HYDROLOGICAL PROCESSES
Hydrol. Process. **24**, 1730–1741 (2010)
Published online 17 February 2010 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.7595

Assessing the impact of mixing assumptions on the estimation of streamwater mean residence time

Fabrizio Fenicia,[1]* Sebastian Wrede,[1,2] Dmitri Kavetski,[3] Laurent Pfister,[1] Lucien Hoffmann,[1] Hubert H. G. Savenije[2] and Jeffrey J. McDonnell[4]

[1] *Centre de Recherche Public—Gabriel Lippmann, Department Environment and Agro-Biotechnologies, L-4422 Belvaux, Grand-Duchy of Luxembourg*
[2] *Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, P.O. Box 5048, NL-2600 GA Delft, The Netherlands*
[3] *Environmental Engineering, University of Newcastle, Callaghan, NSW 2308, Australia*
[4] *Institute for Water and Watersheds, Department of Forest Engineering, Resources and Management, Oregon State University, 015 Peavy Hall, Corvallis, OR 97331-5706, USA*

Abstract:

Catchment streamwater mean residence time (T_{mr}) is an important descriptor of hydrological systems, reflecting their storage and flow pathway properties. T_{mr} is typically inferred from the composition of stable water isotopes (oxygen-18 and deuterium) in observed rainfall and discharge. Currently, lumped parameter models based on convolution and sinewave functions are usually used for tracer simulation. These traditional models are based on simplistic assumptions that are often known to be unrealistic, in particular, steady flow conditions, linearity, complete mixing and others. However, the effect of these assumptions on T_{mr} estimation is seldom evaluated. In this article, we build a conceptual model that overcomes several assumptions made in traditional mixing models. Using data from the experimental Maimai catchment (New Zealand), we compare a complete-mixing (CM) model, where rainfall water is assumed to mix completely and instantaneously with the total catchment storage, with a partial-mixing (PM) model, where the tracer input is divided between an 'active' and a 'dead' storage compartment. We show that the inferred distribution of T_{mr} is strongly dependent on the treatment of mixing processes and flow pathways. The CM model returns estimates of T_{mr} that are well identifiable and are in general agreement with previous studies of the Maimai catchment. On the other hand, the PM model—motivated by a priori catchment insights—provides T_{mr} estimates that appear exceedingly large and highly uncertain. This suggests that water isotope composition measurements in rainfall and discharge alone may be insufficient for inferring T_{mr}. Given our model hypothesis, we also analysed the effect of different controls on T_{mr}. It was found that T_{mr} is controlled primarily by the storage properties of the catchment, rather than by the speed of streamflow response. This provides guidance on the type of information necessary to improve T_{mr} estimation. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS mean residence time; mixing models; partial-mixing; parameter inference

Received 31 August 2009; Accepted 30 November 2009

INTRODUCTION

In recent years, catchment streamwater mean residence time T_{mr} [T], also known as 'transit time' or 'water age', is increasingly used as a compact descriptor of catchment behaviour (McGuire *et al*., 2005; Hrachowitz *et al*., 2009). T_{mr} describes hydrological responses across a wide range of spatial and temporal scales, complementing traditional hydrometric streamflow information (McDonnell *et al*., 2007; Soulsby and Tetzlaff, 2008). In addition, it has been recently used as a constraint for model structural development (Seibert and McDonnell, 2002; Uhlenbrook and Leibundgut, 2002), reduction of parameter uncertainty (Vaché and McDonnell, 2006b) and multi-criteria model calibration (Vaché and McDonnell, 2006a; Dunn *et al*., 2007).

T_{mr} is commonly inferred from measurements of stable water isotope (oxygen-18 and deuterium) compositions in rainfall and discharge. Traditional estimation techniques use lumped parameter models, e.g. based on convolutions or sinewave functions for tracer simulation (McGuire and McDonnell, 2006). These simple approaches require several assumptions that are almost always violated in real-world catchment systems (McGuire and McDonnell, 2006), in particular: (i) steady flow conditions (constant rainfall and discharge), (ii) linear tracer input–output relations and (iii) instantaneous mixing of tracers within the entire catchment storage (Zuber and Maloszewski, 2000). These assumptions are not supported by process understanding. For example, water fluxes within a catchment are generally unsteady and often follow preferential pathways between different storage compartments such as groundwater, surface streams, etc. (e.g. Beven and Germann, 1982). In addition, runoff systems are generally threshold-like and their response to storm rainfall is highly nonlinear (e.g. Tromp van

* Correspondence to: Fabrizio Fenicia, Centre de Recherche Public—Gabriel Lippmann, Department Environment and Agro-Biotechnologies, L-4422 Belvaux, Grand-Duchy of Luxembourg.
E-mail: fenicia@lippmann.lu

Meerveld and McDonnell, 2006a,b; Zehe and Sivapalan, 2009).

Alternatively, T_{mr} may be estimated using more complex flow and transport models. These include conceptual models (e.g. Seibert *et al*., 2003; Page *et al*., 2007; Fenicia *et al*., 2008), compartmental mixing-cell models (e.g. Harrington *et al*., 1999; Vaché and McDonnell, 2006a; Sayama and McDonnell, 2009) and solute transport models using the advection–dispersion equation (e.g. Konikow and Reilly, 1998). These approaches overcome most assumptions of black-box lumped parameter models. However, they are seldom applied for T_{mr} estimation, arguably due to more complicated structure and data requirements.

Despite a growing number of models, current understanding of how different mixing models affect the estimation of T_{mr} remains limited. How strongly are the estimates of T_{mr} dependent on the underlying model hypotheses? Are traditional methods practically reliable and useful, notwithstanding being based on unrealistic assumptions? Are isotopic tracer composition measurements in rainfall and discharge sufficient for estimating T_{mr}, or is additional process information about the catchment necessary? Insights into these questions are a prerequisite for meaningful use of T_{mr}, both as a catchment descriptor and as a constraint during model evaluation.

This work advances the case for using conceptual models to explore the effects of mixing assumptions on the estimation of T_{mr}. Conceptual models allow systematic hypothesis-testing and improvement guided by experimental data (Fenicia *et al*., 2008). In addition, their computational speed allows a comprehensive treatment of uncertainty using Monte Carlo methods. This facilitates a better understanding of catchment behaviour and may help overcome inconsistencies arising from the use of black-box tracer simulation models (Kirchner, 2006), for example, the paradox that water released within minutes or hours of rainfall events is often months to years old (Kirchner, 2003).

In this article, we use simultaneous flow and tracer simulation to explore how the conceptualization of mixing mechanisms affects the inferred distribution of T_{mr}, and hence the inferred flow and tracer dynamics of a catchment. In doing so, we also develop an improved mechanistic understanding of the catchment behaviour through uncertainty estimation and hypothesis-testing. Using data from the well-studied Maimai catchment in New Zealand, we construct two distinct mixing models—one with complete mixing, where water mixes instantaneously within the whole catchment storage, and the other one with partial mixing, where water follows preferential pathways and is separated into 'active' and 'dead' storage compartments. This active–dead storage distinction closely resembles the complex, yet highly qualitative descriptions of McDonnell (1990) for the Maimai site, where the rationale for old water movement through macropores was based on unrequited storage within the soil profile and non-participatory portions of the hillslope material.

The two conceptual models have the same hydrological structure and differ solely in the way water (and hence tracers) mix within catchment compartments. Both models allow simultaneous flow and solute simulation, and therefore represent an improvement with respect to traditional steady-state-flow convolution-based models for tracer simulation (this is verified by direct comparison).

Our main focus is on the influence of mixing assumptions on the estimates and associated uncertainties in T_{mr}. We also investigate the dominant controls on T_{mr}. This is an interesting open question, since currently there is a poor understanding of whether T_{mr} is mostly affected by 'speed of response' aspects such as hydrograph recession characteristics (e.g. Vitvar *et al*., 2002), or by 'storage' properties such as the capacity of the unsaturated zone (e.g. Dunn *et al*., 2007).

STUDY SITE

The Maimai study area on the South Island of New Zealand was chosen for this study due to its long history of intensive hydrological research (see McGlynn *et al*., 2002). The availability of comprehensive and high-resolution data sets was essential for a meaningful understanding of underlying catchment processes and facilitated model development, application and interpretation. Importantly, the comparatively simple and spatially homogeneous physical characteristics and hydrological responses of the Maimai area catchments make them ideal 'benchmark basins' for model evaluation and testing (e.g. Seibert and McDonnell, 2002; Vaché and McDonnell, 2006a; Dunn *et al*., 2007; Fenicia *et al*., 2008).

This study focuses on the small forested M8 catchment, characterized by steep short slopes and deeply incised channels that drain an area of 3·8 ha (for details see Pearce *et al*., 1986). Its soils are generally shallow (about 60 cm on average) and remain highly saturated over the entire year. These highly transmissive soils are underlain by cemented conglomerate bedrock that appears largely impermeable (Mosley, 1979). Its climate is humid, with an average precipitation of about 2600 mm/a and a highly responsive 1550 mm/a stream runoff with almost no seasonal variation.

Under these climatic and physiographic conditions, lateral subsurface flow over the impervious bedrock is the dominant runoff mechanism in the catchment. While slow continuous down-slope drainage prevails during low-flow conditions, storm-flow events are characterized by significant preferential flows along cracks and lateral root networks, as well as along soil horizons and the soil bedrock interface (McDonnell, 1990). Thus, the M8's T_{mr}, estimated at 4 ± 1 months using oxygen-18 data (Pearce *et al*., 1986), is one of the shortest observed in natural catchments.

This study uses hourly rainfall, discharge, potential evaporation data, as well as deuterium composition measurements in the rainfall and the catchment outlet stream. The modelling period spans 1986–1987, while

continuous isotopic rainfall time series and event-based stream isotope data were available from September to December 1987. Isotope data covered the entire rainfall period and selected discharge events. A detailed description of the isotope collection approach is given by McDonnell *et al.* (1991).

METHODOLOGY

We compare two different model hypotheses: complete-mixing (CM) and partial-mixing (PM) model structures. The assumption of partial mixing is often perceived to be more realistic by the experimentalist (e.g. Seibert *et al.*, 2003), and, as discussed later, corresponds closer to the current perception of water movement in the Maimai catchment. Note that the two model structures have an identical water balance conceptualization and differ solely in the way water and isotopes are mixed internally. For completeness, we also evaluate the performance of a traditional steady-state exponential model (EM) on the same dataset (e.g. Zuber and Maloszewski, 2000).

Hydrological model

The water fluxes are described using a single reservoir characterized by an active storage S_a. The storage–discharge relationship of the reservoir was estimated from multiple recession segments using Master Recession Curve (MRC) analysis (Figure 1(a and b)) (Lamb and Beven, 1997; Fenicia *et al.*, 2006).

As seen in Figure 1(b), the dependence of discharge on storage appears to be piecewise linear with a smooth transition. This behaviour was represented using two linear segments connected by a logistic smoother (Kavetski and Kuczera, 2007, see Appendix for details). The method yields a smooth overall functional relationship, with continuous derivatives of all orders.

The water balance of the reservoir is

$$\frac{\mathrm{d}S_a}{\mathrm{d}t} = P - Q - E \tag{1}$$

where S_a is the active storage [L], P is precipitation [L/T], Q is discharge [L/T] and E is evaporation [L/T].

The storage–discharge relationship of the reservoir model is defined as

$$Q = G(S_a|k_1, k_2, S_{bq}, m_q) \tag{2}$$

where Q is discharge and G is a smoothed piecewise-linear function (see Appendix). The parameters of the relationship are as follows: k_1 and k_2 are the slopes of the linear segments separated by a breakpoint at $S_a = S_{bq}$ (Figure 1(b)) and m_q is a dimensionless smoothing parameter (less smoothing as $m_q \to 0$). For clarity, we use the vertical bar | to separate the dependence of G on the time-dependent state variable S_a from its dependence on parameters that are fixed within a given run.

Reservoir evaporation E is assumed to be proportional to the potential evaporation E_p, with a scaling parameter E_s [−]. We used a smooth evaporation function such that $E(S_a = 0) = 0$ and $E(S)$ reaches a constant value $E_p E_s$ as S increases:

$$E(S_a|S_{be}, m_e) = E_p E_s \; G(S_a|1/S_{be}, 0, S_{be}, m_e) \tag{3}$$

where S_{be} and m_e are the breakpoint and smoothing parameters, respectively.

The water balance (Equation (1)) is integrated using the implicit Euler method with fixed time step and a Newton–Raphson solver (with the latter typically converging to double precision in two to five iterations) (see Kavetski *et al.*, 2006a for details).

The water balance model has four calibration parameters: k_1 [1/T], k_2 [1/T], S_{bq} [L] and E_s [−]. The smoothing parameter $m_q = 2$ was determined from Figure 1. The parameters S_{be} and m_e in the evaporation function (Equation (3)) were considered part of the numerical description and were fixed at $S_{be} = 3$ mm and $m_e = 0{\cdot}5$.

Mixing model

In most catchments, the dynamics of the tracer response is strongly damped with respect to the input signal (Kirchner, 2003). In a catchment model, this behaviour can be reproduced by introducing an additional storage that does not alter the hydrological behaviour (streamflow), but influences the way water and tracers mix within it. This approach was suggested by Barnes and Bonell (1996) and finds support in recent studies (e.g.

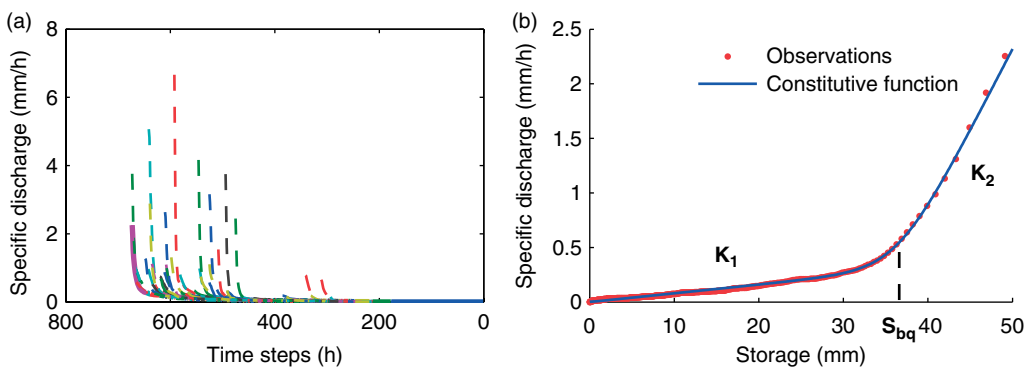

Figure 1. Master recession curve ((a) continuous line) and storage–discharge relation inferred from it (b)

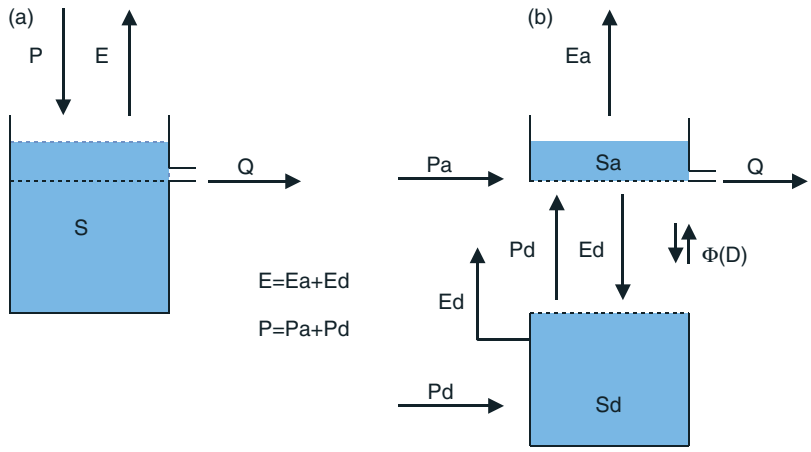

Figure 2. Water balance for the CM and PM models (a) and flux exchanges for the PM model (b)

Vaché and McDonnell, 2006a; Dunn *et al*., 2007; Son and Sivapalan, 2007; Fenicia *et al*., 2008). This dead storage can be interpreted as representing the unsaturated zone, or a groundwater system that is not directly connected to the stream.

This study conceptualizes the dead storage S_d as an additional storage below the threshold that produces runoff (Barnes and Bonell, 1996) (Figure 2). We note that this dead storage represents all non-participatory storage compartments, including soil water, water stored in bedrock depressions, etc.

Complete mixing. The CM model assumes instantaneous mixing between the active and dead compartments. It is represented by the tracer balance equation:

$$\frac{\mathrm{d}(cS)}{\mathrm{d}t} = c_p P - cE - cQ \tag{4}$$

where $S = S_a + S_d$ is the total storage, c is the isotopic tracer composition [M/M] in the reservoir and c_p is the isotopic tracer composition [M/M] in the rainfall P. Expanding the derivative and rearranging yields:

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \frac{1}{S_a + S_d}\left(c_p P - c\left(Q + E + \frac{\mathrm{d}S_a}{\mathrm{d}t}\right)\right) \tag{5}$$

The simulation of tracer dynamics in the CM model adds the additional calibration parameter S_d [L]. To avoid confusion, note that S_a is a state variable, whereas S_d is a calibrated parameter.

Partial mixing. The complete-mixing hypothesis is not realistic in most natural catchments. Even in the Maimai catchment, which is characterized by shallow soil and impermeable bedrock, it has been shown experimentally that water is poorly mixed during storm events (McDonnell, 1990). This is due to a two-component flow system of rapid macropore flow and slow matrix flow. McDonnell (1990) noted that rainfall water infiltrates into macropores, quickly reaching the soil–bedrock interface. As a result, free (new) water develops at soil–bedrock interface and backs-up into the soil matrix, where it mixes with much larger (old) water storage.

This rationale suggests a partial-mixing hypothesis, with at least two storage compartments. Based on these arguments, we developed a mixing model that hypothesizes:

1. instantaneous mixing within the active and dead storage compartments, and
2. partial mixing between the two compartments (e.g. Seibert *et al*., 2003; Page *et al*., 2007).

Similary to the CM model, the dead storage is representative of all water that is not directly connected to the stream. This may include soil water, as well as water stored in bedrock depressions. While this conceptualization remains quite simplistic, it services our primary aim of investigating how different water-mixing assumptions influence the inferred distribution of T_{mr}.

The water fluxes are depicted in Figure 2. We distribute the precipitation between the active and dead storages using a dimensionless coefficient F_r, so that $P_d = F_r P$ and $P_a = (1 - F_r)P$. In the context of water balance, the precipitation P_a that enters the dead storage is released into the active storage. We also assume that total evaporation is partitioned between the two compartments according to their relative storage fraction, i.e. $E_a/E = S_a/S$.

Unlike the CM model, the PM model separately evolves isotopic tracer compositions in the active and dead storages and hence uses a total of three, rather than two, state variables. It conceptualizes the mixing between these compartments using an internal flux $\Phi = D(c_d - c_a)S_a$, where D is a dispersion parameter [1/T], and c_a and c_d are the isotopic tracer compositions [M/M] in the active and dead storages, respectively.

We hypothesize that the mixing flux Φ depends on the active storage S_a because, at least *a priori*, we expect larger exchanges during wetter conditions—in particular, during storm events. The linear relation is used for simplicity; nonlinear dependencies could be readily included, if warranted by the data. Note that the water balance for the dead storage is 0, whereas the water

balance for the active storage is still represented by Equation (1).

The tracer balance equation for the active storage is:

$$\frac{\mathrm{d}(c_a S_a)}{\mathrm{d}t} = c_p P_a + c_d P_d - c_a E - c_a Q + D(c_d - c_a) S_a \tag{6}$$

Rearranging for $\mathrm{d}c_a/\mathrm{d}t$ yields:

$$\frac{\mathrm{d}c_a}{\mathrm{d}t} = \frac{1}{S_a}\left(c_p P_a + c_d P_d - c_a\left(Q + E + \frac{\mathrm{d}S}{\mathrm{d}t}\right) + D(c_d - c_a) S_a \right) \tag{7}$$

Analogously, the tracer balance equation for the dead storage is:

$$\frac{\mathrm{d}c_d}{\mathrm{d}t} = \frac{1}{S_d}\left(P_d c_p - P_d c_d + (c_a - c_d)\frac{S_d}{S_d + S_a} E + D(c_a - c_d) S_a\right) \tag{8}$$

The PM includes the additional calibration parameters F_r [−], D [1/T] and S_d [L]. Again, note that S_a is a state variable, whereas S_d is a calibrated parameter.

It could be argued that the PM model hypothesizes a more physically motivated mixing mechanism than the CM model. In particular, the coefficient F_r represents the effects of preferential flows that explain the rapid response of this catchment to rainfall (Mosley, 1979), whereas the flux Φ reflects the macropore–matrix interactions responsible for the slow mixing within events observed by McDonnell (1990).

Numerical solution. The CM model (Equation (5)) and the PM model [coupled Equations (7 and 8)] are solved numerically using an explicit Runge–Kutta solver with a scaled error tolerance of 10^{-4}. For computational simplicity and efficiency, the water fluxes from the water balance model (Equation (1)) are distributed uniformly over each timestep of the mixing model (numerical errors arising from this linearization are minor).

Model evaluation

Bayesian inference. We use a standard Bayesian approach to estimate the hydrological and mixing parameters of the CM and PM models and compare the performance of these approaches.

The model parameters and response error variances are inferred from the observed discharge and isotopic tracer compositions, $\widetilde{\boldsymbol{Q}}$ and $\tilde{\boldsymbol{c}}$, respectively, using Bayes equation

$$p(\boldsymbol{\theta}, \sigma_q, \sigma_c | \widetilde{\boldsymbol{Q}}, \tilde{\boldsymbol{c}}) = p(\widetilde{\boldsymbol{Q}}, \tilde{\boldsymbol{c}} | \boldsymbol{\theta}, \sigma_q, \sigma_c) p(\boldsymbol{\theta}, \sigma_q, \sigma_c) \tag{9}$$

where $p(\boldsymbol{\theta}, \sigma_q, \sigma_c | \widetilde{\boldsymbol{Q}}, \tilde{\boldsymbol{c}})$ is the posterior distribution of the model parameters $\boldsymbol{\theta}$, the standard deviation of flow errors σ_q and the standard deviation of isotopic composition errors σ_c, while $p(\boldsymbol{\theta}, \sigma_q, \sigma_c)$ is the prior information on these quantities and $p(\widetilde{\boldsymbol{Q}}, \tilde{\boldsymbol{c}} | \boldsymbol{\theta}, \sigma_q, \sigma_c)$ is the likelihood function. In Equation (9), the tildes indicate quantities that are observed and hence subject to sampling and measurement uncertainties.

In the absence of any additional independent knowledge, we used uniform (bounded) priors for all model parameters and Jeffreys' uninformative prior for σ_q and σ_c (Kavetski *et al*., 2006b). Note that both σ_q and σ_c, which reflect the residual uncertainties in the discharge and isotopic compositions arising from data and model errors, are inferred using Equation (9).

Likelihood function. The likelihood function used for the inference is

$$p(\widetilde{\boldsymbol{Q}}, \tilde{\boldsymbol{c}} | \boldsymbol{\theta}, \sigma_q, \sigma_c) = \prod_{i=1}^{N_q} N(\xi_i(\boldsymbol{\theta}, \widetilde{Q}_i) | 0, \sigma_q^2) \times \prod_{j=1}^{N_c} N(\xi_j(\boldsymbol{\theta}, \tilde{c}_j) | 0, \sigma_c^2) \tag{10}$$

where $N(x|m, s^2)$ is the probability density function of a Gaussian deviate x with mean m and variance s^2, ξ_i is the residual error (difference between observed and simulated responses), whereas N_q and N_c are the numbers of discharge and isotopic tracer composition samples, respectively.

The least-squares likelihood function, i.e. Equation (10), assumes that the model residuals, which lump the effects of data and model errors, are independent and identically distributed Gaussian. It also assumes that the model and observation errors in the two responses are uncorrelated.

While these statistical assumptions can and have be questioned (e.g. see discussion in Kavetski *et al*., 2002), deriving a more complicated likelihood function lies outside the scope of this study, which focuses on the physical motivation for the PM model and its comparison with the traditional CM approach. In future work, we will report a more general Bayesian approach for estimating and accounting for the auto- and cross-correlations in the residuals when jointly calibrating the hydrological and tracer models to streamflow and tracer observations.

Inference of mean residence time. The mean residence time in a multi-compartmental system with unsteady flow and full tracer recovery can be obtained from the tracer breakthrough curve. Given a pulse injection of tracer into the system at time t_0, and assuming that the tracer can leave the system either through discharge Q or through evaporation E, the mean residence time can be defined as follows:

$$T_{mr} = \frac{\int_A \int_{t_0}^{\infty} (c_q Q + c_e E) t \, \mathrm{d}t \, \mathrm{d}a}{\int_A \int_{t_0}^{\infty} (c_q Q + c_e E) \, \mathrm{d}t \, \mathrm{d}a} \tag{11}$$

where c_q is the isotopic tracer composition in Q, c_e is the isotopic composition in E, t is time, $\mathrm{d}a$ is an areal element and A is the control surface (catchment area) of the system.

Equation (11) differs from traditional formulations (e.g. Goode, 1996) because it accounts for unsteady streamflow and internal fluxes. Since these fluxes originate in different compartments (e.g. evaporation from the active and dead storages in the PM model) and hence may have different isotopic compositions, the integral over the catchment area is used.

Since Equation (11) implies that $T_{mr} = f(\boldsymbol{\theta}, c_q)$, it was used to obtain the posterior distribution of T_{mr}, $p(T_{mr}|\widetilde{\boldsymbol{Q}}, \tilde{\boldsymbol{c}})$, corresponding to the posterior distribution $p(\boldsymbol{\theta}, \sigma_q, \sigma_c|\widetilde{\boldsymbol{Q}}, \tilde{\boldsymbol{c}})$ given by Equation (9).

RESULTS

The models were run on an hourly time step over the entire year of 1987. The first 10 days of simulation were treated as a warm-up period and discarded from the computation of the performance statistics. The posterior distribution Equation (9) was sampled using the Markov Chain Monte Carlo (MCMC) strategy described by Thyer *et al.* (2009) with a total of 39 000 model runs and 5 parallel chains. During the first 2000 samples, the jump distribution was tuned one parameter at a time. During the next 2000 samples, the jump distribution was tuned by scaling its entire covariance matrix. Following this, the jump distribution was fixed and 35 000 samples were collected. The first 25 000 samples were treated as a warm-up period and only the final 10 000 samples were used to report the parameter distributions. In all inferences, the Gelman-Rubin convergence test was very close to unity, which suggests adequate convergence of the MCMC chains to the stationary distribution.

Since the tracer data contained long gaps between events, we used eight separate events within the observation period for calibration. Before each event, the isotopic tracer composition of the model compartment(s) was reinitialized to the value before (when available) or after the event. However, the performance of each parameter sample was evaluated on all events simultaneously.

T_{mr} was calculated using Equation (11), running the models for the two full years 1986–1987 and adding a constant isotopic tracer composition to all model reservoirs on 1st March 1986. While injecting the tracer on a different date may lead to slightly different estimates of T_{mr}, seasonality effects at the Maimai site are minor because the rainfall regime is relatively stable (see Section on Study Site).

Parameter distributions

CM model. The Nash–Sutcliffe performance (NS) of the CM model with respect to flows is quite high, 0·85. The NS of the tracer output is much lower, i.e. 0·37. However, it stressed that: (1) the tracer signal is strongly damped, and therefore the average of the observation is already a relatively good representation of the data and (2) the sampling uncertainty is relatively high in relation to the dynamics. Reassuringly, the inferred values of σ_c are close to the sampling uncertainty in the deuterium composition measurements, which is commonly between 1‰ and 2‰ (Pearce *et al.*, 1986). This suggests that it would be difficult to expect much higher model performances.

The parameter distributions of the CM model are shown in Figure 3. All model parameters appear to be well identifiable and display little correlation (apart from a mild correlation between S_{bq} and k_2). Figure 3 also shows the estimated distribution of T_{mr} and its correlation with model parameters. Note that T_{mr} itself was not calibrated—it is a derived quantity computed from the samples. It can be seen to be very strongly dependent on S_d, with minimal statistical dependence on other model parameters. Given the good identifiability of S_d for the CM model, the marginal distribution of T_{mr} is itself well identifiable, with likely values in the range of 60–120 days. This result agrees with previous studies of the Maimai catchment, which estimated T_{mr} as approximately 4 months (Pearce *et al.*, 1986). Interestingly, T_{mr} depends primarily on the dead storage S_d and displays very little correlation with other model parameters. We will comment on this result in the Discussion.

Traditional steady-state model. To provide a comparison with current approaches for T_{mr} estimation, we also evaluated a traditional EM (e.g. Zuber and Maloszewski, 2000). The EM is essentially a single-parameter linear reservoir that convolves the isotopic tracer composition through an exponential function. The EM was calibrated and evaluated on the selected events using the same approach as for the CM and PM models. The EM provided a very poor fit to the tracer data, with NS = 0·05. This low value is likely due to the steady-state hypothesis underlying the traditional EM model, which completely neglects the rainfall-runoff dynamics. This is unsurprising and simply confirms that traditional 'steady-state' tracer models are unsuitable for high temporal resolution simulations (such as hourly in this case).

PM model. The PM model performs slightly better than the CM model, raising the NS efficiency of the tracer fit from 0·37 to 0·45. This difference can be attributed to the PM model having three parameters (S_d, F_r and D) controlling the tracer response, whereas the CM model has only one parameter (S).

The hydrological parameters inferred from the PM model have almost identical distributions as for the CM model (hence not shown). This is a direct consequence of the CM and PM models sharing the same hydrological conceptualization.

The distributions of the mixing model parameters are shown in Figure 4. D displays large variability, whereas F_r appears well identified. The upper bound of S_d is poorly identifiable (or at least not within the prior upper bound of 4000 mm allowed during calibration). Due to a strong correlation between T_{mr} and S_d (same as in the CM model), the poor identifiability of S_d translates into poor identifiability of T_{mr}. The lower bound of T_{mr} is

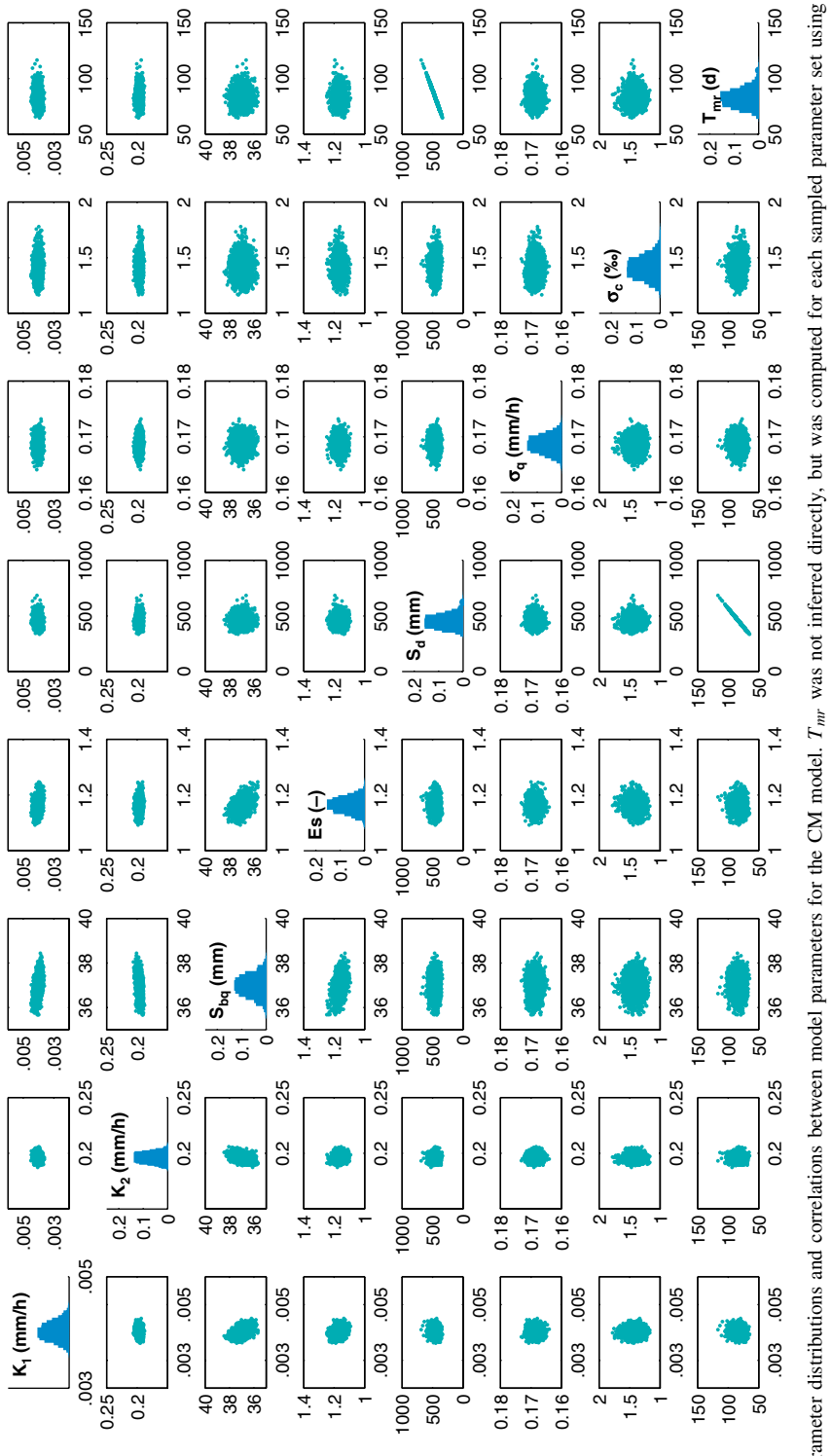

Figure 3. Parameter distributions and correlations between model parameters for the CM model. T_{mr} was not inferred directly, but was computed for each sampled parameter set using Equation (11)

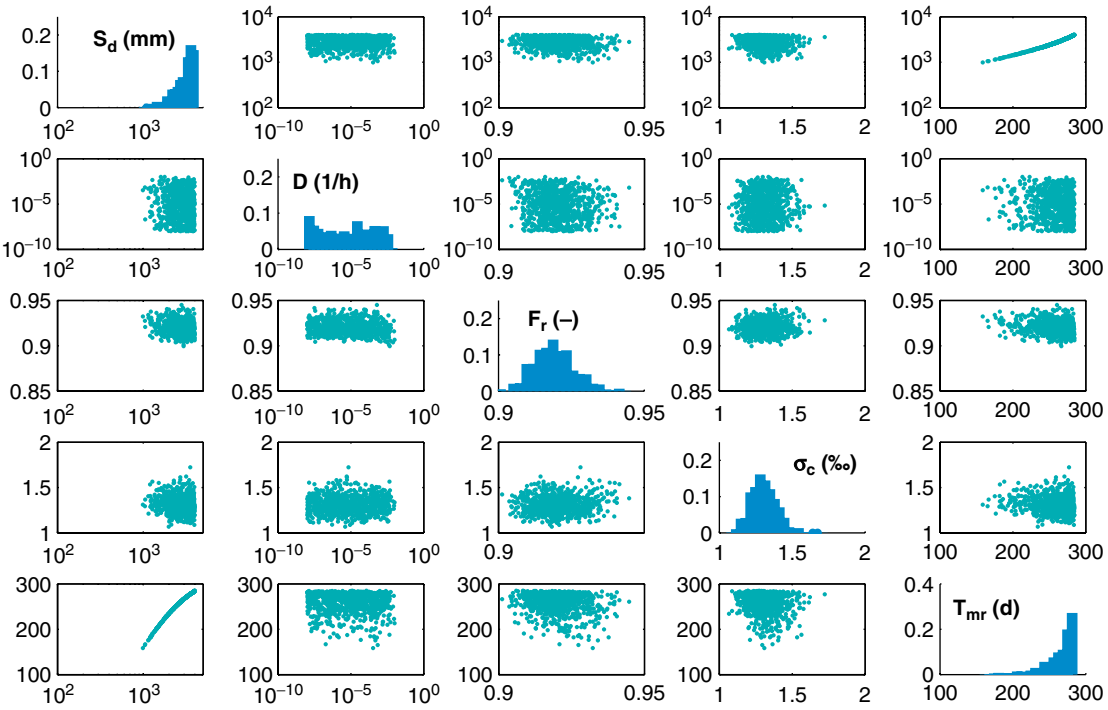

Figure 4. Distribution and correlation between parameters of the CM model (hydrology-related parameters not shown). The T_{mr} distribution derived from parameter samples is also shown

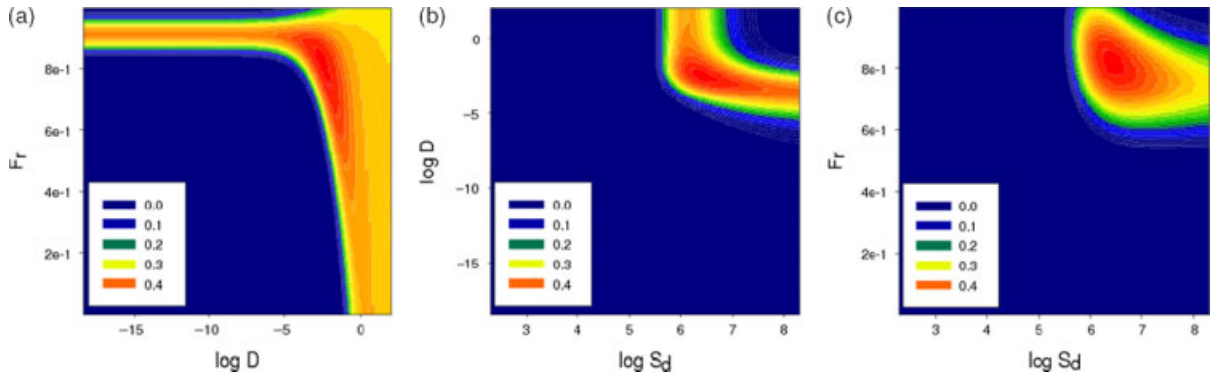

Figure 5. Cross-cuts of the NS index calculated for deuterium composition data alone

close to that estimated with the CM model, but the upper bound is much higher.

The strong correlation between D, F_r and S_d was investigated using cross-sections of the NS measure for the isotope data fit (Figure 5). In this analysis, parameters were varied, two at a time, while keeping all others fixed at near-optimal values identified during the model calibration.

Figure 5(a) shows that when the exchange flux associated with D is small, the damping effect can be largely explained by a fixed partitioning of rainfall between active and dead storages (F_r). When D is large, F_r becomes less identifiable, since the damping can be explained either by a dispersion flux, or by partitioning of rainfall, or by a combination of both. There is insufficient information to identify the process solely from the given data and model conceptualization.

Figure 5(b) demonstrates a negative correlation between S_d and D. This implies that, keeping all other parameters fixed, the same damping of tracer signal can be achieved either by a reduction of the (mixing) dispersion flux, or by increasing the dead storage, or by a combination of both.

Finally, Figure 5(c) shows that S_d tends to be poorly identifiable for large values of F_r. This occurs because S_d behaves like a large reservoir that receives a certain fraction of the precipitation and releases it strongly damped. Therefore, S_d provides a constant 'baseline' contribution to the isotope signal, whereas S_a provides the variable time-dependent contribution. These components are weighted by the parameter F_r to fit the degree of damping in the observed data. This formulation of the system does not require S_d to have an upper bound. Also, note that S_d does not participate in the water balance (except for the minor effect through E_d) and hence is not directly informed by observed streamflow. The combination of these factors may explain why, in the absence of additional

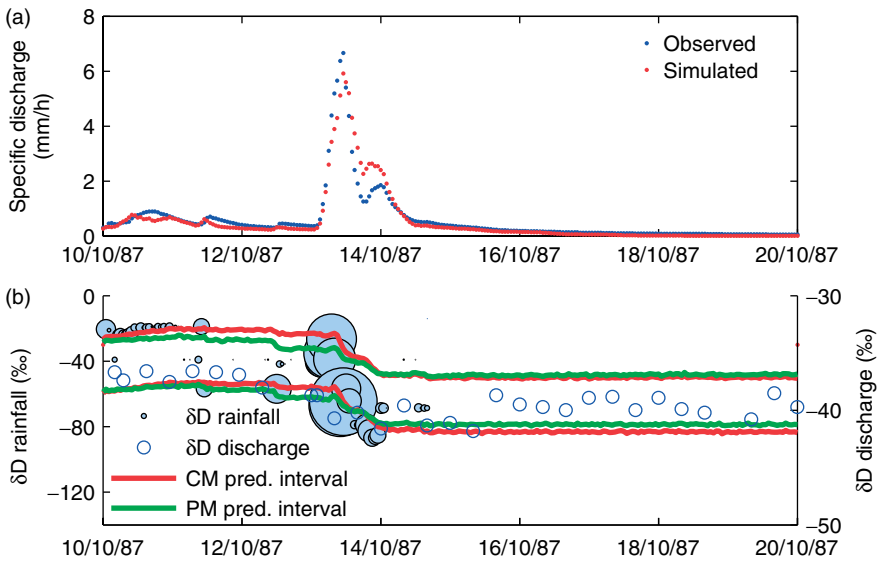

Figure 6. Simulated and observed specific discharge (a) and deuterium compositions (b) for a selected event. The size of the bubbles in (b) is proportional to the observed rainfall

information or constraints, T_{mr} is pushed against its prior upper bound (Figure 4).

System response and predictive limits

Figure 6 shows the model performance (streamflow and tracer predictions) for the largest event in the observation period. Figure 6(a) shows the hydrograph, whereas Figure 6(b) shows the deuterium composition in the rainfall and in the discharge. The isotopic composition in the discharge is strongly damped with respect to the rainfall (note the different scales of the *y*-axis in Figure 6(b)).

Figure 6(b) shows the 5–95% predictive intervals including parametric and residual uncertainty of isotopic tracer compositions simulated by the two models. These bounds are essentially the same for the CM and PM models. Hence, while the two models provide very similar fits to the streamflow and tracer data, they yield very different estimates of T_{mr}.

DISCUSSION

While the CM model provides a well-defined estimate of T_{mr}, the PM model provides a much more uncertain estimate. In addition, the PM hypothesis leads to T_{mr} estimates that are much larger than those corresponding to the CM hypothesis. However, both models provide very similar predictions and uncertainty estimates of the isotope signal. This suggests that the estimation of T_{mr} and its posterior uncertainty can be very sensitive to the assumptions underlying the hypothesized mixing model.

The interpretation of internal mixing processes based on input–output conservative tracer data is ambiguous because different model components may have similar effects on the output (see Renard *et al.* (2010), for a rigorous discussion of model identifiability). Our findings are consistent with those of Page *et al.* (2007), who attempted to simulate observed stream chloride concentrations at Plynlimon (Wales). Based on a similar model setup (mobile–immobile storage mixing), they determined that parameter equifinality prevented the identification of effective catchment mixing volumes and coefficients. Since the way water mixes within the catchment is likely to have a strong effect on the age of water, estimates of T_{mr} calibrated solely to input–output data—and especially under CM assumptions—should be interpreted and used with considerable care.

Even at the Maimai site, which appears to be a relatively well-mixed catchment compared to the range of systems encountered elsewhere, complete mixing is considered unrealistic by the experimentalist. In particular, McDonnell (1990) has proposed a rationale for old water movement through macropores based on unrequited storage within the soil profile and non-participatory portions of the hillslope material. As a result, T_{mr} can be expected to be larger than estimated assuming complete mixing.

Generally, the uncertainty in T_{mr} estimates could be reduced using additional independent insights into the functioning of the catchment. In particular, a better qualitative understanding of storage and flow characteristics and pathways can improve the selection of mixing models and constrain the uncertainty in model parameters. Obtaining and using this type of knowledge will require a close exchange between modeler and experimentalist (Seibert and McDonnell, 2002). For example, previous studies have shown that the Maimai catchment is characterized by shallow soil and impermeable bedrock (Pearce *et al.*, 1986). This may *a priori* limit the depth of the dead storage and constrain the estimates of T_{mr}.

Another important result is that, in this analysis, T_{mr} appears independent from the hydrology-related parameters, but instead is controlled by the total storage of the catchment that participates in the mixing process. This result contradicts some previous studies, which proposed estimating T_{mr} from hydrograph recessions

(Wolock *et al*., 1997; Vitvar *et al*., 2002). Referring to Figure 1(b), in the absence of significant dead storage, T_{mr} is largely determined by recession constants. However, as the dead storage becomes dominant with respect to the active storage, it becomes the main control on T_{mr}.

In contrast, other studies (e.g. Vaché and McDonnell, 2006a; Dunn *et al*., 2007; Son and Sivapalan, 2007; Fenicia *et al*., 2008) found a strong dependency of T_{mr} (or of the isotope response in general) on unsaturated and other storage. These studies support our conclusion that T_{mr} is correlated much stronger to storage properties, rather than to measures of streamflow response speed.

Finally, the traditional convolution-based mixing model based on steady-state flow assumptions provided an unacceptably poor representation of the tracer signal (see Section on Parameter Distributions). While perhaps providing some limited utility for long-term averaged data (e.g. monthly), such models ignore essential catchment dynamics and therefore cannot yield useful insights into the short-term mixing mechanisms operating in a realistic catchment.

While this study focused solely on the Maimai catchment and therefore the generality of its conclusions needs further investigation, we argue that physically motivated models for flow and tracer simulation, such as the conceptual mixing models investigated in this study, are more useful than black-box approaches. In particular, they can provide more integrated and physically realistic approaches for understanding and predicting pollutant transfer in watershed systems, and are more likely to explain apparent hydrological paradoxes, such as rainfall forcings resulting in rapid stream responses with water that is months to years old (Kirchner, 2003).

Furthermore, it is clearly preferable to use physically motivated models, rather than simplistic or black-box approaches, to predict quantities such as T_{mr} or new-water/old-water fractions when evaluating the performance of more complex flow and transport models (Vaché and McDonnell, 2006a; Son and Sivapalan, 2007). It is stressed that, when the model is poorly identifiable, quantities inferred from data, such as T_{mr} or new-water/old-water fractions, can be much more uncertain than the data itself.

CONCLUSION

This study investigated the impact of different water mixing assumptions on the estimation of streamwater mean residence time T_{mr}. While mean residence time is increasingly used for catchment characterization and model evaluation, in practice, its estimation is often performed using models and underlying hypotheses that are demonstrably unsupported by physical and/or statistical evidence.

In a case study based on the experimental Maimai catchment in New Zealand, we compared the mean residence time T_{mr} estimated using a complete-mixing (CM) versus a partial-mixing (PM) conceptual model. The latter was motivated by *a priori* physical insights into the Maimai catchment and allowed for water exchange between an active and a dead storage compartment that does not contribute to the overall catchment water balance. The hydrological and mixing parameters of both models were inferred statistically. Both the PM and CM models account for transient dynamics during storm events and fit the tracer observations much better than convolution-based models based on steady-state hydrology.

While the CM and PM models provided similar fits to the stream tracer signal, with NS indices of 0·37 versus 0·45, respectively (and with predictive bounds of similar width), their T_{mr} estimates varied considerably. This emphasizes that model evaluation and comparison based solely on output series may be highly unreliable, and that detailed evaluation of internal states and associated model characteristics can detect critical differences in model behaviour and interpretation.

Although CM-based T_{mr} estimates are well identified, those derived from the PM model were considerably larger and more uncertain. This occurs because of the mathematical behaviour of the PM conceptualization and formulation. Hence, although the PM model agrees better with experimental perceptions of the mixing mechanisms in the Maimai catchment, its poor identifiability suggests that further independent understanding of mixing processes is needed to constrain the model. It also follows that T_{mr} estimates derived without a careful independent testing of underlying mixing assumptions may be unreliable. Importantly, this highlights that improving the physical realism of a model does not necessarily improve its statistical identifiability, and indeed the contrary may be quite common.

Model analysis also suggested that T_{mr} is controlled primarily by storage properties, rather than by speed-of-response characteristics. This finding implies that estimating T_{mr} from recession constants is unreliable, and supports previous work linking T_{mr} to unsaturated and other storage.

Further analysis is warranted for other catchments to investigate the generality of our conclusions. Nevertheless, we argue that conceptual models provide deeper insights into catchment dynamics than traditional convolution-based mixing models requiring steady-state assumptions, linearity, complete mixing, etc. Conceptual models are more flexible and, especially when combined with additional data and insights, can be used within systematic hypothesis-testing and improvement frameworks to improve our understanding of catchment dynamics and the interaction between different catchment processes.

ACKNOWLEDGEMENT

This work was supported by the National Research Fund of Luxembourg (Grants FNR/08/AM2c/50 and TR-PHD BFR07-047-Investigating scale dependency of runoff generation).

APPENDIX: LOGISTIC SMOOTHER FOR PIECEWISE-LINEAR SEGMENTS

The logistic smoother used in this work is defined as follows (Kavetski and Kuczera, 2007):

$$\Phi(x|m) = m(x + \ln[1 + \exp(-x)]) \tag{A.1}$$

where x is the independent variable and m is a shape parameter.

Next, define F as

$$F(x|k_1, k_2, x_b, m) = \begin{cases} y_b + k_1(x - x_b), \text{ if } x_m < \ln(\varepsilon) \\ y_b + k_2(x - x_b), \text{ if } x_m > -\ln(\varepsilon) \\ y_b + k_1(x - x_b) + \\ (k_2 - k_1)\Phi(x_m|m), \text{ otherwise} \end{cases} \tag{A.2}$$

where k_1 is the slope of the first linear segment, k_2 the slope of the second linear segment, $y_b = k_1 x_b$, $x_m = (x - x_b)/m$ and ε is the machine precision (e.g. 10^{-16} for double precision computation).

To ensure $F(x = 0) = 0$, we define

$$G(x|k_1, k_2, x_b, m) = F(x|k_1, k_2, x_b, m) - F(0|k_1, k_2, x_b, m) \tag{A.3}$$

Both F and G are infinitely differentiable, which is a necessary (and usually sufficient) condition for yielding smooth models with smooth objective functions. The smoothness of the model and its objective function is very beneficial in both calibration and prediction: it allows using fast derivative-based optimization methods, simplifies sampling, yields more stable and well-behaved model predictions, etc. (Kavetski and Kuczera, 2007).

REFERENCES

Barnes CJ, Bonell M. 1996. Application of unit hydrograph techniques to solute transport in catchments. *Hydrological Processes* **10**: 793–802.

Beven K, Germann P. 1982. Macropores and water-flow in soils. *Water Resources Research* **18**: 1311–1325.

Dunn SM, McDonnell JJ, Vache KB. 2007. Factors influencing the residence time of catchment waters: a virtual experiment approach. *Water Resources Research* **43**: W06408. DOI: 10.1029/2006wr005393.

Fenicia F, McDonnell JJ, Savenije HHG. 2008. Learning from model improvement: on the contribution of complementary data to process understanding. *Water Resources Research* **44**: W06419. DOI: 10.1029/2007wr006386.

Fenicia F, Savenije HHG, Matgen P, Pfister L. 2006. Is the groundwater reservoir linear? Learning from data in hydrological modelling. *Hydrology and Earth System Sciences* **10**: 139–150.

Goode DJ. 1996. Direct simulation of groundwater age. *Water Resources Research* **32**: 289–296.

Harrington GA, Walker GR, Love AJ, Narayan KA. 1999. A compartmental mixing-cell approach for the quantitative assessment of groundwater dynamics in the Otway Basin, South Australia. *Journal of Hydrology* **214**: 49–63.

Hrachowitz M, Soulsby C, Tetzlaff D, Dawson JJC, Malcolm IA. 2009. Regionalization of transit time estimates in montane catchments by integrating landscape controls. *Water Resources Research* **45**: W05421. DOI: 10.1029/2008wr007496.

Kavetski D, Franks SW, Kuczera G. 2002. Confronting input uncertainty in environmental modelling. In *Calibration of Watershed Models, Water Science and Application series*, Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds). American Geophysical Union Washington, DC **6**: 49–68.

Kavetski D, Kuczera G. 2007. Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration. *Water Resources Research* **43**: W03411. DOI: 10.1029/2006wr005195.

Kavetski D, Kuczera G, Franks SW. 2006a. Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology* **320**(1–2): 173–186.

Kavetski D, Kuczera G, Franks SW. 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research* **42**: W03407. DOI: 10.1029/2005wr004368.

Kirchner JW. 2003. A double paradox in catchment hydrology and geochemistry. *Hydrological Processes* **17**: 871–874. DOI: 10.1002/Hyp.5108.

Kirchner JW. 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* **42**: W03s04. DOI: 10.1029/2005wr004362.

Konikow LF, Reilly TE. 1998. Groundwater modeling. In *The Handbook of Groundwater Engineering*, Delleur JW (ed). CRC: Boca Raton, MA; 1–20.

Lamb R, Beven K. 1997. Using interactive recession curve analysis to specify a general catchment storage model. *Hydrology and Earth System Sciences* **1**: 101–113.

McDonnell JJ. 1990. A rationale for old water discharge through macropores in a steep, humid catchment. *Water Resources Research* **26**: 2821–2832.

McDonnell JJ, Sivapalan M, Vache K, Dunn S, Grant G, Haggerty R, Hinz C, Hooper R, Kirchner J, Roderick ML, Selker J, Weiler M. 2007. Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. *Water Resources Research* **43**: W07301. DOI: 10.1029/2006wr005467.

McDonnell JJ, Stewart MK, Owens IF. 1991. Effect of catchment-scale subsurface mixing on stream isotopic response. *Water Resources Research* **27**: 3065–3073.

McGlynn BL, McDonnel JJ, Brammer DD. 2002. A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand. *Journal of Hydrology* **257**: 1–26.

McGuire KJ, McDonnell JJ. 2006. A review and evaluation of catchment transit time modeling. *Journal of Hydrology* **330**: 543–563.

McGuire KJ, McDonnell JJ, Weiler M, Kendall C, McGlynn BL, Welker JM, Seibert J. 2005. The role of topography on catchment-scale water residence time. *Water Resources Research* **41**: W05002. DOI: 10.1029/2004wr003657.

Mosley MP. 1979. Streamflow generation in a forested Watershed, New-Zealand. *Water Resources Research* **15**: 795–806.

Page T, Beven KJ, Freer J, Neal C. 2007. Modelling the chloride signal at Plynlimon, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty). *Hydrological Processes* **21**: 292–307. DOI: 10.1022/Hyp.6186.

Pearce AJ, Stewart MK, Sklash MG. 1986. Storm runoff generation in humid headwater catchments 1. Where does the water come from. *Water Resources Research* **22**: 1263–1272.

Renard B, Kavetski D, Kuczera G, Thyer M, Franks SW. 2010. Understanding predictive uncertainty in hydrologic modeling: Le challenge of identifying input and structural errors. *Water Resources Research* (Accepted).

Rowe LK. 1979. Rainfall interception by a Beech-Podocarp-Hardwood forest near Reefton, North Westland, New Zealand. *Journal of Hydrology New Zealand* **18**: 63–72.

Sayama T, McDonnell JJ. 2009. A new time-space accounting scheme to predict stream water residence time and hydrograph source components at the watershed scale. *Water Resources Research* **45**: W07401. DOI: 10.1029/2008wr007549.

Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resources Research* **38**: DOI: 10.1029/2001WR000978.

Seibert J, Rodhe A, Bishop K. 2003. Simulating interactions between saturated and unsaturated storage in a conceptual runoff model. *Hydrological Processes* **17**: 379–390. DOI: 10.1002/Hyp.1130.

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**: W01415. DOI: 10.1029/2006wr005032.

Soulsby C, Tetzlaff D. 2008. Towards simple approaches for mean residence time estimation in ungauged basins using tracers and soil distributions. *Journal of Hydrology* **363**: 60–74. DOI: 10.1016/j.jhydrol.2008.10.001.

Thyer M, Renard B, Kavetski D, Kuczera G, Franks SW, Srikanthan S. 2009. Critical evaluation of parameter consistency and predictive

uncertainty in hydrological modeling: a case study using Bayesian total error analysis. *Water Resources Research* **45**: W00b14. DOI: 10.1029/2008wr006825.

Tromp-van Meerveld HJ, McDonnell JJ. 2006a. Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope. *Water Resources Research* **42**: W02410. DOI: 10.1029/2004wr003778.

Tromp-van Meerveld HJ, McDonnell JJ. 2006b. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research* **42**: W02411. DOI: 10.1029/2004wr003800.

Uhlenbrook S, Leibundgut C. 2002. Process-oriented catchment modelling and multiple-response validation. *Hydrological Processes* **16**: 423–440.

Vaché KB, McDonnell JJ. 2006a. A process-based rejectionist framework for evaluating catchment runoff model structure. *Water Resources Research* **42**: W02409. DOI: 10.1029/2005WR004247.

Vaché KB, McDonnell JJ. 2006b. Process-based strategies for model structural improvement and reduction of model prediction uncertainty. In *Predictions in Ungauged Basins: Promise and Progress*, Sivapalan M (ed). Wallingford, Oxon; 30–37.

Vitvar T, Burns DA, Lawrence GB, McDonnell JJ, Wolock DM. 2002. Estimation of baseflow residence times in watersheds from the runoff hydrograph recession: method and application in the Neversink watershed, Catskill Mountains, New York. *Hydrological Processes* **16**: 1871–1877. DOI: 10.1002/Hyp.5027.

Wolock DM, Fan J, Lawrence GB. 1997. Effects of basin size on low-flow stream chemistry and subsurface contact time in the Neversink River Watershed, New York. *Hydrological Processes* **11**: 1273–1286.

Zehe E, Sivapalan M. 2009. Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications. *Hydrology and Earth System Sciences* **13**: 1273–1297.

Zuber A, Maloszewski P. 2000. Lumped parameter models. In *Environmental Isotopes in the Hydrological Cycle—Principles and Applications*, Mook WG (ed). IAEA and UNESCO: Vienna; 5–35.

