A process-based rejectionist framework for evaluating catchment runoff model structure

Kellie B. Vaché[1,2] and Jeffrey J. McDonnell[1]

Received 11 May 2005; revised 5 October 2005; accepted 1 November 2005; published 21 February 2006.

[1] Complex hydrological descriptions at the hillslope scale have been difficult to incorporate within a catchment modeling framework because of the disparity between the scale of measurements and the scale of model subunits. As a result, parameters represented in many conceptual models are often not related to physical properties and therefore cannot be established prior to a model calibration. While tolerable for predictions involving water quantity, water quality simulations require additional attention to transport processes, flow path sources, and water age. This paper examines how isotopic estimates of residence time may be used to subsume flow path process complexity and to provide a simple, scalable evaluative data source for water quantity- and quality-based conceptual models. We test a set of simple distributed hydrologic models (from simple to more complex) against measured discharge and residence time and employ a simple Monte Carlo framework to evaluate the identifiability of parameters and how the inclusion of residence time contributes to the evaluative process. Results indicate that of the models evaluated, only the most complex, including an explicit unsaturated zone volume and an effective porosity, successfully reproduced both discharge dynamics and residence time. In addition, the inclusion of residence time in the evaluation of the accepted models results in a reduction of the a posteriori parameter uncertainty. Results from this study support the conclusion that the incorporation of soft data, in this case, isotopically estimated residence times, in model evaluation is a useful mechanism to bring experimental evidence into the process of model evaluation and selection, thereby providing one mechanism to further reconcile hillslope-scale complexity with catchment-scale simplicity.

Citation: Vaché, K. B., and J. J. McDonnell (2006), A process-based rejectionist framework for evaluating catchment runoff model structure, *Water Resour. Res.*, *42*, W02409, doi:10.1029/2005WR004247.

1. Introduction

[2] Many different conceptual models of catchment hydrology and runoff formation have been developed during the last few decades [*Singh and Frevert*, 2002]. Increasingly, runoff models form the basis of simulations that address complex environmental problems concerning surface water acidification, soil erosion, pollutant leaching, and possible consequences of land use or climatic changes. These linkages mean that realistic simulations of internal catchment processes determining runoff age, origin and pathway become essential. Notwithstanding, a clear process to define and incorporate relevant experimental observation of dominant internal processes into the runoff model structure runoff remains to be defined. The experimentalist would like to see all of his/her ‘‘perceptual’’ (as defined by *Beven* [2001d]) understanding incorporated into the runoff model. However, a fundamental paradox in catchment modeling is that incorporation of more complex processes requires the addition of model parameters that in turn increase the degrees of freedom and problems of parameter identifiability (discussed in many papers by Keith Beven, including *Beven* [2001a, 2001b, 2001d, 2002]). This paradox often thwarts the best intentions of both modelers and experimentalists to bring more realism into the model structure. While many have argued recently that a dominant processes approach is the best way to define the appropriate level of model complexity [*Grayson and Western*, 2001; *Sivakumar*, 2004], objective determination of dominant processes in the field is very difficult to achieve. Even if these processes could be captured in the model structure, there are few measures beyond the catchment hydrograph that are used typically to evaluate the model performance.

[3] The difficulty in establishing dialog between experimentalist and modeler means that we continue to be mired in model problems such as the need for calibration [*Duan et al.*, 1992], the disparity between the scale of measurements and the scale of model subunits [*Beven*, 1989; *Blöschl*, 2001], and the equifinality of different model structures and parameter sets (i.e., the phenomenon that equally good model simulations might be obtained in many different ways [*Beven*, 2001a]). The main limitation on the experimental side is that we still cannot quantify the complex

[1]Department of Forest Engineering, Oregon State University, Corvallis, Oregon, USA.

[2]Now at Institute for Landscape Ecology and Resources Management, Giessen, Germany.

0043-1397/06/2005WR004247

nature of catchment flow paths. Studies abound in the literature with anecdotal evidence of flow paths on certain hillslope segments [*Tsuboyama et al.*, 1994; *Freer et al.*, 2002] or predicted bulk flow components in the stream hydrograph [e.g., *Soulsby et al.*, 2003]. However, robust descriptions of flow path dynamics beyond a soil core or trenched hillslope are impossible to distinguish with today's measurement technology. Notwithstanding, model structures designed to produce estimates of water quality must define, to some degree, realistic internal flow paths, fluxes, and stores of water upon which geochemical and ecological processes depend. This additional realism is necessary because residence time in the subsurface is a controlling feature of water quality [*Burns et al.*, 2003].

[4] So how do we build catchment model structures with water quality sensitive flow paths if we cannot define the flow paths experimentally? This paper examines how mean residence time (MRT) may be used to subsume flow path process complexity and provide a simple, scalable evaluative data source for water quantity- and quality-based conceptual models at the catchment scale. We argue that this approach is one mechanism to evaluate the degree of complexity warranted in a model structure with water quality sensitive flow paths. To date, a small number of catchment models have incorporated residence time into the model structure [*Barnes and Bonell*, 1996; *Hooper et al.*, 1998; *Robson et al.*, 1992; *DeGrosbois et al.*, 1988; *Beck et al.*, 1990] and fewer still have used tracer simulations to produce additional evaluative criteria or complementary constraints on water quantity/quality simulation [*Seibert and McDonnell*, 2002; *Weiler and McDonnell*, 2004; *Vaché et al.*, 2004]. No studies to date have utilized observational estimates of residence time as a posteriori model calibration criteria. These complementary measures contrast with the now-standard streamflow signal (pressure propagation) by representing quantities moving with the water (tracer movement). Flow has been the most popular (and in many cases the only!) evaluation criteria for many modeling studies because it is integrative of the whole-system response and a continuous signal of pressure propagation throughout the catchment. Like discharge, stream water residence time is also integrative and meaningful across all space scales, from the mean residence time of soil water as deduced from a suction lysimeter on a hillslope [e.g., *Asano et al.*, 2002] to small watershed [*Maloszewski and Zuber*, 1993] or large watershed scale [*Michel*, 2004]. Recent work has suggested that residence time in a watershed may be distributed spatially using topographic data from widely available DEM information [*McGlynn et al.*, 2003; *McGuire et al.*, 2005].

[5] This paper builds upon recent process work and brings together flow, stream water MRT and distributed soil water MRT as complementary evaluation criteria for simple models of catchment runoff that include water quality sensitive flow paths. We use experimentally determined stream MRT as a posteriori model calibration criteria. These data are compared against MRT estimates distilled from the results of a conservative tracer component of the hydrologic model. In addition, we derive for the first time, spatially distributed soil water residence time estimates as a third constraining, evaluative measure. Our goal in this paper is not to build a model with residence time per se, but rather use stream water and soil water residence time as additional measures to reduce uncertainty and guide decisions on the degree of complexity that is warranted in a rainfall runoff model. This approach is aimed at rejection of model structures that do not meet minimum requirements for simulation of flow and transport.

[6] The concept of rejection as a positive result of modeling studies has been gaining ground in the recent literature [*Zak and Beven*, 1999; *Freer et al.*, 2003]. It represents a different way to approach the model enterprise, where the model structure is malleable and evolves with process-based tests. Our approach is motivated by *Hooper* [2001] and *Beven* [2001c], both of whom advocate better use of the scientific method (i.e., hypothesis generation, subsequent rejection, and from that the generation of new hypotheses) in catchment modeling. In this paper, we propose a suite of plausible model structures, starting with the most basic configurations for flow and transport. We allow for the rejection of these model structure "hypotheses", and use rejection as a basis for the inclusion of further model complexity. Our philosophy embraces the methodology outlined initially outlined by Peter Young [e.g., *Young et al.*, 1996] and recently revisited by *Wagener et al.* [2001] and *Atkinson et al.* [2002] who advocate a modeling framework that balances model complexity, the level of available data, and the purposes of the modeling exercise. Here we use experimentally derived residence times to characterize and constrain parameterizations for storage across the prior parameter distribution, in an extension of the conceptual rainfall runoff modeling framework of *Vaché et al.* [2004]. This provides a process basis for decisions regarding the complexity incorporated into the structure of the model. We argue that the incorporation of MRT estimates as additional evaluative criteria may lead to better predictions of the elusive combination of water and solute flux from hydrological models.

[7] The Maimai watershed on the south island of New Zealand is utilized as a test bed for these examinations. Our specific objectives are (1) to develop and apply a set of model structures designed to reflect the most basic controls on water flux and transport at Maimai; (2) to evaluate various residence time measures to balance model simplicity with the incorporation of observed process heterogeneity; and (3) to establish the utility of tracer simulation in identifying the residence time distribution of different catchment models, and constraining acceptable model parameters and structures.

2. Study Site Characteristics

[8] The Maimai research catchments are a set of highly responsive, steep, wet, watersheds on the west coast of the South Island of New Zealand. Maimai has a long history of hillslope hydrological research (see *McGlynn et al.* [2002] for a complete review). More importantly and unlike other sites where we have done experimental work, Maimai shows striking simplicity in its hydrologic response. Several recent model studies have used the Maimai data set as the basis for the development and testing of new model structures [*Seibert and McDonnell*, 2002] and multicriteria model calibration techniques [*Freer et al.*, 2004; *Vaché et al.*, 2004]. Similarly, the steep, wet catchments have been a focal point for experimental studies and conceptual model development since the mid-1970s [*Mosley*, 1979, 1982;

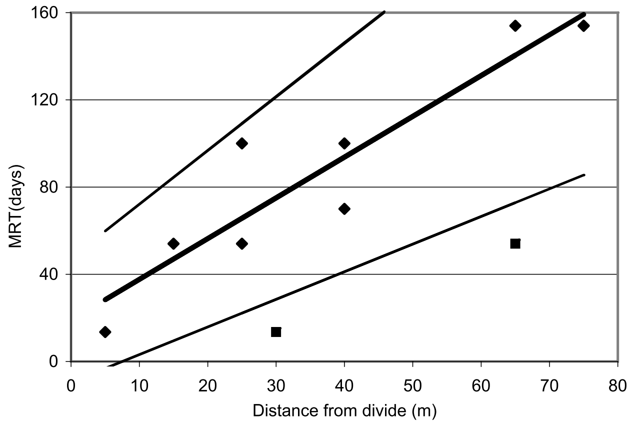

Figure 1. Ordinary least squares regression of isotopic MRT versus flow path distance to the stream at the Maimai M8 catchment. MRT is taken from *Stewart and McDonnell* [1991] and represents soil water. Values reported were derived from the steady state exponential models, except for longer MRT waters, which were not well defined by the steady state exponential. In these cases we report values from the nonsteady exponential model. Squares represent near-surface lysimeters that were excluded from the regression. Upper and lower 95% confidence intervals are included for reference.

McDonnell, 1990; *McDonnell et al.*, 1991; *Stewart and McDonnell*, 1991; *Woods and Rowe*, 1996; *McGlynn et al.*, 2003]. The simplicity in catchment response is determined largely by the lack of seasonality and chronically wet state of the system. Soils rarely fall below 90% of saturation [*Mosley*, 1979] and overlie effectively impermeable compacted and cemented conglomerate [*McKie*, 1978]. Quickflow (QF as defined by *Hewlett and Hibbert* [1967]) comprises 65% of the mean annual runoff and 39% of annual total rainfall (P) [*Pearce et al.*, 1986].

[9] Streamflow mean residence times at Maimai are very short. *Pearce et al.* [1986] report values of 4 months for the M6 catchment. In terms of soil water mean residence time, the Maimai isotopic data show clear patterns of downslope aging. Both of these observations reflect the combination of highly transmissive soil over largely impermeable bedrock. *Stewart and McDonnell* [1991] used water collected from suction lysimeters at 60 Kpa to estimate MRT at a variety of locations within the M8 catchment. Results indicate that between-storm matrix water varied in age from approximately 14 days at the catchment divide to over 100 days in downslope locations near the main M8 channel. Even the longest of these values are some of the shortest hillslope water residence times recorded in the literature and reflect the responsive nature of the catchment. More importantly, these data embody fundamental information regarding hillslope complexity and its coupling to the catchment scale since the residence time distribution is a direct reflection of the diversity of the flow paths. Figure 1 shows the relation between measured surface flow path distance from the ridge top divide to the nearest perennial stream and soil water MRT from a subset of suction lysimeters reported by *Stewart and McDonnell* [1991]. These data include only those lysimeters near the soil bedrock interface that might more readily respond to upslope contributions of water, rather than rainfall. While there is some scatter about the trend, the downslope aging of water, noted by *Stewart and McDonnell* [1991], is clearly illustrated. Other landscape properties, including slope, soil characteristics, as well as sample depth, also contribute to the distribution of fluxes that results in the spatial variation of groundwater residence times. Nevertheless, above channel slopes vary only marginally throughout M8, and show no relationship with estimated MRT values, soil properties are quite homogeneous [*McKie*, 1978], and only those lysimeters near the soil bedrock interface were utilized in Figure 1. Given this, the regression relationship outlined in Figure 1, was utilized directly in a regionalization procedure designed to provide an indication of how MRT may vary spatially in the M8 catchment. While additional measurements would be necessary to corroborate Figure 2 (including uncertainty), our use is consistent with the goal to evaluate the model on the basis of available soft data sources.

3. Hydrologic Model

3.1. Conceptual Framework

[10] We use a conceptual modeling framework to develop a set of four potential models designed to correspond

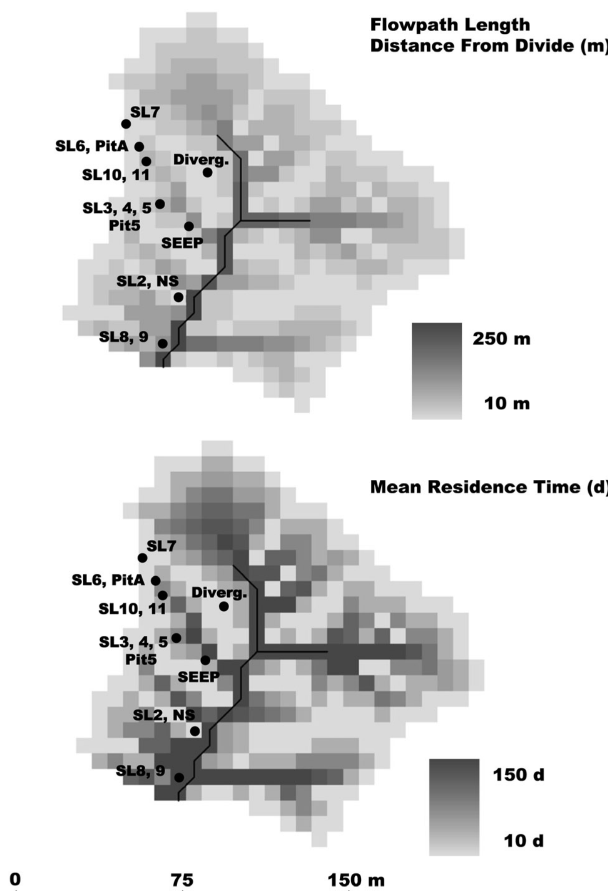

Figure 2. Regionalized MRT based upon the regression outlined in Figure 1: (top) calculated distance to the divide and (bottom) estimated MRT. While highly simplified, these regionalized data clearly indicate the relatively large range in residence times and the downslope aging. Approximate perennial channel network and suction lysimeter locations are indicated.

Table 1. Major Differences Between Each of the Four Models[a]

	Saturated Zone	Effective Porosity	Explicit Unsaturated Zone	Number of Tuned Parameters
Model 1	Yes	No	No	3
Model 2	Yes	Yes	No	4
Model 3	Yes	No	Yes	5
Model 4	Yes	Yes	Yes	6

[a]Note that the number of tuned parameters increases with model complexity.

closely to the dominant runoff generation processes at Maimai. The framework incorporates a variety of different model structures (both state variables and rates), which we evaluate to find a balance between the need to reduce model complexity, and yet adequately capture the process complexity. In all cases, the saturated zone is treated using a distributed model that corresponds in time and space to the data against which we seek to measure – a 10 m grid representing the 3.2 ha catchment. The subsurface flow routines follow from *Wigmosta et al.* [1994], where potential kinematic shocks (overtaking waves) are effectively avoided because of the independence of the solution for each unique location [*Beven*, 2001d]. This gridded scheme was selected as the basis for all the evaluated models because it provides an explicit mechanism to incorporate transient subsurface flows and the land surface slopes that play an important role in driving lateral flow and aging at the Maimai catchment. The hypothesis that transient subsurface flows dominate flux was originally proposed by *McDonnell* [1990] and has been corroborated by the many process studies completed at the site. These observations indicate that topography well approximates the gradient driving lateral water flow that occurs predominantly at the interface between bedrock and soil (as transient saturation) [*Freer et al.*, 2004]. Our model(s) also correspond well to the downslope aging indicated in both Figures 1 and 2.

[11] Four model structures are tested, ranging in number of tuned parameters from 3 to 6 (Table 1). Each of the four models explicitly capture the volume of water in the saturated zone, downslope fluxes of that water, as well as estimates of evapotranspiration and groundwater losses. Given the steep, responsive nature of the catchment, only advective transport of water and tracer is explicitly simulated, although as with all numerical models, a component of dispersion is introduced by the solution procedure itself. The models are differentiated on the basis of the inclusion or exclusion of an explicit representation of the unsaturated zone dynamics and a tracer specific retention term. Model 1 (3 parameters) excludes both the explicit unsaturated zone and the tracer parameter. Model 2 (4 parameters) includes the tracer parameter, but no explicit formulation of an unsaturated zone. Model 3 (5 parameters) includes an explicit formulation of an unsaturated zone but the tracer parameter. Finally, Model 4 (6 parameters) includes both additional complexities.

[12] Each of the independent model structures is evaluated under a Monte Carlo framework using a uniform distribution to randomly sample parameter values from within the feasible parameter range (Table 2). Our use of a uniform sampling strategy, as well as our framework for considering the prior information follows from *Beven and Freer* [2001] who advocated uniform sampling when detailed measurements of covariance were lacking, and also for the idea that observed and modeled parameters are scale-dependent, and that the scales of use may, in fact, not be the same. Our practical response to the latter point was to encompass measured values of observable parameters within the sampling distribution, but to expand the sampled range to allow for exploration of a somewhat wider variety of potential parameter values. Table 2 indicates those parameters for which measurements were utilized to constrain the prior parameter range.

[13] The volume of water within the saturated zone of each reservoir is accounted for as

$$\frac{dV_s}{dt} = I + SS_{in} - SS_{out} - SOF_{out} - K_d + EX_w \tag{1}$$

where V_s is the specific volume of water in each reservoir (m), t is current time (days), and I is the vertical exchange rate of water across the upper surface of the saturated zone. When the unsaturated zone is not explicitly accounted for (models 1 and 2), I is equivalent to the difference between the rainfall rate and an a priori estimate of the evapotranspiration rate (m/d). The ET estimates used in this study were originally available as daily totals based on the average monthly mean values derived from 5 different standard methods using meteorological data for 1987. The mean of these PET values were modified using a sine curve distribution between the hours of 06:00 and 18:00 (J. Freer, personal communication, 2005). For the set of models with an explicit unsaturated zone storage, I is equivalent to the recharge rate from the unsaturated zone, $K(\theta)$ defined below in (9). K_d is the loss to groundwater (m/d), here set to the measured yearly average of 100 mm on the basis of data presented by *Pearce et al.* [1986], SS_{out} is the rate of subsurface outflow from each reservoir (m/d), SS_{in} the rate of subsurface inflow (m/d), and SOF_{out} is the output rate of saturation excess overland flow (m/d). EX_w is active only as a component of models 3 and 4, which include an explicit unsaturated zone, and is simply set to zero in models 1 and 2. The term represents the volumetric exchange between the saturated and the unsaturated zones that occur as the water table adjusts, and its definition follows closely from *Seibert et al.* [2003] and *Weiler and McDonnell* [2004]. An increase in water volume results in an increase in the depth of the saturated zone, and a corresponding decrease in storage of the unsaturated zones. These depths are characterized by model parameters representing soil depth (SD) and porosity

Table 2. Minimum and Maximum Value for the Prior Distribution of Each of the Tuned Model Parameters[a]

Parameter	Units	Model Equation	Minimum Value	Maximum Value	Reference
PLE	-	(5)	1	20	-
θ_s, m/m	m/m	(4), (9)	0.1	0.6	*McDonnell* [1989]
K_s, m/d	m/d	(5), (9)	100	250	*McDonnell* [1989]
θ_r, m/m	m/m	(9)	0.01	0.15	-
λ	-	(9)	0.1	1	*McDonnell* [1989]
θ_{eff}, m	m	(10), (13)	0.01	0.7	-

[a]Monte Carlo sampling assumed a uniform distribution between these values.

(θ_s). It is defined in a slightly different fashion depending upon the direction of change. For rising water tables, water from the unsaturated zone is added to the saturated zone as

$$EX_w = \Delta l * \theta_u \tag{2}$$

where Δl is the change in the water table depth (m) during the time step and θ_u is the calculated water content in the unsaturated zone calculated as

$$\theta_u = \frac{V_u}{A * l_{unsat}} \tag{3}$$

where A is the surface area (m^2) and l_{unsat} is the depth of the unsaturated zone (m). For the case of a falling water table, water is taken from the saturated zone and added to the unsaturated zone as

$$EX_w = \Delta l * \theta_s \tag{4}$$

Transmissivity (T) is assumed to decline with depth as a power law, with the degree of decline modulated by a power law exponent (PLE). The decline is defined following from *Iorgulescu and Musy* [1997] as

$$T = \frac{K_s SD}{PLE} * \left(1 - \frac{z}{SD}\right)^{PLE} \tag{5}$$

where z is the depth to the water table measured from the soil surface (m). Subsurface inflows and outflows (SS_{in} and SS_{out}) follow from (5) and are independently defined for each grid cell. SS_{in} is based upon upslope neighbors, while SS_{out} is based upon downslope neighbors. In a single pass through the model grid (after each time step's solution to the mass balance differential equations), the values of SS_{in} and SS_{out} are calculated at each location as

$$SS_{in,i,j} = \sum_{k=0}^{k<9} T_{i,j,k} * Slope_{i,j,k} \qquad \text{for } \mathrm{Slope}_{i,j,k} > 0 \tag{6}$$

$$SS_{out,i,j} = \sum_{k=0}^{k<9} T_{i,j,k} * \left|Slope_{i,j,k}\right| \qquad \text{for } \mathrm{Slope}_{i,j,k} < 0 \tag{7}$$

where subscripts i and j indicate individual grid cells and k indicates the direction of the neighboring grid cell. Slope is referenced from the i,j cell, resulting in a negative value for downslope neighbors. It is calculated using water surface elevations. The i,j,k value of (5) is used to calculate SS_{in} and the i,j value of (5) is used to calculate SS_{out}. Infiltration excess overland flow has not been observed in the catchment, and so infiltration is simply assumed to occur when the soil is not saturated. A saturated overland flow mechanism is invoked for those areas and time periods where saturation occurs. In these instances, excess precipitation and SS_{in} is ponded and delivered directly to the stream network as SOF.

[14] In models 3 and 4, the unsaturated zone mass balance is defined as

$$\frac{dV_u}{dt} = I - K(\theta) - EX_w \tag{8}$$

where V_u is the specific volume of water in the unsaturated zone (m), I is the difference between the precipitation rate and the previously defined evapotranspiration rate, EX is the volumetric exchange as defined above, but with an opposite sign, and K(θ) is the unsaturated zone hydraulic conductivity (m/d), assumed to move in the vertical direction to recharge the saturated zone. This vertical movement is treated according to the Brooks-Corey relationship as

$$K(\theta) = K_s \left(\frac{\theta - \theta_s}{\theta_s - \theta_r}\right)^{\lambda} \tag{9}$$

where θ is the volumetric water content, θ_s is the residual water content, θ_r is the wilting point and λ is the pore size index.

[15] Modeled tracer simulations were used to develop the concentration breakthrough necessary to estimate the mean residence times of the simulations. The tracer model is defined by a set of mass balances representing an arbitrary conserved tracer. These equations are solved along with (1) and (8) using the Runge-Kutta procedure. In the case of the saturated zone, the mass balance is

$$\frac{dM_s}{dt} = IC_e + SS_{in}C_{in}\theta_{eff} - SS_{out}C_{out}\theta_{eff} + EX_m \tag{10}$$

where M_s is the tracer mass (kg), I is the rainfall rate for models 1 and 2 or transfer from the unsaturated zone (K(θ)) for models 3 and 4, C_e is the concentration of tracer (kg/m^3) in rainfall (models 1 and 2) or the concentration in the unsaturated zone (models 3 and 4), SS_{out} and SS_{in} are the subsurface water flux rates out of and into the reservoir from equations (6) and (7), C_{out} is represented as the concentration in the saturated zone (kg/m^3) from the previous time step, C_{in} is the concentration of tracer into the reservoir (kg/m^3), taken from the previous time from upslope grid cells, and θ_{eff} is the effective porosity fraction indicating that solute fluxes interact with a volume of water smaller than the total available volume. This simple parameterization relaxes the assumption that effective mixing volumes of tracer and water are the same, and plays a potentially important role in the estimation of MRT because the assumption is generally not correct [*Iorgulescu et al.*, 2004]. EX_m is the mass exchange between the saturated and unsaturated zone that occurs as the water table fluctuates (kg/d). The term is analogous to the water mass balance exchange term (EX_w) defined above. For rising water tables,

$$EX_m = EX_w * \frac{M_u}{V_u} \tag{11}$$

and for falling water tables,

$$EX_m = EX_w * \frac{M_s}{V_s} \tag{12}$$

For the models that include an explicit unsaturated zone mass balance, the unsaturated zone tracer mass balance is defined as

$$\frac{dM_u}{dt} = IC_e - K(\theta)C_u\theta_{eff} - EX_m \tag{13}$$

where M_u is the mass of tracer in the unsaturated zone (kg) and C_u is the concentration of tracer in the unsaturated zone (kg/m^3).

3.2. How MRT Is Calculated Using Field Data and Within the Model

[16] *Stewart and McDonnell* [1991] used environmental tracers (i.e., ^{2}H) of input and output to estimate the residence time (using standard techniques described by *Maloszewski and Zuber* [1996], *Turner and Barnes* [1998], and *McGuire and McDonnell* [2006]). This approach assumed that tracer composition of precipitation that falls on a catchment would be delayed by some timescale(s) according its physical properties and current state. The stream outflow composition at any time consists of past inputs lagged according to their travel time distribution (see review of this methodology by *Maloszewski and Zuber* [1996]). The travel time or residence time distribution (RTD) then describes the fractional weighting of how mass (i.e., tracer) exits the system, which is equivalent to the probability density function or transfer function of the sampled tracer. If the tracer is conservative and accesses the entire catchment volume (an assumption of complete mixing) then the tracer RTD is equivalent to the water RTD. The definition of residence time that we use in this work is "the time elapsed since the water molecule entered the catchment as recharge to when it exits at some discharge point" (i.e., catchment outlet, monitoring well, soil water sampler, etc.) [*Eriksson*, 1971; *McGuire and McDonnell*, 2006].

[17] The tracer-based convolution integral approach is the standard tool for the calculation of residence time distributions and values representing the MRT of catchments. However, the catchment modeler who often works within the confines of something similar to equations (1) and (8), requires a mechanism to evaluate residence times of different individual conceptual simulations, unrelated to the isotope-based approach. The direct simulation of MRT is well established in the groundwater literature [*Goode*, 1996; *Etcheverry and Perrochet*, 2000]. Its calculation requires the inclusion of a conservative tracer model as outlined in equations (10) and (13). For the special case of a spatially uniform impulse injection of tracer, the concentration breakthrough is a representation of the residence time distribution of molecules in the system [*Danckwerts*, 1953]. In the case of a completely mixed system, this residence time distribution is an exponentially decaying function where the concentration of the outflow is equivalent to the concentration within the reservoir, which decreases monotonically because of mass loss and the input of zero concentration fluid. For a piston flow system, the residence time distribution is a Dirac delta function. For the case of a distributed system of instantly mixed boxes (which is the case here) the distribution is expected to deviate from these extremes [*Haitjema*, 1995].

[18] The mean residence time can be derived by this concentration breakthrough and is defined as

$$MRT = \frac{\int_0^\infty tcdt}{\int_0^\infty cdt} \tag{14}$$

where c is breakthrough concentration and t is time. The numerator is the first moment (concentration weighted average) of the tracer distribution and the denominator is the zeroth moment, or total mass [*Goode*, 1996]. In our simulations, we apply (14) to estimate the MRT of each unique location in the saturated zone (where the solution to equations (1) and (10) are combined to produce the concentration breakthrough at each location) and within the catchment steam network. In the case of the saturated zone, the calculated MRT of any single grid element reflects the flow path distribution of all upslope grid elements because of the assumption that water moves laterally in the saturated zone. While concentration breakthrough curves for each unsaturated zone element in models 3 and 4 could also be used to estimate unsaturated zone MRT, we explicitly assume vertical water movement in the unsaturated zone, and therefore no upslope connection. As this work focuses on the observed downslope aging of water near the soil bedrock interface, no analysis of unsaturated zone MRT is included.

[19] Strictly speaking, the MRT defined by (14) is equivalent to that defined through convolution only when the direct simulation incorporates the same flow path distribution as is incorporated by the isotope-based procedure, a top-down estimate of the true flow path complexity within the catchment. The environmental tracers used in the convolution approach access the full catchment volume and, more importantly, potentially reflect zones of immobility. Our use of an effective tracer porosity fraction is designed to reflect this immobility, however the catchment model is clearly a simplification that is not designed to incorporate the full flow path distribution. Our goal is evaluate the degree to which the simplification affects model residence time. If we can establish that the differences are large, we can then reject the model and use that as a sound basis to iteratively incorporate additional complexity.

3.3. Evaluative Criteria

[20] We evaluate the models using stream discharge and stream water residence times at the outlet of the M8 catchment. In addition, we utilize spatially distributed estimates of soil water residence time estimates as further evaluative criteria. Unlike the catchment MRT, soil water MRT is represented by a distribution, where the values that make up the distribution are MRT estimates calculated at each unique landscape location (or grid square). Here we simplify the comparison between measured and modeled soil water MRT by examining only the maximum range for each simulation. Discharge efficiency is defined as the Nash-Sutcliffe measure (R_{eff})

$$R_{eff} = -\frac{\frac{1}{n}\sum_{t=0}^{t=n}\left(d_t - o_t(\theta)\right)^2}{\frac{1}{n}\sum_{t=0}^{t=n}\left(d_t - \overline{d}\right)^2} \tag{15}$$

where n is the number of observations, t is time, d is the measured value of discharge, $o_t(\theta)$ is the modeled value of discharge, given the parameter array θ, and $\overline{d}$ is the average measured runoff over the observed time period. Both stream water MRT and the range in soil water MRT are represented

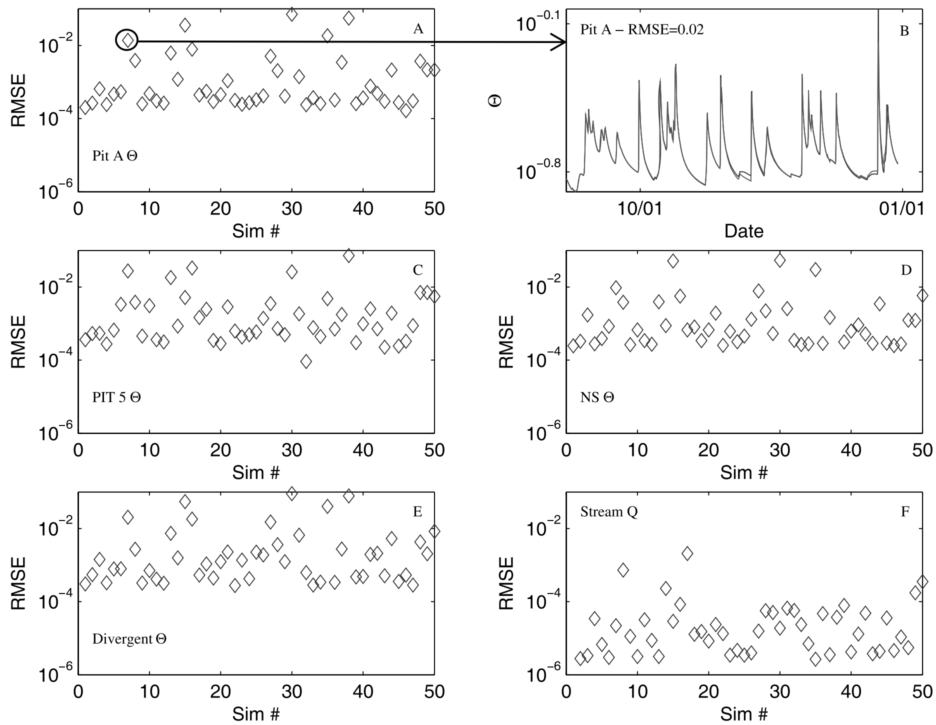

Figure 3. Stability analysis comparing results from five locations for a time step of 1.44 min against a time step of 14.4 min for 50 randomly sampled parameter vectors. Results are based upon simulated average water content (averaged across the saturated and unsaturated zone) (a, b, c, d, and e) at different grid cells within the model domain and (f) on stream discharge at the catchment outlet. The reader is referred to Figure 2 (top) for the locations of each point. Figure 3b represents the average water content for each of the two time steps during simulation 9 at Pit A (indicated by the arrow). It is included as a reference to assist in interpretation of the magnitude of the RMSE.

as scalar quantities, and are therefore reported as a percentage of measurement value.

4. Results

4.1. Stability Analysis

[21] Rather than developing a finite difference solution to a PDE representing the water in time and space, the model is cast as a set of mass balance equations that vary only in time, but utilize the grid to independently estimate incoming and outgoing horizontal fluxes. The resulting set of equations is solved simultaneously at each time step to update the volume of water (V_s), and mass of tracer (M_s) in the saturated zones, at each location. For two of the model structures, these volumes and masses are also maintained for the unsaturated zone (V_u and M_u respectively). Given the updated volumes, water table depths, material transfers between the saturated and unsaturated zones, saturation deficits, exfiltration, as well as unsaturated zone fluxes are then updated to reflect current conditions. A fifth-order adaptive time step Runge-Kutta solution following from *Press et al.* [1992], was utilized to solve the system of equations. As with all numerical models the temporal dynamics, spatial discretization and model time steps must balance with computational limitations and numerical stability. We accomplished this through an evaluation of model results across the simulation time period and given different model time steps. Initial exploration using the variable time step procedure indicated a time step of 0.01 days resulted in a stable solution, across a sample of the prior parameter range and model structures. Further exploration of the stability of this time step is presented in Figure 3. Here we present results for the hydrologic components of models 3 and 4, which given the increased vertical discretization, would likely be the least stable of the model structures. Fifty parameter vectors were randomly selected (on the basis of Table 2) and for each, the model was run three times, with a progressively smaller time step (0.01, 0.001 and 0.0001 days). Comparisons between the longer time step and 0.001 days were made at different spatial locations, and summarized as a root mean square error (RMSE) (Figure 3). To assist in the interpretation of the RMSE, the time series responsible for a relatively large error of 0.02 (in 3A) are included (in 3B). Differences between these simulations are apparent, but relatively minor, with the majority of the error associated with the December peak flow event. Given these analyses we elected to utilize a constant time step of 0.01 days as a reasonable balance between the requirement for numerical stability and the computational limitations imposed through our use of Monte Carlo simulations involving multiple distributed models. This constant value was used in place of the variable step as it removed the potential for the development of large time steps over drier periods, and the potential for numerical instability during subsequent rain events.

4.2. General Findings

[22] Input data at Maimai were available from 3 September to 30 December 1987 (Figure 4). While the period of

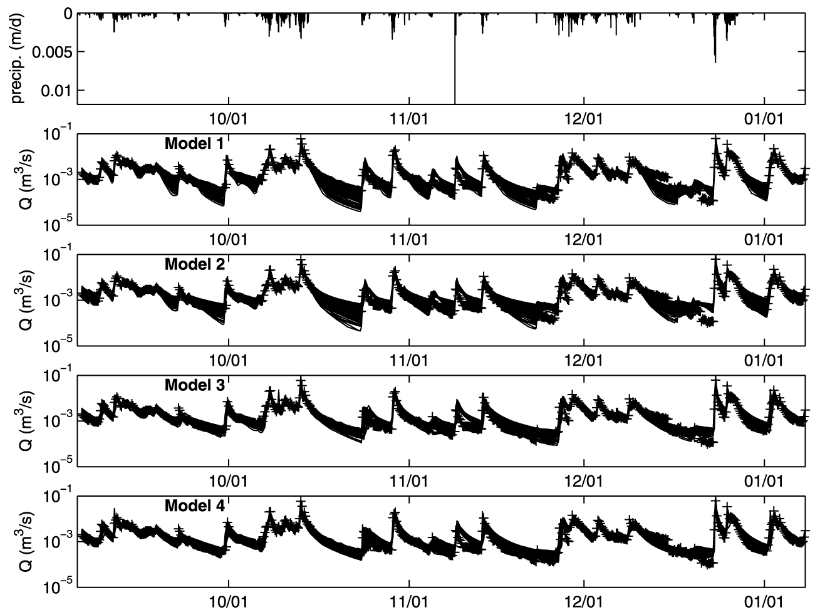

Figure 4. Simulation results for stream discharge. The *y* axis is log transformed to outline more clearly model inconsistencies at lower discharge. The calibration strategy focused on untransformed R_{eff}, and peak flows are correspondingly better captured. The plotted simulations are those found with Nash-Sutcliffe efficiencies over 0.75. Measured values are plotted as crosses.

record was suitable to establish reasonable discharge efficiencies, it was not enough time to result in complete tracer recovery, a necessary condition to calculate MRT values. To accommodate this longer-timescale process, and given the clear lack of seasonality at Maimai, we doubled the record length by appending each data set (20 min rainfall and discharge and hourly ET estimates) to its end (Figure 4). The resulting 7 month simulation period provided enough time for complete tracer recovery, and estimates of MRT. The total number simulations for each model structure varied—models with larger numbers of parameters included a larger number of total runs. For the saturated zone models (models 1 and 2), initial soil saturation was assumed to be 50%. In the case of the coupled saturated-unsaturated models (models 3 and 4), the water table was placed initially at 50% of the soil depth and the unsaturated zone water content was set to 50%. In both instances the first 30 days of simulation were allocated toward the stable redistribution of the initial conditions, with model efficiencies calculated on the basis of results after that point in time. Tracer application occurred at the first time step on the 30th day of simulation. At that point in time, the concentration of tracer in soil water and stream water was set to 50 mg/L and no additional tracer was added to the system.

[23] Parameter data, efficiencies, stream water mean residence times and minimum and maximum soil water residence were collected at 2.4 hour intervals for simulations with discharge efficiencies of untransformed discharge over 0.0. Time series data representing modeled discharge and tracer breakthrough for those simulations over 0.75 discharge efficiency were also collected (Figures 4 and 5, respectively). The more restrictive threshold ($R_{eff} = 0.75$) was utilized for time series collection to highlight the relationship between tracer breakthrough and discharge for a relatively small set of highly efficient discharge simulations.

[24] All models were essentially equally capable of simulating discharge dynamics (Figure 4). However, these efficient discharge simulations showed significant differences in tracer breakthrough curves (Figure 5) and hence MRT. For models 1, 2, and 3, the simulations with $R_{eff} > 0.75$ all resulted in consistently fast tracer breakthroughs. However, tracer breakthroughs for the set of high R_{eff} in simulations of model 4 were considerably more variable. For some simulations, tracer breakthroughs had lower peak concentrations and longer tails.

[25] These relationships are explored more clearly through an analysis of parameter space. Parameter scatterplots for each model are shown in Figures 6, 7, 8, and 9. A binary classification scheme indicates parameter vectors that simulate greater than and less than 50% stream water MRT values. Multicriteria efficiency plots, including both stream MRT and the range of soil MRT against R_{eff}, again for each of the four models, are presented in Figure 10. Models plotting in the upper right corner of Figure 10 are those simulations that result in $R_{eff} > 0.6$, and stream and soil MRT values within 50% of measurements. The acceptance of any models with simulated MRT within 50% of measured value is rather generous; however given the uncertainty in these soft data sources, we view this value as appropriate.

4.3. Saturated Zone Models

4.3.1. Model 1

[26] The first two models did not incorporate an explicit unsaturated zone, in each case reducing the number of tuned parameters by 2. Model 1 was able to reproduce the discharge response of the catchment, with maximum R_{eff} values of over 0.85. Parameter plots showed some identifiability in θ_s and the PLE, but indicated that none of the parameter vectors resulted in stream MRT within 50% of the measured values. In the event that discharge alone was

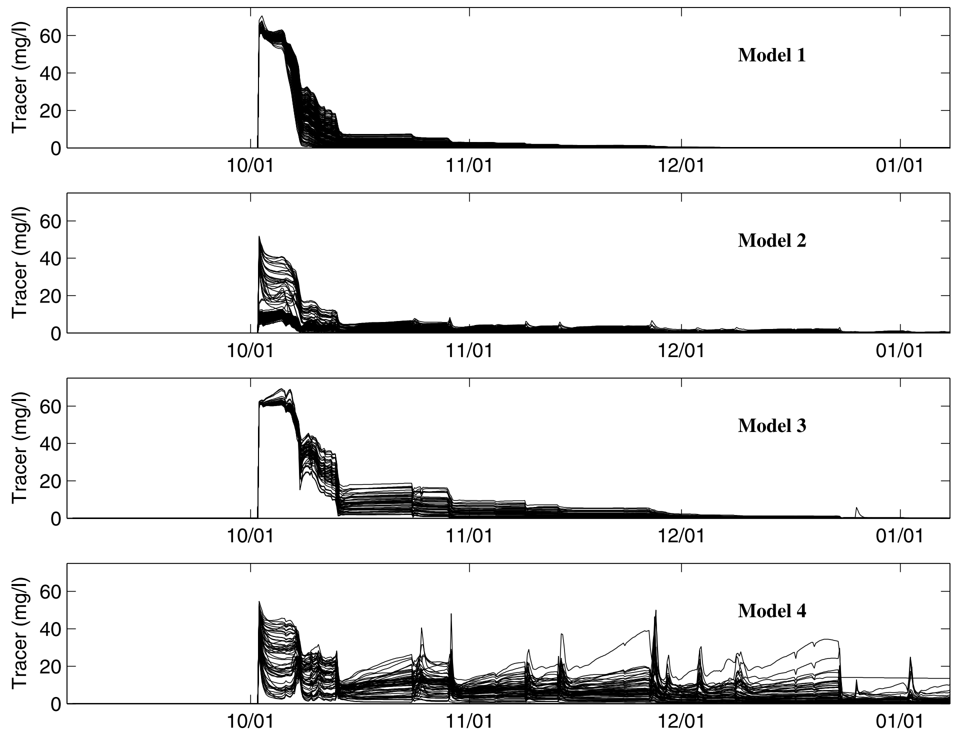

Figure 5. Tracer breakthrough curves used to calculate mean residence time. A constant concentration of 50 mg/L was added to all modeled reservoirs on 1 October. These curves represent the breakthrough concentrations, in stream water, at the catchment outlet. Models were run until 1 May to capture near-complete breakthrough. These plots are truncated at 10 January to better outline differences.

the evaluative criteria, this relatively parsimonious model with its effective simulations of discharge, might be considered an adequate model structure. However, the inclusion of estimates of MRT as evaluative criteria provided another perspective on the transport component of the model. The tracer response of model 1 was rapid (Figure 5), with maximum values of stream MRT for behavioral discharge simulations of 40 days. The range in soil water MRT was 1 to 42 days. This meant that for the longest MRT simulations, the entire catchment volume contributed within 42 days to catchment discharge. Unlike the wave celerity, the transport behavior is fundamental to solute disposition in the

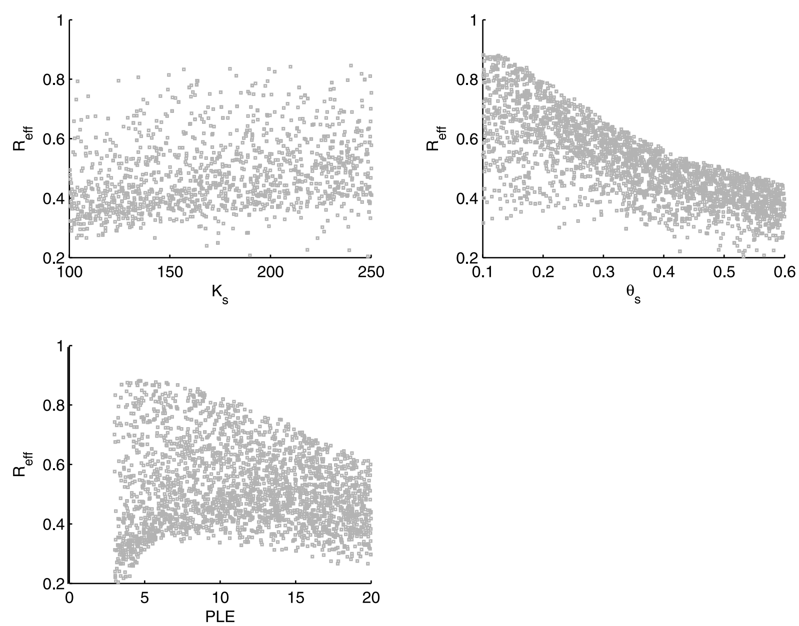

Figure 6. Parameter space for model 1. No simulations resulted in stream MRT that were within 50% of the measured values.

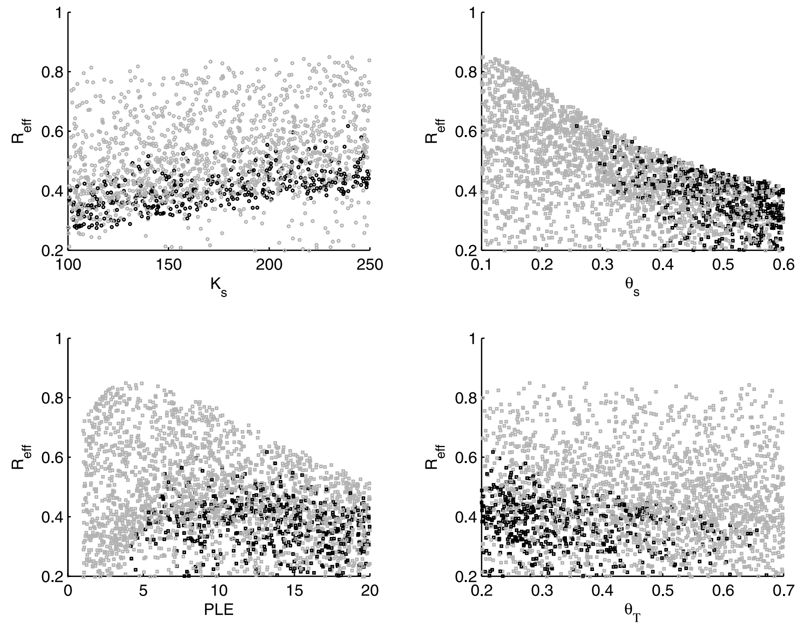

Figure 7. Parameter space for model 2. Dark models have MRT within 50% of measured values.

catchment. The fact that MRT across the prior parameter range did not approximate either the magnitude or range of measured MRTs indicates clearly that the three parameter model, while acceptable for discharge, is generally unacceptable given is inability to capture the observed catchment MRT.

4.3.2. Model 2

[27] We hypothesized that model 2 would be likely to produce MRT values longer than model 1 because of the distinction between dynamic catchment volume responsible for the discharge response and more restrictive volume available in the tracer response with the added parameter. Figure 7 indicates that despite movement toward longer MRT, maximum values of both measures were within only ~65% of the measurement values, and the longest MRT values occurred only for relatively poor simulations of discharge.

[28] A clear tradeoff exists between discharge efficiency and MRT for models 1 and 2. The inability of both models to acceptably capture the dynamics of the soil and stream MRT in the catchment (within a single set of parameters), lead to the conclusion that both models represent an overly gross simplification of catchment processes. Given these results it is clear that within this general gridded model

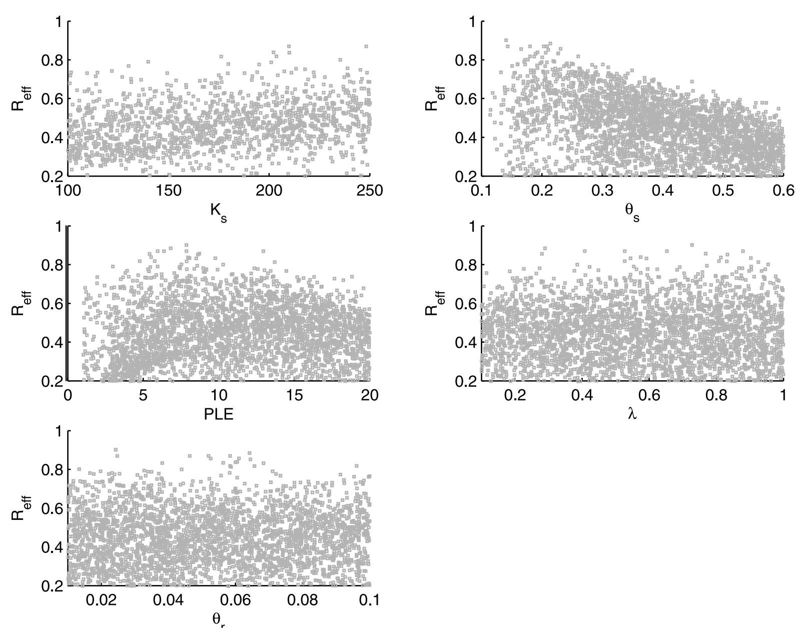

Figure 8. Parameter space for model 3. Similar to results for model 1, no simulations resulted in MRT values within 50% of the measured value.

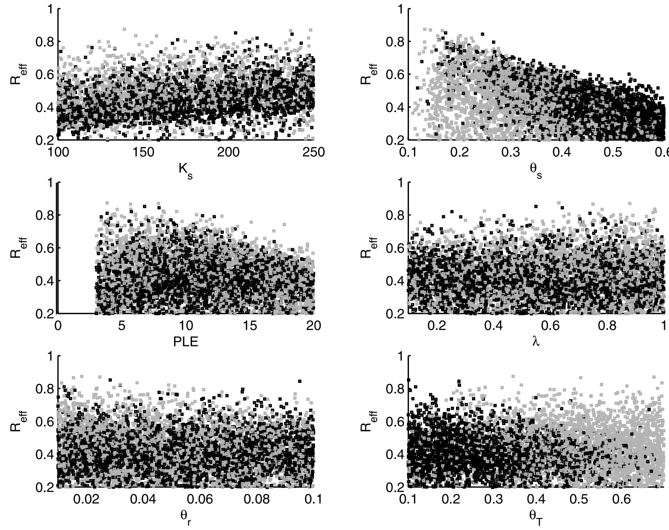

Figure 9. Parameter space for model 4. Dark models have MRT within 50% of measured value. A tradeoff exists between higher values of porosity, longer residence times, and highly efficient discharge simulations.

structure, additional model complexity (and therefore parameters) is necessary to capture the MRT and discharge dynamics of the catchment.

4.4. Coupled Saturated and Unsaturated Zone Models

4.4.1. Model 3

[29] The explicit incorporation of the unsaturated zone using equation (8) increased the number of parameters, but also added a physically realistic volume of storage, the unsaturated zone, with generally reduced transport velocities. This additional volume therefore offers the potential for relatively long residence time. Figure 8 represents the parameter space for this model. The patterns for k_s, θ_s and PLE were similar to models 1 and 2, with at least a clear reduction in the width of the posterior parameter range given the MRT criteria. The two parameters in the unsaturated zone equations were less well identified. The lengthening of the tracer breakthroughs (Figure 5) suggests that

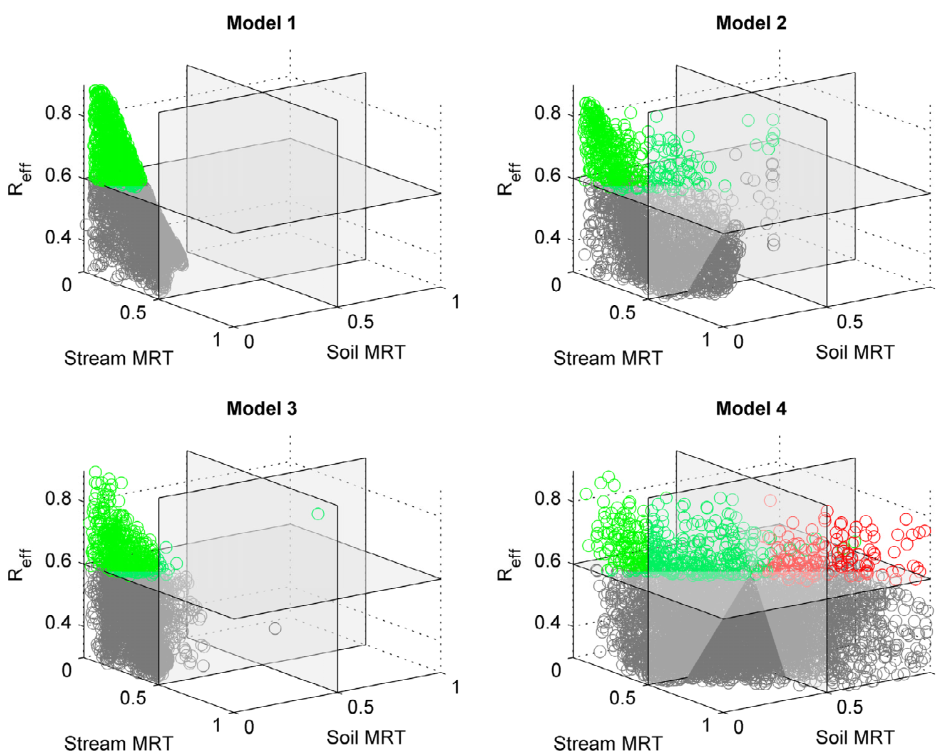

Figure 10. Multicriteria plots of R_{eff}, stream MRT, and the range in soil MRT for each of the four models. Note that only model 4 results in a set of simulations in the upper right octant, indicating $R_{eff} > 0.6$, and stream MRT and soil MRT range within 50% of the isotopically derived values.

the unsaturated zone did result in increased retention of tracer compared to model 1. This result is further outlined in Figure 8 where the number of simulations with longer MRT approximations increases. However, unlike model 2, no simulations resulted in $R_{eff} > 0.6$, and soil and water MRT > 0.5. This decrease in the model structure's capacity to capture the pressure and transport response suggests that while the unsaturated zone is an effective mechanism to increase the MRT of simulations, it is nevertheless insufficient to acceptably differentiate between the dynamic and total storage volumes.

4.4.2. Model 4

[30] The addition of the effective porosity with an explicit representation of the unsaturated zone dynamics (model 4) results in the only model, of the four evaluated here, where discharge, stream MRT and soil MRT are all acceptably reproduced by individual parameter vectors (Figure 9). This result supports the argument that the complexity involved in the inclusion of the both the explicit unsaturated zone and a mechanism to differentiate between dynamic and total storage represents one model structure that successfully reproduces discharge dynamics, stream MRT and an appropriate range in the soil MRT.

4.5. Parameter Uncertainty

[31] While this paper focuses more on the rejection of unacceptable model structures, it is worthwhile to note the potential constraints on model parameter uncertainty, through the incorporation of these complementary data. Models 1 and 3 were unable to successfully reproduce either soil or stream water MRT, leading us to reject them. Models 2 and 4, however, did successfully reproduce the measures. The production of acceptable MRT values is doubly positive. First, the model structures are potentially acceptable. More than that, however, the additional criteria significantly constrain the posterior parameter distribution for some of the parameter values. Using the simple binary threshold to evaluate stream MRT (Figures 7 and 9) usefully indicates that the both the total porosity and the effective porosity parameter are significantly constrained by these additional observations. Larger values of total porosity tend to result in improved estimates of MRT and smaller values of effective porosity increase the domain of immobility, also resulting in improved MRT values.

5. Discussion

5.1. Stream Water Residence Time and Soil Water Residence Time As New Catchment Diagnostics

[32] Multicriteria model calibration studies have been completed in recent years involving saturated area mapping [*Franks et al.*, 1998], groundwater levels [*Kuczera and Mroczkowski*, 1998; *Seibert et al.*, 1997], more complete exploitation of the information content in discharge [*Gupta et al.*, 1998; *Boyle et al.*, 2000], time source components of storm flow [*Seibert and McDonnell*, 2002; *Vaché et al.*, 2004], and geographic source components [*Scanlon et al.*, 2001] as integrative multi criteria tools. However, the development of evaluation criteria comparable to that of a discharge–that are both integrative and scalable—has remained elusive. Also, relatively little guidance has been given in the literature to date on what measures might best constrain a realistic simulation of flow and transport. We argue in this paper that the combination of flow, stream MRT and, quite possibly, soil water residence time provide a meaningful test of hydrologic models that are extended to address water quality related concerns. A small set of models, each progressively more complicated, was tested against these observations. While many of the decisions regarding these structures and their acceptability were subjective, they were designed with a perceptual model and process scale of interest in mind. It is not entirely surprising that the more highly parameterized models were the most successful—the increased degrees of freedom and effective mixing parameter suggest a wider range of potential results. However, our tuned parameter numbers were modest compared with many hydrologic models. More importantly, the simpler models could not reproduce MRT, and we can identify clearly the point at which an increasing degree of complexity could. In this case, model 4 was a successful stopping point because it met our initial criteria: the perceptual model and scale of interest were captured, and the discharge, stream MRT and soil MRT were reproduced to within our criteria of acceptability.

[33] We do not suggest that the mechanism we have outlined to calculate model-simulated MRT is unique—in fact it is very well established in the groundwater field [*Goode*, 1996]. However, since discharge is a weak test of model structure we argue that calculating the MRT of simulations provides a mechanism to characterize the degree to which the effects of process complexity are incorporated into a model structure, in a way that the pressure response data alone does not. The tracer response captured by MRT, like discharge, is an integrative measure. However, unlike discharge, the MRT is sensitive not to the pressure response of a catchment, but rather to linear transport velocities. The two quantities represent very different information and these complimentary measures of catchment behavior are key to our evaluation and rejection of model structures.

5.2. Incorporation of Residence Time Into the Model Structure

[34] Other studies have examined water residence time in a catchment model context. *Lindström and Rodhe* [1986] developed a lumped tracer model calibrated against measured $\delta^{18}O$ in the Gardsjön catchment, and estimated the transit time distribution using calibrated results. *Bergström et al.* [2002] used a very similar model within a multicriteria calibration framework and concluded that the additional information content available from detailed continuous simulations of $\delta^{18}O$ significantly improved model performance. *Uhlenbrook and Leibundgut* [2002] utilized observationally based estimates of residence time in an a priori manner to identify homogenous model units for use within the TAC modeling framework. *Melhorn and Leibundgut* [1999] incorporated estimates of base flow residence time into the calibration of a hydrologic model, but began with model structures that explicitly incorporated residence time and turnover time as model parameters. *Scanlon et al.* [2001] report groundwater residence time values on the basis of simulations using a modified TOPMODEL. They used an estimate of catchment turnover (t = V/Q) in their calculations but did not have measurements against which to test the modeled turnover times and unlike MRT, turn-

over time is insensitive to transport velocities [*Eriksson*, 1971].

[35] In the present paper, we have extended the use of simulated tracers and residence time to evaluate the degree to which a set of models captures both the pressure and transport response of the catchment. Most importantly, our use of a distributed model resulted in the capacity to evaluate stream and soil water MRT independently. These data also correspond to measurements derived through the sampling of both stream and soil water by *Stewart and McDonnell* [1991]. The high degree of uncertainty behind our very basic regionalization procedures kept us from incorporating the spatial distribution of precise soil MRT as additional criteria, but we did attempt to incorporate these estimates through the use of the range of measured soil MRT. While this line of reasoning seems to have potential to add necessary information to the parameterization of upslope regions, our focus was on a relatively simple spatially homogeneous parameterization. Long-stream MRT simulations were very much correlated with the range in soil MRT, as reflected in the lack of simulations within the +SMRT and +STRM zone of Figure 10. More complete advantage of these data might be taken in efforts to spatially distribute parameters, most notably those parameters that significantly affect the spatial distribution of storage – soil depth and porosity. As suggested by *Freer et al.* [2002], the distribution of soil depths may exert a controlling influence on water flux and MRT. While in some sense soil depth and the distribution of porosity might be considered measurable quantities, there is a large degree of uncertainty about them. Given this a priori uncertainty, we see the spatial distribution of MRT as one potential mechanism better constrain these spatially distributed parameter values. Nevertheless, the results presented here suggest that in the common case of spatially homogenous parameterizations (in either lumped or distributed simulations) the incorporation of stream water MRT alone has the potential to constrain significantly the a posteriori parameter distribution. Given, in addition, the relative simplicity of isotopically estimating stream MRT, we suggest that for modeling studies undertaken in the many catchments with existing stream MRT estimates – but without corresponding soil MRT estimates, the lack of the latter should not deter the use of the former.

5.3. Outlook for the Future

[36] This study evaluated four related modeled structures. The idea that modelers should use observation to reject models and suggest alternatives is a very useful concept, but clearly a wide variety of alternatives might be as effective as those we evaluated here. We do not suggest that model 4 is appropriate for all catchments or forcing conditions. Instead, we view these results as relevant to the hydrologic community for two related reasons: (1) not all model structures are appropriate for all situations. The treatment of models as hypotheses, with a specific goal of falsification, is a useful mechanism to incorporate into studies, and it has the potential, as is the case here, to identify a model structure that reproduces observation with a minimal number of parameters, and (2) MRT is an effective metric to add to traditional calibrations—both single and multiple criteria, when water chemistry is of interest.

[37] We selected Maimai because of its simplicity. The very long flow paths, or residence times, that are often estimated from isotope data [*McGuire et al.*, 2005; *Kirchner et al.*, 2001] do not dominate the distribution at Maimai. This is due to the wet nature of region, the steep short slopes, and the effective impermeable bedrock barrier at an average of 0.6 meters. For this reason, the study did not place emphasis on the potential difference in the degree to which the original isotopic sampling focused on low versus high flows, and hence implicitly on long-term or short-term residence times.

[38] Isotopic or chemical estimates of MRT are resource intensive values that are clearly not available at most catchments. Certainly estimates of discharge are more readily available historically and are relatively more straightforward to collect. Here we outline a mechanism for incorporating estimates of MRT into a hydrologic modeling framework. We see the collection of these data as increasingly routine. In fact, inexpensive laser mass spectrometry that can be field deployed for continuous isotope measurement will soon be commercially available (the DLT-100 H_2O isotope analyzer from Los Gatos Research, San Jose CA provides simultaneous measurements of $\delta^{17}O$, $\delta^{18}O$, δD at better than 0.5‰, 0.3‰, and 2.0‰ accuracy respectively using off-axis integrated cavity output spectroscopy). Instruments such as this have the potential to revolutionize the collection of the isotopic data upon which these estimates rely. The regionalization of catchment MRT is feasible, given different relationships with various morphometric landscape descriptions [*McGlynn et al.*, 2003; *McGuire et al.*, 2005]. These relationships provide the potential for MRT estimates for a wide variety of different catchments that might successfully be generated from a relatively few number of increasingly routine measurements.

6. Conclusions

[39] Improvements in mesoscale predictions will come when we devise new ways to capture process detail into more integrative measures [*Sivapalan*, 2003]. To date, experimentalists have not done a very good job at coming up with quantities that are both integrative and scalable, however MRT is one such measure. Our presentation of a set of simple distributed hydrologic models and evaluation of the MRT of the model subunits within a Monte Carlo framework outlines a large set of models that are acceptable for flow, but may be wholly rejected for an inability to capture residence time dynamics. This process-based rejectionist approach utilizes the Monte Carlo framework to evaluate the identifiability of parameters and how values of MRT contribute to the evaluative process and ultimately, the level of model complexity warranted in the model structure. We argue in this paper that residence time may provide a measure of flow path heterogeneity useful in model structural evaluation and testing. Since conceptual, physically based models are designed to reflect, with varying degrees of complexity, the main stocks and flows of water through catchments, a model that correctly captures discharge and water residence time is more realistic than one that captures only the former. More importantly, in some cases a model can perform reasonably well when evaluated for discharge alone, but additional compositional criteria can result in rejection of the model structure itself, as was demonstrated for three of the model structures we

evaluated at Maimai. The incorporation of residence time into evaluation procedures is one mechanism to help understand the limitations of conceptual simulations with water quality sensitive flow paths, and to independently assess the need to incorporate additional process detail or heterogeneity.

[40] **Acknowledgments.** This research was funded in part by USFS contract PNW 02-CA-11261930-011 and by NSF grant EAR-0196381. We thank Jan Seibert, Markus Weiler, and Brian McGlynn for useful discussions on Maimai and our conceptual model and Kevin McGuire, Sarah Dunn, and Mario Martina for dialog related to residence time and its physical meaning. Review comments from Keith Beven and two anonymous referees contributed significantly to the text.

References

Asano, Y., T. Uchida, and N. Ohte (2002), Residence times and flow paths of water in steep unchannelled catchments, Tanakami, Japan, *J. Hydrol.*, *261*, 173–192.

Atkinson, S. E., R. A. Woods, and M. Sivapalan (2002), Climate and landscape controls on water balance model complexity over changing timescales, *Water Resour. Res.*, *38*(12), 1314, doi:10.1029/2002WR001487.

Barnes, C. J., and M. Bonell (1996), Application of unit hydrograph techniques to solute transport in catchments, *Hydrol. Processes*, *10*, 793–802.

Beck, M. B., F. M. Kleissen, and H. S. Wheater (1990), Identifying flow paths in models of surface water acidification, *Rev. Geophys.*, *28*, 207–230.

Bergström, S., G. Lindström, and A. Pettersson (2002), Multi-variable parameter estimation to increase confidence in hydrological modeling, *Hydrol. Processes*, *16*, 413–421.

Beven, K. (1989), Changing ideas in hydrology: The case of physically-based models, *J. Hydrol.*, *105*, 157–172.

Beven, K. (2001a), Calibration, validation and equifinality in hydrological modelling: A continuing discussion, in *Model Validation: Perspectives in Hydrological Science*, edited by M. G. Anderson and P. D. Bates, John Wiley, Hoboken, N. J.

Beven, K. (2001b), How far can we go in distributed hydrological modelling?, *Hydrol. Earth Syst. Sci.*, *5*, 1–12.

Beven, K. (2001c), On hypothesis testing in hydrology, *Hydrol. Hydrol. Processes*, *15*, 1655–1657, doi:10.1002/hyp.436.

Beven, K. (2001d), *Rainfall-Runoff Modeling: The Primer*, 360 pp., John Wiley, Hoboken, N. J.

Beven, K. (2002), Towards a coherent philosophy for modeling the environment, *Proc. R. Soc. London, Ser. A*, *458*, 2465–2484.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*, 11–29.

Blöschl, G. (2001), Scaling in hydrology, *Hydrol. Processes*, *15*, 709–711.

Burns, D. A., et al. (2003), The geochemical evolution of riparian ground water in a forested piedmont catchment, *Ground Water*, *41*, 913–925.

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, *36*, 3663–3674.

Danckwerts, P. V. (1953), Continuous flow systems: Distributions of residence times, *Chem. Eng. Sci.*, *2*, 1–13.

DeGrosbois, D., R. Hooper, and N. Christophersen (1988), A multi-signal automatic calibration methodology for hydrochemical modeling: A case study of the Birkenes model, *Water Resour. Res.*, *24*, 1299–1307.

Duan, Q., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*, 1015–1031.

Eriksson, E. (1971), Compartment models and reservoir theory, *Annu. Rev. Ecol. Syst.*, *2*, 67–84.

Etcheverry, D., and P. Perrochet (2000), Direct simulation of groundwater transit-time distributions using the reservoir theory, *Hydrogeol. J.*, *82*, 200–208.

Franks, S. W., P. Gineste, K. J. Beven, and P. Merot (1998), On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, *Water Resour. Res.*, *34*, 787–797.

Freer, J., J. J. McDonnell, K. J. Beven, N. E. Peters, D. A. Burns, R. P. Hooper, B. Aulenbach, and C. Kendall (2002), The role of bedrock topography on subsurface storm flow, *Water Resour. Res.*, *38*(12), 1269, doi:10.1029/2001WR000872.

Freer, J. E., K. J. Beven, and N. E. Peters (2003), Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 69–87, AGU, Washington, D. C.

Freer, J. E., H. McMillan, J. J. McDonnell, and K. J. Beven (2004), Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, *J. Hydrol.*, *291*, 254–277.

Goode, D. J. (1996), Direct simulation of groundwater age, *Water Resour. Res.*, *32*, 289–296.

Grayson, R., and A. Western (2001), Terrain and the distribution of soil moisture, *Hydrol. Processes*, *15*, 2689–2690.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763.

Haitjema, H. M. (1995), On the residence time distribution in idealized groundwatersheds, *J. Hydrol.*, *172*, 127–146.

Hewlett, J. D., and A. R. Hibbert (1967), Factors affecting the response of small watersheds to precipitation in humid areas, in *Forest Hydrology*, edited by W. E. Sopper and H. W. Lull, pp. 275–291, Elsevier, New York.

Hooper, R. P. (2001), Applying the scientific method to small catchment studies: A review of the Panola Mountain experience, *Hydrol. Processes*, *15*, 2039–2050.

Hooper, R. P., B. T. Aulenbach, D. A. Burns, J. McDonnell, J. Freer, C. Kendall, and K. Beven (1998), Riparian control of stream-water chemistry: Implications for hydrochemical basin models, in *Hydrology, Water Resources and Ecology in Headwaters*, edited by K. Kovar et al., pp. 451–458, IAHS Press, Wallingford, U. K.

Iorgulescu, I., and A. Musy (1997), Generalization of TOPMODEL for a power law transmissivity profile, *Hydrol. Processes*, *119*, 1353–1355.

Iorgulescu, I., K. Beven, and A. Musy (2004), Data-based modeling of runoff and chemical tracer concentrations in the Hayte-Mentue research catchment (Switzerland), *Hydrol. Processes*, *18*, 2557–2573, doi:10.1002/hyp.5731.

Kirchner, J. W., X. Feng, and C. Neal (2001), Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations, *J. Hydrol.*, *254*, 82–101.

Kuczera, G., and M. Mroczkowski (1998), Assessment of hydrologic parameter uncertainty and the worth of multiresponse data, *Water Resour. Res.*, *34*, 1481–1489.

Lindström, G., and A. Rodhe (1986), Modelling water exchange and transit times in till basins using oxygen-18, *Nord. Hydrol.*, *17*(4–5), 325–334.

Maloszewski, P., and A. Zuber (1993), Principles and practice of calibration and validation of mathematical models for the interpretation of environmental tracer data, *Adv. Water Resour.*, *16*, 173–190.

Maloszewski, P., and A. Zuber (1996), Lumped parameter models for the interpretation of environmental tracer data, in *Manual on Mathematical Models in Isotope Hydrogeology*, *Rep. TECDOC-910*, pp. 9–58, Int. At. Energy Agency, Vienna.

McDonnell, J. J. (1989), The age, origin and pathway of subsurface stormflow in a steep humid headwater catchment, Ph.D. dissertation, 270 pp., Univ. of Canterbury, Christchurch, New Zealand.

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

McDonnell, J. J., M. K. Stewart, and I. F. Owens (1991), Effect of catchment-scale subsurface mixing on stream isotopic response, *Water Resour. Res.*, *27*, 3065–3073.

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

McGlynn, B., J. McDonnell, M. Stewart, and J. Seibert (2003), On the relationships between catchment scale and streamwater mean residence time, *Hydrol. Processes*, *17*, 175–181.

McGuire, K. J., and J. J. McDonnell (2006), Stable isotopes as a tool to resolve hydrological processes that relate to ecological questions, in *Stable Isotopes in Ecology and Environmental Science*, 2nd ed., edited by K. Lajtha and R. Michener, Blackwell, Malden, Mass., in press.

McGuire, K. J., J. J. McDonnell, M. Weiler, C. Kendall, B. L. McGlynn, J. M. Welker, and J. Seibert (2005), The role of topography on catchment-scale water residence time, *Water Resour. Res.*, *41*, W05002, doi:10.1029/2004WR003657.

McKie, D. (1978), A study of soil variability within the Blackball Hill Soils, Reefton, New Zealand, M.Ag.Sc. thesis, 180 pp., Univ. of Canterbury, Christchurch, New Zealand.

Melhorn, J., and C. Leibundgut (1999), The use of tracer hydrological time parameters to calibrate baseflow in rainfall-runoff modeling, in *Integrated Methods in Catchment Hydrology: Tracer, Remote Sensing, and New Hydrometric Techniques (Proceedings of IUGG 99 Symposium HS4)*, vol. 258, edited by C. Leibundgut, J. McDonnell, and G. Schultz, pp. 119–125, IAHS Press, Wallingford, U. K.

Michel, R. L. (2004), Tritium hydrology of the Mississippi River basin, *Hydrol. Processes*, *18*, 1255–1269.

Mosley, M. P. (1979), Streamflow generation in a forested watershed, *Water Resour. Res.*, *15*, 795–806.

Mosley, M. P. (1982), Subsurface flow velocities through selected forest soils, South Island, New Zealand, *J. Hydrol.*, *55*, 65–92.

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

Press, W. H., B. Flannery, S. Teukolsky, and W. Vetterling (1992), *Numerical Recipes: The Art of Scientific Computing*, 2nd ed., Cambridge Univ. Press, New York.

Robson, A., K. Beven, and C. Neal (1992), Towards identifying sources of subsurface flow: A comparison of components identified by a physically based runoff model and those determined by chemical mixing techniques, *Hydrol. Processes*, *6*, 199–214.

Scanlon, T. M., J. P. Raffensperger, and G. M. Hornberger (2001), Modeling transport of dissolved silica in a forested headwater catchment: Implications for defining the hydrochemical response of observed flow pathways, *Water Resour. Res.*, *37*, 1071–1082.

Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.

Seibert, J., K. H. Bishop, and L. Nyberg (1997), A test of TOPMODEL's ability to predict spatially distributed groundwater levels, *Hydrol. Processes*, *11*, 1131–1144.

Seibert, J., A. Rodhe, and K. Bishop (2003), Simulating interactions between saturated and unsaturated storage in a conceptual runoff model, *Hydrol. Processes*, *172*, 379–390.

Singh, V. P., and D. Frevert (Eds.) (2002), *Mathematical Models of Small Watershed Hydrology and Applications*, 950 pp., Water Resour. Publ., Highland Ranch, Colo.

Sivakumar, B. (2004), Dominant processes concept in hydrology: Moving forward, *Hydrol. Processes*, *18*, 2349–2353.

Sivapalan, M. (2003), Process complexity at hillslope scale, process simplicity at the watershed scale: Is there a connection?, *Hydrol. Processes*, *175*, 1037–1041.

Soulsby, C., M. J. J. Petry, S. M. Brewer, B. O. Dunn, and I. A. Malcolm (2003), Identifying and assessing uncertainty in hydrological pathways: A novel approach to end member mixing in a Scottish agricultural catchment, *J. Hydrol.*, *274*, 109–128.

Stewart, M. K., and J. J. McDonnell (1991), Modeling base flow soil water residence times from deuterium concentrations, *Water Resour. Res.*, *27*, 2681–2693.

Tsuboyama, Y., R. C. Sidle, S. Noguchi, and I. Hosoda (1994), Flow and solute transport through the soil matrix and macropores of a hillslope segment, *Water Resour. Res.*, *30*, 879–890.

Turner, J. V., and C. J. Barnes (1998), Modeling of isotopes and hydrochemical responses in catchment hydrology, in *Isotope Tracers in Catchment Hydrology*, edited by C. Kendall and J. J. McDonnell, pp. 723–760, Elsevier, New York.

Uhlenbrook, S., and C. Leibundgut (2002), Process-oriented catchment modelling and multiple-response validation, *Hydrol. Processes*, *162*, 423–440.

Vaché, K. B., J. J. McDonnell, and J. P. Bolte (2004), On the use of multiple criteria for a posteriori parameter estimation, *Geophys. Res. Lett.*, *31*, L21504, doi:10.1029/2004GL021577.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, *5*, 13–26.

Weiler, M., and J. McDonnell (2004), Virtual experiments: A new approach for improving process conceptualization in hillslope hydrology, *Hydrol. Processes*, *285*, 3–18.

Wigmosta, M. S., L. W. Vail, and D. P. Lettenmaier (1994), A distributed hydrology-vegetation model for complex terrain, *Water Resour. Res.*, *30*, 1665–1679.

Woods, R., and L. Rowe (1996), The changing spatial variability of subsurface flow across a hillside, *Hydrol. Processes*, *35*, 51–86.

Young, P. C., S. Parkinson, and M. J. Lees (1996), Simplicity out of complexity in environmental modeling: Occam's razor revisited, *J. Appl. Stat.*, *23*, 165–210.

Zak, S., and K. Beven (1999), Equifinality, sensitivity and uncertainty in the estimation of critical loads, *Sci. Total Environ.*, *236*, 191–214.

J. J. McDonnell, Department of Forest Engineering, Oregon State University, Corvallis, OR 97331-5706, USA.

K. B. Vaché, Institute for Landscape Ecology and Resources Management, Heinrich-Buff-Ring 26, D-35392 Giessen, Germany. (kellie.b.vache@agrar.uni-giessen.de)