Iterative approach to modeling subsurface stormflow based on nonlinear, hillslope-scale physics

J. H. Spaaks¹, W. Bouten¹, and J. J. McDonnell²

¹University of Amsterdam, Institute for Biodiversity and Ecosystem Dynamics, Computational Geo-Ecology – Nieuwe Achtergracht 166, 1018WV Amsterdam, The Netherlands
²Oregon State University, Department of Forest Engineering – 015 Peavy Hall, Corvallis OR 97331-5706, USA

Received: 16 July 2009 – Accepted: 21 July 2009 – Published: 3 August 2009

Correspondence to: J. H. Spaaks (jspaaks@uva.nl)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Soil water transport in small, humid, upland catchments is often dominated by subsurface stormflow. Recent studies of this process suggest that at the plot scale, generation of transient saturation may be governed by threshold behavior, and that transient saturation is a prerequisite for lateral flow. The interaction between these plot scale processes yields complex behavior at the hillslope scale. We argue that this complexity should be incorporated into our models. We take an iterative approach to developing our model, starting with a very simple representation of hillslope rainfall-runoff. Next, we design new virtual experiments with which we test our model, while adding more structural complexity. In this study, we present results from three such development cycles, corresponding to three different hillslope-scale, lumped models. Model\textsubscript{1} is a linear tank model, which assumes transient saturation to be homogeneously distributed over the hillslope. Model\textsubscript{2} assumes transient saturation to be heterogeneously distributed over the hillslope, and that the spatial distribution of the saturated zone does not vary with time. Model\textsubscript{3} assumes that transient saturation is heterogeneous both in space and in time. We found that the homogeneity assumption underlying Model\textsubscript{1} resulted in hillslope discharge being too steep during the first part of the rising limb, but not steep enough on the second part. Also, peak height was underestimated. The additional complexity in Model\textsubscript{2} improved the simulations in terms of the fit, but not in terms of the dynamics. The threshold-based Model\textsubscript{3} captured most of the hydrograph dynamics (Nash-Sutcliffe efficiency of 0.98). After having assessed our models in a lumped setup, we then compared Model\textsubscript{1} to Model\textsubscript{3} in a spatially explicit setup, and evaluated what patterns of subsurface flow were possible with model elements of each type. We found that Model\textsubscript{1} tended to generate relatively smooth, steady state-like spatial patterns. Model\textsubscript{3} generated more complex patterns, in which lateral flow could be concurrently increasing and decreasing in different parts of the hillslope. We realize that the concepts proposed in this manuscript do not represent the only way in which nonlinear dynamics may be implemented in a model of subsurface stormflow. However,
we believe that development of new model structures and the subsequent confrontation of model results with existing preconceptions will lead to a better understanding of subsurface stormflow and catchment runoff dynamics.

1 Introduction

Over the past decades, much time and effort has been spent in researching how water moves through the soil in small, humid, upland catchments. Under these circumstances, we often have to deal with the mostly unknown effects of a shallow soil, that is usually underlain by an impermeable layer of (for example) bedrock or glacial till. Due to the consolidated nature of this material, these regions are usually dominated by steep slopes of 30° or more. Together, these conditions often lead to subsurface stormflow being the main contributor to storm discharge in these areas (Hursh and Brater, 1941; Hursh, 1944; Hewlett and Hibbert, 1967; Weiler et al., 2005). Since the 1930s, various mechanisms of flow have been proposed to describe water transport in these catchments, e.g. overland flow (Horton, 1933), partial contributing areas (Benson, 1964), variable source areas and translatory flow (Hewlett and Hibbert, 1967), return flow and direct precipitation onto saturated areas (Dunne and Black, 1970), saturated wedge development (Weyman, 1973), flow through pipes and macropores (Jones, 1971; Mosley, 1979), transmissivity feedback (Rodhe, 1987; Seibert et al., 2003), flow at the soil-bedrock interface (McDonnell, 1990; Peters et al., 1995; Tani, 1997; Freer et al., 1997, 2002), hydrogeomorphic linkages (Sidle et al., 2000), and the fill-and-spill concept (Tromp-van Meerveld and McDonnell, 2006b).

In spite of this apparent abundance of ideas, many of our models still rely on a number of assumptions such as (McDonnell, 2003): (i) surface topography driven drainage direction; (ii) gradually declining lateral hydraulic conductivity with increasing depth; (iii) spatially more or less uniform water table response to precipitation and (as a result) development of a saturated wedge from the valley upwards. However, convincing field evidence has been presented in recent years which suggests that, for subsurface
stormflow dominated catchments, these assumptions may not be valid. For example, local drainage direction of subsurface flow seems to be driven by subsurface topography and spatial variations in soil depth, rather than soil surface topography (McDonnell, 1990; Peters et al., 1995; Tani, 1997; Freer et al., 1997, 2002; Buttle et al., 2004; Tromp-van Meerveld and McDonnell, 2006a,b). In most cases, soil matrix hydraulic conductivity decreases with increasing depth, but preferential flow may strongly affect how water travels through the soil, as shown by Noguchi et al. (1999) and Sidle et al. (2000). Covering a range of scales, these preferential flow pathways include, but are not limited to, root channels, organically rich horizons, and channels in the bedrock. The presence of these features increases the complexity of the hillslope’s hydrological functioning, in particular with respect to feedback mechanisms, storage effects, self-organization, hysteresis and threshold behavior (Phillips, 2003). Also, hydrologic conductivity values derived from soil core experiments can become almost useless if preferential flow pathways exist in the hillslope; transient water table response is not uniform over the hillslope, but is governed by more complex behavior – see for example Seibert et al. (2003) and Tromp-van Meerveld and McDonnell (2006b).

Gaining a better understanding of subsurface stormflow generation is very important because practical issues such as flood forecasting, ecosystem management, water quality control and contaminant transport are all very much related to subsurface stormflow in this environment. For stable systems, a high degree of accuracy can be achieved with model types incorporating very little process knowledge (metric models, Wheater et al., 1993) with regard to the internal functioning of a hillslope or catchment (Andrews et al., 1995). However, improved representation of first-order controls on subsurface stormflow in our quantitative models will be crucial for making good predictions when conditions such as climate and landuse shift beyond the range of prior experience (Kirchner, 2006), as well as with problems for which the flow path of water is important. In such cases, accurate predictions can only be realistically expected from models based on process knowledge of subsurface stormflow.

In an effort to help advance the understanding of subsurface stormflow, Kirchner
(2006) suggests five directions of research that, in his view and ours, offer good chances of success. In this paper, we pay special attention to two of his suggestions: (i) development of models that describe more consistently the nonlinear behavior observed in hillslope hydrologic systems; (ii) development of physically-based gray-box models for describing the hydrological system at the hillslope scale, while recognizing that the governing equations at this scale may look different from those which govern the small-scale physics.

Following the downward approach (Klemes, 1983; Sivapalan and Young, 2005), this study implements an iterative research cycle consisting of: (1) explicit formulation of a hypothesis in a computer model structure; (2) designing and performing a virtual experiment, with which we test our hypothesis; (3) confronting the simulation results with our perceptual model (Beven, 2001); (4) introduction of additional complexity into our hypothesis, and going back to step 1.

2 Methods

Our philosophy is that we should carry out the steps of the iterative research cycle until our model is consistent with our collective field knowledge, or until the introduction of additional complexity is no longer warranted by the available data (Jakeman and Hornberger, 1993). To avoid problems with overparameterization and equifinality (e.g. Beven and Freer, 2001; Beven, 2006), each model must be fairly simple in terms of the model structure, its spatial discretization and the number of parameters, so that a meaningful assessment of its dynamics can be made. In a series of virtual experiments (Weiler and McDonnell, 2004), we ask the following questions:

1. For a lumped model, how is discharge affected by assuming that the pattern of transient groundwater within a model element is spatially continuous?

2. For a lumped model, how is discharge affected by assuming that the pattern of transient groundwater within a model element is spatially discontinuous?
3. For a lumped model, how is discharge affected by assuming that the pattern of transient groundwater within a model element is spatially discontinuous and variable over time?

4. If we take two spatially distributed transect models of subsurface stormflow, one in which transient groundwater within a model element is spatially continuous, and one in which it is spatially discontinuous and variable over time, how do the patterns of lateral flow compare?

2.1 Data

To address these questions, we parameterize our evolving model with data from the extensively studied hillslope at the Panola Mountain Research Watershed (Georgia, USA). The data include highly detailed information on precipitation and streamflow fluxes, as well as a map of soil depth derived from about 250 measurements covering the approximately 50×20 m hillslope. We selected a major storm event, during which 59 mm of precipitation fell over a time span of about 1.5 days. Our streamflow data has been recorded in the trench that forms the lower boundary of the research area, about 30 m upslope from an ephemeral stream. Though the original data discriminates between matrix flow and macropore flow for individual trench sections, we chose to use the total flow rate, calculated as the sum of matrix flow and macropore flow from all sections. For further details on the study site, the measuring devices, recorded variables, etc., please refer to Freer et al. (2002) and Tromp-van Meerveld et al. (2008), and references therein. Calibration of our models was done using the SCEM-UA algorithm (Vrugt et al., 2003; Spaaks, 2009). The total sum of squares between the observed and simulated discharge was used as the objective function.

2.2 Model evaluation

Following Wagener et al. (2004), we assess our evolving model based on its ability to reproduce past observations (performance), its uncertainty, and its underlying as-
sumptions (*model realism*, Seibert and McDonnell, 2002). Our optimization procedure facilitates an easy assessment of both performance and uncertainty. With regard to model realism, we analyze to what extent our models are capable of reproducing the behavior of subsurface stormflow dominated hillslopes. Process studies from around the world have established that this behavior is typically characterized by: a connectivity threshold-dominated flow regime, with significant flows occurring only after a certain threshold is exceeded (Tani, 1997; Tromp-van Meerveld and McDonnell, 2006a); highly skewed hydrographs with a steep increase in discharge when the threshold is exceeded (Tromp-van Meerveld and McDonnell, 2006a); a break in slope on the recession limb (Hewlett and Hibbert, 1963; Harr, 1977); convergence of subsurface flow being driven by topography of a hydrologically impeding layer in the shallow subsurface (Freer et al., 1997, 2002); transient saturation occurring first in areas of shallow soil, and extending in the downslope direction (Tromp-van Meerveld and McDonnell, 2006b); transient saturation being a prerequisite for significant lateral subsurface flow (Weiler et al., 2005); non-steady state groundwater table dynamics (Seibert et al., 2003).

### 3 Virtual experiment 1

#### 3.1 Rationale

We model the Panola hillslope as a lumped tank model, in which the discharge rate is proportional to the pressure head in the tank. Although many hydrologically important variables are heterogeneous spatial fields, we assume that this averages out at the scale of the hillslope.
3.2 Model: homogeneous transient saturation, lumped

3.2.1 Tension storage

The hillslope is treated as a spatial element of a certain area \( A_{\text{elem}} \) \([L^2]\), soil depth \( d \) \([L]\) and effective porosity \( \phi \) equal to the volumetric water content at saturation \( \theta_{\text{sat}} \) \([L_3 \cdot L^{-3}]\), which together determine the element’s total pore space where soil water can be stored \( V_{\text{max}} \) (see Fig. 1):

\[
V_{\text{max}} = A_{\text{elem}} \cdot d \cdot \theta_{\text{sat}}
\]  

(1)

When a completely saturated element is allowed to drain freely, part of the soil water that was initially present, will remain in the soil due to capillary forces. After all of the excess soil water has been drained, the soil still contains a volumetric fraction \( \theta_{\text{fc}} \) \([L_3 \cdot L^{-3}]\) of water (fc for “field capacity”). We divide the total pore space into 2 conceptual stores based on the value of \( \theta_{\text{fc}} \). As long as the element’s actual volumetric water content \( \theta_{\text{act}} \) \([L_3 \cdot L^{-3}]\) is below \( \theta_{\text{fc}} \), all water present in that element is retained by the soil. If the below-\( \theta_{\text{fc}} \) store is completely full, any extra infiltration or lateral flow coming into the element will lead to emergence of excess water at the soil-impeding layer interface. The rate at which this happens is the volumetric emergence rate \( E \) \([L_3 \cdot T^{-1}]\).

3.2.2 Emergence of tension-excess soil water

When the below-\( \theta_{\text{fc}} \) store is full, any extra water added to the spatial element will generate transient saturation:

\[
E = \begin{cases} 
0 & 0 \leq \theta_{\text{act}} \leq \theta_{\text{fc}} \\
\rho + Q_{\text{in}} & \theta_{\text{fc}} < \theta_{\text{act}} \leq \theta_{\text{sat}} 
\end{cases}
\]

(2)
where \( E \) is the volumetric emergence rate \([L^3 \cdot T^{-1}]\), \( P \) is the volumetric precipitation rate \([L^3 \cdot T^{-1}]\) falling onto the spatial element and \( Q_{\text{in}} \) is the lateral flow \([L^3 \cdot T^{-1}]\) coming into the element from neighboring elements.

### 3.2.3 Lateral flow

When \( \theta_{\text{act}} \) exceeds \( \theta_{\text{fc}} \), a transient water table is assumed to be present within the spatial element. The spatial gradient of transient water table drives the outflow into one of the neighboring downslope elements according to:

\[
Q = -K_{\text{sat}} \cdot A_{\text{flow}} \cdot \frac{\Delta H}{\Delta s}
\]  
(3)

where \( Q \) is the lateral flow \([L^3 \cdot T^{-1}]\), \( K_{\text{sat}} \) is the saturated hydraulic conductivity \([L \cdot T^{-1}]\), \( A_{\text{flow}} \) is the cross-sectional area of flow \([L^2]\), and \( \Delta H \cdot \Delta s^{-1} \) is the spatial gradient in water table potential \([L \cdot L^{-1}]\). \( H \) is calculated according to:

\[
H = z + V_m \frac{A_{\text{elem}} \cdot (\theta_{\text{sat}} - \theta_{\text{fc}})}{A_{\text{elem}} \cdot (\theta_{\text{sat}} - \theta_{\text{fc}})}
\]  
(4)

where \( H \) is the hydraulic head \([L]\) and \( z \) is the elevation head of the impeding layer surface \([L]\). The subsurface drainage direction is determined using a steepest-descent algorithm like that of O'Callaghan and Mark (1984), applied to \( H \).

### 3.2.4 Leakage of excess soil water into the bedrock

Transport of transient groundwater from the above-\( \theta_{\text{fc}} \) store to the bedrock is governed by:

\[
V_m = (\theta_{\text{sat}} - \theta_{\text{fc}}) \cdot A_{\text{elem}} \cdot (H - z)
\]  
(5)
L = -K_l \cdot V_m \quad (6)

where $V_m$ is the volume of water [$L^3$] in the above-$\theta_{fc}$ store ($m$ for mobile), $L$ is the volumetric leakage rate of transient groundwater to the bedrock [$L^3 \cdot T^{-1}$] and $K_l$ [$T^{-1}$] is a proportionality constant.

### 3.2.5 Recharge

In the current formulation of our hypothesis, recharge is equal to emergence:

$$R = E \quad (7)$$

where $R$ represents the volumetric recharge rate [$L^3 \cdot T^{-1}$] of the saturated zone in a spatial element.

### 3.2.6 Governing equation

The volume of mobile water in each spatial element changes as a result of lateral flow, recharge, and leakage to the bedrock according to:

$$V_m|_{t+\Delta t} = V_m|_t + (Q_{out} + R + L) \cdot \Delta t \quad (8)$$

The current model formulation is hereafter referred to as Model$_1$.

### 3.3 Analysis

Figure 2 shows rainfall-runoff observations and simulations. The discharge curve generated using the best parameter set is presented in subplot B. The best performing
parameter set has been included for convenience in Table 1, along with the boundaries of the parameter search.

The figure shows that the model dynamics are different from the observed dynamics. More specifically, when transient saturation occurs at the beginning of Sector B (6 February, 06:30–20:15), the flow response is immediate and coincides with the observed timing of first flow. However, the simulated hydrograph is steeper compared to the observations made throughout Sector B. When the observed discharge eventually starts to rise steeply at the beginning of Sector C (6 February, 20:15–22:15), the slope of observed discharge in that sector exceeds that of the simulated discharge from both Sector B and C. Also, simulated peak flow precedes observed peak flow by about 3.6 h, and when it occurs, it is only 56% of observed discharge (0.164 and 0.291 mm hr$^{-1}$ for the “best simulation” run and observed discharge, respectively).

The observed falling limb can be divided into two parts based on local slope; during the first part (Sector D: 6 February, 22:15–7 February, 01:30), the observed slope is more negative compared to the second part (Sectors E and F: 7 February, 01:30–9 February, 23:45). This difference is not present in the simulated discharge. Summarizing the overall fit in a performance statistic, the best parameter set is associated with a relatively low Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) of 0.60. This indicates that the current model formulation does not accurately capture the observed hydrological behavior. Table 2 presents an overview of typically observed dynamics and what aspects are captured by the current model formulation.

By making use of SCEM-UA’s ability to provide information about parameter uncertainty, we also constructed a Bayesian probability interval using the last 1000 model parameterizations (shaded area in Fig. 2b). From the interval shown in Fig. 2b it is clear that there is very little uncertainty caused by the parameters, as was expected given the relative simplicity of the model and the low number of parameters. Though the uncertainty may be small, the [5,95] percentile contains only a few measurements besides those in Sector A.
3.4 Discussion

From Eqs. (7) and (8) it can be inferred that the model assumes that the transient saturated layer is present everywhere in the spatial element (see Fig. 3a). This implies that once excess soil water starts to emerge at the soil-impeding layer interface, hillslope-scale discharge will occur. This effect can be seen in the simulated discharge curve presented in Fig. 2b; as soon as the soil is at $\theta_{\text{act}}$ by the end of Sector A, discharge occurs. The overestimation of discharge in Sector B and the underestimation in Sectors C, D and E is due to our calibration procedure: since we minimized the sum of squared deviations on all observation points, we ended up with what might be considered an average value for $K_{\text{sat}}$. This value does indeed minimize the global misfit, but it also yields a discharge curve that does not fit the observations in any part of the data set.

4 Virtual experiment 2

4.1 Rationale

In the previous experiment, it became apparent that the heterogeneity which is present at the sub-element scale does not average out at the hillslope scale. Representing transient saturation as a continuous layer is therefore not appropriate for hillslope-scale lumped modeling, since it leads to discharge that rises too steeply during the initial stages of the event and fails to reach the correct peak height. In the current experiment, we bring our implementation more in line with the updated perceptual model. To do so, we assume that the zone of transient saturation does not cover all parts of the soil-impeding layer interface (see Fig. 3b).
4.2 Model 2: heterogeneous transient saturation, lumped

In this model, we reject the idea that a spatial element is internally homogeneous, or that the heterogeneity averages out at the scale of the hillslope. Instead, we embrace the idea of having a spatially discontinuous layer of transient saturation within an element (see Fig. 3b). As a result, emerging excess water leaving the below-$\theta_{fc}$ store cannot be added immediately to the above-$\theta_{fc}$ store. Assuming that excess soil water emerges homogeneously from the bottom of the soil column, an instantaneous pulse of emerging excess soil water will not reach the above-$\theta_{fc}$ store instantaneously, as was implied by Eq. (7). Instead, arrival of this water at the zone of transient saturation will be delayed according to the frequency distribution of travel times $g(t) \left[ T^{-1} \right]$ (e.g. Mazor and Nativ, 1992; Maloszewski and Zuber, 1993):

$$g(t) = \frac{M(t)}{M_{t_0}} = \frac{M(t)}{\int_{t_0}^{+\infty} M(t) \, dt}$$  \hfill (9)

where $M_{t_0}$ is the mass emerging as an instantaneous pulse at time $t_0$ and $M(t)$ represents the mass reaching the transient saturation zone at time $t$. In order to conserve mass, the integral of $g(t)$ must equal unity:

$$\int_{t_0}^{+\infty} g(t) \, dt = 1$$  \hfill (10)

In contrast to the conventional travel time distribution methodology, as described by McGuire and McDonnell (2006) among others, we do not use the convolution integral. Instead, we use a numerical approach to integrate Eq. (9) over the model time step $\Delta t$.

For the current model formulation we replace Eq. (7) by:

$$R = \sum_{\forall i \in I} \int_{t-\tau_i}^{t-\tau_i+\Delta_t} g(t - \tau_i) \, dt \cdot E_i$$  \hfill (11)
where \( I \) is the set of iterations during which excess water emerged, \( \tau_i \) is a shift in time pertaining to iteration \( i \), \( g(t - \tau_i) \) is the travel time distribution in the spatial element, and \( E_i [L^3 T^{-1}] \) is the volumetric rate at which excess water emerges during iteration \( i \).

For simplicity, we assume a uniform travel time distribution according to:

\[
g(t - \tau_i) = \begin{cases} T^{-1} & t - \tau_i < T \\ 0 & t - \tau_i \geq T \end{cases}
\]

(12)

where \( T \) is the duration of the uniform travel time distribution. The current model formulation is hereafter referred to as Model 2.

### 4.3 Analysis

Figure 4 shows rainfall-runoff observations and simulations. The discharge curve generated using the best parameter set is presented in subplot B. The best performing parameter set has been included for convenience in Table 1, along with the boundaries of the parameter search.

For the best performing simulation, discharge behavior was not in accordance with observations in most sectors. For example, simulated flow in Sector B increased more steeply than what was observed, even though there was some improvement in that sector relative to the previous experiment. The model dynamics in Sectors C and D are completely different from the observations. The simulated peak is very smooth and has a relatively low magnitude and a long duration, rather than the high, short, and steep peak which was observed. Simulated discharge in these sectors actually becomes less steep rather than steeper, and therefore does not produce an accurate estimate of the timing of peak flow. Peak flow magnitude is just 68% of observed peak flow (0.197 and 0.291 mm hr\(^{-1}\) for the “best simulation” run and observed discharge, respectively). The best parameter set is associated with a relatively low Nash-Sutcliffe efficiency of 0.85, which shows that the current model formulation does not accurately...
capture the observed hydrological behavior. Table 2 presents an overview of typically observed dynamics and what aspects are captured by the current model formulation.

Bayesian probability intervals were constructed in the same way as in the previous experiment. The parameters could be well identified from the discharge data. As in the previous experiment, the probability interval is small. Compared to Model\textsubscript{1}, it contains more observations throughout Sectors E and F.

4.4 Discussion

At first sight, the evaluation criteria in Table 2 do not suggest that the introduction of a heterogeneous distribution of transient saturation improved our representation of subsurface stormflow. However, if we focus on sector B, it appears that the simulated dynamics are similar to those of the observations in that sector, even though the fit is not as good as it could be. As in the first experiment, this is caused by our choice of objective function, which represents the model misfit as a single statistic based on data from all sectors, rather than a subset thereof. Had we run a set of simulations based on Sectors A and B only, we would have found a parameter set that fitted the data very well. One could argue that Model\textsubscript{2} therefore represents the processes at low flow (Sectors A and B) fairly accurately, but that the model does not yet incorporate processes acting under high flow conditions. In the next section, we will therefore investigate if the different behavior at low flow and high flow could be the result of differences in the spatial organization of mobile water.

5 Virtual experiment 3

5.1 Rationale

We interpret the observed discharge as being generated by a mechanism that encompasses impedance to flow during the initial stages of the event, perhaps as a result of
spatial organization of excess soil water (in particular with regard to the connectivity of wet patches). If the hillslope is indeed governed by such a mechanism, it could store incoming precipitation without generating much discharge during the initial stages of the event, while at the same time allowing for quick discharge of excess water once the spatial element exceeds a certain wetness threshold. After this has occurred, the water that has accumulated in the spatial element under the inefficient regime is subsequently discharged, resulting in a large rise in subsurface stormflow.

5.2 Model 3: spatially and temporally heterogeneous transient saturation, lumped

Our numerical approach to solving Eq. (9) enables the use of different travel time distributions, with which different states of the spatial element can be expressed. Figure 3c shows the surface of an impermeable layer, as well as a branched, spatially discontinuous saturated zone, which crosses the boundary of the spatial element and enters the downstream neighbor. Any soil water excess emerging in the spatial element is assumed to leave the element through this preferential flow path. As the element becomes wetter, some preferential flow features that were previously inactive may become part of the subsurface flow network, effectively increasing the subsurface drainage density (see Fig. 3d). With the expansion of the subsurface flow network, the unsaturated soil volume around the zone of transient saturation becomes more efficiently connected to the downstream element. Consequently, transport of soil water internal to the element is governed by a shorter travel time distribution, resulting in more dynamic input-output behavior of the element as a whole, or equivalently, a reduced dampening effect of the input signal. We adapt Eq. (12) to represent unsaturated
flow as a threshold-based travel time distribution:

\[
g(t - \tau_i) = \begin{cases} 
T_{\text{dry}}^{-1} & V_m \leq WT \land t - \tau_i < T_{\text{dry}} \\
0 & V_m \leq WT \land t - \tau_i \geq T_{\text{dry}} \\
T_{\text{wet}}^{-1} & WT < V_m \land t - \tau_i < T_{\text{wet}} \\
0 & WT < V_m \land t - \tau_i \geq T_{\text{wet}} 
\end{cases}
\] (13)

with \(T_{\text{dry}}\) and \(T_{\text{wet}}\) \([T]\) the duration of the uniform distribution under relatively dry and relatively wet conditions, respectively, and \(WT\) the threshold wetness \([L^3]\) at which the spatial element changes from dry to wet and vice versa. Further details on the setup of the model can be found in Table 1. The current model formulation is hereafter referred to as Model3.

### 5.3 Analysis

Figure 5 shows rainfall-runoff observations and simulations. The discharge curve generated using the best parameter set is presented in subplot B. The best performing parameter set has been included for convenience in Table 1, along with the boundaries of the parameter search.

The figure shows that the best performing model simulation follows the observed behavior more closely and consistently than the two previous models. For example, the timing of first flow coincides with the observed timing, and simulated discharge remains relatively low throughout Sectors A and B, even though more than 90% of total precipitation falls during this period. For the best performing simulation, peak discharge is 86% of observed discharge (0.251 and 0.291 mm hr\(^{-1}\) for the “best simulation” run and observed discharge, respectively). The simulated peak flow precedes the observed peak flow by 0.75 h.

The model behavior on the falling limb is very close to observed behavior except that the discharge curves generated by the current model do not have an inflection point.
similar to that of the observed data. As a result of the closely matching behavior of simulated and observed dynamics, the Nash-Sutcliffe efficiency for the best performing discharge curve is 0.98. Table 2 presents an overview of discharge dynamics as they are typically observed and the aspects which are captured by Model3.

The introduction of a threshold-based travel time distribution resulted in a very small Bayesian probability interval for this experiment. The parameters were identified without much difficulty.

5.4 Discussion

Despite the good performance of discharge realizations on the falling limb (Sectors D, E, and F), a parameter set that could mimic the point of inflection occurring between Sectors D and E was not found. This may be the result of some unknown process that the model does not account for, such as hydrophobicity or the occurrence of air pockets in the soil. Alternatively, one could surmise that the hillslope acts as a number of separate units. Each unit may have a smooth falling limb, but depending on how they interact, a point of inflection may be present in the hillslope-scale hydrograph. We will investigate this further in the next section.

6 Virtual experiment 4

6.1 Rationale

If the hillslope acts as a number of elements, the hillslope-scale discharge dynamics must be the result of how these elements interact. The purpose of the current experiment is therefore to assess whether model elements of different types give rise to specific spatial patterns of subsurface stormflow.
6.2 Spatially and temporally heterogeneous transient saturation, spatially distributed

We divide a transect of the Panola hillslope into eight spatial elements, with each element draining into its downstream neighbor. Local bedrock slope along the transect ranges from 0.155 to 0.390 $m^2 m^{-1}$ and soil depth (see yellow bars in Fig. 6) varies from 0.180 to 1.39 m. We again use SCEM-UA to calibrate the parameters of each element, using the same calibration data as in the first experiment. We assume that each spatial element has the same hydrologic properties so any particular set of parameters is applied to all spatial elements. In addition to the recorded precipitation, we applied a small artificial event approximately two days after natural precipitation ceased (see Fig. 6). The precipitation intensity of the second, artificial storm was 3.00 mm hr$^{-1}$ throughout its 4-hr duration. The simple temporal structure of the artificial storm allowed us to make a better assessment of any differences between the models with regard to flow timing and magnitude.

6.3 Analysis

6.3.1 Spatial elements of type Model$_1$

At the beginning of the experiment, the soil moisture deficit is proportional to soil depth (see Table 3), and there is no mobile water since $\theta_{act}<\theta_{fc}$ for every element. Because of this, hillslope internal flow occurs first in the most shallow elements at $y=[42,30,36,6]$. Even after emergence of excess water occurs in some elements, the resulting lateral flow is not large enough to disrupt the soil depth-driven pattern of first flow. Because the soil depth is relatively high at $y=18$, this element can store the largest volume of incoming water before excess soil water emerges. As a result, only a small volume of water is expelled as lateral flow.

$\footnote{Spatial elements and their outflow curves are indicated by the same y-value to improve legibility}$
For this type of element, the start of the second precipitation event coincides exactly with the increase in hillslope internal flow, while the end of the precipitation event coincides exactly with maximum flow for all elements. Spatial differences in distributed flow are limited to the rate at which flow increases. Table 2 presents an overview of what aspects of subsurface stormflow dominated systems are captured by the spatially distributed Model1.

### 6.3.2 Spatial elements of type Model3

For Model3, the spatial pattern of first flow is identical to that of Model1: spatial elements with shallow soils have the smallest absolute soil moisture deficit, so emergence occurs first in these elements. Since none of the elements contains mobile water at the start of the experiment (θ<sub>act</sub>&lt;θ<sub>fc</sub>), all elements are in the same state of wetness and therefore use the same travel time distribution. As a consequence, the spatial pattern of emergence also applies to recharge and therefore to timing of first flow.

At the onset of the second event, the upslope elements at \( y = [24,30,36,42] \) as well as the downslope elements at \( y = [0,6] \) have switched back again to the longer “dry” travel time distribution, while the midslope elements at \( y = [12,18] \) are still in the responsive “wet” mode. This pattern of travel time distribution is related to the local impedance to lateral flow. The midslope elements have a very deep soil, hence their bedrock gradient is less steep compared to other elements. Because of this, water drains more slowly from these elements and they adhere longer to the “wet” travel time distribution. Because of the shorter travel time distribution, precipitation is more directly linked to lateral flow in these elements, therefore lateral flow starts to increase from the beginning of the second precipitation event, and continues rising until the end of the artificial storm, when peak flow occurs. All in all, the hydrological dynamics of the midslope elements during the second event are very similar to that of Model1.

By the time the artificial precipitation event begins, upslope elements have already switched back to the “dry” distribution. As a result, the emergence of excess soil water does not lead to a large increase in lateral flow from the element, although a low-flow
level is sustained throughout most of the simulation.

Contrary to this behavior, the downslope elements respond to the precipitation event in an entirely different, nonlinear fashion. At the beginning of the artificial precipitation event, these elements have drained enough water to be in the “dry” state. Therefore, the precipitation event results in a sustained flow of low intensity for about 11.75 h after precipitation ceases; after that, enough water has accumulated in \( V_m \) to make the element switch to the “wet” travel time distribution, which in turn leads to a steep increase in flow for these elements. Table 2 presents an overview of what aspects of subsurface stormflow dominated systems are captured by the spatially distributed Model\(_3\).

### 6.4 Discussion

In the concept underlying Model\(_1\), discharge from a model element is directly related to the volume of emerging water. When applied to a transect of elements, as in the current experiment, this leads to patterns of subsurface stormflow that are less complex in structure compared to patterns commonly encountered in the field (e.g. Seibert et al., 2003). Lateral flow in this concept acts as a smoothing function. Contrary to this, a model element governed by a nonlinear storage-flow relation, as in Model\(_3\), can generate lateral flow curves of different shapes as well as magnitudes. When such a concept is applied to a series of elements, the interaction between elements may yield a pattern as simple as that of the more straightforward Model\(_1\), but it may also yield a complex, heterogeneous pattern. This wider range of complexity can be achieved without compromising the model’s parsimony by resorting to spatially distributed parameters.

The pattern of subsurface flow presented in Fig. 6b should not be interpreted as an accurate prediction of the real-world pattern; this study has been focused on the development of a model that would in principle be able to generate subsurface stormflow dynamics that are more in accordance with the heterogeneity typically observed in the field. Since the spatially distributed Model\(_3\) is very sensitive to changes in soil depth,
initial conditions and so on, relatively small errors imposed on the measurements can result in a rather different result. Moreover, our choice of where to position the transect and what grid cell size to use, also affects the results. Because of this, there is no purpose in drawing conclusions about how well the simulated patterns fit the real-world patterns, other than in terms of the dynamics.

The results presented in Fig. 6 also constitute a plea for a more effective experimental design of distributed observation networks, and for stricter testing when discriminating between model structures, as argued by Kirchner et al. (1996), Mroczkowski et al. (1997), Kirchner (2006), and others. During the second event for example, the flow curves in the midslope area (y=[12,18]) are similar for both models, even though the hydrological dynamics within the hillslope are very different for each model. When provided with data from the midslope section only, it is easy to see how a researcher could be lured into accepting either one of the models as an appropriate representation of reality. Even though the models cannot coexist, they can generate similar flow curves in parts of the hillslope. In order to distinguish one model from the other, we would need to introduce additional, spatially explicit measurements. With the data that has been presented for this study, however, it is unjustified to continue the Iterative Research Cycle.

7 Summary and conclusions

Many of the physically-based hydrological models that are currently in use, are based on an often implicit scaling assumption (Kirchner, 2006). This assumption states that small-scale heterogeneity upscales to the scale of a model grid cell, in such a way that the model’s governing equation will differ in terms of its parameters, but not in terms of its structure. This, however, is not in line with typically observed behavior (see Table 2), as soil core scale behavior differs from hillslope scale behavior. To acknowledge the nonlinear properties associated with the hillslope hydrological system, we developed three hillslope-scale models of subsurface stormflow within an iterative research
cycle. In our first experiment, we assumed that transient saturation occurs as a spatially continuous layer on top of the hydrologically impeding layer. When this concept is applied at the hillslope scale, the timing of first hillslope-scale discharge coincides with the emergence of transient saturation. Due to the implicit connectivity within the hillslope, transport of excess water is relatively unobstructed. As a result, discharge rises too steeply during the early stages of an event, which leads to peak flow reaching only 56% of observed discharge.

In our second experiment, we rejected the idea that transient saturation occurs as a spatially continuous layer within the hillslope. Instead, water that emerges at the impeding layer must travel laterally through the unsaturated zone before it reaches the saturated zone, from where it can be discharged. The transport time through the unsaturated zone was modeled by calibrating a one-parameter frequency distribution of travel times. To maintain flexibility, this travel time distribution was solved in a numerical scheme rather than by using the convolution integral. The global fit improved relative to that of Model 1, but in Sectors C and D the simulated dynamics still deviated substantially from the observed dynamics. In Sectors B, E and F, the model successfully captured the discharge dynamics, but the fit was negatively affected by our use of a single objective function operating on all data points.

In experiment 3, we built on the experience gained from virtual experiment 2 by introducing a threshold-based travel time distribution. In this way, we were able to separate the parameters which were relevant under relatively dry conditions from those which were relevant under relatively wet conditions. We hypothesized that at low flow, patches of excess water were present within the hillslope, but that they were not sufficiently connected to each other and to the trench face to yield large flow rates. After a certain threshold was exceeded, however, these connections were sufficiently active to account for the steep, large rise in subsurface stormflow. With this model formulation, we achieved a good fit as well as good dynamics in all sectors, since the parameters that govern the behavior at low discharge rates were separated from those at high discharge. Although the model captured almost all of the aspects of the observed discharge.
hydrograph, the break in slope on the falling limb was not reproduced by the model.

After having assessed these three models in a lumped setup, we then compared Model_1 to Model_3 in a spatially explicit setup, to determine what patterns of subsurface stormflow were possible with each model. We found that Model_1 tended to generate relatively smooth, steady state-like spatial patterns. With Model_3, we were able to generate more complex patterns, in which lateral flow could be concurrently increasing and decreasing in different parts of the hillslope. The pattern generated by Model_3 was more in concordance with our perceptual model of subsurface stormflow than that of Model_1. Also, the observed dynamics are better represented by Model_3 compared to Model_1. Even so, it is unlikely that the simulated pattern precisely fits patterns observed at Panola, because it remains difficult to relate observations to model entities.

We realize that our approach is not the only way in which nonlinear processes can be incorporated into a hydrological model. However, we believe that the development and subsequent testing of various nonlinear models will lead to an inspiring confrontation of model results with preconceptions about hydrological mechanisms. As such, these models can help us explore new ways of thinking about hillslope hydrology, which will ultimately lead to improved understanding of the subsurface stormflow process.

Acknowledgements. This study was conducted within the Virtual Laboratory for e-Science project (http://www.vl-e.nl) which is supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science and the ICT innovation program of the Dutch Ministry of Economic Affairs. The Panola data was kindly provided by Ilja Tromp-van Meerveld. We thank Jake Peters for the continued support of the Panola research site. The authors wish to thank Hanne Laine for reviewing an earlier draft of this manuscript. Chris Graham, Luisa Hopp and Takahiro Sayama are thanked for the useful discussions along the way.

References

Beven, K. J.: Rainfall-runoff modelling – the primer, John Wiley and Sons Ltd., Chichester, 2001. 5209
Horton, R. E.: The rôle of infiltration in the hydrological cycle, in: Transactions of the American Geophysical Union, 14, 446–460, 1933. 5207
Noguchi, S., Tsuboyama, Y., Sidle, R. C., and Hosoda, I.: Morphological characteristics of macropores and the distribution of preferential flow pathways in a forested slope segment,


### Table 1. Model configuration for virtual experiments 1, 2 and 3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{\text{start}}$</td>
<td>Start date and time of simulation</td>
<td>6 February, 2002 00:00</td>
<td>N/A</td>
</tr>
<tr>
<td>$t_{\text{end}}$</td>
<td>End date and time of simulation</td>
<td>9 February, 2002 23:45</td>
<td>N/A</td>
</tr>
<tr>
<td>$P_{\text{total}}$</td>
<td>Total precipitation depth</td>
<td>58.9</td>
<td>mm</td>
</tr>
<tr>
<td>$I_{\text{max}}$</td>
<td>Maximum precipitation intensity</td>
<td>7.11</td>
<td>mm hr$^{-1}$</td>
</tr>
<tr>
<td>$\nabla z$</td>
<td>Spatial gradient of the impeding layer</td>
<td>0.272</td>
<td>–</td>
</tr>
<tr>
<td>$\theta_{\text{sat}} = \phi$</td>
<td>Effective porosity</td>
<td>0.500</td>
<td>–</td>
</tr>
<tr>
<td>$\theta_{\text{fc}}$</td>
<td>Water content at field capacity</td>
<td>0.150</td>
<td>–</td>
</tr>
<tr>
<td>$\theta_{\text{act}, \text{init}}$</td>
<td>Initial actual volumetric water content</td>
<td>0.135</td>
<td>–</td>
</tr>
<tr>
<td>$d$</td>
<td>Soil depth</td>
<td>0.628</td>
<td>m</td>
</tr>
<tr>
<td>$K_{\text{sat}}$</td>
<td>Hydraulic conductivity for soil water transport between spatial elements</td>
<td>287 $\left(8.33 \times 10^{-5}\right)$</td>
<td>mm hr$^{-1}$</td>
</tr>
<tr>
<td>$K_f$</td>
<td>Factor controlling efficiency of flow to the bedrock sink.</td>
<td>$2.83 \times 10^{-2}$ $\left(4.17 \times 10^{-4}, 8.33 \times 10^{-2}\right)$</td>
<td>hr$^{-1}$</td>
</tr>
<tr>
<td>$T_{\text{wet}}$</td>
<td>Duration of the travel time distribution under wet conditions</td>
<td>0.851 $\left(0.00, 6.43\right)$</td>
<td>hr</td>
</tr>
<tr>
<td>$T_{\text{dry}}$</td>
<td>Duration of the travel time distribution under dry conditions</td>
<td>26.3 $\left(2.41, 965\right)$</td>
<td>hr</td>
</tr>
<tr>
<td>$WT$</td>
<td>Wetness threshold at which a spatial element switches from one travel time distribution to another</td>
<td>9.92 $\left(1.00, 15.0\right)$</td>
<td>m$^3$</td>
</tr>
</tbody>
</table>

* Parameter space search limits in brackets.
Table 2. Typically observed dynamics and to what extent they are captured by the model formulations presented in this study.

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>Model₁</th>
<th>Lumped Model₂</th>
<th>Model₃</th>
<th>transect Model₁</th>
<th>transect Model₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connectivity-threshold dominated flow regime</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Steep increase in discharge after threshold is exceeded</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Break in slope on the recession limb</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Subsurface topography driven convergence of subsurface stormflow</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Transient saturation occurring first in areas of shallow soil, extending in the downslope direction</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Transient saturation as a prerequisite for lateral subsurface flow</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Non-steady state groundwater table dynamics</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 3. Model configuration for virtual experiment 4.

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>value</th>
<th>units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{\text{start}}$</td>
<td>Start date and time of simulation</td>
<td>6 February, 2002 00:00</td>
<td>N/A</td>
</tr>
<tr>
<td>$t_{\text{end}}$</td>
<td>End date and time of simulation</td>
<td>9 February, 2002 23:45</td>
<td>N/A</td>
</tr>
<tr>
<td>$t_{\text{end}}$</td>
<td>End date and time of simulation</td>
<td>12 February, 2002 23:45</td>
<td>N/A</td>
</tr>
<tr>
<td>$P_{\text{total}}$</td>
<td>Total precipitation depth</td>
<td>58.9</td>
<td>mm</td>
</tr>
<tr>
<td>$P_{\text{total}}$</td>
<td>Total precipitation depth</td>
<td>58.9 + 12.0</td>
<td>mm</td>
</tr>
<tr>
<td>$I_{\text{max}}$</td>
<td>Maximum precipitation intensity</td>
<td>7.11</td>
<td>mm/hr</td>
</tr>
<tr>
<td>$V_{\text{z}}$</td>
<td>Spatial gradient of the impeding layer</td>
<td>0.155–0.390</td>
<td>–</td>
</tr>
<tr>
<td>$\theta_{\text{sat}} = \Phi$</td>
<td>Effective porosity</td>
<td>0.500</td>
<td>–</td>
</tr>
<tr>
<td>$\theta_{\text{fc}}$</td>
<td>Water content at field capacity</td>
<td>0.150</td>
<td>–</td>
</tr>
<tr>
<td>$\theta_{\text{act,init}}$</td>
<td>Initial actual volumetric water content</td>
<td>0.135</td>
<td>–</td>
</tr>
<tr>
<td>$d$</td>
<td>Soil depth</td>
<td>0.180–1.39</td>
<td>m</td>
</tr>
</tbody>
</table>

Virtual experiment 4, model element type 1

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>value</th>
<th>units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{\text{sat}}$</td>
<td>Hydraulic conductivity for soil water transport between spatial elements</td>
<td>9.43 ($4.17 \times 10^{-2}, 417$)</td>
<td>mm/hr</td>
</tr>
<tr>
<td>$K_{f}$</td>
<td>Factor controlling efficiency of flow to the bedrock sink.</td>
<td>2.67 ($4.17 \times 10^{-5}, 0.417$)</td>
<td>hr^-1</td>
</tr>
</tbody>
</table>

Virtual experiment 4, model element type 3

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>value</th>
<th>units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{\text{sat}}$</td>
<td>Hydraulic conductivity for soil water transport between spatial elements</td>
<td>350 ($10.2, 1.08 \times 10^3$)</td>
<td>mm/hr</td>
</tr>
<tr>
<td>$K_{f}$</td>
<td>Factor controlling efficiency of flow to the bedrock sink.</td>
<td>3.25 x 10^-2 ($1.44 \times 10^{-3}, 1.44$)</td>
<td>hr^-1</td>
</tr>
<tr>
<td>$T_{\text{dry}}$</td>
<td>Duration of the travel time distribution under dry conditions</td>
<td>25.1 (0.00, 184)</td>
<td>hr</td>
</tr>
<tr>
<td>$T_{\text{wet}}$</td>
<td>Duration of the travel time distribution under wet conditions</td>
<td>0.827 (0.00, 27.6)</td>
<td>hr</td>
</tr>
<tr>
<td>$WT$</td>
<td>Wetness threshold at which a spatial element switches from one travel time distribution to another</td>
<td>4.91 x 10^-2 (0.00, 1.00)</td>
<td>m^3</td>
</tr>
</tbody>
</table>

Parameter space search limits in brackets.
* See Fig. 6 for the spatial distribution of soil depth.
+ Calibration event.
◊ Evaluation event.
Fig. 1. General model structure.
Fig. 2. Virtual experiment 1. (A) Observed precipitation. (B) Discharge observations, best simulation, and Bayesian probability interval for Model\textsubscript{1}. (C) Sector identifier.
Fig. 3. Top views of grid cells showing the hydrologically impeding layer with saturated zone (shaded area) and unsaturated zone (white) for Model₁, Model₂ and Model₃. Dashed arrows in plots (B), (C) and (D) represent transport through the unsaturated zone. Note the relative abundance of short arrows in plot (D), compared to plot (C).
Fig. 4. Virtual experiment 2. (A) Observed precipitation. (B) Discharge observations, best simulation, and Bayesian probability interval for Model₂. (C) Sector identifier.
Fig. 5. Virtual experiment 3. (A) Observed precipitation. (B) Discharge observations, best simulation, and Bayesian probability interval for Model$_3$. (C) Sector identifier.
Fig. 6. Virtual experiment 4. Upper-left back pane, black: observed and artificial precipitation intensity (0.05 x). Upper-right back pane, yellow: hillslope soil depth (0.20 x). Bottom pane, blue: distributed flow curves of subsurface stormflow along the hillslope transect, as simulated by model elements of type (A) Model$_1$, and (B) Model$_3$. Green bars coming from x=0 represent the time to first flow. Magenta dots are observations of hillslope-scale discharge as used in the calibration.