Do time-variable tracers aid the evaluation of hydrological model structure? A multi-model approach

Hilary McMillan[1*], Doerthe Tetzlaff[2], Martyn Clark[3], Chris Soulsby[2]

[1] National Institute of Water and Atmospheric Research, New Zealand

[2] School of Geosciences, University of Aberdeen

[3] National Center for Atmospheric Research, Boulder, Colorado

*h.mcmillan@niwa.co.nz

For submission to Water Resources Research

25 November 2011

Revisions Submitted

29 February 2012

Abstract

In this paper we explore the use of time-variable tracer data as a complementary tool for model structure evaluation. We augment the modular rainfall-runoff modelling framework FUSE (Framework for Understanding Structural Errors) with the ability to track the age distribution of water in all model stores and fluxes. We therefore gain the novel ability to compare tracer/water age signatures measured in a catchment with those predicted using hydrological models built from components based on 4 existing popular models. Key modelling decisions available in FUSE are evaluated against streamflow tracer dynamics using weekly observations of tracer concentration which reflect the tracer Transit Time Distribution (TTD). Model structure choice is shown to have a significant effect on simulated water age characteristics, even when simulated flow series are very similar. We show that for a Scottish case study catchment, careful selection of model structure enables good predictions of both streamflow and tracer dynamics. We then use FUSE as a hypothesis testing tool to understand how different model characterisation of TTDs and MTTs affect multi-criteria model performance. We demonstrate the importance of time-variation in TTDs in simulating water movement along fast flow pathways, and investigate sensitivity of the models to assumptions about our ability to sample fast, near-surface flow.

1. Introduction

A wide range of lumped, conceptual, rainfall-runoff model structures are currently used for hydrological modelling applications (e.g. Singh, 1995). The model parameters are typically set by calibration which continues to be an important research strand within hydrology (e.g. Kavetski *et al.*, 2011; McMillan and Clark, 2009; Reichert and Mieleitner, 2009). However, a current shift in thinking is leading the hydrological community to re-emphasise the importance of model structure over and above model calibration (Beven, 2010; Clark *et al.*, 2011b; Krueger *et al.*, 2010; Savenije, 2009; Sivapalan, 2009). Model structure is critical because if model representations of the dominant runoff generation mechanisms of a catchment are not consistent with reality, the predictive power of the model may be reduced, especially outside the range of calibration conditions (Kirchner, 2006).

The challenge of selecting appropriate model structure for a given catchment is substantial. Aggregated performance measures such as the Nash-Sutcliffe may fail to distinguish between model structures (Clark *et al.*, 2008). This may be due to the compression of the error series into a single-valued measure (Gupta *et al.*, 2008; Schaefli and Gupta, 2007), to the choice of performance measure which may be sensitive to model structural complexity (e.g. Akaike, 1974), or to flexibility in parameterisation meaning that very similar flow predictions may be obtained from multiple model structures. Multi-response data have the potential to reduce ambiguity between competing model structures via evaluation of individual model components. This was shown in diagnostic tests proposed recently by McMillan *et al.* (2011) and Clark *et al.* (2011a), building on the concept of diagnostic signatures for model evaluation (Gupta *et al.*, 2008) and previous research into the benefits of auxiliary data to improve process understanding (e.g. Fenicia *et al.*, 2007; 2008; Seibert and McDonnell, 2002; Son and Sivapalan, 2007). Further challenges to selecting model structure include the common finding that increased model complexity is needed as extra data sources become available for evaluation (Vache and McDonnell, 2006) and the inability of standard data sources of rainfall and flow to discriminate between some aspects of model structure.

In this paper we explore the use of environmental tracer data as a complementary response dataset for model structure evaluation. Tracers are used to investigate geographical source areas and runoff pathways (e.g. Bergstrom *et al.*, 1985; Rodgers *et al.*, 2005a; Soulsby *et al.*, 2003; Soulsby *et al.*, 2006; Tetzlaff *et al.*, 2007b). Diagnostic tests using hydrometric data in conjunction with time domain or geographic source tracers, offer an alternative view on model performance (Birkel *et al.*, 2011a; Birkel *et al.*, 2011b; Botter *et al.*, 2008; Iorgulescu *et al.*, 2005). For example, Uhlenbrook and Leibundgut (2002) carried out a multi-response validation of a process-orientated catchment model, using measured runoff together with silica, ^{18}O, tritium and CFC tracers, and showed how the auxiliary data sources enabled a more realistic conceptualisation of runoff generation in their catchment. An important additional benefit of validating a hydrological model against both flow and tracer dynamics is that it could be used for integrated water quantity and quality applications (Krueger *et al.*, 2009).

When evaluating a hydrological model using environmental tracer data, two characterisations of transit time, i.e. the time water spends travelling through a catchment to the stream, are commonly used for comparison. These are the Mean Transit Time (MTT) and the Transit Time Distribution (TTD) of the tracer (which is assumed to be identical to that of the water). The TTD is the probability density function (pdf) of the time taken for water (or tracer) falling at a given moment to exit the catchment (i.e. the breakthrough curve). The MTT is the mean of this distribution. Estimates of MTT from observed data rely on an underlying model of tracer transport, often a simple pre-specified time-invariant TTD with calibrated parameters. Popular distributions include gamma, exponential, or exponential-piston flow; a review is given by McGuire and McDonnell (2006). The gamma distribution with shape parameter ≈0.5 has been shown to be appropriate for many catchments by analysis of the power spectra of conservative tracers in rainfall and streamflow (Godsey *et al.*, 2010; Kirchner *et al.*, 2000), implying the general need for a more peaked initial response and more sustained tail than a exponential distribution, i.e., as derived from a completely mixed reservoir.

For two reasons, the approach of a pre-specified time-invariant transit time distribution has recently been put under scrutiny. Firstly, work by Rinaldo *et al.* (2011; 2006) and Botter *et al.* (2011) has emphasised the differences between water ages in different storages and fluxes in a generalised theoretical model of a catchment, leading to inherent time-variation in TTDs. Secondly, Beven (2010) highlighted the need to apply a hypothesis testing framework to the estimation of TTDs and not to assume a particular form without evidence. Working within a multi-modelling framework allows exploration of these assumptions. The model performance can be evaluated using the tracer concentrations in the stream, requiring the model to reproduce the observed tracer dynamics, with the assessment made either graphically or using a performance measure (Fenicia *et al.*, 2010; Vache and McDonnell, 2006). The model simulations can then be used to derive and investigate the MTT, the shape of the TTD, and its variation with time and catchment wetness conditions. These characteristics can also be compared to possible TTD shapes and previous estimates of the MTT.

In this study we augment the modular modelling system FUSE (Framework for Understanding Structural Errors; Clark *et al.*, 2008) with the ability to track the age distribution of water in all model storages and fluxes. FUSE is a rainfall-runoff model building toolkit which allows the user to investigate hydrological modelling decisions, in particular the choice of state variables and flux equations to simulate water flow through a catchment. A complete model can be constructed with components based on well-known rainfall-runoff models: ARNO/VIC (Wood *et al.*, 1992), PRMS (Leavesley *et al.*, 1983), Sacramento (Burnash *et al.*, 1973) and Topmodel (Beven and Kirkby, 1979). The FUSE concept is designed to allow testing of competing modelling hypotheses of similar complexity but alternative structures, with individual control of each model component allowing systematic testing. We therefore gain the novel ability to track conservative tracers and compare tracer/water transit time signatures measured in a catchment with predictions made using this flexible modelling system. Our aims are as follows: [1] To compare the ability of competing model structures to predict stream tracer response, while retaining similar stream flow behaviour [2] To use the FUSE models as a tool to explore how different model characterisations of TTDs and MTTs (including time-variability) affect model behaviour and multi-criteria model performance [3] To use sensitivity analyses to show how simulated tracer response is affected by the interaction of model structure with parameter values and mixing assumptions.

2. Study Site

2.1 Catchment Characteristics

The Loch Ard Burn 10 (B10) catchment (0.9 km^2) lies in the Central Scottish Highlands (Figure 1), and was chosen due to availability of long-term hydrochemical tracer data. Average annual precipitation is 1980 mm and average runoff is 1660 mm. Slopes are gentle (generally less than 10º) and mean elevation is 170 m. The catchment is forested with plantations of Sitka Spruce (*Pitea Sitchensis*). Forest operations occurred between 1990-2002 with 39% of forest cover felled, however there is little evidence for any major change in average or high flows after the felling (Tetzlaff *et al.*, 2007a). The geology is dominated by low permeability metamorphic rocks (Miller *et al.*, 1990); bedrock outcrops occur on interfluves of the steep northwestern slopes. The most common soils are thin, poorly drained minerogenic gleyed soils.

Runoff generation processes are relatively well understood in the catchment (Dawson *et al.*, 2008; Hrachowitz *et al.*, 2009a; Tetzlaff *et al.*, 2010; Tetzlaff *et al.*, 2007a). The catchment is highly responsive, with low baseflow levels compared to stormflow (the ratio of low flows to flood flows may be up to 10^4) (Tetzlaff *et al.*, 2007a). The catchment maintains low soil moisture deficits and most parts of the catchment are highly connected to the stream network via a series of drainage ditches and saturated riparian zones, leading to high runoff:rainfall ratios (varying between 0.64 and 0.98; Dawson *et al.*, 2008). Storm runoff is thought to be dominated by flow paths in the upper soil horizons, influenced by high vertical gradients in the saturated hydraulic conductivity of the soil. Conductivity was found to vary from 0.3 cmh^{-1} in lower layers to 600 cmh^{-1} in surface layers in similar forested gley soils elsewhere (Soulsby and Reynolds, 1993). However tree roots and areas of exposed bedrock provide pathways to fracture systems in the bedrock, allowing some deeper recharge to occur (Tetzlaff *et al.*, 2010). Although hydrograph separations based on stream alkalinity are uncertain, average groundwater contributions to annual streamflow were estimated to be in the range 35 – 47 %.

2.2 Hydrometric Data

Daily rainfall totals were available using records from three gauges close to the catchment. Due to the flashy nature of the small catchment, rainfall totals at a sub-daily timestep were required in order to capture the fast runoff generation mechanisms and ensure correct timing of runoff in the model. Hourly rainfall data was available from four stations (Sloy, Loch Venachar. Abbotsinch and Bishopton) at 18 to 30 km from Loch Ard. The hourly data were expressed as a fraction of daily precipitation total at each hourly station, and the hourly ratios were interpolated (using inverse distance weighting) to the basin centroid. This timing information was then used to disaggregate the daily rainfall totals. Potential evapotranspiration (PET) was calculated based on daily temperature data using the Hamon method which is recommended for cases where radiation data is not available (Lu *et al.*, 2005). Flow data has been collected since 1989, using a concrete crump weir maintained by the Scottish Environment Protection Agency (SEPA). Flow data was extracted at a daily time step from the UK National River Flow Archive.

2.3 Hydrochemical Data

During the period 1990-2002, a consistent set of hydrochemical data including weekly precipitation and streamflow samples was available, and hence this time period was used for analysis (Figure 2). The precipitation samples (collected using open funnel bulk deposition samplers) and streamflow dip samples were filtered through a 0.45 μm polycarbonate membrane filter. Ion chromatography was then used to determine Chloride (Cl^-) concentrations. Chloride quantities in the catchment are increased due to dry and occult deposition, and hence the input concentrations were rescaled to ensure mass balance using an adjustment factor, assumed constant with time. Some previous studies suggest a range of models of dry and occult deposition including dependence on wind speed, wind direction and land use changes (e.g. Page *et al.*, 2007; Oda *et al.*, 2009). However, in the Atlantic-maritime Scottish context, dry and occult deposition is generally highest when sea-salt concentrations in the atmosphere are highest, which is also when wet deposition tends to be highest, hence a constant correction is a reasonable assumption. Kirchner *et al.* (2010) showed that when using a constant correction assumption in Scottish catchments, the use of chloride vs. isotope tracers led to consistent process identification, and therefore concluded that the unmodelled depositional processes do not materially affect inferences drawn from the data. For further details on the hydrochemical data collection, processing, and mass balance adjustment refer to Hrachowitz et al. (2009).

3. Methods

3.1 Tracking water through hydrological models

This paper uses the FUSE multi-model framework to enable individual control of hydrological model components, based on a variety of popular models. The modelling choices available include the choice of state variables in the unsaturated and saturated zones, and the choice of flux equations for surface runoff, interflow, vertical drainage, baseflow and evaporation. In order to compare modelled and measured tracer dynamics, in addition to flow dynamics, capability was added to the models to simulate routing and transit times of individual water 'parcels' through conceptual model stores.

We identified two possible strategies to achieve this capability, distinguished by the additional state variables used to track water movement. The first strategy uses state variables which quantify tracer concentrations in each conceptual store. The evolution of tracer concentration is controlled by input precipitation depth and tracer concentration, and flux equations describing tracer movement between storages. This is the method most commonly used in previous studies which integrate tracer information into hydrological models (e.g. Birkel *et al.*, 2010; Birkel *et al.*, 2011b; Dunn *et al.*, 2010; Fenicia *et al.*, 2010; Vache and McDonnell, 2006).

The second strategy uses state variables which quantify the distribution of water ages (defined as the elapsed time since a particle of water fell as rainfall) in each store, at a given time (i.e. the state variables are multi-dimensional and specify an empirical histogram of water ages). The evolution of the distributions is controlled by input precipitation depths, aging of the water in each store, and flux equations describing water movement between stores. This strategy is a generalisation of the previous method, as tracer concentrations in any store or flux can be directly calculated using convolution of the water age distribution with the corresponding input tracer concentrations (Figure 3). It also allows additional information to be easily derived such as mean and shape of the simulated water age distribution. This strategy relies on the underlying equations for conservative tracers derived by Botter *et al.* (2010; in particular Eq 17 for tracer mass flux) and summarised in Botter *et al.* (2011; Table 1). However the numerical implementation used in this paper differs as we solve concurrently for both soil water dynamics and age distributions.

An example of the implementation of the second strategy is given here for demonstration. Consider a simple model with variable S_1 representing water volume in the soil zone. The equation controlling evolution of S_1 may be as follows:

$$\frac{dS_1}{dt} = \left(p - q_{sx}\right) - e - q \qquad \text{Eq 1}$$

Where p is precipitation, q_{sx} is saturation excess runoff, e is evaporation and q is drainage. Now define a histogram (i.e. numerical vector representation of the pdf) $\mathbf{S_1^t}$ partitioning the volume S_1 by age. The equivalent differential equation for $\mathbf{S_1^t}$ is as follows:

$$\frac{d\mathbf{S_1^t}}{dt} = \left(\mathbf{p^t} - \mathbf{q_{sx}}^{\mathbf{t}}\right) - \mathbf{e^t} - \mathbf{q^t} \qquad \text{Eq 2}$$

Eq 2 relies on similar histogram distributions $\mathbf{p^t}$, $\mathbf{q_{sx}}^{\mathbf{t}}$, $\mathbf{e^t}$ and $\mathbf{q^t}$ of the fluxes p, q_{sx}, e and q. However, these histograms are known: water in rainfall (p) and q_{sx} is all of age 1; water in e and q has age distributions equal to that of $\mathbf{S_1^t}$ at the start of the timestep under the complete mixing assumption (refer to Section 3.3 on mixing assumptions), and the magnitude of these fluxes is given by the model equations. Therefore Eq 2 can be solved for $\mathbf{S_1^t}$ at the next timestep. The same strategy can be used for each model state equation, giving a complete solution for water age evolution in each store and flux. Finally, the method requires an initial histogram form (exactly as an initial value for all model states is required). A uniform distribution is used, followed by a spin up period as for the other model states.

In this study, the second strategy was preferred for its generality. An important aim of the study is to understand how different model characterisations of MTTs and TTDs affect model performance, and this information can be estimated more completely using the second method (see Section 3.2 for description of the relationship between TTD and water age). Hence, the additional capability was added to a FUSE prototype.

3.2 Model Output

The water-tracking model framework was designed to allow output of various aspects of simulated water age and transit times. Time series of the model state variables provide the age distribution in all stores, at each timestep (1 day increments were used here, matching the flow data resolution, but the timestep could be varied). Age distributions of all fluxes, including the catchment outlet flow, are also calculated. Time-varying statistics of the distributions, e.g. mean water age, can easily be derived. The TTD is calculated for each timestep in a secondary step which links each input quantity of rainfall to its age at the time it exited the catchment as streamflow. The TTD depends on both antecedent and current catchment wetness conditions, which determine how quickly water is driven through the catchment system. The TTDs may also be averaged over all timesteps to create a ‘master TTD’ (Botter *et al.*, 2011; Rinaldo *et al.*, 2011). The tracer volume or flux is given by the convolution of the water age distribution with the time series of input tracer concentrations. The model can be evaluated by its ability to simulate tracer dynamics by direct comparison of modelled and measured tracer outflow concentrations. This is a more direct and powerful test than invoking the MTT as a comparison tool, as any calculation of MTT relies on some underlying model of TTD.

3.3 Mixing Assumptions

Simulated water ages within a hydrological model are strongly dependent on the mixing assumptions used. Within a conceptual model store, instantaneous and complete mixing is the most usual assumption (e.g. Fenicia *et al.*, 2010; Vache and McDonnell, 2006). A justification for this may be that by stipulating the store as the fundamental unit of model design, complete mixing within that store is implicit: otherwise the store would represent an amalgamation of lower-level stores in which complete mixing did occur.

Recent work has however suggested that partial mixing behaviour may provide a more accurate representation of observed tracer concentrations (Barnes and Bonell, 1996; Dunn *et al.*, 2007; Fenicia *et al.*, 2010). Partial mixing refers to a water store in which some fraction of the volume controls hydrological response, with the remaining inert volume contributing only to tracer mixing. This concept is equivalent to a modification of the storage-discharge behaviour of the water store, i.e. that no discharge occurs below some threshold. Such behaviour is commonly assumed in hydrological models, e.g. that modelled percolation only occurs when soils are above field capacity (e.g. in the PRMS and Sacramento models underlying FUSE). In this study, mixing behaviours will only be changed in this way, i.e. through alternative storage-discharge parameterisations for both unsaturated and saturated model zones. The relevant model choices are as follows: In the upper zone, use of a single state variable simulates partial mixing, whereas use of split state variables simulates total mixing within the free storage reservoir. In the lower zone, the parallel linear reservoirs options simulate total mixing, but the Topmodel option simulates a hybrid method whereby discharge is greatly reduced but not zero as the volume of stored water decreases (for information on these model options refer to Section 3.4 and Figure 4).

An important aspect of mixing behaviour is the extent to which precipitation is assumed to mix with shallow soil water before flowing into the channel as saturation excess or other overland flow representations. Although saturation excess flow might be visualised as unmixed with soil water, empirical evidence using geochemical tracers in Scottish catchments suggests that surface runoff does in often partially acquire the chemical signature of soil water (Birkel *et al.*, 2011b). If model simulation of mixing is required, its occurrence and extent must be exactly specified, possibly through introduction of calibrated parameters if sufficient process knowledge is not available. In this study, the simplest option was used whereby saturation excess flow was treated as unmixed, in common with previous studies (e.g. Botter *et al.*, 2008). To explore the impact of this assumption, a sensitivity analysis was carried out to investigate the effect of flow partitioning between surface (unmixed) and subsurface (mixed) pathways (refer to Section 4.5).

3.4 Model Implementation

The FUSE framework provides hundreds of possible model combinations using different combinations of components from 4 popular hydrological models (Clark et al., 2008). In this study, to provide a manageable scope we investigate the effect of key decisions of upper and lower layer architecture on the simulated streamwater transit time (Conceptual diagrams including the outflow pathways for each model component are shown in Figure 4).

In all cases the following decisions are treated as fixed. [a] Evapotranspiration is satisfied from the single upper soil layer: this is the simplest option available. [b] Percolation is parameterised as a linear function of upper zone storage above field capacity: again the simplest option. Note also that the alternative formulation of percolation as a power function of total upper zone storage was found to give poor results in initial trials. [c] Surface runoff is parameterised as a power function of total upper zone storage, except when using the Topmodel formulation where it is controlled directly from lower zone storage. The state and flux equations defining each of the resulting 6 models are given in Table 1, with fluxes defined in Table 2. The alternative choices provided for in FUSE could be investigated for their effect on transit time in future work.

FUSE is formulated as a state-space model and enables several classes of time stepping schemes to control model numerical behaviour (Clark and Kavetski, 2010; Kavetski and Clark, 2010). The additional model equations required to track water age are similarly written in state-space form. The numeric scheme chosen was a fixed-step Explicit Euler for simplicity, using short 15 minute sub-steps to ensure numerical stability and accuracy. The model used input precipitation data at hourly resolution. Model flow simulations were evaluated at a daily timestep, commensurate with flow data availability and which minimises the effect of any rainfall timing errors introduced by the interpolation method used for rainfall disaggregation. In our study, evaluation at daily timestep seemed sufficient to capture the flow generation processes of interest (i.e. the effect of upper and lower zone model architecture choices), and is at higher resolution than processes captured by tracer measurements which relate to (slower) water transit times rather than the sub-daily dynamic response.

3.5 Model Parameters

When comparisons are made between hydrological model structures, there is interplay between the choice of model structure and the choice of model parameters: both can influence flow and transit time predictions and each can compensate for deficiencies in the other, though not necessarily in agreement with reality. In this study the focus was on model structure. Therefore default parameter values for the FUSE models were used where possible, as recommended by Clark et al. (2011a). Measured information or process knowledge from the Loch Ard catchment was also used to set parameter values where appropriate; this method assumes a translation from field to model scale but given the process-orientated nature of the models it was considered preferable to setting the parameters via calibration. The depth of the upper humic/peaty soil layer contributing to shallow subsurface slow is approximately 400 mm (Tetzlaff *et al.*, 2007a); assuming a typical porosity for peat of 0.8 allows the upper store depth to be set as 320 mm. Typical field capacity for peat of 0.35 enables the fraction of total storage as tension storage to be set at 0.44 (=0.35/0.8). Known values of the fractional groundwater contribution to streamflow were also used as 'soft data' (Seibert and McDonnell, 2002) to guide the parameter choice. A digital terrain model (EDINA Digimap) of the catchment was used to estimate the topographic index distribution parameters required for the Topmodel component of FUSE.

The remaining 1 or 2 parameters relating to the lower zone storage (storage depth, baseflow exponents, baseflow depletion rate(s)) were chosen using a simple calibration procedure by exhaustive search (accompanying visualisation by contour plot) of model performance in relation to parameter value (Figure 5). As shown, the single linear reservoir model is not sensitive to the lower zone storage depth (this parameter only influences model predictions in the rare case that the tank fills completely) and hence this is set to infinite depth in the model (this is also true for the stores in the model with two parallel linear reservoirs). The Topmodel nonlinear reservoir model shows dependency between the lower zone storage size and baseflow exponent, which could therefore be varied jointly in the model to improve tracer simulations if necessary. The dependence is indicated by the form of the baseflow equation (Table 1). The same parameter sets were used for both single and split variable upper zone structures. The complete parameter sets thus derived provide a robust baseline calibration for comparisons between structures (Table 3). The fitted models all give very similar predictions of flow dynamics, with only very minor differences in the flood peaks and low recessions. Nash Sutcliffe scores were all in the range 0.75 – 0.80 when validated over a 12 year period.

4. Results

This section is organised as follows. First the 6 different FUSE model structures (2 options for upper zone architecture * 3 options for lower zone architecture) are evaluated against the tracer measurements from the B10 catchment using direct comparison using tracer output series [Section 4.1]. A comparison with the results of previous studies is also made using MTTs [Section 4.2].

Secondly, we use the FUSE models as a tool for hypothesis testing by comparing characteristics of simulated TTDs and MTTs between models with differing performance. (1) Models are run in steady state (i.e. constant precipitation input) to study time-invariant representations of the TTD [Section 4.3.1] (2) Models are run dynamically (i.e. measured precipitation input) to study time-varying behaviour on MTT and TTD caused by seasonal/event-scale changes in wetness conditions [Section 4.3.2] (3) A sensitivity analysis of effect of model calibration [Section 4.4] and mixing behaviour [Section 4.5] on the shape of the modelled TTD

4.1 Model structure evaluation: Output tracer dynamics

The models were driven using measured precipitation depths and weekly precipitation chloride concentrations for the years 1990 – 2002. Observed chloride concentrations in streamwater were then compared with the model simulations. The results are shown in Figure 6 (Panels A & B), with close-ups (Panels C & D) of the largest peak in the tracer concentration series, from Dec 1992 – Oct 1993.

Figure 6 shows the clear differences in simulated tracer response between models using single vs. split upper state variables. The models using a single variable simulate greater mixing of soil water and hence a more damped tracer response, which corresponds more closely to the measured streamwater chloride concentrations. The split upper state variable approach produces simulated spikes in tracer concentration (due to reduced mixing within the model leading to faster tracer breakthrough) which do not occur in the measured data. Hence to provide a model which can simulate both flow dynamics and tracer response in the Loch Ard catchment, the single state variable formulation would be the preferred choice.

Within those models using the single upper state variable, the choice of lower zone formulation makes a smaller but evident difference in simulated tracer response. The single linear reservoir model simulates extended peaks of tracer concentration higher than those measured, and concentrations which are too low during recession periods. This indicates that water is routed too quickly through the model, with insufficient depth of stored water for realistic mixing behaviour. The parallel linear reservoir and Topmodel formulations simulate less sustained peak concentrations which more closely match the measured values (e.g. Figure 6C). In recession periods however, the parallel linear reservoir model simulates too low concentrations, and hence this model has insufficient mixing in the lower reservoirs. The Topmodel architecture (i.e. a single nonlinear reservoir) most closely simulates tracer recession behaviour, and is overall most successful in reproducing the tracer dynamics.

Both the parallel linear reservoir and Topmodel architectures produce unobserved short-duration fluctuations in tracer concentration, and all models simulate unrealistic periods of constant tracer concentration. Recessions in the chloride concentrations are also too rapid in some cases (e.g. 1997-1998). These weaknesses are caused by limitations in all the structures tested which assume a maximum 3 flow pathways, often decreasing to 1 flow pathway during recession periods when surface and subsurface stormflow pathways are not active. The short-duration fluctuations are largest in the Topmodel architecture because water ages differ most strongly between the upper and lower reservoirs, the same characteristic which produces realistic extended recession curves. In reality, chloride concentrations represent an aggregation of pathways derived from the spatial and temporal heterogeneity of the catchment (as shown by Rinaldo *et al.*, 2006). This aggregated solute mixing behaviour is analogous to that found for flow recessions at the catchment scale which integrate the behaviour of many hillslopes (Harman *et al.*, 2009).

4.2 Model Structure Evaluation: Mean Transit Times

In this section we investigate the effect of model structure on MTT and compare the 6 FUSE model estimates of MTT with those previously derived for the Loch Ard B10 catchment. The MTTs predicted by the FUSE models are all relatively short, less than 150 days (Figure 7). There is a marked split whereby models which use a single upper state variable [S/1Linear, S/2Linear, S/Topmodel] have longer MTTs than those which use split upper state variables for tension and free storage [Sp/1Linear, Sp/2Linear, Sp/Topmodel]; resulting from the different mixing characteristics as described in the previous section. Short MTTs are consistent with the dominant responsive soils (peats, gleys) that generate a quickflow response in the Loch Ard catchment. Indeed, previous work has shown dominant soil cover to be the best single landscape predictor of catchment MTTs in the Scottish Highlands (Hrachowitz *et al.*, 2009a; Rodgers *et al.*, 2005b; Speed *et al.*, 2010; Tetzlaff *et al.*, 2009). Previous estimates of the MTT (Table 4 and Figure 7) are typically longer than the FUSE estimates, and have a wide range due to the range of models used (refer to Table 4), highlighting the difficulty of choosing an appropriate TTD shape, particularly under an assumption of time invariance. The time invariance assumption may also lead to an under-representation of fast flow pathways and hence a longer MTT (refer to Section 4.3.2 for a discussion). The FUSE models demonstrate that the range in MTT due to dynamic wetness conditions can be greater than the range due to choice of model structure.

4.3 Synthetic Experiments

The FUSE models can be used to investigate the relationship of model structure to simulated water age characteristics. The performance of the models in reproducing tracer concentrations (Figure 6) can then be used to judge which types of water age dynamics are most realistic. The models can be used to investigate aspects of water age which we are not currently able to measure directly, such as TTDs.

4.3.1 Steady State Models

In real conditions, TTDs can change seasonally or by event (McGuire and McDonnell, 2010; Weiler *et al*., 2003) in any catchment due to varying catchment wetness (Birkel *et al.*, 2012; Hrachowitz *et al.*, 2010; McGuire *et al.*, 2007; Nyberg *et al.*, 1999); this correspondingly causes temporal variation of MTTs (Lindstrom and Rodhe, 1992; Turner *et al.*, 1987). We initially avoided this complexity by using a synthetic constant precipitation input, to determine whether different model structures simulate different steady state water TTDs (even when simulated flow dynamics are similar) and how that gives rise to the different tracer dynamics shown in Figure 6.

Each catchment model was run with constant rainfall and PET input set at the average per-timestep depth. The models were spun up to steady state (1 year) and then run for a further 11 years to capture the TTD including the tail, consistent with the findings of Hrachowitz *et al.* (2011) who found that a spin-up period of approximately 3 times the MTT was required. The steady state TTDs are shown in Figure 8 for (A) Total subsurface flow and (B) Deep groundwater only.

The TTDs demonstrate clear differences between model structures. The total flow TTDs show that models with split upper state variables have a more peaked initial response than those with a single variable. This helps to explain differences in simulated tracer dynamics, as the former route storm water more quickly to the channel with less mixing with older water. The poorer performance in tracer simulation for these models shows that this is a less realistic conceptualisation, concurring with previous studies which highlight the importance of deep flow pathways for solute transport (e.g. Botter *et al.*, 2008).

The maximum of the distribution is typically close to zero indicating the dominance of fast flow pathways; although models using a single upper state variable and linear lower zone reservoirs have a slightly later maximum. Non-zero peaks have been found in previous studies e.g. McGuire *et al.* (2007) who simulated bromide tracer flux in a steep hillslope with gravelly clay loam soils over relatively low permeability bedrock and found that modelled TTDs peaked at 10 days rather 0 days. In some drier climates, lags may also be related to inter-arrival times of storms or wet periods when more than one storm event is required to flush the tracer through the catchment (Rinaldo *et al.*, 2011; the climate example used was 180 mm/yr rainfall with 10% rainy days). The models using the Topmodel formulation, most successful in simulating tracer response, have flatter responses than those using linear reservoirs. Note that the TTDs given do not include the saturation excess flow pathway: this pathway provides an unmixed pulse of tracers at transit times of < 1 day. The TTDs for baseflow only (indicative of the behaviour of the catchment in a drier state) are flatter with more delayed responses showing the longer transit times for water following deeper flow pathways.

4.3.2 Effect of rainfall variation and antecedent catchment wetness on water transit time distribution

The previous section examined the case of the catchment in steady state, and hence an invariant transit time distribution. This assumption lies behind the majority of interpretations provided of experimental data for MTTs which use a fixed distribution to model the TTD. In reality, TTDs vary according to the wetness state of the catchment on both a seasonal and event time-scale. Recently, the time variant nature of TTDs has been stressed by Botter *et al.* (2010) who also developed the underlying theory. Complementary work by Hrachowitz *et al.* (2010) demonstrated inter-annual variation in gamma TTDs and showed that the b (scale) parameter could be linked to precipitation intensity. However, when applied to a catchment like Loch Ard B10, time-variance may be weaker due to the year-round wet climate and peaty soils, as has been found in other case studies carried out in wet catchments (Hrachowitz *et al.*, 2010; Rinaldo *et al.*, 2011). The long data record also helps to ensure that the full range of catchment response pathways is captured and hence a stationary TTD more completely represents catchment behaviour.

The FUSE framework allows us to explore the TTD time variation simulated by different model structures, and hence test the hypothesis that these variations are required for realistic tracer simulation. Here, the FUSE models were driven using the recorded precipitation time series (after spin-up to steady state as for Section 4.3.1). Figure 9 demonstrates how MTT varies over a multi-year period, showing strong seasonal variation in 4 of the 6 models. The longer MTTs during dry periods contribute proportionately less to the total MTT due to the weighting effect of the lower fluxes involved. Note that the modelled dry season MTTs are still relatively short, reflecting the small size of the actual groundwater stores at Loch Ard and their rapid turnover. The 2 models with weak seasonality have split upper state variables and linear lower reservoirs, and display very short MTTs (< 10 days) which vary with individual rainfall events rather than the seasonal cycle. Models with a single upper state variable display longer MTTs as these simulate greater mixing of water within the soil zone (a conceptualisation of mixing of water held in tension in the soil matrix with free water in the matrix or macropores). The structure of the lower zone also affects MTT: in particular the Topmodel formulation leads to longer MTTs since the nonlinear drainage function means that a greater volume of water is retained in the lower store between rainfall events.

By comparing the MTT variability (Figure 9) with model tracer simulations (Figure 6) we see that the models which simulate longer, seasonally-varying MTTs provide most realistic tracer dynamics. However it is not sufficient for a model to reproduce the seasonal cycle in MTT to achieve good performance. For example, the model with split upper state variables, and Topmodel formulation lower architecture, produces a seasonal cycle due to the larger lower store, but produces unrealistic event-scale tracer response due to lack of simulated mixing in the upper soil zone. None of the models tested are able to simulate long MTTs without also producing a seasonal cycle in MTT, because tracers that persist over multiple months are subject to seasonal changes in the model wetness state that are necessary to simulate seasonal differences in the flow dynamics.

In addition to the MTT, the full TTDs for different wetness conditions can be compared with both the master and steady-state TTDs (Figure 10). This helps to determine whether steady state models can produce a good approximation to the master TTD. The answer is likely to be catchment-specific, as catchments with less pronounced fluctuations in their climate (including seasonality and other timescales) will have more similar master and steady-state TTDs. Here we show TTDs for the three model structures which simulated the most realistic tracer series, i.e. upper zone modelled with a single variable, 3 lower zone architectures. In all cases, there is a strong differential between TTD shapes in wet and dry conditions for fast flow pathways (less than 30 days). In particular, the dry TTD is bimodal with peaks at < 5 days and 50-60 days, but a reduction in flow paths compared to the wet TTD in the approximate range 5-30 days. The differences between wet and dry TTDs are due to the initial water depth in the model, the extent to which later rainfall fills model stores and increases flow, and the proportions of runoff from saturation excess flow, interflow and baseflow.

The differences in TTD for timescales up to 30 days for wet and dry conditions (with corresponding differences between the master and steady state TTDs), suggest that steady state models will not simulate realistic tracer transport at short time scales. However, at longer time scales there is less difference between TTDs for wet or dry conditions, especially in the best performing model (Topmodel formulation) where all TTDs have heavier tails. We conclude that for slow flow pathways in the B10 catchment, the dynamic nature of the TTD is less important and could reasonably be approximated by a steady state model. In log space [not shown] the steady state TTDs are approximately linear, suggesting that an exponential model could be used. However the dynamic TTDs show additional fast flow pathways which are not captured by the exponential distribution. This helps to explain why a gamma function is often found to be more successful than an exponential function in reproducing tracer dynamics (Godsey *et al.*, 2010), especially at the event scale (Birkel *et al.*, 2012), although modelled TTDs do not always conform to simple statistical distributions (Dunn *et al.*, 2010).

4.4 Sensitivity of TTDs to model parameters

The model TTD is sensitive to parameter values as well as model structure. Although most parameters were set using field knowledge, there is still uncertainty in the appropriate value at model scale. We therefore undertook a sensitivity analysis to investigate the effect of available depths of upper and lower zone storage on the model TTD, allowing some insights into the interplay of model structure and parameterisation. The storage depths were chosen as parameters to be varied because the depth of water available for mixing is know to be an important control on model ability to simulate tracer dynamics (Fenicia *et al.*, 2010).

The model used was [Upper zone: Single Variable, Lower Zone: Topmodel], as this produced the most realistic simulation of tracer dynamics (Figure 6). The effects of changing upper and lower zone storage depths on TTD and model performance are shown in Figure 11. The results show that the TTD is more sensitive to the size of the upper zone store than the lower zone. We suggest that this is due to the greater nonlinearity of response in the upper store which is controlled by a threshold rather than a power function. The changes resulting from perturbation of upper zone size are of comparable magnitude to those resulting from a change in model structure and should therefore be considered alongside model structure when creating a model which realistically reproduces tracer dynamics.

Figure 11 also shows that model performance is more sensitive to the size of the lower zone store than the upper. Performance falls quickly away from the optimal value. Less sensitivity is found to the size of upper zone store but model performance could be slightly improved by increasing the store size above the value of 320mm set using results from field knowledge, with a corresponding increase in MTT.

4.5 Sensitivity of TTDs to mixing of saturation excess flow

An important decision in the modelling process was whether saturation excess flow should be treated as mixed or unmixed with subsurface stormflow. In some environments, high intensity rainfall may run off quickly and be missed by weekly sampling. However in the Loch Ard wet environment with peaty soils and relatively low intensity frontal rainfall, there is usually ready availability of water in the upper organic horizons for mixing and hence the displacement of resident soil water becomes the dominant source of runoff.

To explore this question a sensitivity analysis was carried out into the effect of flow partitioning between surface (unmixed) and subsurface (mixed) pathways. We used the model with a single upper state variable and 2 parallel linear reservoirs, because it provides a good simulation of tracer dynamics during high flows (when surface pathways are active) and the effect of surface flow mixing can be easily studied by changing the parameter 'ARNO/VIC b exponent' which controls the quantity of surface *vs.* subsurface flow by changing the estimate of saturated area based on upper zone soil water storage (see model equation in Table 1). The results (Figure 12) show that the TTDs are relatively insensitive to the introduction of additional water into the soil zone (i.e. increasing b), when compared to sensitivity to store sizes (Figure 11). We therefore suggest that in this case it is acceptable to make the simplifying assumption the saturation excess flow is unmixed.

5. Discussion

Water transit time characteristics provide a valuable diagnostic tool for evaluation of model structure, to complement the traditional comparison of modelled and measured discharge series, as shown both in this and previous papers (Birkel *et al.*, 2011b). While other data sources such as soil moisture or depth to water table can also be used for multi-response evaluation, they are typically point measurements subject to the 'scaling problem' (Blöschl and Sivapalan, 1995; Sivapalan *et al.*, 2004). Tracer dynamics are particularly useful as they provide an alternative integrated signal to the hydrograph.

In return, hydrological models (including mixing assumptions) provide a tool for investigating scenarios of water TTD shape, and variability with catchment wetness. These characteristics are not directly measurable using environmental tracers, and hence models provide a method for their estimation. An estimate of the TTD shape is required for studies which use inverse modelling to obtain MTT estimates and then apply the results to simulate tracer, chemical or contaminant transport (McDonnell *et al.*, 2010). It is hoped that future work will indicate whether distribution shapes and variability are transferable between neighbouring or hydrologically similar catchments.

This study relied on several assumptions. Firstly, uncertainty in rainfall, climate, streamflow and chloride measurements was not considered, although it is well known that measured hydrological data is subject to many sources of uncertainty (e.g. Andreassian *et al.*, 2004; Beck, 1987; Pelletier, 1988). These uncertainties can have substantial effects on calibrated parameter values (e.g. McMillan *et al.*, 2010) and may therefore indirectly affect water transit time characteristics predicted by the model. A second assumption was that the effects of dry deposition and biogeochemical cycling on chloride concentrations were modelled using a constant, multiplicative adjustment factor to correct the mass balance (refer to Section 2.3).

Our modelled transit times were generally shorter than previous estimates from tracer data, consistent with previous findings that model storage volumes required to capture water quantity dynamics are smaller than those required to reproduce tracer dynamics (e.g. Fenicia *et al.*, 2010). Our work highlighted the value of FUSE to understand which model structure and TTD characteristics (shape, time-variability) enable simulation of both flow and tracer concentrations. For example, at Loch Ard this could be achieved using a Topmodel style nonlinear lower zone store, with a TTD which has a greater weight of fast flow pathways than the exponential distribution and varies with catchment wetness at short time-scales. Although previous studies have shown that water and tracer dynamics can be used to tailor a model for an individual catchment (e.g. Birkel *et al.*, 2011b), the FUSE framework provides much greater flexibility in model structure. It leads towards a robust, transferable method for water and tracer modelling that could be relatively easily used in a wide range of catchments by selection of appropriate FUSE model components according to process knowledge or structural diagnostics, on a per-catchment basis or using a regionalisation method.

The study has also led to recommendations for model structure options that could be added to FUSE to improve the concurrent representation of streamflow and tracer dynamics. For example, subsurface stormflow is currently modelled as a linear function of free storage in the upper zone. When this pathway is used in a model with separate state variables for tension and free storage, the free storage becomes a very fast response store with low transit times. Recent ecohydrologic experiments suggest that in a strongly seasonal, Mediterranean climate where there is significant summer soil drying, water in the soil matrix may be largely decoupled from that in fast flows paths (Brooks *et al.*, 2010; Phillips, 2010). In climates where it occurs, this behaviour would be more closely modelled by the split upper state variables approach. One method to reconcile longer mean transit times with split state variables would be to use a nonlinear response function for interflow (e.g. a power function similar to those used to model percolation).

There are many needs for future research into transit time distribution characterisation; a summary was provided by McDonnell *et al.* (2010). This study highlighted that although MTT provides a very useful summary statistic of catchment behaviour, there is a need for better measurement techniques which work towards characterisation of the complete time-variable TTD: this would reduce ambiguity in transit time estimates and provide extremely valuable data against which to test different model structures. Further, although of lesser importance in a fast-responding catchment such as Loch Ard, conservative/natural tracers are not adequate to capture behaviour in catchments with MTT of greater than a few years (Hrachowitz *et al.*, 2009a; Stewart *et al.*, 2010), meaning that alternative tracers or methods are needed to investigate TTD tails in catchments with long response times. Improved understanding of the true TTD would also help to counter other causes of bias such as streamwater tracer sampling biased towards low flows, or model inability to differentiate multiple deep groundwater stores.

6. Conclusions

In this paper we demonstrated how augmenting the FUSE rainfall-runoff modelling framework with a water-tracking ability provides the opportunity to use tracer data as an additional model structure diagnostic. Using a range of calibrated models for the Loch Ard B10 catchment in Scotland, we showed that different model structures which provide very similar flow dynamics (and hence performance as measured by a sum-of-squared-errors score) can produce very different simulations of water TTD and tracer dynamics. We evaluated different model structures against streamflow tracer dynamics using weekly observations of tracer concentration. In the Loch Ard catchment, a model structure could be selected to provide good simulations of both flow and tracer dynamics. We used the water-tracking models as a hypothesis testing tool to explore the effect of catchment transit time characteristics on model behaviour and performance. Across model structures we showed strong seasonality and event-scale fluctuation in MTT and TTDs; and corresponding differences between dynamic and steady state TTDs. The results suggest that steady-state approximations to the catchment TTD at Loch Ard will not simulate realistic tracer transport at short time scales (< 30 days), although differences are less marked at longer time scales. The FUSE framework with water age characterisation provides a tool to investigate flow and tracer modelling in competing model structures, which could be relatively easily applied to many catchments.

Acknowledgements: The authors are grateful to Dr Iain Malcolm of Marine Science Scotland's Freshwater Laboratory for access to the Loch Ard chloride data. We thank Tobias Krueger and two anonymous referees for their thorough and constructive reviews.

Figure Captions

Figure 1. (A) Loch Ard B10 catchment map and instrument locations (B) Photograph of Loch Ard B10 catchment

Figure 2. Hydrometric and hydrochemical data available for the Loch Ard B10 catchment

Figure 3. Conceptual diagram showing process used to calculated model TTDs and outflow tracer concentrations in a sample FUSE model. (A) Water age distribution of each reservoir (S_1 upper zone, $S_2^{A,B}$ lower zone) stored as histogram. Fluxes (p precipitation, q_{if} interflow, q_{12} drainage, $q_b^{A,B}$ baseflow) have the age signature of their source reservoir. (B) Outflow age distribution for time t is the sum of distributions from component fluxes (q_{if}, $q_b^{A,B}$). (C) TTD of input water is calculated from the corresponding outflow times. (D) Outflow tracer concentration calculated by convolution of outflow age distribution with precipitation tracer concentrations

Figure 4. Simplified wiring diagram showing model architecture options used in this study. Upper Zone: [S] A single state variable S_1 combining tension and free storage [Sp] Separate state variables for tension S_1^T and free S_1^F storage. Lower Zone: [1 Linear] S_2 A single linear reservoir [2 Linear] S_2^A, S_2^B Two parallel linear reservoirs [Topmodel] S_2 A single nonlinear reservoir based on Topmodel concepts (where surface runoff q_{sx} is controlled by the lower zone). Key to soil moisture values: θ_W wilting point (here 0), θ_F field capacity $\left(\phi \cdot S_{1,\max}\right)$, θ_S saturation point $\left(S_{1,\max}\right)$.

Figure 5. Calibration results for lower zone storage parameters for 3 lower zone model architectures. The objective function is the sum of squared errors between modelled and measured discharge series, after Box-Cox transformation to normalise error variance. The calibration period was over two hydrological years (1998-1999).

Figure 6: Time series of measured Chloride input and output concentrations and comparisons with model predictions. (A) Models with a single upper zone storage variable (B) Models with split upper zone storage variables (C) Close-up of A for largest event (Dec 1992 – Oct 1993). (D) Close-up of B for largest event

Figure 7: Comparisons of MTT estimates between models (run in dynamic and steady state mode) and from previous studies (Table 4) of the B10 catchment.

Figure 8 Steady state transit time distributions for a range of model structures. (A) Combined flow: Subsurface stormflow + groundwater flow (B) Groundwater flow only

Figure 9: Variation of Mean Transit Time with time for a range of model structures (A) Models with single upper state variable, (B) Models with split upper state variables. (C) Measured Flow is plotted for comparison

Figure 10: Variation in transit time distribution according to catchment wetness condition for 3 model structures. TTDs are given for (All): All days in record, (Wet): Days in lower quartile of MTT distribution, (Dry): Days in upper quartile of MTT distribution, (Steady State): Steady state TTD for comparison.

Figure 11. The effects of changing upper and lower zone storage depths on Transit Time Distribution (upper panels) and model performance (lower panels). TTDs are shown for equal increments/decrements of store size (thin lines) up to the maximum/minimum values given (thick lines).

Figure 12. The sensitivity of the model to soil water mixing is shown by varying the surface flow b parameter. Effects are shown on Transit Time Distribution (upper panel) and Percentage share of flow volume between pathways (lower panel).

Table 1. State and flux equations for the 6 FUSE models tested in this paper.

Model	1 (Single Upper / 1-Linear Lower)	2 (Single Upper / 2-Linear Lower)	3 (Single Upper / Topmodel Lower)	4 (Split Upper / 1-Linear Lower)	5 (Single Upper / 2-Linear Lower)	6 (Single Upper / Topmodel Lower)
Unsaturated Zone Architecture *	$\frac{dS_1}{dt} = \frac{(p-q_{sx})-e-}{q_{12}-q_{if}-q_{ufof}}$			$\frac{dS_1^T}{dt} = (p-q_{sx})-e-q_{utof}$ $\frac{dS_1^F}{dt} = q_{utof}-q_{12}-q_{if}-q_{ufof}$		
Saturated Zone Architecture *	$\frac{dS_2}{dt} = q_{12}-q_b-q_{sfof}$	$\frac{dS_2^A}{dt} = \frac{q_{12}}{2}-q_b^A-q_{sfofa}$ $\frac{dS_2^B}{dt} = \frac{q_{12}}{2}-q_b^B-q_{sfofb}$	$\frac{dS_2}{dt} = q_{12}-q_b-q_{sfof}$	$\frac{dS_2}{dt} = q_{12}-q_b-q_{sfof}$	$\frac{dS_2^A}{dt} = \frac{q_{12}}{2}-q_b^A-q_{sfofa}$ $\frac{dS_2^B}{dt} = \frac{q_{12}}{2}-q_b^B-q_{sfofb}$	$\frac{dS_2}{dt} = q_{12}-q_b-q_{sfof}$
Evapotranspiration	$e = pet \cdot \min\left(\frac{S_1}{\phi \cdot S_{1\max}},1\right)$					
Drainage	$q_{12} = k_u \cdot \left(\frac{S_1^F}{(1-\phi)S_{1\max}}\right)^c$					
Interflow ***	$q_{if} = k_i \cdot \left(\frac{S_1^F}{(1-\phi)S_{1\max}}\right)$					
Baseflow ***	$q_b = \upsilon S_2$	$q_b = \upsilon_A S_2{}^A + \upsilon_B S_2{}^B$	$q_b = \frac{K_s S_{2\max}}{\lambda^n n} \cdot \left(\frac{S_2}{S_{2\max}}\right)^n$	$q_b = \upsilon S_2$	$q_b = \upsilon_A S_2{}^A + \upsilon_B S_2{}^B$	$q_b = \frac{K_s S_{2\max}}{\lambda^n n} \cdot \left(\frac{S_2}{S_{2\max}}\right)^n$
Surface Runoff ***	$q_{sx} = p\left(1-\left(1-\frac{S_1}{S_{1\max}}\right)^b\right)$		$q_{sx} = p\int_{\zeta crit}^{\infty} f(\zeta)d\zeta$ ** $\zeta_{crit} = \lambda\left(\frac{S_2}{S_{2\max}}\right)^{-1}$	$q_{sx} = p\left(1-\left(1-\frac{S_1}{S_{1\max}}\right)^b\right)$		$q_{sx} = p\int_{\zeta crit}^{\infty} f(\zeta)d\zeta$ ** $\zeta_{crit} = \lambda\left(\frac{S_2}{S_{2\max}}\right)^{-1}$

* Overflows from tension (q_{utof}), free (q_{qufof}) and lower (q_{sfof}) reservoirs represent addition flow into the free storage, surface runoff and baseflow, respectively. Logistic functions are used to smooth the threshold relating to the fixed storage capacities (following Clark *et al*., 2008; Section 4.8).

** The variable ζ for surface runoff parameterization of Models 3 and 6 describes the spatial distribution of the topographic index (Beven and Kirkby, 1979). The distribution used is $\Gamma(\lambda,\chi)$ fitted to data from the digital elevation model (following Clark *et al*., 2008; Section 4.6).

*** The time delay in runoff is modelled using a gamma distribution $\Gamma(\mu,3)$ of routing times applied to all fluxes (following Clark *et al*., 2008; Section 4.9)

Table 2. Model Fluxes (all unit are mm d^{-1})

Variable Name	Description
p	Precipitation
e	Evapotranspiration
q_{sx}	Saturation Excess Runoff
q_{if}	Interflow (Subsurface Stormflow)
q_{12}	Drainage from upper to lower zone
q_b q_b^A q_b^B	Baseflow (from single, primary, secondary reservoir)
q_{utof} q_{ufof}	Overflow from upper zone (from tension, free reservoir)
q_{sfof} q_{sfofa} q_{sfofb}	Overflow from lower zone

Table 3. Parameters used for different FUSE models. Parameter values are identified as [`Field`] identified from field knowledge, [`Default`] Default recommended FUSE values, or [`Calibrate`] Calibrated. Refer to Section 3.5 for details.

		Lower Zone Formulation			
Parameter	**Description**	**Single Linear**	**Parallel Linear**	**Topmodel**	**Parameter Type**
$S_{1,\max}$	Maximum storage in unsaturated zone (mm)	320.0	320.0	320.0	Field
$S_{2,\max}$	Maximum storage in saturated zone (mm)	Inf	Inf	91.3	Calibrate
ϕ	Fraction total storage as tension storage	0.440	0.440	0.440	Field
k_u	Vertical drainage rate (mm/day)	750.0	750.0	750.0	Default
c	Vertical drainage exponent	1.0	1.0	1.0	Default
k_i	Interflow rate (mm/day)	1000.0	1000.0	1000.0	Default
k_s	Baseflow rate (mm/day)	1000.0	1000.0	1000.0	Default
n	Baseflow exponent	N/A	N/A	12.18	Calibrate
υ	Baseflow depletion rate (single reservoir) (/day)	0.176	N/A	N/A	Calibrate
υ_A	Baseflow depletion rate (primary reservoir) (/day)	N/A	0.840	N/A	Calibrate
υ_B	Baseflow depletion rate (secondary reservoir) (/day)	N/A	0.0317	N/A	Calibrate
b	ARNO/VIC 'b' exponent	0.500	0.500	0.500	Default
λ	Mean of log topographic index distribution (m)	N/A	N/A	5.91	Field
χ	Shape parameter of topographic index distribution	N/A	N/A	2.57	Field
μ	Time delay in runoff	0.3	0.3	0.3	Calibrate

Table 4. Previous estimates of MTT in the Loch Ard B10 catchment

Reference	**Model**	**MTT (days).**
Tetzlaff *et al.* (2007a)	Exponential	120-180
	Exponential-Piston flow	180-270
	Sine wave	60
Godsey *et al.* (2010)	Gamma ($\alpha = 0.56$)	29.2
Hrachowitz *et al.* (2009b)	Exponential	93
	Gamma	62-203
	Two parallel linear reservoirs	54-254

References

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19** (6): 716–723.

Andreassian V, Perrin C, Michel C. 2004. Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. *Journal of Hydrology* **286**: 19-35.

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

Beck MB. 1987. Water-Quality Modelling - a Review of the Analysis of Uncertainty. *Water Resources Research* **23**: 1393-1442.

Bergstrom S, Carlsson B, Sandberg G. 1985. Integrated modelling of runoff, alkalinity and pH on a daily base. *Nordic Hydrology* **16**: 89-104.

Beven KJ. 2010. Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models Preface. *Hydrological Processes* **24**: 1537-1547.

Beven KJ, Kirkby MJ. 1979. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.*, **24**, 43–69.

Birkel C, Dunn SM, Tetzlaff D, Soulsby C. 2010. Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall-runoff model. *Hydrological Processes* **24**: 2335-2348.

Birkel C, Tetzlaff D, Dunn SM, Soulsby C. 2011a. Using lumped conceptual rainfall-runoff models to simulate daily isotope variability with fractionation in a nested mesoscale catchment. *Advances in Water Resources* **34**: 383-394.

Birkel C, Tetzlaff D, Dunn SM, Soulsby C. 2011b. Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models. *Water Resources Research* **47**: W02515

Birkel C, Soulsby C, Tetzlaff D, Dunn S, Spezia L. 2012. High-frequency storm event isotope sampling reveals time-variant transit time distributions and influence of diurnal cycles *Hydrological Processes*. DOI: 10.1002/hyp.8210.

Blöschl G, Sivapalan M. 1995. Scale Issues in Hydrological Modeling - A Review. *Hydrological Processes* **9**: 251-290.

Botter G, Bertuzzo E, Rinaldo A. 2010. Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox. *Water Resources Research* 46: W03514.

Botter G, Bertuzzo E, Rinaldo A. 2011. Catchment residence and travel time distributions: The master equation. *Geophysical Research Letters* **38**: L11403

Botter G, Peratoner F, Putti M, Zuliani A, Zonta R, Rinaldo A, Marani M. 2008. Observation and modeling of catchment-scale solute transport in the hydrologic response: A tracer study. *Water Resources Research* **44**: W05409.

Brooks JR, Barnard HR, Coulombe R, McDonnell JJ. 2010. Ecohydrologic separation of water between trees and streams in a Mediterranean climate. *Nature Geoscience* **3**: 100-104.

Burnash RJC, Ferral RL, McGuire RA. 1973. A generalized streamflow simulation system: Conceptual modeling for digital computers, technical report, U.S. Natl. Weather Serv., Sacramento, Calif.

Clark M, McMillan H, Collins D, Kavetski D, Woods R. 2011a. Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses. *Hydrological Processes* **25**: 523-543.

Clark MP, Kavetski D. 2010. Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. *Water Resources Research* **46**: W10510.

Clark MP, Kavetski D, Fenicia F. 2011b. Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research* **47**: W09301.

Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV, Wagener T, Hay LE. 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research* **44**: W00B02.

Dawson JJC, Soulsby C, Tetzlaff D, Hrachowitz M, Dunn SM, Malcolm IA. 2008. Influence of hydrology and seasonality on DOC exports from three contrasting upland catchments. *Biogeochemistry* **90**: 93-113.

Dunn SM, Birkel C, Tetzlaff D, Soulsby C. 2010. Transit time distributions of a conceptual model: their characteristics and sensitivities. *Hydrological Processes* **24**: 1719-1729.

Dunn SM, McDonnell JJ, Vache KB. 2007. Factors influencing the residence time of catchment waters: A virtual experiment approach. *Water Resources Research* **43**: W06408.

Fenicia F, McDonnell JJ, Savenije HHG. 2008. Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research*, **44**(6): W06419, doi:06410.01029/02007WR006386.

Fenicia F, Savenije HHG, Matgen P, Pfister L. 2007. A comparison of alternative multiobjective calibration strategies for hydrological modeling. *Water Resources Research*. **43**: W03434, doi:03410.01029/02006WR005098.

Fenicia F, Wrede S, Kavetski D, Pfister L, Hoffmann L, Savenije HHG, McDonnell JJ. 2010. Assessing the impact of mixing assumptions on the estimation of streamwater mean residence time. *Hydrological Processes* **24**: 1730-1741.

Godsey SE, Aas W, Clair TA, de Wit HA, Fernandez IJ, Kahl JS, Malcolm IA, Neal C, Neal M, Nelson SJ, Norton SA, Palucis MC, Skjelkvale BL, Soulsby C, Tetzlaff D, Kirchner JW. 2010. Generality of fractal 1/f scaling in catchment tracer time series, and its implications for catchment travel time distributions. *Hydrological Processes* **24**: 1660-1671.

Gupta HV, Wagener T, Liu YQ. 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes* **22**: 3802-3813.

Harman CJ, Sivapalan M, Kumar P. 2009. Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics. *Water Resources Research* **45**: W09404.

Hrachowitz M, Soulsby C, Tetzlaff D, Dawson JJC, Dunn SM, Malcolm IA. 2009a. Using long-term data sets to understand transit times in contrasting headwater catchments. *Journal of Hydrology* **367**: 237-248.

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

Hrachowitz M, Soulsby C, Tetzlaff D, Malcolm IA. 2011. Sensitivity of mean transit time estimates to model conditioning and data availability. *Hydrological Processes* **25**: 980-990.

Hrachowitz M, Soulsby C, Tetzlaff D, Malcolm IA, Schoups G. 2010. Gamma distribution models for transit time estimation in catchments: Physical interpretation of parameters and implications for time-variant transit time assessment. *Water Resources Research* **46**: W10536.

Iorgulescu I, Beven KJ, Musy A. 2005. Data-based modelling of runoff and chemical tracer concentrations in the Haute-Mentue research catchment (Switzerland). *Hydrological Processes* **19**: 2557-2573.

Kavetski D, Clark MP. 2010. Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. *Water Resources Research* **46**: W10511.

Kavetski D, Fenicia F, Clark MP. 2011. Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment. *Water Resources Research* **47**: W05501.

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.

Kirchner JW, Feng XH, Neal C. 2000. Fractal stream chemistry and its implications for contaminant transport in catchments. *Nature* **403**: 524-527.

Kirchner JW, Tetzlaff D, Soulsby C. 2010. Comparing chloride and water isotopes as hydrological tracers in two Scottish catchments. *Hydrological Processes* **24**: 1631-1645.

Krueger T, Freer J, Quinton JN, Macleod CJA, Bilotta GS, Brazier RE, Butler P, Haygarth PM. 2010. Ensemble evaluation of hydrological model hypotheses. *Water Resources Research* **46**: W07516.

Krueger T, Quinton JN, Freer J, Macleod CJA, Bilotta GS, Brazier RE, Butler P, Haygarth PM. 2009. Uncertainties in data and models to describe event dynamics of agricultural sediment and phosphorus transfer. *Journal Of Environmental Quality* **38**: 1137-1148.

Leavesley GH, Lichty RW, Troutman BM, Saindon LG. 1983. Precipitation-runoff modeling system: User's manual, U.S. Geol. Surv. Water Invest. Rep., 83-4238, 207 pp.

Lindstrom G, Rodhe A. 1992. Transit times of water in soil lysimeters from modeling of oxygen-18, *Water Air Soil Pollut.*, **65**(1-2), 83-100.

Lu JB, Sun G, McNulty SG, Amatya DM. 2005. A comparison of six potential evapotranspiration methods for regional use in the southeastern United States. *Journal of the American Water Resources Association* **41**: 621-633.

McDonnell JJ, McGuire K, Aggarwal P, Beven KJ, Biondi D, Destouni G, Dunn S, James A, Kirchner J, Kraft P, Lyon S, Maloszewski P, Newman B, Pfister L, Rinaldo A, Rodhe A, Sayama T, Seibert J, Solomon K, Soulsby C, Stewart M, Tetzlaff D, Tobin C, Troch P, Weiler M, Western A, Worman A, Wrede S. 2010. How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. *Hydrological Processes* **24**: 1745-1754.

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

McGuire KJ, McDonnell JJ. 2010. Hydrological connectivity of hillslopes and streams: Characteristic time scales and nonlinearities. *Water Resources Research*. **46**(10), W10543, doi:10510.11029/12010WR009341.

McGuire KJ, Weiler M, McDonnell JJ. 2007. Integrating tracer experiments with modeling to assess runoff processes and water transit times. *Advances in Water Resources* **30**: 824-837.

McMillan H, Clark M. 2009. Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme. *Water Resources Research* **45**: W04418.

McMillan H, Clark M, Bowden WB, Duncan MJ, Woods R. 2011. Hydrological field data from a modeller's perspective. Part 1: Diagnostic tests for model structure. *Hydrological Processes* **25**: 511-522.

McMillan H, Freer J, Pappenberger F, Krueger T, Clark M. 2010. Impacts of uncertain flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes* **24**: 1270-1284

Miller JD, Anderson HA, Ferrier RC, Walker TAB. 1990. Hydrochemical fluxes and their effects on stream acidity in two forested catchments in central Scotland. *Forestry* **63**: 311-331.

Nyberg L, Rodhe A, Bishop K. 1999. Water transit times and flow path from two line injections of 3H and 36Cl in a microcatchment at Gardsjön, Sweden. *Hydrological Processes.* **13**, 1557-1575.

Oda T, Asano Y, Suzuki M. 2009. Transit time evaluation using a chloride concentration input step shift after forest cutting in a Japanese headwater catchment. *Hydrological Processes*. **23**(19), 2705-2713.

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**(3), 292-307.

Pelletier MP. 1988. Uncertainties in the determination of river discharge: a literature review. *Canadian Journal of Civil Engineering* **15**: 834-850.

Phillips FM. 2010. Soil-water bypass. *Nature Geoscience* **3**: 77-78.

Reichert P, Mieleitner J. 2009. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. *Water Resources Research* **45**: W10402.

Rinaldo A, Beven KJ, Bertuzzo E, Nicotina L, Davies J, Fiori A, Russo D, Botter G. 2011. Catchment travel time distributions and water flow in soils. *Water Resources Research* **47**: W07537.

Rinaldo A, Botter G, Bertuzzo E, Uccelli A, Settin T, Marani M. 2006. Transport at basin scales: 1. Theoretical framework. *Hydrology and Earth System Sciences* **10**: 19-29.

Rodgers P, Soulsby C, Waldron S. 2005a. Stable isotope tracers as diagnostic tools in upscaling flow path understanding and residence time estimates in a mountainous mesoscale catchment. *Hydrological Processes* **19**: 2291-2307.

Rodgers P, Soulsby C, Waldron S, Tetzlaff D. 2005b. Using stable isotope tracers to assess hydrological flow paths, residence times and landscape influences in a nested mesoscale catchment. *Hydrology and Earth System Sciences* **9**: 139-155.

Savenije HHG. 2009. HESS Opinions "The art of hydrology". *Hydrology and Earth System Sciences* **13**: 157-161.

Schaefli B, Gupta HV. 2007. Do Nash values have value? *Hydrological Processes* **21**: 2075-2080.

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* **3** (11), 1241, doi:10.1029/2001WR000978.

Singh VP. 1995. *Computer Models of Watershed Hydrology*. Water Resources Publications: Highlands Ranch, Colorado; 1130.

Sivapalan M. 2009. The secret to 'doing better hydrological science': change the question! *Hydrological Processes* **23**: 1391-1396.

Sivapalan M, Grayson R, Woods R. 2004. Scale and scaling in hydrology. *Hydrological Processes* **18**: 1369-1371.

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:01410.01029/02006WR005032.

Soulsby C, Reynolds B. 1993. Influence of soil hydrological pathways on stream aluminium chemistry at Llyn Brianne, Mid Wales. *Environmental Pollution* **81**: 51-60.

Soulsby C, Rodgers P, Smart R, Dawson J, Dunn S. 2003. A tracer-based assessment of hydrological pathways at different spatial scales in a mesoscale Scottish catchment. *Hydrological Processes* **17**: 759-777.

Soulsby C, Tetzlaff D, Rodgers P, Dunn S, Waldron S. 2006. Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment: An initial evaluation. *Journal of Hydrology* **325**: 197-221.

Speed M, Tetzlaff D, Soulsby C, Hrachowitz M, Waldron S. 2010. Isotopic and geochemical tracers reveal similarities in transit times in contrasting mesoscale catchments. *Hydrological Processes* **24**: 1211-1224.

Stewart MK, Morgenstern U, McDonnell JJ. 2010. Truncation of stream residence time: how the use of stable isotopes has skewed our concept of streamwater age and origin. *Hydrological Processes* **24**: 1646-1659.

Tetzlaff D, Brewer MJ, Malcolm IA, Soulsby C. 2010. Storm flow and baseflow response to reduced acid deposition-using Bayesian compositional analysis in hydrograph separation with changing end members. *Hydrological Processes* **24**: 2300-2312.

Tetzlaff D, Malcolm IA, Soulsby C. 2007a. Influence of forestry, environmental change and climatic variability on the hydrology, hydrochemistry and residence times of upland catchments. *Journal of Hydrology* **346**: 93-111.

Tetzlaff D, Seibert J, Soulsby C. 2009. Inter-catchment comparison to assess the influence of topography and soils on catchment transit times in a geomorphic province; the Cairngorm mountains, Scotland. *Hydrological Processes* **23**: 1874-1886.

Tetzlaff D, Waldron S, Brewer MJ, Soulsby C. 2007b. Assessing nested hydrological and hydrochemical behaviour of a mesoscale catchment using continuous tracer data. *Journal of Hydrology* **336**: 430-443.

Turner JV, Macpherson DK, Stokes RA. 1987. The Mechanisms of Catchment Flow Processes Using Natural Variations in Deuterium and O-18. *Journal of Hydrology*, **94**(1-2), 143-162.

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

Vache KB, McDonnell JJ. 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. *Water Resources Research* **42**: W02409.

Weiler M, McGlynn BL, McGuire KJ, McDonnell JJ. 2003. How does rainfall become runoff? A combined tracer and runoff transfer function approach. *Water Resources Research*, **39**(11), 1315, doi:1310.1029/2003WR002331.

Wood EF, Lettenmaier DP, Zartarian VG. 1992. A land-surface hydrology parameterization with subgrid variability for general-circulation models. *J. Geophys. Res.*, **97**(D3), 2717–2728.

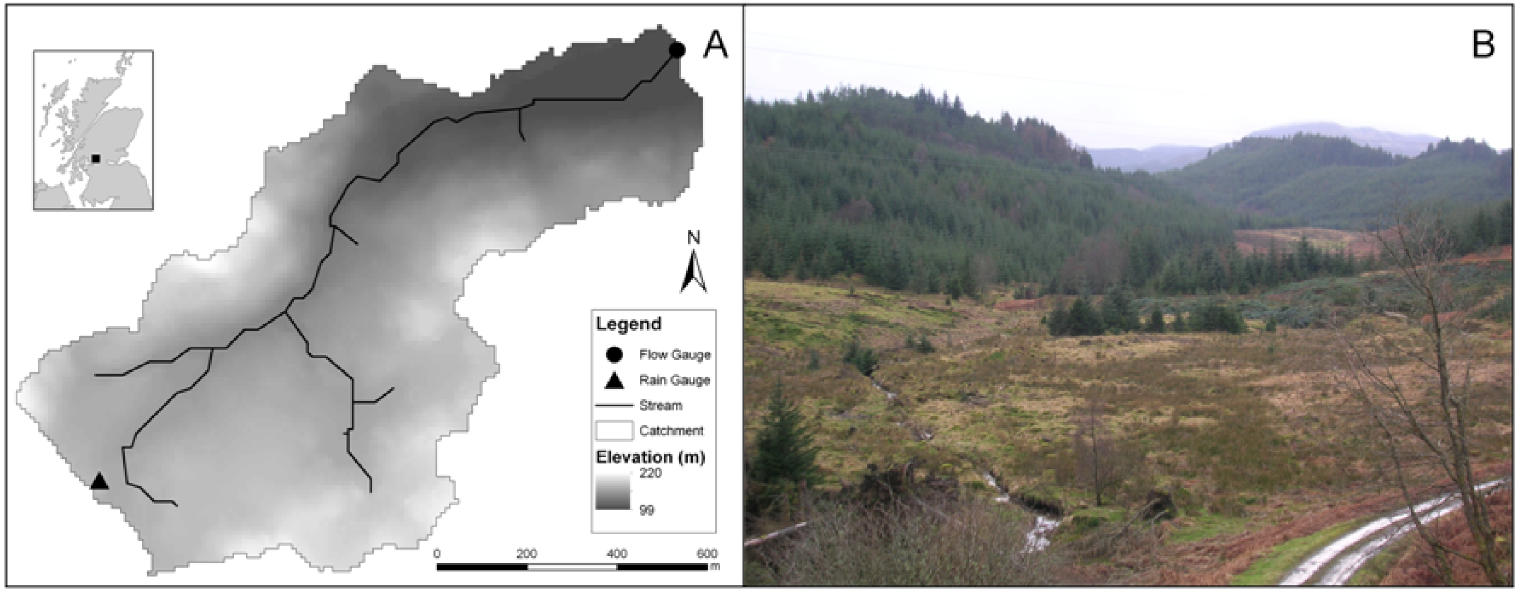

Figure 1. (A) Loch Ard B10 catchment map and instrument locations (B) Photograph of Loch Ard B10 catchment

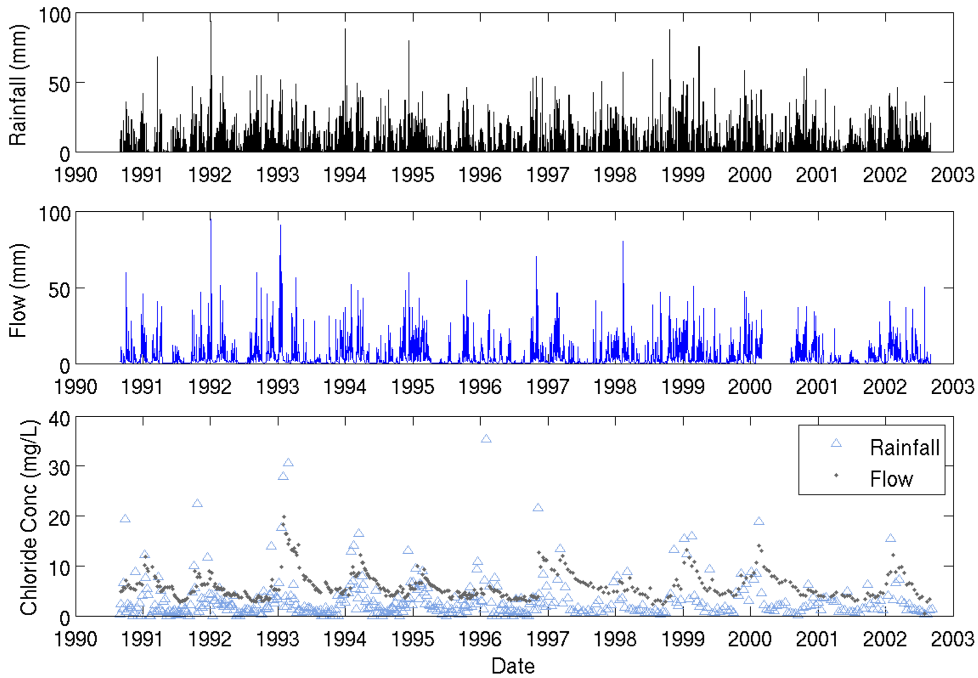

Figure 2. Hydrometric and hydrochemical data available for the Loch Ard B10 catchment

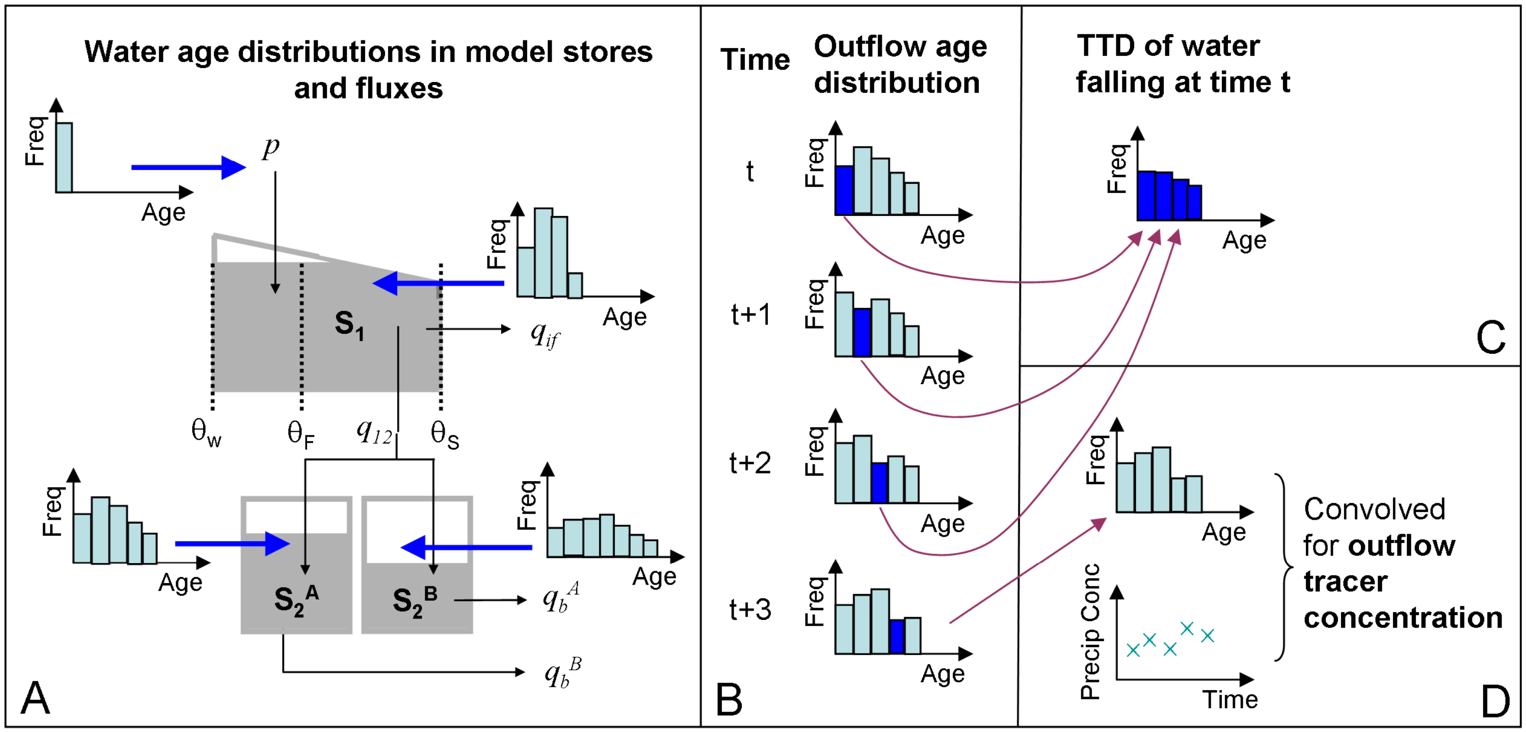

Figure 3. Conceptual diagram showing process used to calculated model TTDs and outflow tracer concentrations in a sample FUSE model. (A) Water age distribution of each reservoir (S_1 upper zone, $S_2^{A,B}$ lower zone) stored as histogram. Fluxes (p precipitation, q_{if} interflow, q_{12} drainage, $q_b^{A,B}$ baseflow) have the age signature of their source reservoir. (B) Outflow age distribution for time t is the sum of distributions from component fluxes (q_{if}, $q_b^{A,B}$). (C) TTD of input water is calculated from the corresponding outflow times. (D) Outflow tracer concentration calculated by convolution of outflow age distribution with precipitation tracer concentrations

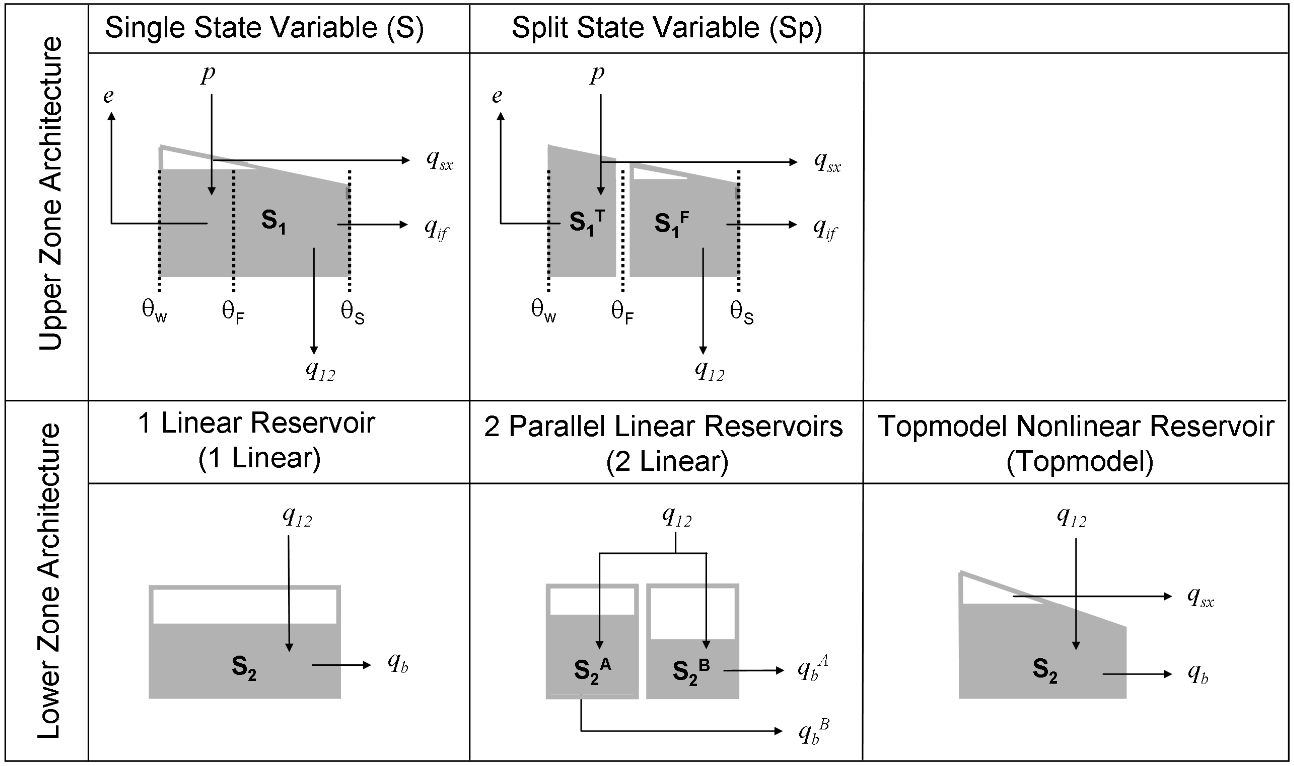

Figure 4. Simplified wiring diagram showing model architecture options used in this study. Upper Zone: [S] A single state variable S_1 combining tension and free storage [Sp] Separate state variables for tension S_1^T and free S_1^F storage. Lower Zone: [1 Linear] S_2 A single linear reservoir [2 Linear] S_2^A, S_2^B Two parallel linear reservoirs [Topmodel] S_2 A single nonlinear reservoir based on Topmodel concepts (where surface runoff q_{sx} is controlled by the lower zone). Key to soil moisture values: θ_W wilting point (here 0), θ_F field capacity $(\phi \cdot S_{1,\max})$, θ_S saturation point $(S_{1,\max})$

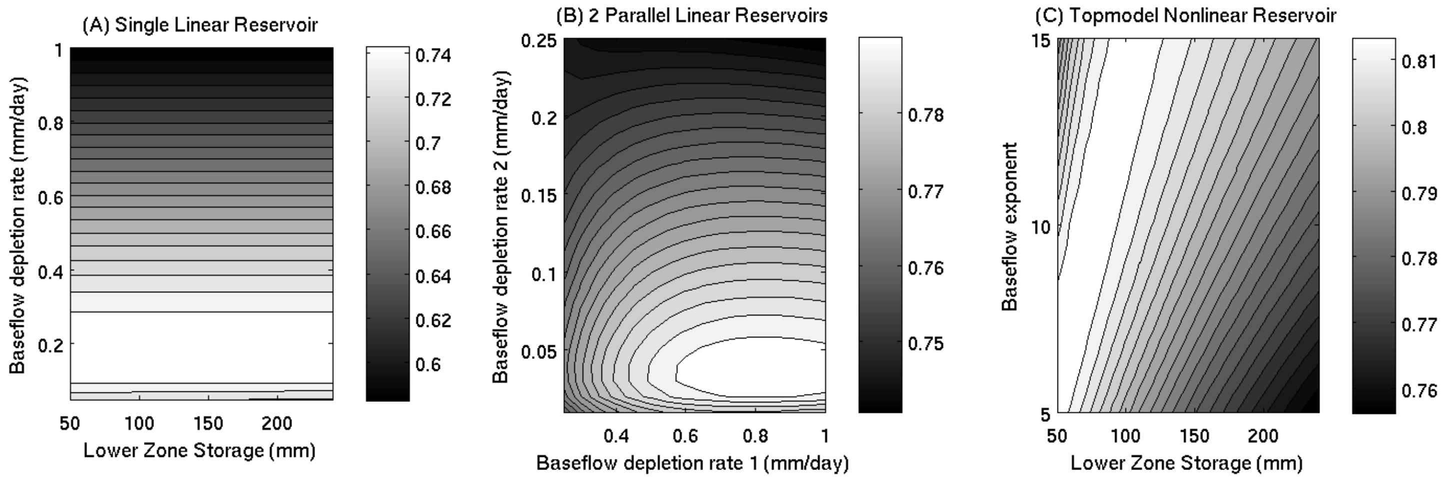

Figure 5. Calibration results for lower zone storage parameters for 3 lower zone model architectures. The objective function is the sum of squared errors between modelled and measured discharge series, after Box-Cox transformation to normalise error variance. The calibration period was over two hydrological years (1998-1999).

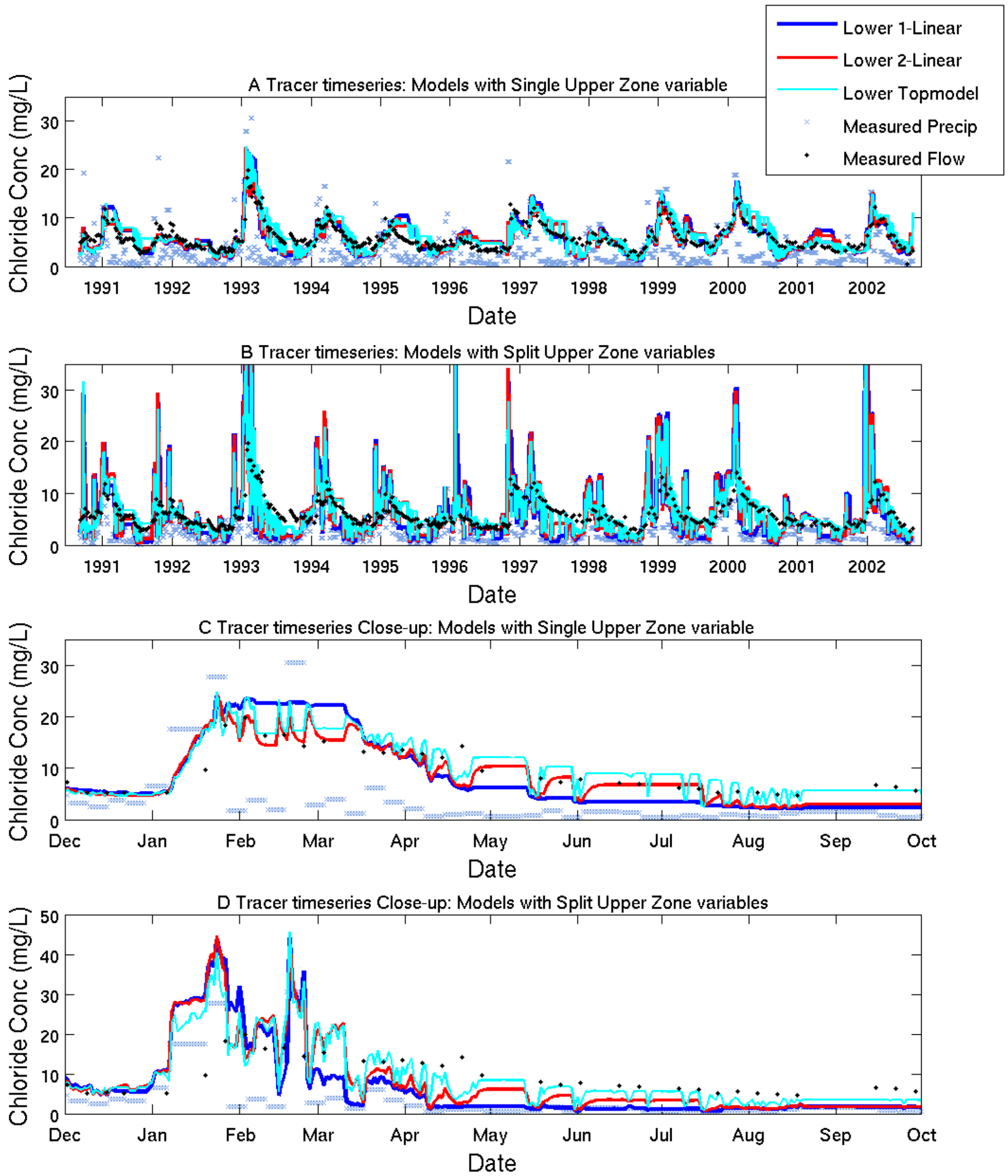

Figure 6. Time series of measured Chloride input and output concentrations and comparisons with model predictions. (A) Models with a single upper zone storage variable (B) Models with split upper zone storage variables (C) Close-up of A for largest event (Dec 1992 – Oct 1993). (D) Close-up of B for largest event

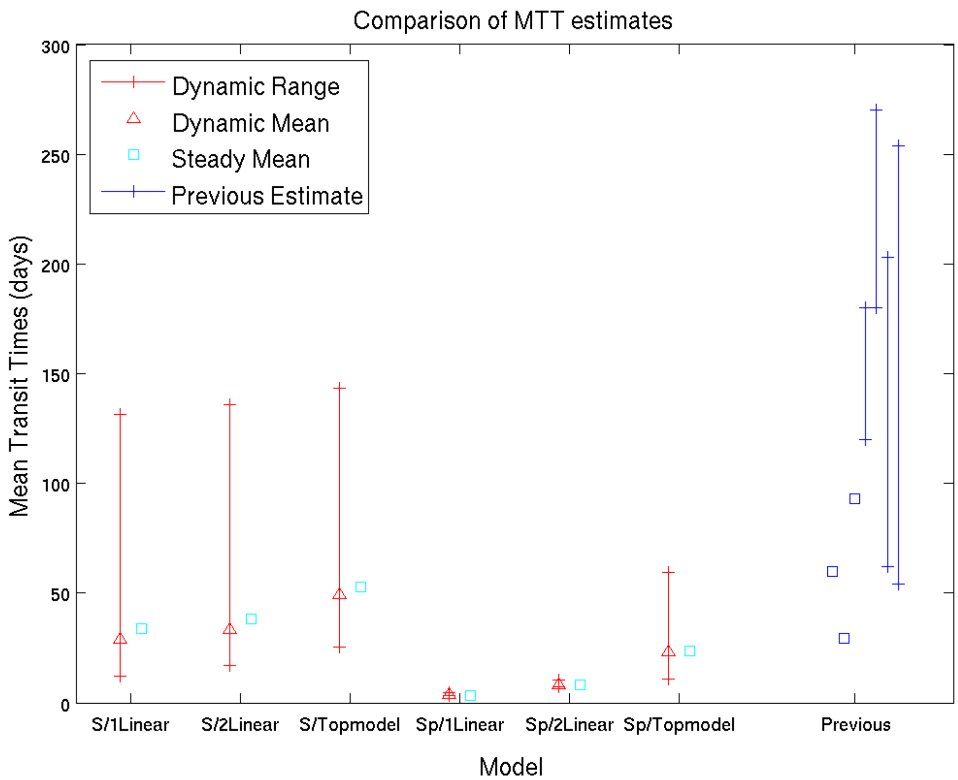

Figure 7. Comparisons of MTT estimates between models (run in dynamic and steady state mode) and from previous studies (Table 4) of the B10 catchment.

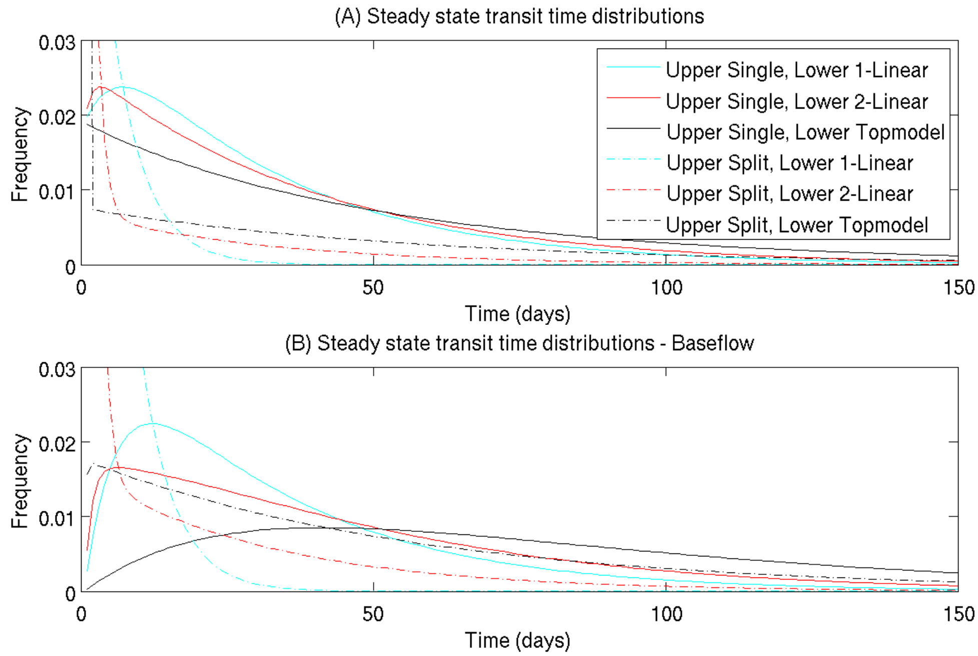

Figure 8. Steady state transit time distributions for a range of model structures. (A) Combined flow: Subsurface stormflow + groundwater flow (B) Groundwater flow only

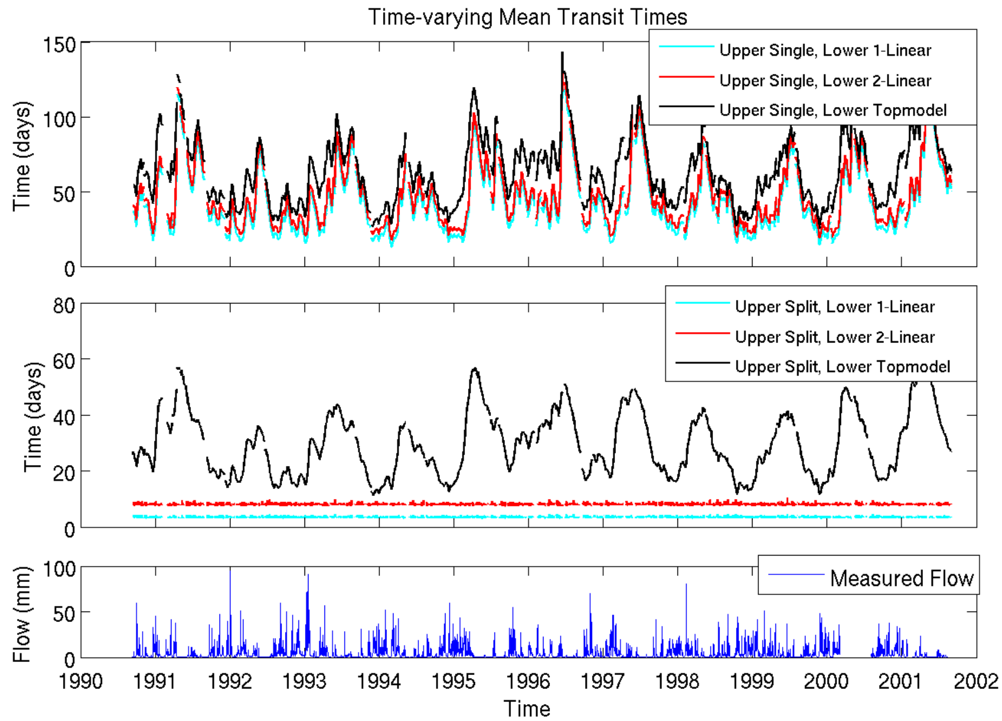

Figure 9. Variation of Mean Transit Time with time for a range of model structures (A) Models with single upper state variable, (B) Models with split upper state variables. (C) Measured Flow is plotted for comparison

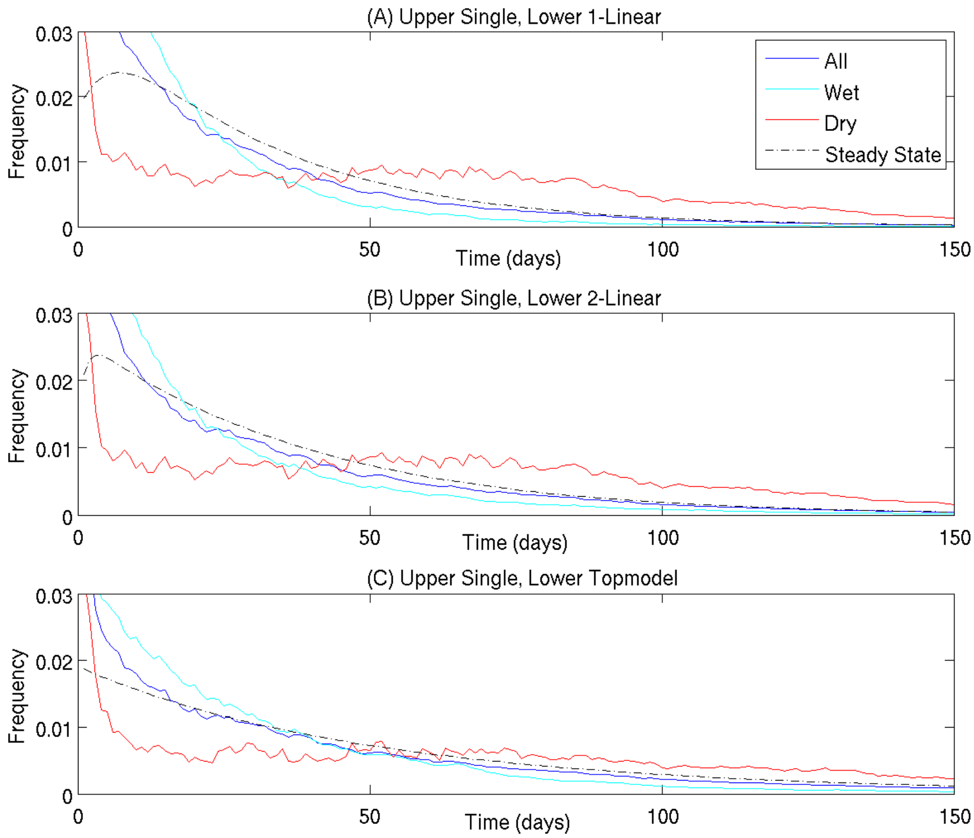

Figure 10. Variation in transit time distribution according to catchment wetness condition for 3 model structures. TTDs are given for (All): All days in record, (Wet): Days in lower quartile of MTT distribution, (Dry): Days in upper quartile of MTT distribution, (Steady State): Steady state TTD for comparison.

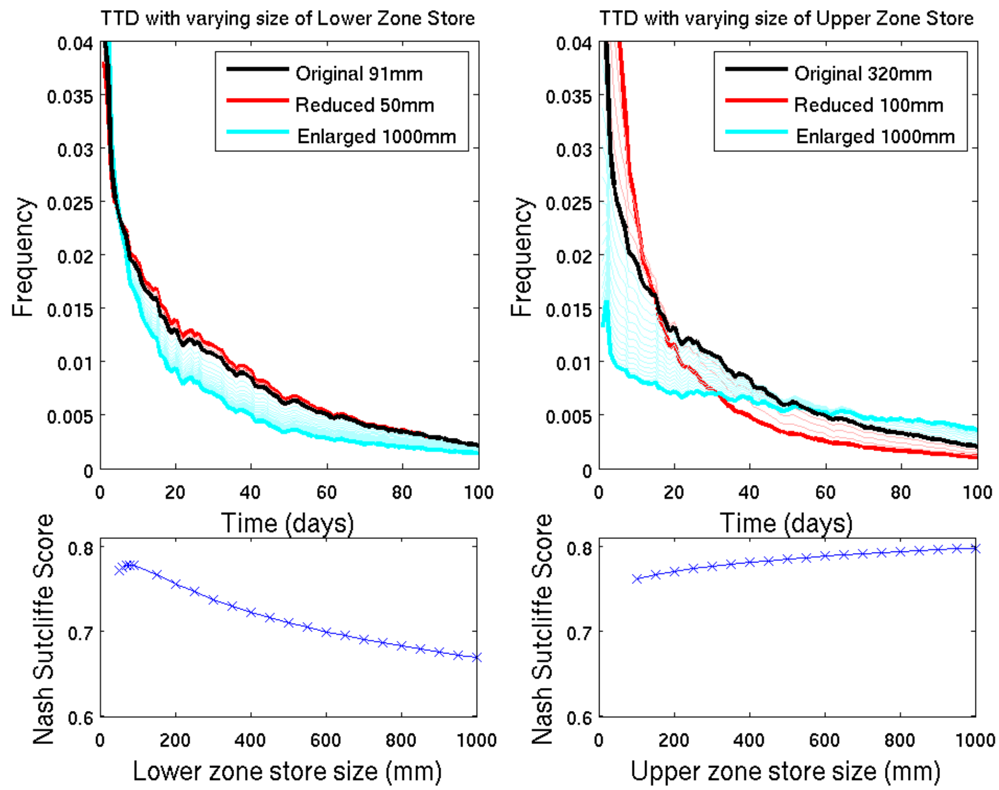

Figure 11. The effects of changing upper and lower zone storage depths on Transit Time Distribution (upper panels) and model performance (lower panels). TTDs are shown for equal increments/decrements of store size (thin lines) up to the maximum/minimum values given (thick lines).

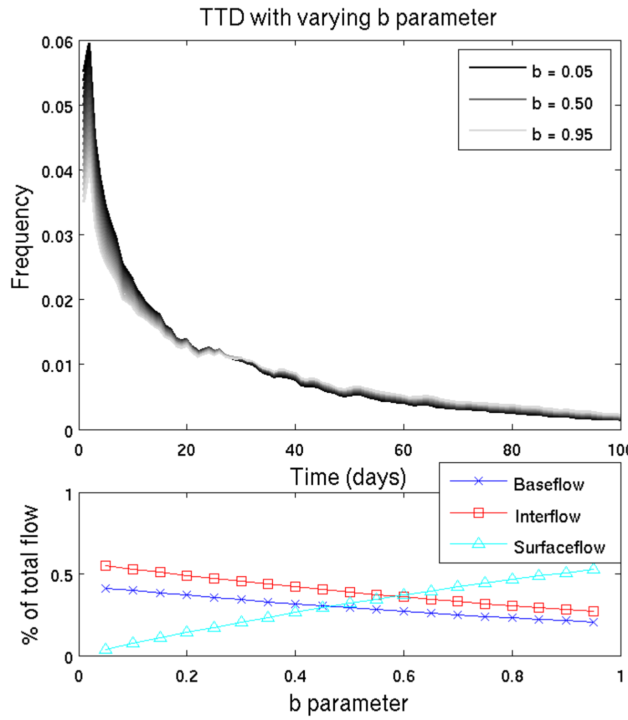

Figure 12. The sensitivity of the model to soil water mixing is shown by varying the surface flow b parameter. Effects are shown on Transit Time Distribution (upper panel) and Percentage share of flow volume between pathways (lower panel).