)
\]

\[
q(F, L) = v(F) \cdot w(F, L) = \frac{g}{3\eta} \cdot L \cdot F^3. \quad (A3)
\]

\[
c = \frac{dq}{dw} = \frac{g}{3\eta} \cdot F^2 = 3 \cdot v(F), \quad (A4)
\]

where \( F \) (m) and \( L \) (m m⁻²) are the thickness of the moving water film and its effective contact length per unit cross-sectional area (or vertical area of momentum dissipation per unit volume) of all sessile parts of the three-phase system, \( zW(t) \) is the time-dependent depth of the moving wetting shock-front, \( T_B \) indicates the beginning of water input to the soil surface, \( t \) (s) is time, \( g \) is acceleration due to gravity, and \( \eta \) (= 10⁻⁶ m² s⁻¹) is the kinematic viscosity of water. \( F \) and \( L \) are considered the result of the water’s spontaneous reaction to the input rate \( q_S \) (m s⁻¹) and the internal geometry of the permeable medium. Note that in this framework, velocity and celerity are functions only of film thickness and consequently require only a very local micrometer-scale equilibrium assumption. A constant flux should then lead to a constant film thickness and a constant wetting front velocity, consistent with evidence presented in the main text.

[75] The end of input at \( T_E \) releases a draining front that moves with celerity \( c \). After its arrival at \( T_f(Z) = T_E + Z/L \) a trailing wave forms according to

\[
w(Z, t) = F \cdot L \cdot \left( \frac{Z}{c \cdot (t - T_E)} \right)^{1/2}. \quad (A5)
\]

[76] Because from equations (A2) and (A4) \( c = 3 \cdot v \), the wetting front intercepts the draining at time \( T_{wc} \) and depth \( Z_t \) given by

\[
T_t = \frac{3 \cdot T_E - T_B}{2}, \quad (A6)
\]

\[
Z_t = F^2 \frac{g}{3\eta} \cdot \frac{3 \cdot T_E - T_B}{2}. \quad (A7)
\]

[77] After \( T_f \) and beyond \( Z_t \) the wetting front will move with a celerity proportional to the cubic root of time [Germann, 1985]. When input rates vary, there will be multiple wetting and drying waves moving through the profile, but this should follow the flux, velocity, celerity relationships of these equations.

[78] Equations (A1)–(A7) assume free-surface flow of a planar water film. Other geometries result in comparable relationships. Germann et al. [2007] assessed the impact of presumed flow geometries on the relevant flow parameters. They found a ratio of 1:1.22 for the contact lengths per cross-sectional area of free-surface flow compared with the circumference of flow in the corresponding size and numbers of Hagen-Poiseuille cylinders. Likewise, a ratio of 1:1.63 resulted when comparing the film thickness of free-surface flow with the pore radii of equivalent cylindrical pores. Thus, for a first-order representation of preferential flow, considering the correct geometry of the conduits seems
of minor importance vis-à-vis all other uncertainties and particularly in view of the intriguing variety of flow geometries demonstrated in numerous 3-D CAT scans as shown, for example, in Luo et al. [2010] and Sichota et al. [2010].

[80] Hincapié and Germann [2009a] referred to time series of water contents during constant flux infiltration as a water-content wave, WCW. Figure A1 illustrates its typical features of a steep increase of mobile water content between the arrival of the wetting front at time $T_B(Z)$, while the amplitude $w_S (m^3 m^{-3})$ of the WCW is set equal to the mobile water content. The parameters $F$ and $L$ follow from $v$, equation (A2), and $w_S$ in equation (A1). Neither an antecedent steady state nor a homogeneous distribution of any property is required, thus leaving to the infiltrating water the freedom to find its own pathways within the viscosity constraints of film thickness provided by equations (A1)–(A4). Equation (A4) also leads to

$$T_D(Z) = T_E + \frac{T_W(Z) - T_B}{3}. \quad (A8)$$

[80] The WCW-approach can be directly related to a stable Stokes flow representation. The limit for such a flow occurs at a Reynolds number of $Re = F / \eta = 3$ [Lin and Wang, 1986], implying a maximum film thickness of $100 < F_{\text{max}} < 150 \mu m$ and an associated range of wetting front velocities of $0.03 < v_{\text{max}} < 0.08 \text{ m s}^{-1}$ in vertical flow [Germann et al., 2007]. This range of $F_{\text{max}}$ suggests that there should still be some effect of capillary forces. Madgahi-Alaoui et al. [1997] have confirmed this with a series of in situ infiltrations into a clay soil with gradually increasing antecedent moisture from one experimental run to the next, thus gradually reducing the effect of capillarity on preferential flow. Water abstraction from the WCW into the surrounding smaller yet not completely water-saturated pores led Germann and Beven [1985] and Di Pietro et al. [2003] to add a sink term to the moving WCW. From observations of WCW propagation, Hincapié and Germann [2009a] estimated that such abstraction could apply over depths of between 0.2 and about 10 m, i.e., typical depths of soil profiles and vadose zones.

[81] Germann et al. [1997] assessed the lower limit for the dominance of viscous over capillary flow based on the notion that diffusivity $D$ represents the dissipation of the capillary potential gradient in the Richards [1931] equation, where $D(\theta$ or $h) = K(\theta$ or $h) / C(\theta$ or $h)$. $C = -d\theta / dh$ (m$^2$), and $\theta$ (m$^2$) and $h$ (m) are the volumetric water content and the capillary head. The assumption that $D$ represents dissipation of any flow-driving force, including the shear force that originates from momentum dissipation expressed with kinematic viscosity $\eta$, leads to exclusively capillary flow when $D < \eta$. From modeling $D(h_j) = \eta (= 10^{-6} \text{ m$^2$ s}^{-1})$ to represent different soil textures Germann et al. [1997] found at the threshold of $D = \eta$ a range of capillary heads of $-7 \leq h_j \leq -2 \text{ m}$. From the Laplace-Young equation it follows that this is equivalent to pore widths of between 4 and 15 $\mu m$ which includes the thinnest water films of 6 $\mu m$ showing constant wetting front velocities as reported by Hincapié and Germann [2009a].

[82] Preferential flow according to equations (A1)–(A8) offers a variety of relationships worth experimental testing. Evidence for observed constant wetting front velocities, equation (A2), within the time and depth ranges of $[T_B(t) \leq t \leq Z]$ and $[0 \leq z \leq Z]$ was already presented in the main text (Table 1). The advancement of wetting fronts at $z(t)$ at a rate proportional to $t^{1/3}$ beyond $Z_l$ hints at this depth's significance as parameter for scaling flow processes. The trailing wave according to equation (A4) models the data well as depicted in Figure A1. It also demonstrates the plausibility of the ratio of $c/v = 3$ from equations (A2) and (A4) that led to equation (A7).

A2. Derivation of Stokes Flow Characteristics

[83] Given information on the progression of wetting fronts into the soil profile, the characteristics of film thickness $F$ and the specific contact length $L$ that expresses the solid-water interface per unit volume of the permeable medium onto which momentum dissipates, can be calculated using equations (A1)–(A8) above. Time-domain reflectometry is well suited to investigate preferential flow at the soil profile scale as the numerous references demonstrate. Best results are achieved with horizontally installed wave guides that are at least 0.2 m long, and the system should allow for data recording at close to 1 Hz. A depth-sequence of TDR waveguides permits water content waves to be followed in the vadose zone [al Hagrey et al., 1999]. The ground penetrating radar technique, GPR, allows the repeated recording of wetting front depths [al Hagrey et al., 1999; Hauerer et al., 2011].

![Figure A1. Water-content wave, WCW, recorded with TDR-equipment at the depth of $Z = 0.1 \text{ m}$ due to sprinkler infiltration of $q_S = 1.25 \times 10^{-5} \text{ m s}^{-1}$ (= 45 mm h$^{-1}$) during the period $(T_B-T_B) = 3600 \text{ s}$. Data recording was with 6 s intervals. The gray line shows the measurements while the black line indicates the approach according to equations (A1)–(A7). The points in time indicate $T_B$ beginning of sprinkling, $T_W$ arrival of the wetting front at $Z$, $T_D$ end of sprinkling, and $T_D$ the arrival of the draining front, while $w_S$ is the amplitude of the wave of mobile water, WCW. The concave decrease of the water content after $T_B$ is referred to as trailing wave.](image-url)
A3. Stokes Flow as a Representation of Downslope Water Flows

[84] As described in the main text, Stokes flow might also be used to describe a lateral flux on a slope of angle $\alpha$. In this case the downslope volume flux density is given by

$$q_{\text{lat}} = \frac{g \cdot \sin(\alpha)}{3 \cdot \eta} \cdot L \cdot F_{\text{lat}}^3$$

(A9)

with water content

$$w_{\text{lat}} = LF_{\text{lat}} =$$

(A10)

lateral velocity

$$v_{\text{lat}} = \frac{q_{\text{lat}}(F, L)}{w_{\text{lat}}(F, L)} = \frac{g}{3 \cdot \eta} \cdot \sin(\alpha) \cdot F_{\text{lat}}^2$$

and celerity

$$c_{\text{lat}} = \frac{g}{3 \cdot \eta} \cdot \sin(\alpha) \cdot F_{\text{lat}}^2$$

(A12)

equations (A8)–(A11) are used to provide the results shown in Figures 5 and 6 in the main text.

A4. Stokes Flow as a Kinematic Wave

[85] Beven and Germann [1981], while numerically exploring the special cases of laminar Hagen-Poiseuille flow in cylindrical pores and plane-Poiseuille flow along walls of planar cracks, realized that Stokes flow is equivalent to a form of kinematic wave theory [Lighthill and Whitham, 1955] that provides a set of mathematical expressions for generalizing infiltration, including preferential flow pathways. It turned out that DeQuervain [1972] and Colbeck [1972] had already applied kinematic wave theory to flow in isothermal snow, while Sisson et al. [1980] were probably the first to apply it to infiltration. Smith [1983] and Charbe- neau [1984] applied it as an approximation to the Richards [1931] equation for infiltration and solute transport, respectively. Beven [1981b, 1982a, 1982b] used the kinematic wave approach for both vertical infiltration and downslope flows [see also Williams et al., 2002]. Germann [1985], following Petrascheck [1973] and Smith [1983], analyzed the tailing of the kinematic wave. More recently Jarvis et al. [1997] included kinematic wave theory in their MACRO model, while the particle tracking approach of Davies et al. [2011] is also essentially kinematic. Nimmo [2010] presented an approach that seems partially based on Stokes flow and kinematic wave theory but not all the relationships and assumptions withstand rigorous evaluation [Germann, 2010]. A variation on the approach, interpreted as a soil water balance model, has been used in groundwater recharge estimation by Cuthbert et al. [2013].

[86] Note that in downslope flows, the theoretical ratio of $c/v = 3$ in equations (A8) and (A9) may expand to $c/v \gg 3$ [Rasmussen, 2001] when a water content wave gets dissipated, for instance, due to water abstraction into pools that react much slower than an unrestricted wave of mobile water content would propagate. Such pools may consist of groundwater bodies, perched water tables or soil matrices that need to be substantially supplied before a wave may proceed further downslope.

[87] At the surface a one input pulse is defined by the volume flux density $q_0$ at its beginning and ending at times $T_0$ and $T_E$. Equations (A1)–(A8) describe the propagation of a single input pulse to the depth $Z_t$ while beyond that depth the wetting front decelerates and the mobile water content at the arrival of the draining front decreases according to Hincapié and Germann [2009a] as

$$z_{w}(t) = \left(\frac{3 \cdot V_{WCW}}{2 \cdot L}\right)^{2/3} \cdot \left(\frac{g}{\eta}\right)^{1/3} \cdot (t - T_E)^{1/3}. \quad \text{(A13)}$$

The first derivative of equation (A13) produces the velocity of the wetting front as

$$v(t)|_{zw} = \left(\frac{V_{WCW}}{2 \cdot L}\right)^{2/3} \cdot \left(\frac{g}{3 \cdot \eta}\right)^{1/3} \cdot t^{-2/3}, \quad \text{(A14)}$$

and it follows that

$$w(t)|_{zw} = \left(\frac{w}{g}\right)^{1/3} \cdot \left(\frac{3 \cdot V_{WCW}}{2 \cdot L}\right)^{1/3} \cdot (t - T_E)^{-1/3} \cdot L^{2/3}, \quad \text{(A15)}$$

where $V_{WCW} = q_0$ ($T_E - T_0$) is the total and presumed constant mobile water content of a WCW. Multiplication of equation (A14) with equation (A15) yields the volume flux density at the wetting front as

$$q(t)|_{zw} = \frac{V_{WCW}}{2} \cdot \left(\frac{3}{t^2 \cdot (t - T_E)^{5/3}}\right)^{1/3}. \quad \text{(A16)}$$

[88] Any input function $q(0, t)$ can be represented as a series of pulses that can be routed according to Lighthill and Whitham [1955]. Assuming for simplicity that constant $L$ applies to all pulses, the celerity of an increasing jump is

$$c_{ij+1} = \frac{q_{i+1} - q_i}{w_{i+1} - w_i}, \quad \text{(A16)}$$

where $q_{i+1} > q_i$. Multiple wetting and drying waves can easily be propagated through the profile.

[89] Acknowledgments. Thanks are due to all the graduate students and colleagues we have worked with on preferential flows in soils since 1982. Comments on a draft of this paper from Andrew Binley and Nick Chappell were particularly helpful. Andy also contributed to the section on new measurement techniques. The comments of three referees, including Markus Weiler, greatly improved the presentation of the material in this paper. KB was supported by NERC Grant NE/G0177123/1.

References


Blackwell, Chichester.


Hincaip, I., and P. Germain (2010), Water content wave approach applied in neutron methods of finger flow, Vadose Zone J., 9, 278–284.
Hornerber, G.M., K.J. Beven, and P.F. Germain (1990), Inferences about solute transport in macroporous forest soils from time series models, Geoderma, 46, 249–262.
Hunt, A.G. (2009), Relevance of percolation theory to power-law behavior of dynamic processes including transport in disordered media, Complex- ity, 15, 13–27.
Irgauskus, I., K.J. Beven, and A. Musy (2005), Data-based modelling of runoff and chemical tracer concentrations in the Hauze-Menthue (Swit- zerland) research catchment, Hydrolo. Process., 19, 2557–2574.


Peterscheck, A. (1973), Über die Wirkung systematischer Entwässerungen in Hanglagen, Diss. number 5079, ETH-Zürich, Switzerland.


van Schaik, N.L. M.B. (2009), Spatial variability of infiltration patterns related to site characteristics in a semi-arid watershed, Catena, 78, 36–47.


