This file was created by scanning the printed publication. Text errors identified by the software have been corrected: however some errors may remain.

Journal of Hydrology 393 (2010) 77–93

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Hillslope threshold response to rainfall: (2) Development and use of a macroscale model

Chris B. Graham [a,*], Jeffrey J. McDonnell [b,c]

[a] Department of Crop and Soil Sciences, The Pennsylvania State University, University Park, PA, USA
[b] Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR, USA
[c] University of Aberdeen, School of Geosciences, Aberdeen, Scotland, UK

ARTICLE INFO

Keywords:
Preferential flow
Hillslope hydrology
Numeric models
Model calibration
Virtual experiments

SUMMARY

Hillslope hydrological response to precipitation is extremely complex and poorly modeled. One possible approach for reducing the complexity of hillslope response and its mathematical parameterization is to look for macroscale hydrological behavior. Hillslope threshold response to storm precipitation is one such macroscale behavior observed at field sites across the globe. Nevertheless, the relative controls on the precipitation–discharge threshold poorly known. This paper presents a combined model development, calibration and testing experiment study to investigate the primary controls on the observed precipitation–discharge threshold relationship. We focus on the dominant hydrological processes revealed in part one of this two-part paper and with our new numerical model, replicate the threshold response seen in the discharge record and other hydrometric and tracer data available at the site. We then present a series of virtual experiments designed to probe the controls on the threshold response. We show that the threshold behavior is due to a combination of environmental (storm spacing and potential evapotranspiration) and geologic (bedrock permeability and bedrock topography) factors. The predicted precipitation–discharge threshold subsumes the complexity of plot-scale soil water response. We then demonstrate its use for prediction of whole-catchment storm discharge at other first order catchments at Maimai and the HJ Andrews Experimental Forest in Oregon.

© 2010 Elsevier B.V. All rights reserved.

Introduction

Hillslope hydrology still lacks the compact organization of empirical data and observations of hydrological response to precipitation events that might facilitate extrapolation to and prediction of hillslope behavior in different places. Hillslope hydrology models based on our current small scale theories emphasize the explicit resolution of more and more of the unknown and unknowable heterogeneities of landscape properties and the resulting process complexities (McDonnell et al., 2007). While the utility of a search for macroscale laws was enunciated over 20 years ago (Dooge, 1986), few studies have been able to even observe macroscale behavior given the enormous logistical challenge for characterizing whole-hillslope response. The heterogeneity in hillslope soil, bedrock, and topographic conditions and complexity of the spatial and temporal rainfall and throughfall input are still extraordinarily difficult to quantify and include in macroscale descriptions of hillslope and catchment behavior.

Graham et al. (this issue) presented a new macroscale perceptual model of subsurface flow processes at the well studied Maimai experimental watershed (McGlynn et al., 2002). This work was based on whole-hillslope excavation of subsurface flow paths and detailed hillslope scale irrigation aimed at identifying the dominant subsurface flow pathways and the role of bedrock topography and bedrock permeability on hillslope scale hydrological processes. The complexities of hillslope response and heterogeneity of the hillslope site at Maimai could be summarized by three key process statements: (1) A connected preferential flow network located at the soil/bedrock interface dominates lateral water and solute transport (with very high flow and transport velocities ranging from 6 to 21 m/h). (2) The bedrock surface controls the subsurface flow routing (where the filling of small depressions along the bedrock surface results in threshold lateral subsurface flow). (3) Vertical loss to the permeable bedrock is large (up to 35% of the precipitation input) delaying lateral flow initiation and reducing lateral flow volumes.

Here we take the perceptual model of hillslope behavior developed by Graham et al. (this issue) and apply the dominant processes modeling concept of Grayson and Blöschl (2000) to construct, test and use a macroscale rainfall–runoff model for the Maimai hillslope. Within the dominant processes philosophy, only

* Corresponding author. Address: The Pennsylvania State University, Department of Crop and Soil Sciences, 116 Agricultural Sciences and Industry Building, University Park, PA 16802-3504, USA. Tel.: +1 814 867 3074; fax: +1 814 863 7043.
E-mail address: cbg12@psu.edu (C.B. Graham).

0022-1694/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2010.03.008

the dominant flow processes, in this case the three listed above are incorporated into model structure. This philosophy is motivated by the difficulty in identifying and quantifying the myriad complex and heterogeneous hydrological processes at a given site. Our dominant processes approach is also motivated by the finding that only a small number of processes may dominate lateral subsurface flow and transport at the hillslope scale. We translate the hydrological processes identified at Maimai into a simple, low dimensional conceptual mathematical model. This follows similar model development work at Maimai and elsewhere (Seibert and McDonnell, 2002; Son and Sivapalan, 2007; Weiler and McDonnell, 2007). In this way the experimentalist works directly with the modeler, both in the experimental design to determine the dominant flow processes, and in model design to accurately implement the experimental findings.

We evaluate our new model using a multiple objective criteria framework (Gupta et al., 1998) incorporating extensive hydrometric and tracer data available from at Maimai site. We then use this new model as a learning tool to shed new light on whole-hillslope threshold responses to storm rainfall. Analysis of long term data records of flow at several field sites around the world has shown that such hillslope threshold response (i.e. the precipitation threshold before significant lateral subsurface flow is initiated) is a fundamental constitutive relation in hydrology (Buttle et al., 2004; Mosley, 1979; Peters et al., 1995; Weiler et al., 2006; Whipkey, 1965). While this threshold behavior is a potential macroscale descriptor of hillslope response to storm precipitation, the dominant controls on the magnitude of the threshold are not well known. While catchment geologic factors (e.g. soil depth, bedrock permeability, etc. (Tromp-van Meerveld and McDonnell, 2006b; Uchida et al., 2005)) and catchment environmental factors (e.g. antecedent moisture conditions (Tani, 1997; Tromp-van Meerveld and McDonnell, 2006a)) have been proposed as possible controls, the relative importance of each remains unclear and unresolved.

Two specific hypotheses have been previously proposed to explain the threshold relationship between rainfall and resulting subsurface stormflow: (1) fill and spill, and (2) pre-storm soil moisture deficit. In the fill and spill hypothesis, subsurface storage at the base of the soil profile must be filled (often in saturated patches) to connect the upslope areas with the base of the hillslope (Spence and Woo, 2002; Tromp-van Meerveld and McDonnell, 2006b). Accordingly, the permeability of the bedrock and the volume of subsurface storage that must be filled are the primary controls on the initiation of lateral subsurface flow. Alternatively, the pre-storm soil moisture deficit hypothesis (Tani, 1997; Tromp-van Meerveld and McDonnell, 2006b) suggests that filling of the moisture deficit in the soil profile is a prerequisite for lateral subsurface flow. This hypothesis is supported by an apparent change in the threshold under different antecedent moisture conditions. While both factors may operate in concert with one another, the relative influence of fill and spill and soil moisture deficit factors on the threshold response to precipitation has not been tested to date—largely because of the extremely small sample size of experimental hillslopes and limited range of climate and geology conditions explored to date.

Here we develop and then use our new model to test alternative hypotheses of controls on the threshold response to precipitation for diverse climate and geology. We use our model as a learning tool to explore how subsurface processes represented in our model structure may link to those properties that can be extracted from a long terms data record, such as the threshold for initiation of storm runoff, and the relationship between the excess precipitation and runoff. The new understanding of the controls of the threshold relationship is then tested on a number of different first order catchments at Maimai and at the HJ Andrews Experimental Forest in Oregon, USA. We use readily available data for storm spacing, evaporative demands and storm size extracted from the long term record at these sites to demonstrate how the complexity of catchment response to precipitation can be collapsed to the threshold metric to allow for simple macroscale model prediction of catchment discharge.

Study site and model development

Site physical and process description

We use the experimental work of Graham et al. (this issue) at the Maimai Experimental Catchments as the basis for model development and the virtual experiments aimed at understanding the controls on thresholds. The Maimai Experimental Catchments, South Island, New Zealand, have been a site of continuing hydrological research for over 30 years (see review in McGlynn et al. (2002)). While isotopic work has shown that the majority of hillslope discharge and streamflow at Maimai is pre-event water stored for weeks to months (McDonnell, 1990; Mosley, 1979; Pearce et al., 1986; Sklash et al., 1986), tracer experiments have demonstrated the ability of the hillslopes to rapidly transmit quantities of applied water at high velocities over long distances (Brammer, 1996; Mosley, 1979, 1982). Graham et al. (this issue) showed that lateral preferential flow is confined to the soil bedrock interface where flow velocities are very high (up to 21 m/h), routed by the bedrock topography. Filling subsurface storage in topographic pools upon the bedrock surface is a prerequisite for downslope connection and significant lateral subsurface flow. Once storage is filled, preferential flow paths seen on the bedrock surface have been shown to be connected upslope for distances up to 8 m, and appear to be stationary in time and space (Graham et al., this issue). The bedrock, while previously considered effectively impermeable (McDonnell, 1990; Mosley, 1979), was shown to be semi-perveous, with bedrock hydraulic conductivity on the order of 1–3 mm/h, leading to the potential of substantial fluxes of water and nutrients through the bedrock (Graham et al., this issue). Overland flow has not been observed at this site except in limited areas near the stream channel. Vertical preferential flow from the soil surface to depth during rainfall events has been hypothesized to occur in vertical cracks seen throughout the catchment dissecting the soil profile (Graham et al., this issue; McDonnell, 1990). Mixing of old and new water is thought to occur in both the soil column as well as in transient groundwater that forms at the soil bedrock interface, leading low amounts of new water observed in trench discharge and streamflow (Pearce et al., 1986; Sklash et al., 1986).

Description of the numerical model

The numerical model (called MaiModel) was built to incorporate the dominant processes that control subsurface flow at the Maimai hillslope as described by Graham et al. (this issue). Key components of MaiModel are

- Preferential flow pathways are connected, and located at the soil bedrock interface.
- Lateral subsurface travel velocities are high.
- Subsurface storage on the bedrock surface is explicitly designated.
- The bedrock is permeable.

In general terms, MaiModel consists of two reservoir types, soil storage and bedrock pool storage, which are fully distributed across the model domain (Fig. 1). Two bulk reservoirs are included for system losses of evapotranspiration and bedrock leakage. Water is transmitted vertically from the soil surface, through the

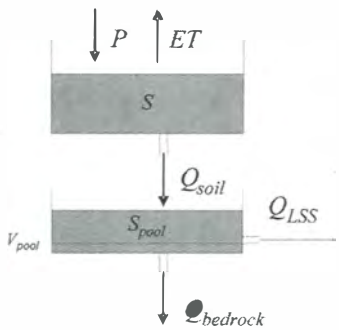

Fig. 1. Schematic of model structure. Fluxes include precipitation (P), evapotranspiration (ET), vertical percolation through the soil profile (Q_{soil}), leakage into the bedrock ($Q_{bedrock}$) and lateral subsurface flow along the bedrock surface (Q_{LSS}). Q_{LSS} is a function of the volume of water in the cell (S_{pool}) and the subsurface storage volume (V_{pool}). Q_{soil} is a function of the soil saturation (S), while $Q_{bedrock}$ is a function of S_{pool}.

soil storage reservoir to the subsurface storage reservoir, with no lateral communication between adjacent soil reservoirs. Lateral subsurface flow is restricted to flow along the bedrock surface among the bedrock pools, consistent with the current experimental evidence of water routing at the soil bedrock interface. Bedrock leakage is driven by water table height, and there is no reemergence of water at the hillslope once it has percolated into the bedrock. Evapotranspiration is driven by the potential evaporation rate, and limited by soil moisture.

Model structure

In MaiModel precipitation is split into canopy interception and soil reservoir recharge. A map of tree locations by Woods and Rowe (unpublished data, 1996) was used to identify areas of interception. Interception rates are based on published values from a similar aged radiata pine (*Pinus radiata*) forest, showing an interception rate of 38% by the canopy (Putuhena and Cordery, 2000). Interception was confined to areas of crown cover, which were estimated as the area within 3 m of the tree stems.

Throughfall enters each soil reservoir and is fully mixed with pre-event soil moisture, following Weiler and McDonnell (2007) and Vaché and McDonnell (2006). Assuming a unit head gradient, vertical drainage to the subsurface storage reservoir (Q_{soil}) [$L^3\ T^{-1}$] is equal to the soil relative hydraulic conductivity ($k(\theta)$), using the formula (Brooks and Corey, 1964):

$$Q_{soil} = k(\theta)A = k_{soil}S^{\beta}A \tag{1}$$

where k_{soil} is the saturated hydraulic conductivity [$L\ T^{-1}$], A the grid cell area [L^2], β a texture dependent exponent [dimensionless], and S is water saturation [dimensionless]:

$$S = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{2}$$

where θ is profile average water content [$L^3\ L^{-3}$], and θ_r and θ_s are the residual and saturated water content, respectively [$L^3\ L^{-3}$]. $Q_{soil} = 0$ if θ is less than the residual water content. Water drains vertically from the soil reservoirs to subsurface storage reservoirs, and does not drain downslope into adjacent soil reservoirs.

Evapotranspiration (ET) is modeled as a boundary flux, rather than as root uptake, and is restricted to the soil reservoir. ET is a function of both the reservoir water storage and the potential evaporation rate (PET) (Seibert, 1997):

$$ET = PET(tod)SA \tag{3}$$

where PET [$L\ T^{-1}$] is a function of the time of day (tod) [T] and the daily average PET:

$$PET(tod) = PET_{daily} \sin\left(2\pi\left(tod - \frac{2}{24}\right)\right) \tag{4}$$

We assume that PET peaks at 14:00 h and reaches a minimum at 02:00 h of each day, and PET is removed evenly throughout the soil profile.

Water drains vertically from the soil reservoir to subsurface storage reservoirs representing topographic pools on the bedrock surface. Water in the subsurface storage reservoir will either drain into the bedrock or be routed downslope along the bedrock to adjacent subsurface storage reservoirs, following the fill and spill hypothesis of Tromp-van Meerveld and McDonnell (2006b). All subsurface storage reservoirs are connected, and flow routing is based on topography, with lateral subsurface flow (Q_{LSS}) split between the (up to) eight adjacent downslope reservoirs, weighted by downslope gradient following the M8 flow routing algorithm (Quinn et al., 1991).

Lateral subsurface flow (Q_{LSS}) is governed using the Dupuit–Forchheimer assumption for a sloping aquifer (Freeze and Cherry, 1979), assuming that the aquifer surface is parallel to the bedrock surface. Q_{LSS} is a function of the exchange length (in this case equivalent to grid reservoir width (w) [L]), local slope (s) [$L\ L^{-1}$], the lateral hydraulic conductivity k_{LSS} [$L\ T^{-1}$], the volume of stored water (S_{pool}) [L^3] greater than the subsurface storage volume (V_{pool}) [L^3]:

$$Q_{LSS} = \frac{wsk_{LSS}(S_{pool} - V_{pool})}{A} \tag{5}$$

Q_{LSS} is defined as zero when $S_{pool} > V_{pool}$. The term $\frac{S_{pool} - V_{pool}}{A}$ is equivalent to the water table height above the threshold of subsurface storage.

As per field measurement of infiltration into the bedrock at the site (Graham et al., this issue), bedrock leakage is a function of water table height and a bedrock leakage coefficient ($c_{bedrock}$) [T^{-1}]:

$$Q_{bedrock} = c_{bedrock}S_{pool} \tag{6}$$

For the lateral and bedrock leakage flux rates, it is important to note that in the model formulation the subsurface storage reservoirs are perfect boxes, and the pool volume is equal to the grid cell cross sectional area multiplied by the water table height. In this way, Eq. (7) is equivalent to:

$$Q_{bedrock} = c_{bedrock}h_{pool}A \tag{7}$$

where h [L_{pool}] is the water table height.

The modeled tracer is assumed to be instantaneously fully mixed in the injection soil reservoirs, and were added as a line source in all cells 35 m upslope of the hillslope base. Tracer fluxes are limited to advective transport vertically between soil and subsurface storage reservoirs, and laterally between subsurface storage reservoirs. The full mixing assumption of the soil reservoirs also holds for the subsurface storage reservoirs as the advected tracer and water move downslope. While tracer can percolate into the bedrock, no tracer is lost to evapotranspiration. While the lack of a specific diffusion/microscale dispersion component to the tracer is a simplification, we believe that at the spatial and temporal scale of the model (hillslope, event based) these processes have a relatively minor impact of solute transport. A similar simplification is of full mixing of tracer in the soil and subsurface storage reservoirs. Again, this simplification was deemed necessary to retain the simplicity of the model and low number of parameters. Subsequently, caution must be used when applying this model to reactive transport.

The grid size of the model is flexible, in this case discretized using a 1 m^2 elements, with the soil depth prescribed by the user. Currently, a fixed time step is used in the model code, in this case 2 min. The space and time discretization are user inputs, and both

were determined via sensitivity analyses before the calibration was begun.

Multi-criteria model calibration and uncertainty analysis

We capitalize on two extensive data sets from field campaigns at the site for model parameterization and calibration. Woods and Rowe (1996) built a 1 m grid Digital Elevation Model (DEM) based on 755 survey points over an area of 2830 m^2, which was used in the model for flow routing. In addition, the interception module is based on a map of tree locations by Woods and Rowe (unpublished data, 1996). For 65 days beginning March 10, 1995 (Brammer, 1996), monitored hillslope discharge at the trench system built by Woods and Rowe (1996). A larger section of hillslope is modeled than drains into the collection trench, to minimize edge effects. In addition to monitoring hillslope discharge, Brammer (1996) added a Br^- tracer solution 35 m upslope of the trench as a 20 m wide line source injected directly into the soil profile 10 cm below the soil surface. Precipitation, hillslope discharge and tracer breakthrough were monitored at the trench for 45 days after tracer application. Rainfall and trench discharge were recorded in 10 min intervals, while tracer breakthrough at the trench was measured in grab samples during and between storms. Reanalysis of the Brammer (1996) tracer concentrations and trench discharge show that 15% of the tracer was recovered over 45 days, and the runoff ratio for the duration of the monitoring was 21%. The tracer breakthrough and trench hydrograph were used for model calibration. For more details about the hillslope gauging system, see Woods and Rowe (1996). For more details on the tracer injection, see McGlynn et al. (2002).

MaiModel was calibrated using Monte Carlo analysis with multiple criteria including the hydrometric and tracer breakthrough data. Using the 40 day Brammer (1996) hyetograph as input, 10,000 simulations were run with five model parameters varied in calibration: soil hydraulic conductivity (k_{soil}), bedrock leakage coefficient ($c_{bedrock}$), lateral hydraulic conductivity (k_{LSS}), active pore space (f_{active}, or the product of the soil depth and the active porosity (θ_s–θ_r), and the subsurface storage volume (V_{pool}). Changing the residual water content, saturation water contents and soil depth had the same impact on the active pore space, so one factor, θ_r, was chosen for calibration. The active pore space is presented as the variable in further analysis. Monte Carlo analyses were performed varying each parameter randomly across ranges of 0–1000% of field measurements or the physically possible range, to ensure that the entire parameter space was interrogated. Field parameter measurements and ranges used in the model calibration are presented in Table 1.

The second subset of model parameters was assigned to field measurements due to either parameter uncertainty or model insensitivity. A spatially detailed soil depth map was unavailable, so soil depth was set to the average soil depth (0.6 m; McGlynn et al., 2002) for the modeled domain. While modeling variations in soil depths has been shown to be important for prediction of hillslope dynamics at other field sites (e.g. Tromp van Meerveld and Weiler (2008), previous work at Maimai has indicated that the soil surface and bedrock surface topography are similar at the hillslope scale, and the soil depth is relatively uniform across the hillslope (Woods and Rowe, 1997). In preliminary calibration runs, MaiModel was found to be insensitive to the Brooks and Corey moisture release coefficient (β), so it was set at a value appropriate for the silt loam soil texture measured in the field (Carsel and Parrish, 1988). As mentioned above, porosity, residual and saturation water contents are interrelated with respect to model function, so residual water content was set as a variable during the calibration, and the porosity and saturation water content were set to field measured values (from McDonnell, 1990). No measurements of *PET* were available at the site for the time of record. At the nearby town of Reefton (10 km northwest of the hillslope), measured evaporation rate of 714 mm/year has been reported (Baker and Hawke, 2007). A *PET* value of 6 mm/day was chosen to result in a modeled actual evapotranspiration of 714 mm/year (1.95 mm/day).

Model performance with respect to the hillslope discharge hydrograph was assessed by the Nash Sutcliffe efficiency factor (E) (Nash and Sutcliffe, 1970):

$$E = 1 - \frac{\sum(Q_o - Q_m)^2}{\sum(Q_o - \overline{Q_o})^2} \tag{8}$$

where Q_o is observed discharge, Q_m is modeled discharge and $\overline{Q_o}$ is the mean observed discharge. An E value greater than 0 indicates the modeled results fit measured discharge better than the mean discharge. An E of 1.0 is a perfect fit. For calibration purposes, an E of over 0.8 was considered an acceptable fit. E calculations were made for six subsets of the time series, including the entire 40 days after tracer application, and for the five largest storms of the data record (storms B1–B5). Only parameter sets with acceptable E for both the 40 day record and each individual storm were considered behavioral.

Due to temporally irregular measurements of tracer breakthrough at the hillslope, model tracer breakthrough was compared on a storm by storm basis during the Brammer (1996) experiment timeframe. Both the spatial pattern of tracer breakthrough along the trench face and storm cumulative breakthrough were compared for each of the five recorded storms. Parameter sets with a correlation coefficient greater than 0.8 for both the spatial and temporal breakthrough comparisons were considered a good fit. Cumulative tracer breakthrough for the 40 day time series was also used as a model evaluation criterion. Due to uncertainties in tracer recovery, modeled tracer recovery of within 2.5% of measured values was deemed behavioral. After Monte Carlo calibration, the parameter sets deemed behavioral were analyzed, and the model run with the highest minimum storm E was chosen for additional virtual experiments.

Table 1
Calibration parameter ranges and sources for parameter ranges. Calibration parameters include bedrock leakage coefficient ($c_{bedrock}$), soil and bedrock hydraulic conductivity (k_{soil} and k_{LSS}), subsurface pool volume (V_{pool}) and active pore space (f_{active}) V_{pool} was not measured in field and was constrained by a pre-calibration sensitivity analysis.

Parameter	Range	Source
$c_{bedrock}$	0–0.0284 1/s	10,000% maximum observed in field (Graham et al., this issue)
k_{soil}	0–3 m/h	10,000% maximum observed in field (McDonnell, 1990)
k_{LSS}	0–30 m/h	Greater than range observed in field (Graham et al., this issue)
V_{pool}	0–0.01 m^3	Sensitivity analysis
f_{active}	0–45%	Spans range of field measured porosity (McDonnell, 1990)

Virtual experiment design

We ran all virtual experiments using a rainfall time series with multiple replications of storm B5 (April 26 through April 27, 1995) from the calibration hyetograph as input. Storm B5 was chosen for the virtual experiments due to its moderate size, variable intensity, and the relatively good model fit from calibration. This 50.6 mm storm had a duration of 24 h, average intensity of 2.1 mm/h, and maximum 10 min intensity of 30 mm/h occurring 70 min after the start of the storm (the peak 60 min intensity of 11.8 mm/h

occurred at 240 min). This storm was the largest of the five storms recorded after tracer application and exhibited the highest rainfall intensity. Measured discharge was 23.0 mm (runoff ratio = 44%), with peak discharge of 3.4 mm/h, 400 min after the start of the storm. The hyetograph and hydrograph were skewed to the left (skew = 3.4 and 2.9, respectively). Rainfall (87.2 and 87.4 mm) fell in the previous 7 and 14 days. Analysis of 2 years of precipitation records at the site (from Beven and Freer (2001)) indicates that this storm falls in the upper 25% and 5% of storms with respect to rainfall total precipitation and average intensity, respectively. The calibrated model had a Nash Sutcliffe efficiency of 0.95 for this storm.

To allow for the impact of antecedent moisture conditions in the virtual experiment hyetographs, storm B5 was replicated nine times, bracketed between 1 and 21 days antecedent drainage time before each replication. With the soil reservoirs initially saturated, the rainfall time series consisted of 10 days drainage followed by the B5 hyetograph. The B5 hyetograph was then repeated nine times with 1, 1, 3, 3, 5, 7, 7, 14, and 21 days drainage between storms (storms V1–V9). The model was run with the virtual experiment hyetograph 11 times, with the total storm precipitation scaled by a factor of 0.1, 0.2, 0.5, 0.75, 0.9, 1, 1.1, 1.5, 2, 5, and 10 (5–506 mm per event). Total storm precipitation was scaled by altering the duration of the storm to reach the desired total storm rainfall amount. For events smaller than the base case storm (51 mm), the hyetograph was truncated once the desired rainfall amount was reached. For storms larger than the base case storm, the storm time series was repeated until the desired precipitation amount was reached. Depending on the size (duration) of the virtual experiment events, the duration of the nine storm VE hyetograph ranged from 95 to 185 days. For each simulation, recorded output included water balance components (discharge, bedrock leakage, *ET* and soil moisture storage), and tracer fluxes.

The virtual experiment hyetograph was used with the parameter set from the best fit model calibration. The model was then run with each of the scaled hyetographs (between 5 and 506 mm precipitation per event). Total event precipitation and total storm discharge were calculated, binned by antecedent moisture and plotted. The threshold and excess precipitation/discharge slope for the calibrated model were determined based on all storms with a runoff ratio greater than 1% using a least squares regression.

For soil moisture deficit experiments, the potential evaporation rate was varied, and the drainage time between storms from the virtual experiment hyetograph was analyzed as the second variable. In 10 separate simulations for each of the 9 scaled precipitation hyetograph, the potential evaporation rate was scaled by a factor of 0, 0.1, 0.2, 0.5, 0.75, 1, 1.5, 2, 5, and 10, resulting in $9 * 10 = 90$ simulations. The process used to determine the threshold and slope for the calibrated model was then repeated using model runs with the varied antecedent drainage time and potential evaporation.

For the virtual experiments that focused on the effect of fill and spill factors, only storm V5, with 3 day drainage was used for analysis. Again, the scaled virtual experiment hyetographs were repeated while also scaling the two fill and spill parameters: bedrock permeability ($c_{bedrock}$) and subsurface storage volume (V_{pool}). Both bedrock permeability and subsurface storage volume were scaled by a factor of 0, 0.1, 0.2, 0.5, 0.75, 1, 1.5, 2, 5, and 10 (10 factors), and scaled concurrently, to create 100 simulations for each of 11 event sizes. $11 * 100 = 1100$ simulations were run covering the ranges of parameter values and event sizes. The process used to determine the threshold and slope for the calibrated model was then repeated using model runs with the scaled bedrock permeability and subsurface storage volumes.

A final set of virtual experiments tested the hypotheses that neither or both fill and spill or soil moisture deficit are the cause of the threshold. *PET*, bedrock leakage coefficient and subsurface storage volumes were set to zero, to determine if a third mechanism beyond "fill and spill" and "soil moisture deficit" could be responsible for the thresholds.

Results

Multi-criteria model calibration

The calibrated model reproduced both the hydrometric and tracer response to precipitation. The calibrated model fit the measured hydrograph well at both the 40 day and individual event time scale (Fig. 2). While the smaller events were generally overpredicted, the large events were well modeled. The hydrograph recessions were generally underpredicted, with the model exhibiting a faster recession than measured discharge. Peak discharge for each event was well represented. While observed patterns of soil moisture were unavailable for comparison, the spatial pattern of hillslope discharge was well correlated with observed patterns. Fig. 3 shows six representative periods of high soil moisture, in this case occurring at peak modeled discharge during storm B4 and in 1 h intervals after peak. Soil moisture and trench discharge, are

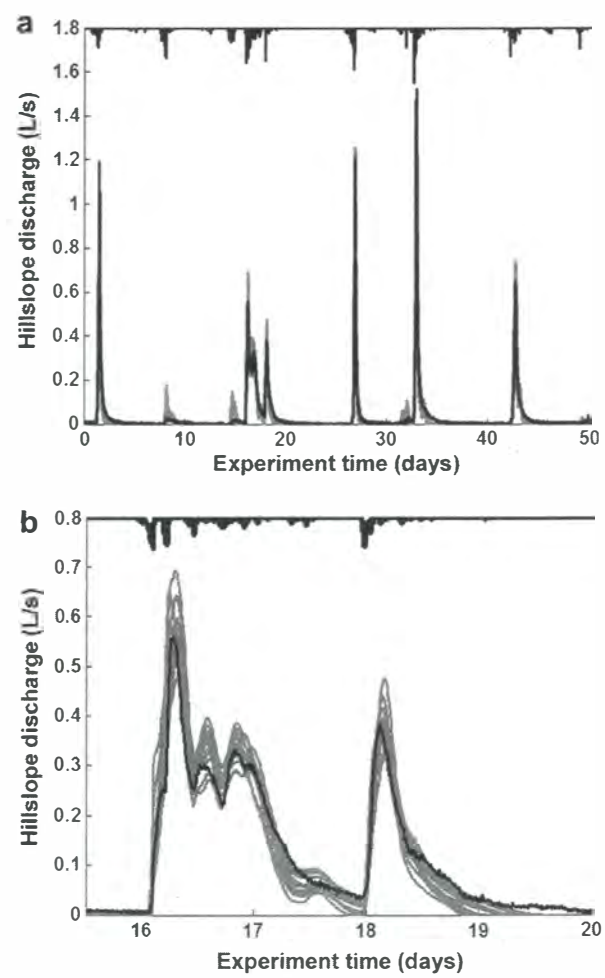

Fig. 2. (a) Measured (black) and modeled (grey) hydrograph (including storms B1–B5) with 13 simulations that matched all objective criteria. The small events were generally overpredicted, while peak discharge for the five events were well fit. For individual events, the rising limb was well modeled, while the modeled recession was generally steeper than the measured. (b) A close up of event B2 shows the range of modeled responses bracket the measured response. B2 was the most difficult to simulate, likely due to the complex double hydrograph.

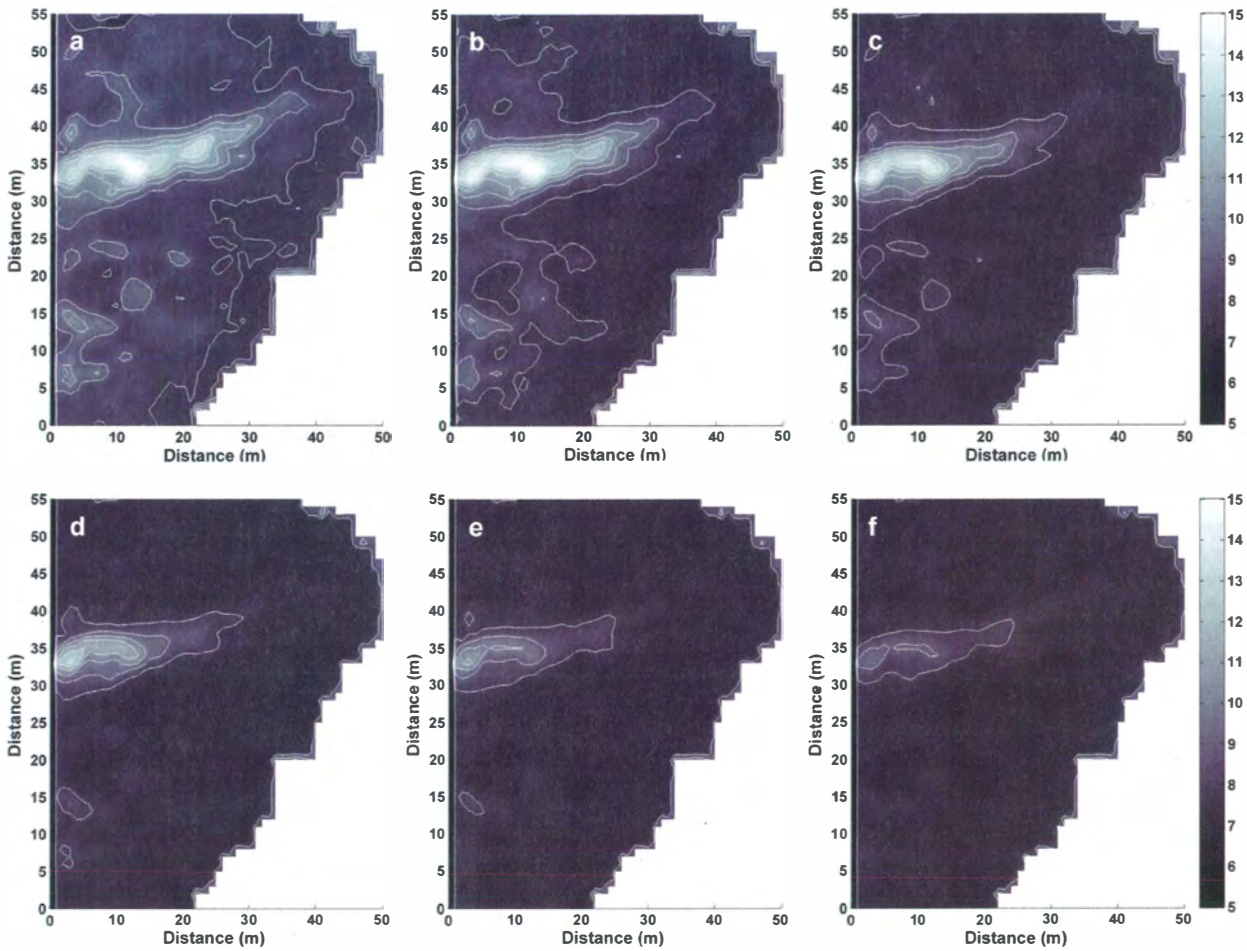

Fig. 3. Modeled spatial pattern of soil moisture (sum of soil and subsurface bedrock storage expressed in cm water). Pictured are peak modeled hillslope discharge (a) and 1 h time increments until 5 h after peak (b–f). Contour intervals are in 1 cm increments. Hillslope base is at left side of figure.

concentrated in the topographic hollow in the center of the modeled area of the hillslope.

Our 9 calibration objective criteria provided different levels of model discrimination (Table 2, Fig. 4). The hydrographs and cumulative tracer breakthrough were effective in reducing the number of behavioral parameter sets between 96% and 99%. The spatial and temporal tracer breakthroughs were not as effective at reducing the parameter range. Of the 10,000 model runs, 294 (2.9%) parameter sets led to acceptable model fits of trench discharge during the entire 40 day record (where an acceptable fit was an E greater than 0.8). The best fit parameter set had an E of 0.95. For storms B1, B3, B4 and B5, a similar number of model runs (parameter sets) were found to be acceptable (260–416 (2.6–4.2%)), with maximum E of 0.97–0.98. Fewer parameter sets led to acceptable model fits of storm B2 ($E > 0.8$ for 102 (1.0%) parameter sets), though a maximum E of 0.97 was obtained for one parameter set. 300 (3.0%) model runs had between 11.5% and 16.5% (±2.5% of measured) of tracer breakthrough during the flow time series. The experimental tracer temporal and spatial breakthrough were acceptably reproduced by a larger fraction of parameter sets; 1462 (14.6%) correctly modeled temporal tracer breakthrough, while 4827 (48.3%) correctly modeled the spatial breakthrough.

Of the 10,000 parameter sets, 13 (0.1%) met all nine objective criteria. Four criteria (storms 1, 2, and 3, and the cumulative tracer

Table 2
Parameter range reduction (in % of initial range) and number of behavioral parameter sets after calibration using each objective criterion (10,000 calibration simulations).

	$c_{bedrock}$	k_{soil}	k_{LSS}	V_{pool}	f_{active}	Criteria	Behavioral parameter sets
40 day discharge	58	2	51	45	62	$E > 0.8$	294
Storm 1 discharge	72	2	47	2	63	$E > 0.8$	260
Storm 2 discharge	75	6	56	83	68	$E > 0.8$	102
Storm 3 discharge	60	2	43	24	47	$E > 0.8$	416
Storm 4 discharge	56	2	47	18	58	$E > 0.8$	402
Storm 5 discharge	56	2	30	34	47	$E > 0.8$	342
Total cumulative tracer breakthrough	44	1	5	6	0	$11.5\% \leqslant T \leqslant 16.5\%$	4827
Tracer breakthrough temporal	0	1	1	1	1	$R^2 > 0.8$	1462
Tracer breakthrough spatial	0	0	1	0	0	$R^2 > 0.8$	300
All criteria	86	6	56	85	78		13

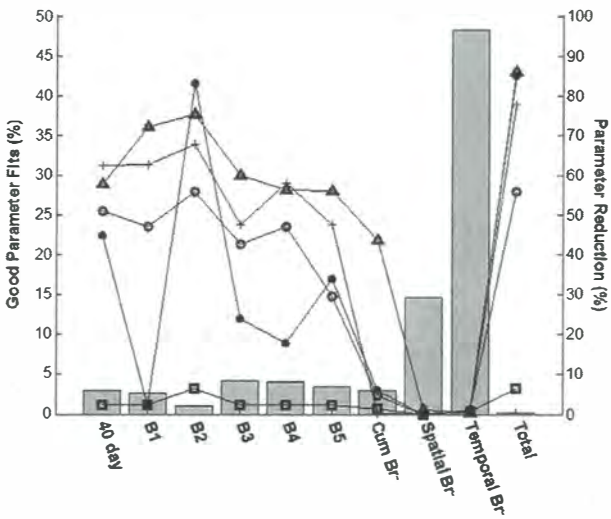

Fig. 4. Summary of MaiModel calibration. Acceptable parameter sets (bars) and the reduction in parameter uncertainty for each model criteria. Parameters: k_{soil} (□); k_{LSS} (○); V_{pool} (●); $c_{bedrock}$ (▲); f_{active} (+). Storm B2 had the lowest number of acceptable parameter sets, and the highest reduction in the parameter space for each variable. While the temporal tracer breakthrough was $c_{bedrock}$.

breakthrough) were sufficient to determine the final group of acceptable parameter sets. Of the 294 simulations that acceptably modeled the entire 40 day trench hydrograph, 39 (13%) acceptably modeled all five storms. No parameter sets had acceptable fit for all storms but did not correctly model the 40 day hydrograph. Of the 39 simulations with acceptable fit for all storms, 13 (33%) had modeled the cumulative tracer breakthrough adequately.

Examination of parameter uncertainty within the calibrated runs

Each objective criterion served to reduce the range of each model parameter (Table 2). To compare the reduction in the uncertainty of each parameter after each objective criterion, the parameter reduction ratio computed for each parameter (computed as 1 – the ratio of the range of parameter values in behavioral models vs. the range of the initial parameter distribution). A ratio of 0% for a given parameter and objective criterion indicates the criterion did not reduce the uncertainty in the parameter, while a ratio of 90% indicates the parameter is restricted to 10% of the initial range for behavioral models. These values are relative as the initial parameter ranges chosen were different for each parameter, generally exceeding the range observed in the field (k_{soil}, $c_{bedrock}$, k_{LSS}), the range of physically possible values (f_{active}), or the range possible determined through pre-experiment sensitivity analysis (V_{pool}). Nevertheless, they serve as a method to determine the relative strength of each objective criterion to reduce parameter uncertainty.

The amount of uncertainty reduced by each objective criterion varied for each parameter (Table 2, Fig. 4). For example, parameter sets that led to acceptably modeled tracer fluxes sampled from the entire initial range for drainable porosity. Parameter sets that acceptably modeled the 40 day and individual hydrographs, however, reduced the range of drainable porosity from 47% to 68%. Soil hydraulic conductivity, on the other hand, was insensitive to all of the objective criteria. While relatively few parameter sets resulted in acceptable cumulative tracer breakthrough (3.0%), this objective did little to reduce the uncertainty for any of the calibrated parameters except the bedrock leakage coefficient (43.6%). Using only the 13 parameter sets that fit all criteria, the parameter uncertainty was reduced (35.9–62.6%) for all parameters except the soil hydraulic conductivity (3.7%).

Of the 13 parameter sets deemed behavioral for all objective criteria, one was chosen as the base case for the virtual experiments. The 13 parameter sets that met all of the model evaluation criteria were then ranked according to their goodness of fit to each of the objective criteria. The parameter set with the highest minimum rank was chosen for the virtual experiments. For the best fit parameter set, model efficiency for the five storms ranged from 0.92 (storm B2) to 0.97 (storm B5), with a 40 day efficiency of 0.95. Modeled cumulative tracer breakthrough was within 2% of the measured value. Modeled spatial and temporal patterns of tracer breakthrough all fell in acceptable ranges. Best fit calibrated parameter values were close to field measured parameters. The drainable porosity was 0.1 m^3/m^3 (compared to 0.05 m^3/m^3 measured in the field (McDonnell, 1990)), bedrock leakage coefficient, 4.25E−5 1/s (compared to 2.63E−5 1/s (Graham et al., this issue)) and lateral hydraulic conductivity, 25.5 m/h (compared to 7.5–26 m/h based on measurements of tracer velocities between 6 and 21 m/h, porosity of 0.45 m^3/m^3 and gradient of 56% (Graham et al., this issue)). Soil hydraulic conductivity was calibrated to 2.67E−4 m/s, nearly an order of magnitude greater than observed values between 2.7E−6 and 8.3E−5 m/s (Mosley, 1979), likely due to the mixing of the preferential and matrix flow in the model structure. The average subsurface storage volume (calibrated to 1.7E−3 m^3, equivalent to 1.7 mm deep pools evenly over the bedrock surface) was not measured in the field.

Virtual experiments with the calibrated model

Calibrated model parameterization

Using the calibrated model parameterization, the 11 scaled virtual experiment hyetographs were applied, with storm total precipitation ranging from 5 to 506 mm (for a subset of model hydrographs, see Fig. 5). For the base case realizations of storm V5 (50.6 mm), with 3 days antecedent drainage (the average interstorm duration at Maimai), bedrock leakage made up the largest part of the water balance. Leakage accounted for 23.7 mm (36% of applied rainfall), with 18.2 mm (36%) as hillslope discharge and 15.9 mm (31%) as evapotranspiration. Soil storage served as a source for this simulated event, supplying 7.2 mm of the water lost through leakage, hillslope discharge and evapotranspiration.

As the storm size was varied in the scaled hyetographs, total discharge ranged from 0 to 272.7 mm with the runoff ratio ranging from 0% to 57% (Fig. 6). Trench flow was not observed for the two smallest events (5 and 10.1 mm rainfall), while the 25 mm storm yielded 3.6 mm of trench discharge (runoff ratio = 14%). Therefore a threshold for lateral subsurface stormflow appeared to exist between 10 and 25 mm for storms with 3 days antecedent drainage time. Total storm discharge increased linearly (R^2 = 0.999) after the threshold, with a slope of 0.45 mm discharge/mm precipitation. The calculated threshold, equal to the x axis intercept, was 17.7 mm precipitation. In the analysis below, the threshold refers to the x axis intercept (reported in mm rainfall), and the slope is the slope of the excess precipitation/discharge line (reported in mm discharge/mm precipitation).

Soil moisture deficit

For the application of the virtual experiment hyetograph using the calibrated model parameterization and the base case storm size (B5, 50.6 mm), total storm discharge was dependent on the antecedent drainage time. Total storm hillslope discharge decreased from 23.5 to 0.0 mm (runoff ratios decreased from 46% to 0%) for the storms with between 1 and 21 days of antecedent drainage (Table 3). For the simulated storm (V1) with the shortest antecedent drainage time, 1 day, the water balance was split

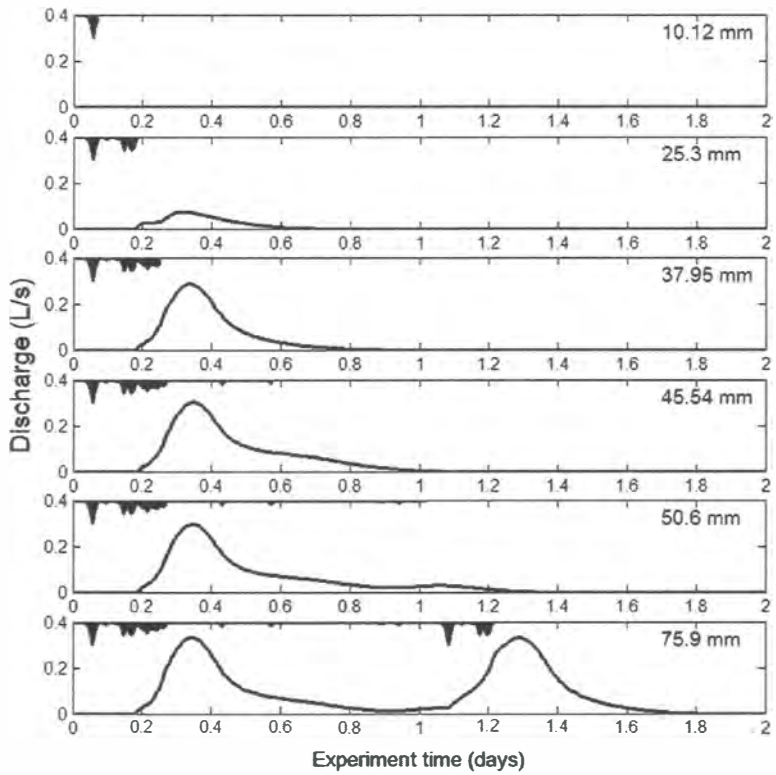

Fig. 5. Modeled hydrographs for virtual experiment. A series of increasing duration hydrographs are applied to MaiModel with calibrated parameters and 3 days antecedent drainage (storm V5). Events with between 10 and 76 mm rainfall are simulated with scaled realizations of the hyetograph from measured storm B5.

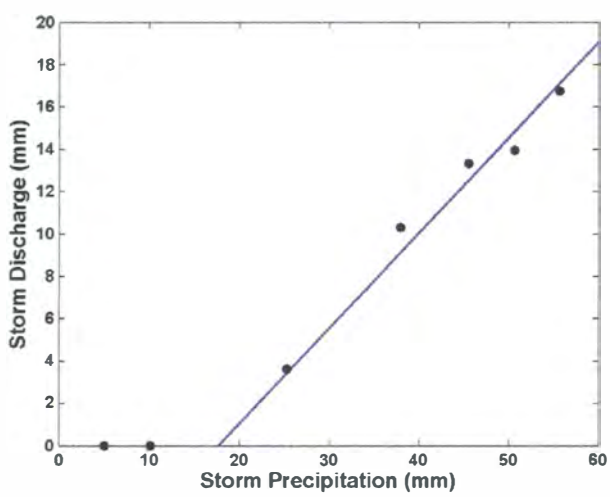

Fig. 6. Whole storm precipitation vs. discharge for modeled events, using calibrated parameters and 3 days antecedent drainage (storm V5). The estimated threshold is 18 mm, and slope is 0.45 mm/mm (points taken from column 5, Table 3).

between discharge (23.5 mm or 46%), bedrock leakage (26.4 mm or 52%), evaporation (10.9 mm or 22%). The soil storage reservoir acted as a source for the additional water for all simulations, as soil moisture storage decreased from event to event. For the storm with 21 days antecedent drainage time (V9), longer than any observed at Maimai in the 2 year data record, discharge was reduced to 0 mm (0%), and bedrock leakage to 5.2 mm (10%). Evaporation increased to 64.4 mm (129%), with the soil storage reservoir again acting as a source. For the storms with long antecedent drainage times, rainfall filled the soil storage deficit and was then lost to evaporation. For storms with shorter antecedent drainage times, soil moisture deficit was quickly filled and precipitation was routed to the bedrock surface and lost to hillslope discharge and bedrock leakage.

The rainfall threshold for producing subsurface stormflow for storm V5 (3 days antecedent drainage) was 17.7 mm (Fig. 6). Calculated thresholds for the other events (time between events) ranged from 9.1 (1 day antecedent drainage time) to 60.8 mm (21 days antecedent drainage time) (Fig. 7, Table 4). The threshold was linearly related to the time between storms of the form $P_o = 9.7$ mm + 2.5 mm/day ($R^2 = 0.984$). The slope did not depend on the time between events, varying from 0.56 to 0.57 mm/mm (Fig. 8, Table 4).

The 11 virtual experiment hyetographs were then run with the potential evaporation rate scaled between 10% and 1000% of the calibration value of 6 mm/day (0.6–60 mm/day). For the following virtual experiments we will focus on storm V5, with 3 days antecedent drainage, as similar patterns were seen for all events.

As the potential evaporation rate was increased, actual modeled evaporation increased, and pre-event soil moisture decreased. These losses were balanced by a decrease in both discharge and bedrock leakage, as less water was available to drain vertically to the subsurface storage and lateral flow pathways. Once the *PET* increased above 200% of the base case (>12 mm/h), evaporative losses from the soil profile were greater than the total storm precipitation, leading to a progressively depleted soil moisture profile. For the simulations with high potential evaporation rates, rainfall went towards filling soil storage and was subsequently lost to evapotranspiration. For simulations with lower potential evaporation rates, soil moisture deficit was quickly filled and precipitation was routed to the bedrock surface and removed from the system as hillslope discharge and bedrock leakage.

Table 3
Storm discharge and calculated threshold and excess precipitation/discharge slope for events V1–V10 using the calibrated model.

Event Precipitation (mm)	Antecedent drainage time (days)									
	10 (V1)	1 (V2)	1 (V3)	3 (V4)	3 (V5)	5 (V6)	7 (V7)	14 (V8)	21 (V9)	28 (V10)
5	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
25	0	5	8	4	4	1	0	0	0	0
38	3	15	15	11	10	8	5	0	0	0
46	5	18	18	13	13	10	8	1	0	0
51	6	18	18	14	14	11	8	1	0	0
56	9	21	21	17	17	14	11	3	1	0
76	17	31	31	26	26	23	20	12	6	3
101	29	41	41	37	37	34	31	23	16	10
253	98	110	110	106	106	103	100	91	85	78
506	212	225	225	220	220	218	215	206	200	193
Threshold	35	9	7	18	18	24	30	48	61	75
Slope	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45

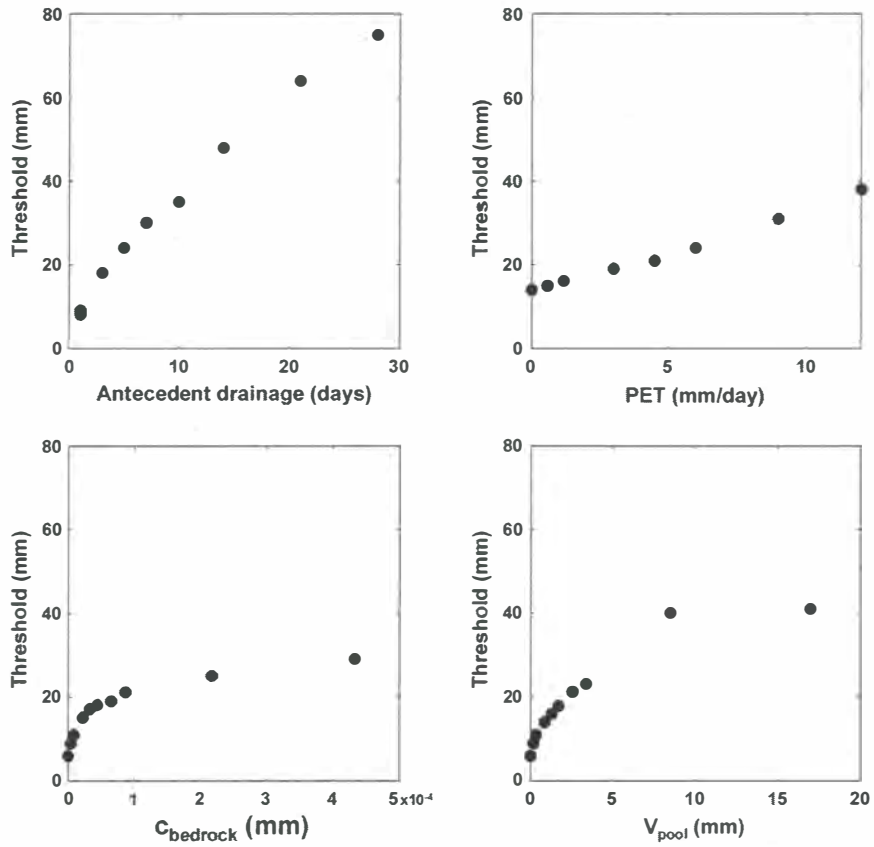

Fig. 7. Dependence of precipitation/discharge threshold on soil moisture deficit (antecedent drainage time and *PET*) and fill and spill (bedrock leakage coefficient and subsurface storage volume).

The precipitation/discharge threshold and the slope of the excess precipitation/discharge line for the soil moisture deficit virtual experiments were calculated in the same way as above, and only the summary results are presented here. The precipitation discharge threshold for the event with 3 days antecedent drainage time (V5) ranged from 11.8 mm (*PET* = 0 mm/day) to 104.8 mm (*PET* = 60 mm/day) (Figs. 7 and 9, Table 4). The slope of the excess precipitation discharge line decreased with increasing *PET*, ranging from 0.21 mm/mm (*PET* = 60 mm/day) to 0.62 mm/mm (*PET* = 0 mm/day) (Figs. 8 and 9, Table 4). The threshold for initiation of hillslope discharge was linearly correlated with the potential evaporation rate, of the form P_o = 11.4 mm + 1.1 mm/(mm/day) - *PET* (R^2 = 0.996). The slope of the excess precipitation discharge line (a) was also linearly correlated of the form a = 0.6279 mm/mm–0.007 (mm/mm)/(mm/day) ∗ *PET* (R^2 = 0.999).

The antecedent potential evapotranspiration (APET), the product of *PET* and the antecedent drainage (a measure of the total potential evaporative demand before the event), was linearly correlated with the observed threshold (P_o = 12.2 mm + 0.35 mm/mm ∗ APET) (R^2 = 0.875). The slope of the excess precipitation/discharge line was a linear function of only *PET* since the antecedent drainage had no effect.

Table 4
Threshold (in mm, upper table) and excess precipitation slope (in mm discharge/mm precipitation, lower table) for each antecedent drainage time and *PET*.

		PET (mm/day)								
		0	0.6	1.2	3	4.5	6	9	12	30
Antecedent drainage time (days)	10	17	18	20	25	30	35	46	57	135
	1	6	6	6	7	8	9	13	16	55
	1	6	6	6	7	7	8	9	11	35
	3	12	12	13	15	16	18	21	25	52
	3	12	12	13	15	16	18	21	25	56
	5	14	15	16	19	21	24	31	38	78
	7	16	17	18	22	26	30	39	48	101
	14	19	21	23	32	40	48	64	76	186
	21	20	23	28	41	54	64	80	101	206
	28	21	26	31	50	64	75	100	137	218
Antecedent drainage time (days)	10	0.5	0.48	0.48	0.47	0.46	0.45	0.43	0.42	0.33
	1	0.5	0.48	0.48	0.47	0.46	0.45	0.44	0.42	0.36
	1	0.5	0.48	0.48	0.47	0.46	0.45	0.43	0.42	0.34
	3	0.5	0.48	0.48	0.47	0.46	0.45	0.43	0.42	0.32
	3	0.5	0.48	0.48	0.47	0.46	0.45	0.43	0.42	0.32
	5	0.5	0.48	0.48	0.47	0.46	0.45	0.43	0.42	0.32
	7	0.5	0.48	0.48	0.47	0.46	0.45	0.43	0.42	0.31
	14	0.5	0.48	0.48	0.47	0.46	0.45	0.43	0.41	0.32
	21	0.5	0.48	0.48	0.47	0.46	0.45	0.43	0.4	0.31
	28	0.5	0.48	0.48	0.47	0.46	0.45	0.42	0.42	0.28

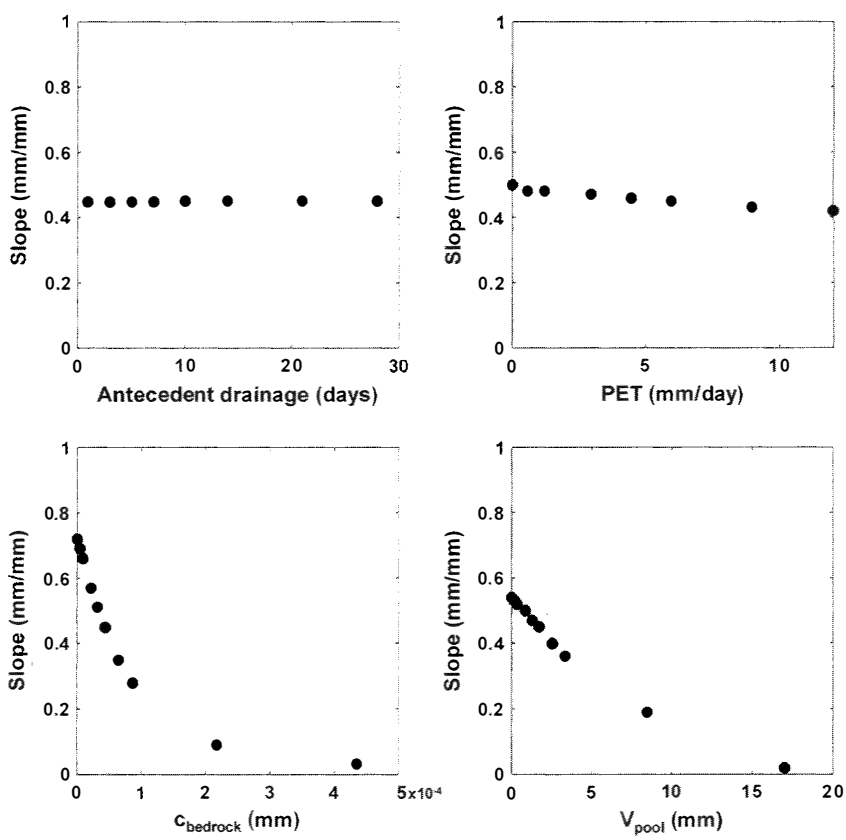

Fig. 8. Dependence of slope of excess precipitation/discharge on soil moisture deficit (antecedent drainage time and *PET*) and fill and spill (bedrock leakage coefficient and subsurface storage volume).

Fill and spill

To determine the influence of fill and spill factors on the precipitation discharge relationship threshold, the bedrock leakage coefficient and subsurface storage volume were scaled. Again, event V5, with 3 days antecedent drainage, was used for the analysis, though the entire 11 storm hyetograph was used. For simulations where the bedrock leakage coefficient was increased, bedrock leakage increased while hillslope discharge decreased. For simulations where the subsurface storage volumes increased, more water was held in the pools for longer periods, allowing for more leakage. Due to the physical disconnection between the soil profile and the subsurface storage on the bedrock surface in the model structure, changing the bedrock permeability and subsurface storage volume did not impact soil moisture storage or

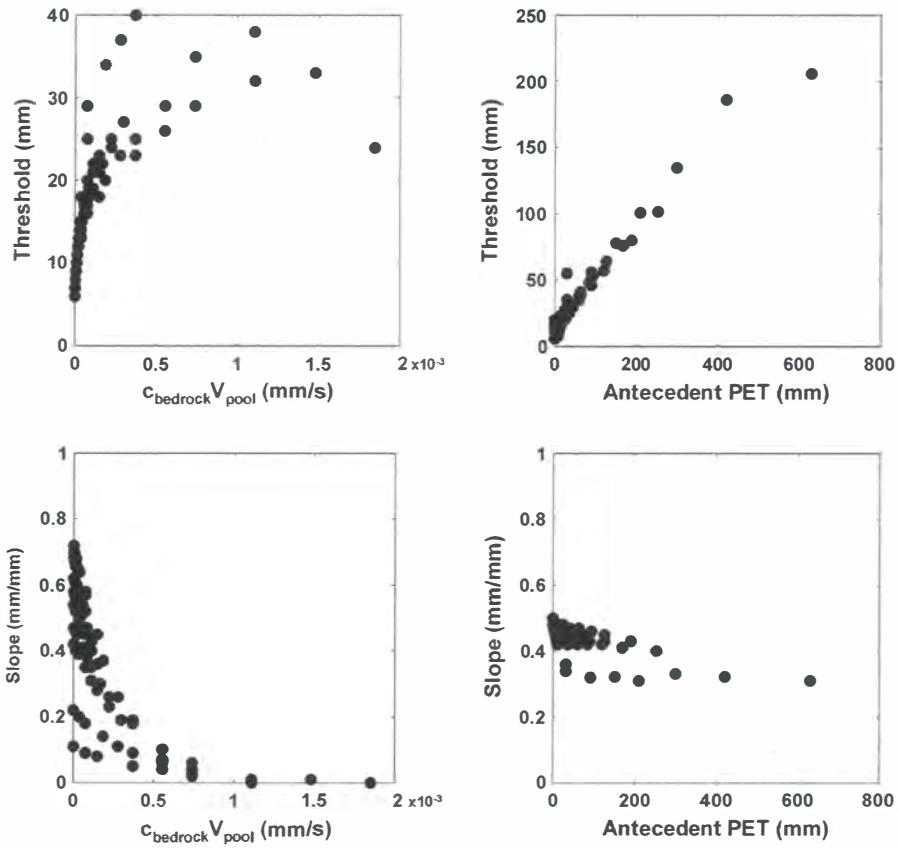

Fig. 9. Threshold and slope vs. products of fill and spill factors, and soil moisture deficit factors. Antecedent *PET* is *PET*˜ antecedent drainage time.

evaporation rates. Therefore, any increase in bedrock leakage flux was balanced by a reduction in hillslope discharge.

The threshold for flow increased nonlinearly with increased $c_{bedrock}$ and V_{pool}, while the slope of the excess precipitation/discharge line decreased linearly with both (Figs. 7 and 9, Table 5). The threshold (slope) varied from a maximum (minimum) of 59.2 mm (0.0002 mm/mm) for the highest values of bedrock leakage coefficient and subsurface storage volumes to a minimum (maximum) of 6.45 mm (0.91 mm/mm) for the simulations with no leakage or subsurface storage (Figs. 8 and 9, Table 5). For simulations with no subsurface storage, the threshold was 6.45 mm and the slope 0.70 mm/mm, while for simulations with no bedrock

Table 5
Threshold (in mm, upper table) and excess precipitation slope (in mm discharge/mm precipitation, lower table) for each V_{pool} and $c_{bedrock}$ combination. For four events with high V_{pool} and $c_{bedrock}$ the discharge was not non-zero for enough events to determine threshold and slope values.

		V_{pool} scaling factor									
		0	0.1	0.2	0.5	0.75	1	1.5	2	5	10
$c_{bedrock}$ scaling factor	0	6	6	6	6	6	6	6	6	6	6
	0.1	6	7	7	8	9	9	10	11	18	29
	0.2	6	7	8	9	10	11	13	15	25	42
	0.5	6	8	9	12	14	15	18	20	34	59
	0.75	6	9	10	13	15	17	19	22	37	49
	1	6	9	11	14	16	18	21	23	40	41
	1.5	6	10	12	16	18	19	22	25	43	38
	2	6	11	13	17	19	21	24	27	45	NA
	5	6	13	16	20	23	25	29	35	24	NA
	10	6	16	18	23	26	29	32	33	NA	NA
$c_{bedrock}$ scaling factor	0	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72
	0.1	0.70	0.70	0.70	0.70	0.69	0.69	0.68	0.68	0.64	0.58
	0.2	0.68	0.68	0.68	0.67	0.66	0.66	0.65	0.64	0.57	0.45
	0.5	0.62	0.62	0.61	0.60	0.58	0.57	0.54	0.52	0.37	0.18
	0.75	0.58	0.57	0.57	0.54	0.53	0.51	0.47	0.43	0.26	0.07
	1	0.54	0.53	0.52	0.50	0.47	0.45	0.40	0.36	0.19	0.02
	1.5	0.47	0.46	0.45	0.41	0.38	0.35	0.30	0.26	0.10	0.00
	2	0.42	0.40	0.39	0.35	0.31	0.28	0.23	0.19	0.06	NA
	5	0.22	0.20	0.18	0.14	0.11	0.09	0.06	0.04	0.00	NA
	10	0.11	0.09	0.08	0.05	0.04	0.03	0.01	0.01	NA	NA

leakage, the threshold was 6.45 mm and the slope 0.95 mm/mm. While both parameters impacted storm response, the bedrock leakage coefficient seemed to have more impact on the slope, while the subsurface storage volume had more impact on the threshold. A similar pattern of bedrock leakage coefficient and subsurface storage volume influence on the precipitation/discharge relationship was seen in the other events, with different antecedent moisture conditions.

Thresholds at the watershed scale

The newfound relationship between the soil moisture deficit and fill and spill factors and the precipitation/discharge threshold was tested against two long term data records. The relationship was tested on a 3.8 ha watershed nearby the modeled hillslope. The precipitation/storm discharge relationship at this catchment, which has similar geology (bedrock permeability and subsurface storage) and climatic conditions (storm spacing and *PET*) should exhibit the fill and spill and soil moisture deficit correlated threshold relationship seen in the virtual experiments. The second test was applied to a set of watersheds at the HJ Andrews Long Term Ecological Research Site (HJA), Western Cascades, Oregon, USA, ranging from 8 to 101 ha. The HJA will be a stronger test of the soil moisture deficit factors due to the higher antecedent drainage times for events towards the end of the summer season.

M8 Catchment, Maimai, New Zealand

Upstream of the instrumented hillslope used for numerical modeling in this paper is the M8 watershed, a first order, 3.8 ha basin instrumented with a *V* notch weir at the outlet. The watershed was gauged for nearly 30 years (1974–2003) and a subset of the data record (1985–1986) was used for watershed scale threshold analysis. Evapotranspiration was derived from estimates of monthly evapotranspiration rates using onsite meteorological data, interpolated first into daily totals, then hourly using a sine curve distribution with a peak at 18:00 (Vaché and McDonnell, 2006).

For the analysis, the 2 year M8 hyetograph was split into 140 storm events. A storm event was defined more than 1 mm precipitation preceded by 24 h of less than 1 mm. Storm runoff was defined as the rise in stream discharge above baseflow (streamflow at the initiation of the rain event) from the initiation of the event until the beginning of the next event. Total storm precipitation during the 2 years of monitoring was 4438 mm, with individual event precipitation ranging from 1 to 220 mm, with an average storm size of 31 mm (Fig. 10b). Total storm runoff for the monitored time period was 1917 mm, where individual event storm runoff ranging from 0 to 110 mm, with an average of 14 mm. Of the 2 years analyzed, the average time between storms was 3.2 days and the maximum was 17.2 days. Estimated *PET* averaged 2.4 mm/day. Antecedent potential evapotranspiration (APET; *PET* $*$ time between storms) ranged from 0 to 45.6 mm, with an average of 7.1 mm.

Discharge was predicted at the M8 watershed using the relationship between the slope and threshold parameters and the antecedent potential evaporation, holding the fill and spill factors at the values determined in the hillslope model calibration described in "Calibrated model parameterization".

$$Q_i = a(P - P_o) \tag{9}$$

where Q_i is predicted catchment storm runoff and P_i is measured precipitation for storm *i*. The precipitation discharge threshold, P_o, is a function of the antecedent potential evapotranspiration (P_o = 12.2 mm + 0.35 mm/mm $*$ *APET*). The slope, *a*, is a function of *PET* (a = 0.628 mm/mm–0.007 (mm/mm)/(mm/day) $*$ *PET*). Predicted

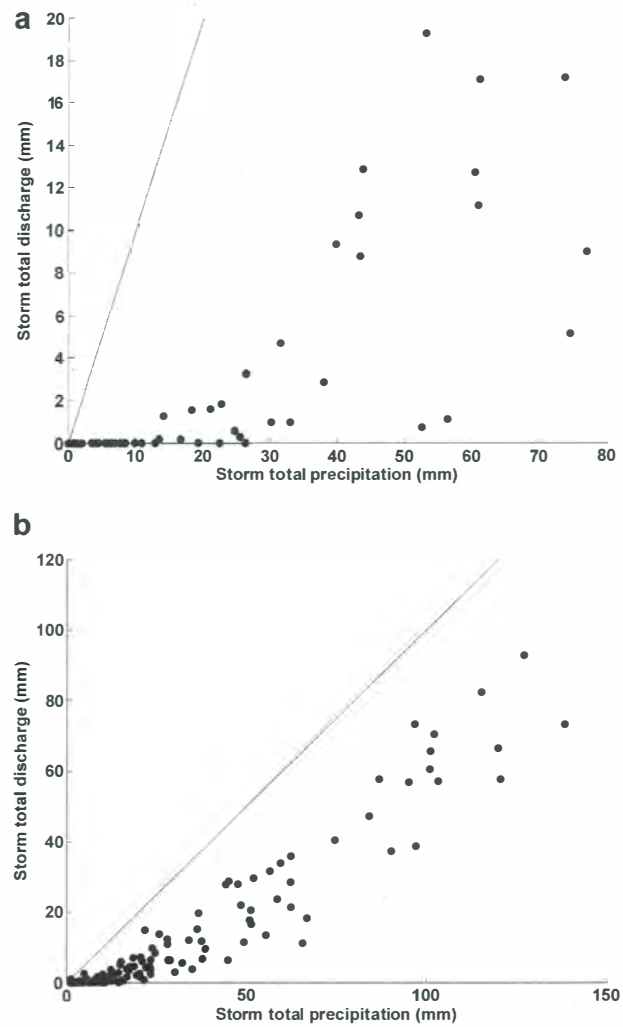

Fig. 10. Measured whole storm precipitation/discharge dynamics at two instrumented field sites: (a) Maimai hillslope (0.09 ha), 2 years of monitoring and (b) M8 catchment (3.8 ha), 2 years of monitoring. Hillslope threshold estimated at 20 mm, while catchment threshold estimated at 8.5 mm.

storm discharge was calculated for each storm using three additional methods: (1) Calculating the threshold and slope calculated from analysis of the entire data record and applying to Eq. (10). The annual threshold and slope were determined by fitting Eq. (10) to the storm precipitation and discharge data using a least squared calibration. (2) Calculating the annual storm runoff ratio R_{yr} (total annual discharge/total annual precipitation) and applying to the equation

$$Q_i = RP_i \tag{10}$$

(3) Calculating the average storm runoff ratio R_{ave} (average of individual event runoff ratios) and applying to Eq. (11).

Storm runoff as predicted from the threshold and slope derived from the virtual experiments fit the measured storm runoff. The root mean square error (RMSE) of the measured vs. modeled discharge was 6.2 mm, and the Nash Sutcliffe efficiency (*E*) was 0.93. Total predicted storm runoff was 1837 mm for the 141 monitored storms over 2 years (96% of the measured storm runoff value of 1916 mm). Using the annual measured slope (0.59) and threshold (8.5 mm) from the discharge/precipitation record resulted in a RMSE of 6.5 mm and *E* of 0.92, with predicted total storm runoff of 2030 mm (106% of measured). Using the annual measured storm

runoff ratio to predict storm discharge led to a poorer fit, with a RMSE of 9.0 mm, E of 0.85, and, by definition, total storm runoff was 1916 mm, or 100% of measured. Using the average storm runoff ratio yielded a higher RMSE (15.5 mm) and lower E (0.55), and low total storm runoff (1109 mm, or 58% of measured).

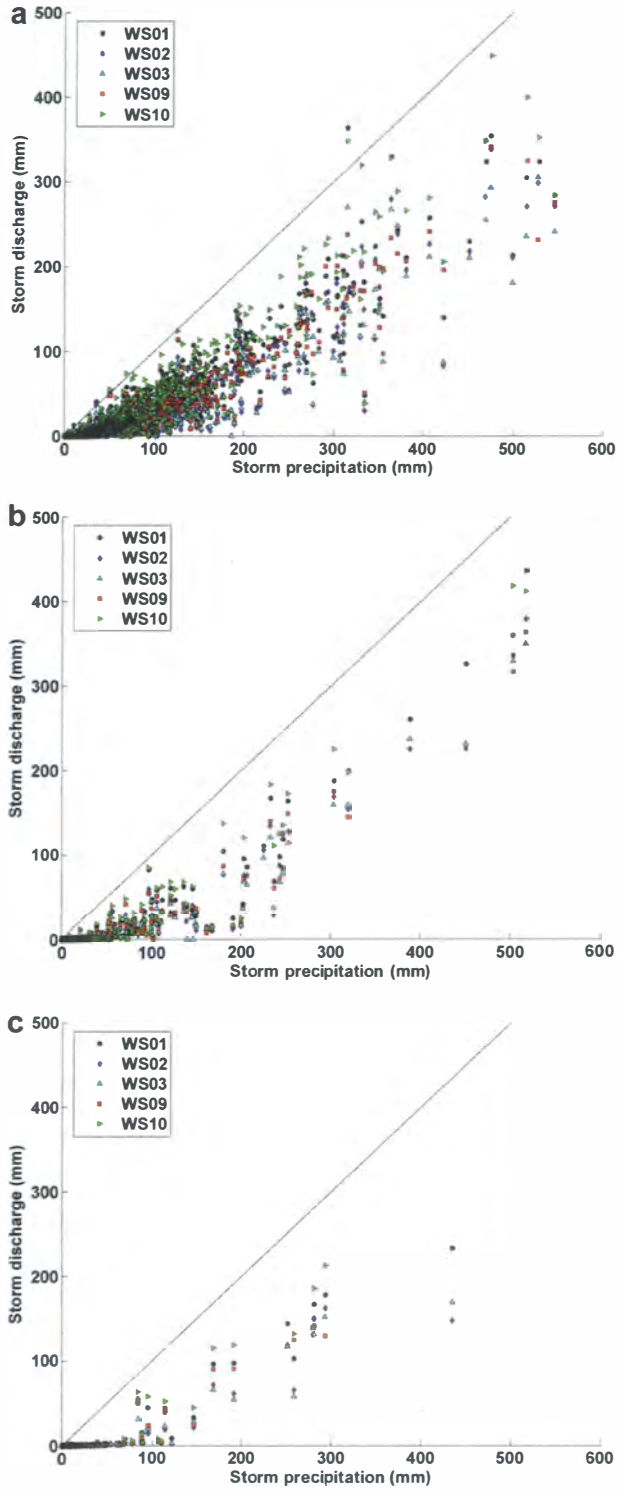

Fig. 11. Measured whole storm precipitation/discharge dynamics at five instrumented research catchments: WS1 (101.3 ha), 2 (96 ha), 3 (60 ha), 9 (9 ha) and 10 (10 ha) at the HJ Andrews Experimental Forest. WS1-3 have 50 years of monitoring, and WS9-10 have 36. Storms are binned according to antecedent drainage time: (a) All events with less than 5 days antecedent drainage; (b) all events with between 5 and 10 days antecedent drainage; and (c) all events with greater than 10 days of antecedent drainage. Estimated thresholds for the three groups are 0 mm, 56 mm, and 83 mm.

HJ Andrews, Oregon, USA

Further analysis of the watershed scale precipitation discharge threshold relationship was performed using the precipitation discharge record at the H J Andrews Experimental Forest (HJA) in western Oregon, USA. At the HJA, continuous discharge and precipitation records have been maintained at ten watersheds, ranging from 9 to over 100 ha for up to 50 years, though we do not have evaporation estimates for the duration. Of the ten gauged catchments, five (WS1, WS2, WS3, WS9, and WS10, from 9 to 101 ha) are at low enough elevation that their annual hydrographs are dominated by rainfall, rather than seasonal snowmelt. These catchments are all steep, forested sites, harvested in the 1960s (WS1, 3), 1970s (WS10) or remaining with an old growth overstory (WS2, 9). More details on the catchments is found in Jones (2000). WS1-3 have been gauged since 1958, while WS9-10 have been gauged since 1969. 2246 (WS1-3) and 1718 (WS9-10) rainfall events were extracted from these records (storms begin when 1 mm rain falls, and end after 24 h of no precipitation), ranging up to 731 mm. Storm runoff and total storm precipitation was extracted from the discharge record as for the M8 watershed procedure. A plot of total storm precipitation vs. discharge shows little evidence of a threshold (Fig. 11a). However, if the storms are binned according to the antecedent drainage time, a threshold appears to exist for events with larger than 5 days, and the threshold increases with increasing antecedent drainage time (Fig. 11b and c). The threshold appears to be consistent between the gauged catchments (~50 mm for greater than 5 days drainage, ~80 for greater than 10 days drainage), despite the wide range of catchment sizes.

Discussion

We used the dominant processes concept of Grayson and Blöschl (2000) to construct a reservoir type numerical model based on the Maimai hillslope. MaiModel, with simple unsaturated storage and flow conceptualization, was able to generally reproduce observed hydrometric and tracer behavior. The calibrated model was able to reproduce the 40 day hydrograph, as well as each individual storm. Additionally, the model was able to reproduce breakthrough of a line tracer application 35 m upslope – one characterized by both rapid initial breakthrough and extended tailing. The model was also able to capture the precipitation–discharge threshold relationship observed in the data record. The model was then used to determine the relative importance of fill and spill and soil moisture deficit factors on the threshold relationship. Below we highlight some of the issues associated with the calibration of the model, the results from the virtual experiments, and the application of the new understanding of threshold controls at the watershed scale, both at a nearby first order watershed and at a different set of watersheds ranging from 8.5 to 101 ha.

On the value of data for model construction and testing

The model objective criteria that we used did not have equal strength in either limiting the range of individual parameters, or in reducing the number of behavioral parameter sets. In general, the model criteria that were effective in reducing the range of individual parameters were also effective in reducing the number of behavioral parameter sets. Of notable exception was the cumulative tracer breakthrough criterion. While able to reject many parameter sets (only 3% of the parameter sets acceptably met the

criterion), it did little to reduce the range of acceptable individual parameters, except for the bedrock leakage coefficient ($c_{bedrock}$). While the range of most parameters were reduced through calibration, soil hydraulic conductivity was not, with acceptable models sampling from 94% of the original parameter space. The other parameters were well identified through calibration, with a reduction of the original parameter space by 56–86% (Table 2).

Modeled hydrograph

The six storm hydrograph criteria were responsible for reduction in both the range of the individual parameters and the total number of acceptable parameter sets (Fig. 4). As in previous studies, different subsections of the hydrograph provided different amounts of power in parameter identifiability (Seibert and McDonnell, 2002; Son and Sivapalan, 2007; Tetzlaff et al., 2008). Breaking up the calibration hydrograph into 5 distinct time periods, centered on the significant rain events, proved to be a strong tool for both parameter identifiability and parameter set rejection. While 2.9% of the parameter sets met the 40 day criterion, only 0.4% of the parameter sets resulted in simulations that met the hydrograph criteria for all five of the events. Seibert and McDonnell (2002) found that a storm event with the largest peak precipitation rate and discharge served as the most stringent criterion in their calibration of a similar reservoir model. In our case, the highest peak of precipitation and discharge occurred in storm B4, which was a relatively weak criterion.

B2, the event with the lowest peak discharge, and longest duration, was the most effective in both narrowing the parameter ranges and rejecting parameter sets. Whereas the other storms were relatively simple, with a single peaked hyetograph and hydrograph, B2 was more complex, with a double peaked hyetograph and hydrograph. B2 was especially effective in reducing the parameter space for two variables: lateral subsurface storage and bedrock leakage. This sensitivity was likely due to the complex filling and draining of subsurface storage, a factor that was masked in the higher shorter, single peaked events, where the subsurface storage is filled early in the event, then monotonically drained. The prolonged nature of the B2 event, along with the refilling during the second peak, required a more precise definition of the subsurface storage processes. This suggests that it is not the size of the event, but perhaps the complexity that is important for model calibration.

Modeled tracer breakthrough

The modeled tracer breakthrough served as another source for parameter identification and parameter set rejection. Other researchers have shown the importance of using tracers (such as isotopic signatures of rainfall) in addition to hydrometric data for model calibration (Fenicia et al., 2008; Son and Sivapalan, 2007; Soulsby and Dunn, 2003; Vaché and McDonnell, 2006). Tracers are attractive as model objective criteria because tracer and pressure response to precipitation are often quite different (i.e. the rapid catchment response dominated by pre-event water (Sklash and Farvolden, 1979). Tracer breakthroughs also serve to integrate hillslope scale response, in contrast to point measurements of water table height, soil moisture status or other similar objective criteria. While isotopic tracers and mean residence times of tracers have been used for model calibration, the use of an applied, chemical tracer is relatively rare for model calibration (although Weiler and McDonnell (2007) analyzed the Brammer tracer injection with a macropore based conceptual model of the Maimai hillslope, demonstrating the importance of the preferential flow network structure).

While the temporal and spatial patterns of tracer breakthrough were not stringent criteria in the MaiModel calibration (eliminating only 48% and 15% of the parameter sets, respectively), the cumulative tracer breakthrough eliminated 97% of the parameter sets, and 66% of the simulations that were deemed behavioral for all storms. The calibration runs that modeled all of the sub-hydrographs and did not match the measured tracer breakthrough had a modeled cumulative tracer breakthrough ranging from 8% to 28%, compared to a measured value of 14%. Of the simulations that had acceptable fits for the hydrographs but missed the cumulative tracer breakthrough, 23% were below the acceptable limits, and 77% were greater. 54% were more than twice the acceptable range from the measured value. This wide range of modeled tracer flux for models that acceptably fit the hydrograph demonstrates the importance of measurements of both particle and pressure response at the hillslope scale for model calibration and validation.

While the cumulative tracer breakthrough was effective in reducing the total number of behavioral parameter sets, it did little to reduce the ranges of the individual parameters, with the exception of $c_{bedrock}$. Two possible explanations of the relative weakness of the tracer breakthrough on the parameter ranges are (1) the tracer breakthrough is due to a combination of parameters, or (2) the cumulative tracer breakthrough is too weak a test, and a time series of tracer breakthrough is needed. Further analysis of the tracer breakthrough against the individual parameters suggests that the first option is more likely. The cumulative tracer breakthrough was compared with the products of each pair of calibrated parameters (10 pairs in total). The cumulative tracer breakthrough was strongly constrained by the product of the bedrock leakage coefficient and the subsurface storage volume, with a reduction of 94% of the widest possible range of the product (Fig. 12). This suggests that it is both the subsurface storage volume and the rate of drainage that controls the cumulative tracer breakthrough, more than each parameter individually that is important. The cumulative tracer breakthrough was not dependent on any other individual parameter, or product of parameters.

Soil hydraulic conductivity

Of the five calibrated parameters, all but the soil saturated hydraulic conductivity (k_{soil}) were significantly better defined through calibration. Of the nine calibration criteria, the number of behavioral parameter sets that matched each criteria was somewhat correlated to the reduction in the parameter space for each criteria ($0.48 < R^2 < 0.67$). The k_{soil}, however was reduced only 6% from the initial parameter range specified (1000% of the maximum measured hydraulic conductivity). This 6% is due more likely through chance than an actual narrowing of the possible parameter set, as acceptable parameter sets were evenly distributed over the calibration range.

The lack of sensitivity of k_{soil} is likely due to a combination of the system dynamics and the model codification of those dynamics. At Maimai, the measured rainfall intensities were rarely greater than the measured soil hydraulic conductivity (5–300 mm/h), and never reached the estimated conductivity of the 18 cm thick surface horizon (>6000 m/h (Webster, 1977)). Maximum hourly rainfall intensity for the calibration period was 19.4 mm/h, and was greater than the lower limit of soil matrix hydraulic conductivity (5 mm/h) for only 27 h over the course of the 40 day calibration period. This is consistent with the lack observations of Hortonian (infiltration excess) overland flow at the site (Mosley, 1979). These observations were codified into the model by excluding an overland flow module. Thus, once initiation of overland flow is eliminated from the possible model scenarios, much of the model sensitivity to k_{soil} is consequently removed. In other catchment situations and hydrological scenarios where overland flow is a more likely hydrological process (sites with higher maximum precipitation intensity or lower soil hydraulic conductivity), sensitivity of the model to k_{soil} would be greater.

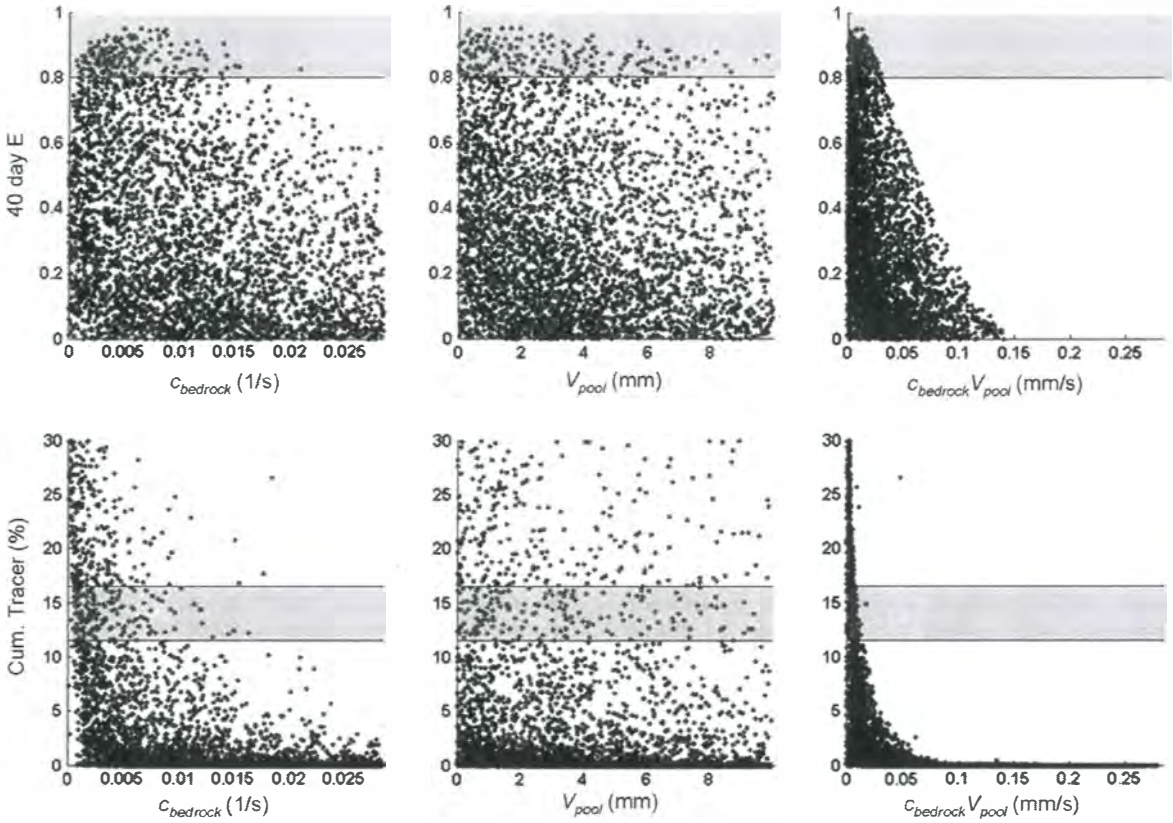

Fig. 12. Dotty plots of $c_{bedrock}$ and V_{pool} vs. 40 day Nash Sutcliffe efficiency (E) and cumulative tracer breakthrough. The grey bars denote the range of acceptable model fits. The x axis spans the initial parameter range. The product of the two parameters, a measure of the speed of subsurface storage drainage, is more identified than either parameter individually.

Improved understanding of thresholds at the hillslope scale

The calibrated model was able to reproduce the precipitation discharge threshold relationship seen at the Maimai hillslope trench. From analysis of measured hillslope discharge by Woods and Rowe (1996) and Brammer (1996), a threshold of between 17 and 21 mm was necessary for flow at the hillslope at Maimai (Fig. 10a). This range bounded the modeled threshold (17.7 mm) for the calibrated model. In the numerical simulations, the threshold was found to be due to both fill and spill factors (subsurface storage and bedrock leakage) as well as soil moisture deficit factors (*PET* and antecedent drainage time) (Fig. 13). A linear relationship between thresholds and the antecedent potential evapotranspiration (APET) was observed, with a slope of 0.38 mm/mm. The relationship between the product of the bedrock leakage coefficient and the subsurface storage volume was nonlinear, with an upper bound on the impact of the fill and spill factors on the threshold, while a similar bound has not been observed in the soil moisture deficit (Fig. 9). Logically, a bound must exist for the soil moisture deficit, once evaporation depletes the entire soil profile and a storm greater than the available storage would overcome the threshold. This bound was not met by the current virtual experiments. The lack of threshold after the fill and spill and soil moisture deficit mechanisms were eliminated indicates that these two are solely responsible for the simulated threshold.

The slope of the excess precipitation/discharge line in the model output was also found to be positively correlated with both fill and spill factors and *PET*, while not the antecedent drainage time (Fig. 9). An increase in both the subsurface storage and bedrock leakage coefficient were shown to increase the slope, as an increase

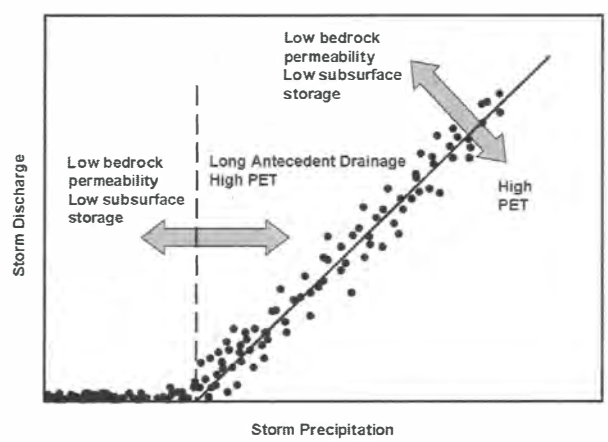

Fig. 13. Schematic of fill and spill and soil moisture deficit control of precipitation discharge threshold and slope of the excess precipitation discharge line.

in each increased bedrock leakage, both directly (increased leakage coefficient = increased leakage rate) and indirectly (increased storage = increased driver on leakage and increased late time storage and leakage). An increase in the *PET* increased the slope, as rainfall stored in the soil profile was lost to evaporation during and after the storm. With fill and spill and soil moisture deficit removed, the slope was unity, indicating these are the only factors affecting the slope in the model. Ninety-four percentage of the reduction in the slope was due to fill and spill mechanisms, while 6% of the reduction is due to the potential evaporation rate. The small

impact of the *PET* on the slope is due to the limited time that *PET* can impact the discharge after the threshold is reached, as hillslope drainage lasted less than 4 days for all simulations.

Improved understanding of thresholds at the watershed scale

Our macroscale hillslope model has shown two causal factors for the storm precipitation–discharge relationship seen at the hillslope and small catchment scale: climatic (event spacing and evaporative losses) and geology (bedrock permeability and subsurface storage). The geologic factors are difficult to determine, with bedrock permeability difficult to measure, and subsurface storage depending on the dominant lateral subsurface flow processes and the bedrock topography, both of which are difficult to measure. The climatic factors, however, are often available in long term data sets. While long term evaporation records remain uncommon, new analysis of long term precipitation records does provide a way forward towards better prediction of catchment storm response.

The soil moisture deficit influence on the precipitation discharge threshold and slope was previously suggested in the analysis of the long term precipitation discharge record of the instrumented hillslope at Panola, Georgia (Tromp-van Meerveld and McDonnell, 2006a). At Panola, it appeared that storms where the soil volumetric water content was lower than 40% prior to the event (at 70 cm depth in the profile) had a higher threshold for flow than those where soils were wetter beforehand. However, the data record at Panola had too few storms with sufficiently dry antecedent moisture conditions to determine the precise relationship between antecedent moisture and thresholds. At Minamitani, Japan, the threshold for flow at an instrumented hillslope and nearby second order catchment was hypothesized to be dependent on the flow rate at the initiation of the event (Tani, 1997). As at Panola, with the limited number of events above and below the threshold for each initial flow rate, the precise nature of this dependency is unclear. In fact, none of the instrumented hillslopes we know of have a sufficient data record with enough storms above and below the threshold to determine the precise relationship between the threshold and antecedent moisture conditions. While it remains difficult and expensive to maintain gauging for a sufficiently long duration at instrumented hillslopes, suitable data records exist at many experimental catchments, such as in the analysis of the M8 and HJA catchments.

Analysis using the predictions for whole storm runoff at the nearby M8 catchment, based on soil moisture deficit factors, was shown to better predict whole storm discharge than the annual threshold or runoff ratio analyses. The root mean square error was minimized and the difference between the measured and modeled annual storm runoff using the soil moisture deficit method of discharge prediction when compared to predictions made using the bulk annual threshold, annual runoff ratio and average runoff ratio methods. These predictions were made over a range of storms with different average and maximum rainfall intensities, durations and precipitation patterns, yet the storm runoff was very well predicted based on two simple factors revealed through the numerical modeling.

Additional analysis at five small research watersheds in western Oregon (10–100 ha) showed a similar dependence on antecedent drainage time for the threshold. While evaporation estimates were not available for the duration of the 50 year data record, events with long antecedent drainage were shown to exhibit a much higher threshold for flow. This threshold appears to be quite high for events with longer than 9 days of antecedent drainage (~80 mm), for a series of watersheds that are very responsive to rainfall (annual storm runoff ratios approach 38% (McGuire et al., 2005)). As predicted in the modeling, the slope of the excess precipitation–discharge line is little changed.

At the M8 and HJA catchments, there is little evidence for a threshold for flow for events with short (<5 days) antecedent drainage time, perhaps due to minimal effect of bedrock leakage. At the Maimai and Panola hillslopes, water that leaks into the bedrock is likely not recovered at the hillslope trench. At the M8 and HJ Andrews catchments, however, bedrock leakage is likely recovered at the watershed outlet. At the catchment scale, the fill and spill mechanism should then not have a large impact on the precipitation discharge threshold. Therefore, during events where the soil has not had a chance to dry due to evaporative losses, a small threshold would be expected at these catchments, as seen in this analysis.

This functional dependence of the threshold on fill and spill and soil moisture deficit factors may be a means for prediction of flow at ungauged hillslopes and basins. At a site where the physical properties are similar to either Maimai, or some basin where the geologic dependent threshold and slope has been determined, the base case threshold can be determined, and the effects of the climatic factors would be determined from the storm spacing and evaporative demand. The geologic factors (precipitation discharge threshold and slope) can be determined by analysis of the system response to precipitation at the lower extreme of *PET* and storm spacing. The analyses of the HJA watersheds suggest that the antecedent drainage dependence of the thresholds may apply to other steep forested hillslopes and catchments. Special attention needs to be placed on locations with different geology, catchment geometry and dominant flow processes.

Conclusions

Graham et al. (this issue) developed a new perceptual model of hillslope subsurface flow processes at a well studied field site. They determined that lateral subsurface flow is dominated by flow in a well connected preferential flow network at the interface between the soil profile and permeable bedrock. This paper used this new perceptual model as the basis for a numerical model designed to model flow and transport based on these dominant processes. The model was able to reproduce both hydrometric and tracer data, using five tunable parameters. A series of virtual experiments aimed at revealing the controls on the threshold response of hillslope discharge to precipitation were performed using the numerical model. We found that both fill and spill (bedrock permeability and subsurface storage) and soil moisture deficit (storm spacing and potential evapotranspiration rates) factors influenced threshold magnitude. While the climatic controls were shown to have a large potential impact on flow dynamics, in a climate like that of the study hillslope, where storm spacing was short (average time between storms = 3 days) and the *PET* demand was low (<6 mm/day), the geologic controls dominated (66% of the threshold and 94% of the slope of the excess precipitation/discharge relationship were determined by the geologic components). The relationship between the climatic factors and the precipitation discharge threshold and slope were applied to a nearby catchment and demonstrated to better predict storm discharge than either the average runoff ration, annual runoff ratio or the bulk threshold relationship.

Acknowledgements

This work was funded through an NSF IGERT Ecosystem Informatics internship. We would like to thank Jurriaan Spaaks, Ciaran Harmon, Luisa Hopp, Taka Sayama, and Kellie Vache for helpful comments on the modeling strategy. Special thanks go to Ross Woods, Lindsay Rowe and Dean Brammer for the use of their experimental data for model development and calibration. Thanks to the reviewers and editors of Journal of Hydrology, who's comments greatly improved this paper.

References

Baker, T.M., Hawke, R.M., 2007. The effects of land application of farm dairy effluent on groundwater quality – West Coast. Journal of Hydrology (New Zealand) 45.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249, 11–29.

Brammer, D., 1996. Hillslope hydrology in a small forested catchment, Maimai, New Zealand. M.S. Thesis, State University of New York College of Environmental Science and Forestry, Syracuse, 153 pp.

Brooks, R.H., Corey, A.T., 1964. Hydraulic properties of porous media. Hydrology Paper 3, Colorado State University, Fort Collins, CO.

Buttle, J.M., Dillon, P.J., Eerkes, G.R., 2004. Hydrologic coupling of slopes, riparian zones and streams: an example from the Canadian Shield. Journal of Hydrology 287 (1–4), 161–177.

Carsel, R.F., Parrish, R.S., 1988. Developing joint probability distributions of soil water retention characteristics. Water Resources Research 24 (5), 755–769.

Dooge, J.C.I., 1986. Looking for hydrologic laws. Water Resources Research 22, 46S–58S.

Fenicia, F., McDonnell, J.J., Savenije, H.H.G., 2008. Learning from model improvement: on the contribution of complementary data to process understanding. Water Resources Research 44, 1–10.

Freeze, R.A., Cherry, J.A., 1979. Groundwater. Prentice-Hall, Englewood Cliffs, NJ. 604 pp.

Graham, C.B., McDonnell, J.J., Woods, R., this issue. Hillslope threshold response to rainfall: (1) a field based forensic approach. Journal of Hydrology.

Grayson, R.B., Blöschl, G., 2000. Summary of pattern comparison and concluding remarks. In: Grayson, R.B., Blöschl, G. (Eds.), Spatial Patterns in Catchment Hydrology: Observations and Modeling. Cambridge University Press, Cambridge, UK, pp. 355–367.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Towards improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resources Research 34 (4), 751–763.

Jones, J.A., 2000. Hydrologic processes and peak discharge response to forest removal, regrowth, and roads in 10 small experimental basins, western Cascades, Oregon. Water Resources Research 36 (9), 2621–2642.

McDonnell, J.J., 1990. A rationale for old water discharge through macropores in a steep, humid catchment. Water Resources Research 26 (11), 2821–2832.

McDonnell, J.J. et al., 2007. Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. Water Resources Research 43, 1–10.

McGlynn, B.L., McDonnell, J.J., Brammer, D.D., 2002. A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand. Journal of Hydrology 257, 1–26.

McGuire, K.J. et al., 2005. The role of topography on catchment-scale water residence time. Water Resources Research 41.

Mosley, M.P., 1979. Streamflow generation in a forested watershed. Water Resources Research 15, 795–806.

Mosley, M.P., 1982. Subsurface flow velocities through selected forest soils, South Island, New Zealand. Journal of Hydrology 55, 65–92.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, I, a discussion of principles. Journal of Hydrology 10, 282–290.

Pearce, A.J., Stewart, M.K., Sklash, M.G., 1986. Storm runoff generation in humid headwater catchments: 1. Where does the water come from? Water Resources Research 22, 1263–1272.

Peters, D.L., Buttle, J.M., Taylor, C.H., LaZerte, B.D., 1995. Runoff production in a forested, shallow soil, Canadian shield basin. Water Resources Research 31 (5), 1291–1304.

Putuhena, W.M., Cordery, I., 2000. Some hydrological effects of changing forest cover from eucalypts to *Pinus radiata*. Agricultural and Forest Meteorology 100 (1), 59–72.

Quinn, P.F., Beven, K.J., Chevallier, P., Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. Hydrological Processes 5, 59–79.

Seibert, J., 1997. Estimation of parameter uncertainty in the HBV model. Nordic Hydrology 28, 4–5.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. Water Resources Research 38 (11), 1–10.

Sklash, M.G., Farvolden, R.N., 1979. The role of groundwater in storm runoff. Journal of Hydrology 43, 45–65.

Sklash, M.G., Stewart, M.K., Pearce, A.J., 1986. Storm runoff generation in humid headwater catchments: 2. A case study of hillslope and low-order stream response. Water Resources Research 22 (8), 1273–1282.

Son, K., Sivapalan, M., 2007. Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data. Water Resources Research 43, 1–10.

Soulsby, C., Dunn, S.M., 2003. Towards integrating tracer studies in conceptual rainfall–runoff models: recent insights from a sub-arctic catchment in the Cairngorm Mountains, Scotland. Hydrological Processes 17, 403–416.

Spence, C., Woo, M.-K., 2002. Hydrology of subartic Canadian shield: bedrock upland. Journal of Hydrology 262, 111–127.

Tani, M., 1997. Runoff generation processes estimated from hydrological observations on a steep forested hillslope with a thin soil layer. Journal of Hydrology 200, 84–109.

Tetzlaff, D., Uhlenbrook, S., Eppert, S., Soulsby, C., 2008. Does the incorporation of process conceptualization and tracer data improve the structure and performance of a simple rainfall–runoff model in a Scottish mesoscale catchment? Hydrological Processes 1.

Tromp-van Meerveld, H.J., McDonnell, J.J., 2006a. Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope. Water Resources Research 42.

Tromp-van Meerveld, H.J., McDonnell, J.J., 2006b. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. Water Resources Research 42.

Tromp van Meerveld, I., Weiler, M., 2008. Hillslope dynamics modeled with increasing complexity. Journal of Hydrology 361, 24–40.

Uchida, T., Tromp-van Meerveld, H.J., McDonnell, J.J., 2005. The role of lateral pipe flow in hillslope runoff response: an intercomparison of non-linear hillslope response. Journal of Hydrology 311, 117–133.

Vaché, K.B., McDonnell, J.J., 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. Water Resources Research 42.

Webster, J.R., 1977. The hydrologic properties of the forest floor under beech/podocarp/hardwood forest, North Westland. Masters in Agricultural Science Thesis, University of Canterbury, Christchurch, NZ.

Weiler, M., McDonnell, J.J., 2007. Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes. Water Resources Research 43, 1–10.

Weiler, M., McDonnell, J.J., Tromp-van Meerveld, I., Uchida, T., 2006. Subsurface stormflow. Encyclopedia of Hydrological Sciences 3, 1–10.

Whipkey, R.Z., 1965. Subsurface storm flow from forested slopes. Bulletin of the International Association of Scientific Hydrology 2, 74–85.

Woods, R., Rowe, L., 1996. The changing spatial variability of subsurface flow across a hillside. Journal of Hydrology (New Zealand) 35 (1), 51–86.

Woods, R., Rowe, L., 1997. Reply to Comment on The changing spatial variability of subsurface flow across a hillside by Ross Woods and Lindsay Rowe. Journal of Hydrology (New Zealand) 36 (1), 51–86.