Conceptual modelling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls non-stationary water age estimates

Journal:	*Hydrological Processes*
Manuscript ID:	HYP-14-0391.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	22-Nov-2014

Complete List of Authors:	Birkel, Christian; University Aberdeen, Northern Rivers Institute Soulsby, Chris; University Aberdeen, Northern Rivers Institute Tetzlaff, Doerthe; University Aberdeen, Northern Rivers Institute

Keywords:	Catchment storage, water age, rainfall-runoff model, tracers, connectivity, stable isotopes

Conceptual modelling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls non-stationary water age estimates

Christian Birkel[1], Chris Soulsby[1], Doerthe Tetzlaff[1]

[1]Northern Rivers Institute, School of Geosciences, University of Aberdeen, Aberdeen, UK.

Correspondence to: c.birkel@abdn.ac.uk

Abstract

Although catchment storage is an intrinsic control on the rainfall-runoff response of streams, direct measurement remains a major challenge. Coupled models that integrate long-term hydrometric and isotope tracer data are useful tools that can provide insights into the dynamics of catchment storage and the volumes of water involved. In this study, we use a tracer-aided hydrological model to characterize catchment storage as a dynamic control on system function related to streamflow generation, which also allows direct estimation of the non-stationarity of water ages. We show that in a wet Scottish upland catchment dominated by runoff generation from riparian peats (histosols) with high water storage, non-stationarity in water age distributions are only clearly detectable during more extreme wet and dry periods. This is explained by the frequency and longevity of hydrological connectivity and the associated relative importance of flow paths contributing younger or older waters to the stream. Generally, these saturated riparian soils represent large mixing zones that buffer the time variance of water age and integrate catchment-scale partial mixing processes. Although storage simulations depend on model performance, which is

influenced by input variability and the degree of isotopic damping in the stream, a longer-term storage analysis of this model indicates a system which is only sensitive to more extreme hydroclimatic variability.

Keywords: Catchment storage, water age, rainfall-runoff model, tracers, stable isotopes, connectivity.

I Introduction

The relationship between catchment water storage and discharge is fundamental to understanding the hydrological response of streams (Kirchner, 2009). Recent years have seen many studies attempting to characterize catchment storage dynamics. These include for example direct hydrometric analysis (McNamara *et al*., 2011; Brauer *et al*., 2013), use of conservative tracers (Soulsby *et al*., 2009), application of hydrological models (Birkel *et al*., 2011a), new field instrumentation (e.g. COSMOS – Zreda *et al*., 2012) and remote sensing (e.g. GRACE - Rodell *et al*., 2009). These alternative approaches have yielded different insights often relevant at contrasting spatial scales. However, given the difficulties in fully characterizing subsurface hydrology (e.g. soil moisture and groundwater variability) at the catchment scale, in most cases, only a partial view of storage dynamics are given by the different methods. For example, hydrometric analysis provides insight into the near-surface storage dynamics that reflect water balance processes (Kirchner, 2009). However, models that concurrently track conservative tracers between rainfall and runoff usually infer much larger volumes of storage, needed to damp and lag tracer input variability in outputs (Birkel *et al*, 2011a). These large storage volumes were first proposed in the form of additional parameters in conceptual models reflecting storage volumes that participate with solute mixing but do not affect streamflow dynamics (Barnes and Bonell, 1996). Even detailed soil moisture and groundwater monitoring at the hillslope scale can miss important dynamics in the weathering zone (Salve *et al.*, 2012). Only in extensively studied small experimental watersheds - usually with a well-characterized, shallow subsurface – can constrained, quantitative estimates of total catchment storage be achieved (e.g. Bishop *et al.*, 2011; Peters *et al.*, 2013).

Investigating both short-term (e.g. hourly and daily) and longer-term (e.g. intra- and inter-annual) storage change is crucial to understanding non-stationarity in catchment hydrological function and stream flow behaviour (e.g. Brooks *et al.*, 2009; Rinaldo *et al.*, 2011). Conservative tracers can play an important role in this regard and integrating tracers in rainfall-runoff models shows promise as learning tools in hypothesis testing (Soulsby *et al.*, 2008). Such integration has helped to elucidate the inter-relationships between the short-term storage dynamics that reflect the catchment water balance and stream flow response with the mixing processes that regulate solute transport and control water transit times (e.g. Fenicia *et al.*, 2010; Heidbüchel *et al.*, 2012). However, the potential of such models to explore non-stationarity in mixing processes and resulting time-variant water age distributions has only recently started to be explored (McMillan *et al.*, 2012; Hrachowitz *et al.*, 2013). Theoretical work by Rinaldo *et al.* (2011) and Botter *et al.* (2011) emphasized the differences between the age of water in transit and in storage, which under non-steady conditions can not be assumed identical. In this context, integrated flow-tracer models have particular potential to better understand catchment functioning (McDonnell and Beven, 2014) in catchments where long-term hydrometric data is complemented by time series of tracer data from the major runoff source areas where mixing occurs (e.g. soil water, groundwater etc.). This allows direct testing of hypotheses about mixing and water age in different storage components (Birkel *et al.*, 2014).

In this study, we build on a previous model of the Girnock catchment (Birkel *et al.*, 2010) by examining long term hydrometric data (> 40 years), tracer data (> 10 years), and ancillary process-based data on soil hydrology and groundwater dynamics (Tetzlaff *et al.*, 2014). The site is also more broadly representative of other northern upland catchments facilitating transferability. Tracer-aided models have been developed here which allow catchment storage in different landscape units to be estimated (Birkel *et al.*, 2011b). However, the literature is marked by inconsistency in the terminology used to describe different elements of catchment storage; with terms like active, dynamic, hydraulic, passive, residual, immobile and dead storage, used amongst other. For clarity in our model, we define catchment storage as the sum of dynamic storage and additional storage available for mixing that can be detected using hydrometric and conservative tracer data. This acknowledges that the total catchment storage is difficult – usually impossible - to determine. In this paper we define storage that actively or hydraulically contributes to discharge as *dynamic storage*. The storage that does not necessarily contribute to discharge but is available for mixing tracer inputs is subsequently termed the *additional storage mixing volume*.

We use a unique six year data set of weekly and event-targeted daily oxygen-18 isotopes of precipitation and streamflow from the Girnock catchment. These years from 2003 to 2009 incorporate periods of marked hydrological contrasts including the latter part of the Europe-wide drought in 2003 and a very wet year in 2006/07. The tracer data was integrated into a low-parameter conceptual model to simulate stream flow on the basis of dynamic storage changes, as well as exploring how hydroclimatic variability results in differences in the proportion of additional storage volumes mobilized for mixing. The variability of mobilized additional storage volumes is hypothesized as the main driver of non-stationary water ages. Our specific objectives to assess the latter are firstly to explore long-term internal storage dynamics using a multi-objective, moving window calibration of a parsimonious, rainfall-runoff model coupled with an isotope transport model to simulate daily discharge and oxygen-18 dynamics. Secondly, we investigate how the different calibration windows of dynamic storage and additional storage mixing volumes relate to hydroclimatic variability, catchment connectivity and flow paths, model performance, the information content of tracer data and resulting stream water age estimates.

II Study catchment

The Girnock Burn is located in the Cairngorm Mountains of Scotland and drains about 30 km^2 (57° 00′ 59′′ North and 3° 07′ 60′′ West, Figure 1). Its catchment characteristics are detailed elsewhere (Soulsby *et al.*, 2007; Tetzlaff *et al.*, 2007). Briefly, average annual precipitation over the study period from 2000 to 2009 is around 950 mm, with annual runoff of 530 mm and potential evapotranspiration of 408 mm. These figures are close to the long-term averages since 1970. The altitudinal range is 631 m with a maximum at 861 m. Granitic rocks can mostly be found at higher-elevation areas while schists predominate at lower elevations (Soulsby *et al.*, 2007). These rocks have generally poor aquifer characteristics, and due to the glacial legacy much of the catchment is covered by low-permeability drift deposits (in parts > 30 m deep). In the valley bottoms histosols (peats and peaty gleys) fringe much of the stream channel network. These riparian soils remain close to saturation all year round and generate saturation overland flow as storm runoff (Ali *et al*., 2014). They receive input from upslope freely draining podzolic soils (which also facilitate deeper groundwater recharge) in the form of quasi-continuous groundwater seepage and threshold-like non-linear near-surface connections from lateral flow during larger storm events (Tetzlaff *et al*., 2014). These connectivity dynamics cause the saturation zone to expand and contract (its observed spatial extent varies between 2 – 40% of the catchment area) and drive the catchment hydrological response (Birkel *et al*., 2010). Snow usually melts quickly (within a matter of days over most of the catchment) and generally comprises of < 5 % of the annual water balance (Capell *et al*., 2013).

III Data and Methods

3.1 Hydrometric and isotopic data

Daily discharge (*Q*) data was derived from recorded water levels at Littlemill (Scottish Environment Protection Agency, SEPA). Precipitation (*P*) was interpolated using a squared elevation inverse distance weighted algorithm similar to Birkel *et al*. (2011a) based on 12 SEPA rain gauges located around the Girnock catchment. Potential evapotranspiration (*PET*) was estimated using Penman-Monteith from an automatic weather station within the Girnock (Marine Scotland Science) for the dominant heather moorland vegetation and then adjusted prior to modelling using an annual scaling coefficient *kET* to match the water balance. This resulted in modifications between 0.94 < kET < 1.09.

Water samples (accumulated rain and instantaneous stream spot samples) were analyzed for oxygen-18 isotopes in ‰. Samples were collected at approximately weekly intervals but where possible, attempts were also made to capture the peaks of storm events between October 2003 and September 2009. Isotope analysis was conducted using gas source isotope mass spectroscopy (precision: ±0.1 ‰) at the Scottish Universities Environmental Research Centre and the James Hutton Institute.

3.2 Model approach

Previous work has reported a suite of rainfall-runoff models developed for the Girnock using tracers to help constrain the model structure. These models simulated stream flow based on the dominant geographic sources of runoff (Tetzlaff *et al*., 2008; Birkel *et al*., 2010) and their temporal dynamics (Birkel *et al*., 2011a, b; Birkel *et al*., 2014). The non-linearities of the runoff response were successfully incorporated in a tracer-conditioned dynamic saturation area model (full details given in Birkel *et al*., 2010). This latter model is based on a dynamic conceptualization of hydrological connectivity associated with quick near-surface runoff generation mechanisms in saturation areas, together with a linked groundwater store (to see Figure 2 for the model structure, connections between stores, and basic equations). The connectivity of the saturation area to the stream was estimated as a function of antecedent wetness and soil moisture capacity. A 7-day antecedent wetness algorithm was calibrated against seven actual saturation area maps (derived from field observation) to best represent the dynamic expansion and contraction of the saturation area (details in Birkel *et al*., 2010; also see Ali *et al*., 2014). This allowed a daily time series of % catchment saturation area ($dSAT$) to be used directly as model input to volumetrically differentiate the hillslope reservoir from the saturation area store. Subsequently, catchment precipitation P and potential evapotranspiration ET are distributed into hillslope (P_{up}, ET_{up}) and saturation area (P_{sat}, ET_{sat}) according to the extent of $dSAT$ (only the equations for precipitation are given for illustration):

$$P_{up} = PdSAT \qquad (1)$$

$$P_{sat} = 1 - P_{up} \qquad (2)$$

Here, the original model of Birkel *et al*. (2010) was modified to conceptualize unsaturated storage units (Figure 2); in other words reservoirs do not have a lower limit and levels below a pre-defined baseline prevents generation of lateral and vertical fluxes except evapotranspiration. The unsaturated storage was conceptualized to reflect soil moisture-related threshold processes of runoff generation (Tetzlaff *et al*., 2014). The storages S are model state variables and we describe the following fluxes and calibrated parameters. The unsaturated hillslope reservoir S_{up} is drained (flux Q_1 in mm d^{-1}) by the linear rate parameter a (d^{-1}) and directly contributes to the saturation area store S_{sat}; the recharge rate R (mm d^{-1}) to the groundwater reservoir S_{low} is linearly calculated using the parameter Re (d^{-1}); the S_{low} generates runoff Q_{low} (mm d^{-1}) contributing to total streamflow Q (mm d^{-1}) using the linear rate parameter b (d^{-1}). The runoff component Q_{sat} (in mm d^{-1}) generated nonlinearly from S_{sat} conceptualizes saturation overland flow using the rate parameter k (d^{-1}) and the nonlinearity parameter α (-) in a power function-type equation (Figure 2); and Q is simply the sum of Q_{sat} and Q_{low}. Each of the three reservoirs (S_{up}, S_{sat} and S_{low} in mm) incorporates an additional calibrated storage parameter (upS_p, $satS_p$ and $lowS_p$, respectively in mm) for isotope transport simulations. This was achieved by creating an additional volume available for isotope storage, mixing and transport that does not affect the dynamic water storage and fluxes:

$$\frac{d(c\ S\ )}{dt} = \sum_j c_{I,j} I_j - \sum_k c_{O,k} O_k \qquad (3)$$

With c being the $\delta^{18}O$ signature of storage components (‰) in j storage inflows I_j (e.g. P, Q_{up}, R) and k outflow O_k components (e.g. ET, Q_{low}, Q_{sat}), which characterizes the catchment storage S dynamics (sum of dynamic and additional storage available for mixing) and associated isotope signature c. Dynamic storage (S_d) was calculated as the sum of the three internal model storage units (S_{up}, S_{sat} and S_{low}). However, the additional storage parameters were subsequently converted into time-variable mixing volumes (MV). This was done with the premise of using the non-linear rainfall-runoff model as a means of relaxing the assumption of complete mixing at the catchment scale without introducing new parameters. Thus, as the rainfall-runoff model dynamically varies the extent of the S_{sat} and S_{up}, the complete mixing in individual reservoirs is integrated at the catchment scale in a non-linear manner. The MVs therefore represent a partial mixing mechanism and were calculated according to the catchment wetness (*dSAT*) state assuming that greater wetness results in greater saturation area extent but also greater potential for mixing (MV_{sat}). In contrast, it is assumed that the hillslope mixing volume (MV_{up}) decreased as the saturation area expands. This is consistent with the hydrological model and the key goal of a minimal number of model parameters:

$$MV_{sat} = satS_p\, dSAT \qquad (4)$$

$$MV_{up} = upS_p\, (1 - dSAT) \qquad (5)$$

Since S_d actively contributes to discharge (Figure 2), we excluded situations when storages below the baseline did not generate outflow. Catchment storage (*S*) estimates (identifiable from hydrometric and isotopic data) use the three mixing volumes (S_{MV}) in addition to the S_d. The simple *S* to *Q* ratio (total streamflow) is used as the mean annual transit time estimate (*TT*) in an attempt to assess inter-annual dynamics and possible non-stationary behaviour (Zuber, 1986; Soulsby *et al*., 2009):

$$TT = \frac{S_d + S_{MV}}{Q} \qquad (6)$$

Recognizing that exponential models resulting from a linear, well-mixed reservoir (Maloszewski and Zuber, 1982) are usually inadequate in fully describing tracer dynamics

(Kirchner *et al.*, 2000), this model effectively integrates three reservoirs. One reservoir is non-linear (S_{Sat}), two reservoirs incorporate partial mixing mechanisms (S_{Sat} and S_{up}) and transport to the stream is calculated by two parallel reservoirs (S_{Sat} and S_{low}). Two linear reservoirs have been shown to give comparable results to the more parsimonious gamma distribution model (Shaw *et al.*, 2008) and two parallel reservoirs are conceptually closer to an advective-dispersion mechanism (Kirchner, 2000). Here we describe a more complex model that might result in a non-smooth transit time distribution as previously reported by Dunn *et al.* (2010) and Van der Velde *et al.* (2011). We therefore track the age of waters in flux using a time stamp tagging each daily incoming and outflowing flux; in addition to the mean annual TT this also allows us to assess the intra-annual variability similar to Hrachowitz *et al.* (2013):

$$p_{F,Q}(t_j - t_i,\ t_j) = \sum_{n=1}^{N} p_{F,Q_n}(t_j - t_i,\ t_j)\frac{Q_n(t_j)}{Q(t_j)} \tag{7}$$

Where $p_{F,Q}$ is the distribution of water age of all contributing fluxes Q_n to total discharge Q with t_j being the time of exit at the catchment outlet and t_i the time of entry with P.

The model was calibrated (parameters a, b, Re, k, α upS_p, $satS_p$ and $lowS_p$) using a multi-objective (flow and tracer) non-dominated sorting genetic algorithm (NSGA2) for optimization applied to each water year starting 01/10/2003 until 30/09/2009 (Deb *et al.*, 2002). Water years were chosen for calibration in an attempt to generate time-variable parameter sets as basis to assess non-stationary catchment behaviour (Seibert and McDonnell, 2010; Westra *et al*., 2014). Preliminary tests varying the length of time series for calibration showed that periods >1 year lost important information on extreme values whilst <1 year were insufficient to pick up seasonality effects. The calibration procedure was based on recommendations by Deb *et al*. (2002) and tested to minimize computation time guaranteeing an exhaustive search across the parameter space. The optimization included 500 parameter sets (for each of the parameters shown in Figure 2) defined as an initial population. This population was tested and repeatedly constrained over 100 iterations each defined as a generation. This resulted in a total of 50000 tested different parameter combinations. Only the final and best parameter population (500 parameter sets) was retained and used for further analysis which included calculation of simulation bounds representing posterior parameter variability. The retained 500 parameter sets were applied to simulate model output and it was assumed that the results provide a range of accepted models equivalent to a formal uncertainty analysis as shown by Andrews *et al*. (2012). Calibration objectives were formulated using the modified Kling-Gupta efficiency (*KGE*) criterion (Kling *et al*., 2012) simultaneously applied to discharge (*KGE_Q*) and oxygen-18 ($KGE\_\delta^{18}O$) time series. The *KGE* is a three-dimensional representation (Euclidean distance) of the widely-used Nash-Sutcliffe criterion overcoming some weaknesses of the latter (Schaefli and Gupta, 2007) balancing dynamics (correlation coefficient r), bias (bias ratio $ß$) and variability (variability ratio γ). The following equations mainly use the mean (μ) and standard deviation (σ) of observed (subscript *obs*) against simulated (subscript *sim)* values:

$$KGE = 1 - \sqrt{(1-r)^2 + (1-\gamma)^2 + (1-\beta)^2} \qquad (8)$$

$$\gamma = \frac{\left(\frac{\sigma_{sim}}{\mu_{sim}}\right)}{\left(\frac{\sigma_{obs}}{\mu_{obs}}\right)} \quad (9)$$

$$\beta = \frac{\mu_{sim}}{\mu_{obs}} \quad (10)$$

The *KGE* ranges from - Infinity to a perfect fit of 1. Modelling was started for each iteration 01/01/2000, but calibration was initiated with available isotope data from 01/10/2003 and then continued separately for the following water years. The preceding 2.5 years plus subsequent water years were used as a warming-up period to fill storages and initiate isotope signatures in storages. Observed soil and ground water isotope values were used to initiate the isotope signatures in the storage units (Tetzlaff *et al*., 2014). The 2.5 years warm-up period was deemed sufficient after tests running the model showed a negligible impact on model calibration.

IV Results

4.1 Long-term hydroclimatic and isotopic variability

The study period covers the driest period 03/04 in the 42 year available record for the Girnock. The summer drought 2003 followed remarkably wet periods in 00/01 and 01/02 which were amongst the wettest years on record (Figure 3). Below average *Q* persisted for over two years following the 2003 drought, a period which was also characterized by below average *P*; low annual runoff coefficients (RC) reflect this. The year 06/07 was again extremely wet, with the following two years only slightly below average. Estimated *ETP* appears quite stable oscillating between 380 and 430 mm yr^{-1} which reflects the energy-limited environment. Uncertainty in the annual water balance is however most likely affected by underestimated *P* inputs and measurement errors in the montane environment despite the interpolation method used. In terms of the tracer measurements which began in October 2003, the greatest isotopic range in *Q* was observed for 03/04 followed by the wet year 06/07 and the lowest range in the year 04/05 (Table 1). This did not, however, always correspond to the input variability. The mean annual isotopic *P* and *Q* values are generally similar, but the mean oxygen-18 in *Q* tended to be more depleted compared to the mean in *P*.

4.2 Model simulations

The models calibrated to individual water years showed varying, but mostly reasonable performance (Figure 4). The Pareto fronts showed sharp gradients indicating performance close to optimal values (Figure 5). The water year 04/05 gave the best discharge performance with KGE_Q = 0.85, but among the lowest isotope performance with KGE_δ^{18}O = 0.7. The discharge and isotopic dynamics were mostly captured by the simulation ranges (Figure 4) although the highest peak discharges tended to be under-predicted and dry period flows are over-estimated (most notably during summer 2004 and 2005). This most likely resulted from underestimated *P* input, carry-over effects of the summer 2003 drought resulting in unsaturated storages and a model calibration that compensated both high and low flow periods. Additionally, in some years (2004, 2005 and 2008), under-prediction of summer stream isotope signatures was evident, which would be consistent with fractionation in the riparian wetlands; a process not considered by the model used here (cf. Birkel *et al.*, 2014). The information content of tracers was variable, being greatest when input variation was most marked and limited if low (Table 1). The information content is probably reflected in the best isotope performance (KGE_δ^{18}O = 0.87) achieved for the water year 07/08 where inputs were most variable.

4.3 Dynamic and additional storage dynamics

The daily mean *Q* over the study period showed clear seasonality which mainly reflected the influence of daily mean *PET* as *P* was fairly evenly distributed throughout the year (Figure 6). This seasonality was also reflected in the mean daily dynamic storage S_d which was derived from the model, though this was slightly lagged with lower (negative storage reflects unsaturated conditions) storage values during summer through to autumn (roughly between May and October). The storage deficits were largely restricted to S_{up} as both S_{low} and S_{sat} on average remained positive due to limited groundwater drawdown and maintenance of saturation in the riparian zone. The S_d changes for individual years exhibited marked differences. For example, the daily S_d for the 2003 drought showed storage deficits from the end of April, and unsaturated conditions persisted well into 2004. In contrast, the wettest years such as 2007 exhibited persistent positive storage in S_{sat} and S_{low}. Periods of short, transient storage deficits which appeared limited to S_{up} occur in May and June, with the wet summer maintaining saturated conditions (and positive storage) throughout the rest of the year. This characteristic of the catchment was reflected in runoff responses to even small rainfall events and the perennial nature of the stream (Tetzlaff *et al.*, 2014).

In the upper hillslopes, the daily S_{up} estimates showed a highly non-linear relationship with MV_{up} while this relationship was more linear in the saturation area (Figure 7). Furthermore, the storage range and variability was much greater in the hillslopes (Figure 7a) compared to the saturation area (Figure 7b) and groundwater storages (not shown). The greatest storage range could be observed for 03/04 in the hillslopes in contrast to the lowest range in the saturation area which resulted in lower S_d. Wetter years (06/07 and 07/08) could also clearly be distinguished by the largest S_{Sat} and MV_{sat} storage ranges in the saturation area. There was less clear distinction for the calibrated *MVs* in the hillslopes showing significant overlap for most years. High mixing volumes were only evident in response to decreased unsaturated conditions during wetter years of 06/07 and 07/08.

4.4 Long-term storage tracking and water age

The inter-annual S_d variability was clearly related to water balance considerations driven by the magnitude and distribution of *P* and resulting *Q* (Figure 8a and b, Table 3) while the much larger (order of magnitude) calibrated mixing volumes *MVs* were more complex. We hypothesized that the *MVs* would vary with hydroclimatic conditions and tested this using the calibrated parameter values from the model. Obviously, this introduced sources of uncertainty; for example the model performance played an important role in the calibration of the catchment storage volumes and derived *TT* (Figure 8a and c). Linked to this was that tracer simulations were generally better when variability of inputs was most marked in a particular water year, which introduced a stochastic element into each year's calibration.

Despite these limitations, the model results were generally reasonable for both flow and tracers in each year (Figure 4). Consequently, cautious interpretation suggested that the calibrated *MVs* do produce plausible mean annual *TT* estimates (Figure 8a and d). However, errors expressed as standard deviations over all simulations were large, which effectively restricts analysis to distinguishing between extreme values. However, this analysis seeks to differentiate relative dynamics rather than absolute values. The *TT* estimates were lowest for the wettest year (1.4 yr for 06/07). This seemed to be positively related to the variability of the saturation area dynamics (*CV_dSAT*) (Figure 8d), which reflects the distribution and frequency of catchment *P* and resulting near-surface flow pathways. The limited *CV_dSAT* (0.69) in the driest year 03/04 was also consistent with a higher *TT* (2.1 yr).

In contrast to the inter-annual assessment using mean annual *TT*, analysis of intra-annual variability used flux water age and the associated controlling factors are conceptualized in Figure 9. The two most contrasting water years (dry 04/05 and wet 06/07) were used to show the interplay of tracer damping (Figure 9a), the frequency of near-surface flow pathway contributions to *Q* (Figure 9b), the frequency of hydrological connectivity in the form of expanded saturation areas adjacent to the stream (Figure 9c) and resulting S_d variability (Figure 9d) on clearly non-stationary water ages (Figure 9e). The non-stationarity of flux water ages is further emphasized if expressed as cumulative distribution functions *CDF*. Shorter water age could be associated with runoff events and mostly younger waters (in the order of weeks to a few months) originating from the connected saturation areas contributing to *Q* via near-surface flow pathways (saturation overland flow). In contrast, contributions of older groundwaters (>4 years in age) during drier conditions and with reduced connectivity resulted in increased water age. In the case of 04/05, the less frequent expansion of the saturated areas resulted in a lower overall contribution of near surface flow paths to the annual runoff, and a greater proportion of deeper groundwater contributions. The resulting greater mixing increased water age and reduced the influence of the younger waters. As a result, the model derived estimate of the daily catchment water age can vary over the course of a year between a couple of months to >4 years. When the daily flux water ages were used to derive mean annual values these resulted in 1.1 years for the wet year 06/07 and in 1.6 years for the drier year 04/05 (taken from Figure 9e).

V Discussion

5.1 Internal model storage dynamics

In this study we integrated isotope oxygen-18 transport into a conceptual rainfall-runoff model to track internal dynamic storage changes and the additional mixing volumes needed to match observed tracer damping in streamwater. This built on earlier work by Birkel *et al.*, 2011b and is conceptually similar to the approach of Hrachowitz *et al*. (2013) which used Chloride as a conservative tracer. However, in this intensively monitored catchment, the model was developed using the dominant process concept reflecting the landscape controls on runoff generation including dynamic saturation areas, freely draining hillslopes and significant groundwater storage. Within this framework, a key goal was minimizing the number of calibrated parameters (Sivakumar, 2004). Central to this landscape-based approach was the conceptualization of dynamic hydrological connectivity that was based on empirical data (Ali *et al*., 2014). This dynamic connectivity incorporates – when necessary - a high-degree of non-linearity specifically in terms of storage dynamics and resulting near-surface runoff generation mechanisms which are broadly consistent with the process-based insights from the catchment (Tetzlaff *et al.*, 2014). In addition, the constrained initial parameter ranges (Table 2) used for calibration match the hydrochemically determined groundwater contributions from previous observations (Birkel *et al.*, 2010). In other words the model performs an internal hydrograph separation consistent with geochemical source tracers (alkalinity) to help ensure simple conceptualisation of process realism as advocated by Kirchner (2006).

The model was calibrated using a state-of-the-art multi-objective optimization algorithm (NSGA2 by Deb *et al.*, 2002) minimizing the modified Kling-Gupta efficiency criterion (Kling *et al.*, 2012). This allowed simultaneous calibration against tracer and discharge data, overcoming reported limitations of single performance measures such as the widely used Nash-Sutcliffe efficiency (Schaefli and Gupta, 2007). Recognizing that no optimal solution can be found with an eight-parameter model (e.g. Beven, 2012) the sharp gradients of the Pareto fronts indicate that solutions fall at least close to optimum (achievable with this model) for the calibration targets and are assumed to be behavioural representations of the system (Figure 4). Nevertheless, we acknowledge the inherent uncertainties and limitations of the data sources, our simple conceptual model approach and the variability introduced using sub-periods for calibration. Nevertheless, these sub-periods allow the generation of time-variable parameter sets and subsequently an insight into non-stationary catchment behaviour similar to Wresta *et al.* (2014).

Model storage dynamics and resulting water age - as expected - depend on model performance and parameter sensitivity (e.g. Dunn *et al*., 2010). Furthermore, concerns have been expressed as to what degree simple conceptual models are able to reproduce complex non-linear behaviour in the form of e.g. hysteresis (Beven, 2012). Nevertheless, model simulations can be useful learning tools for understanding system dynamics and it is in this sense that we argue our approach, although in some ways qualitative, provides useful insights. Results showed, for example, that there is a clear seasonal correspondence between the simulated dynamic storage and observed hydrometric data (Figure 6). This is despite the fact that the Girnock does not exhibit a simple non-linear storage-discharge relationship (Kirchner, 2009) at daily or sub-daily time scales (Birkel *et al*., 2011a). It also becomes apparent from the storage simulations that the drought period 2003 showed a marked persistence well into the second half of 2004 in terms of unsaturated (negative) storage conditions (Figure 6). Other modelling (not shown here) has suggested that the only other occasion when this occurred was for the second most severe drought on record in the 1970s. This is not visible from discharge simulations alone and is an insight from the storage tracking undertaken in this study. Furthermore, simulated dynamic storage was consistent with observed soil moisture and groundwater dynamics recently reported in Birkel *et al*. (2014) and Tetzlaff *et al*. (2014). This included marked non-linear dynamics on the hillslopes and increasingly linear behaviour in the well-buffered saturation areas.

In contrast to the dynamic storage (which is a model state), the additional storage needed for mixing is based on calibrated parameters (e.g. Barnes and Bonell, 1996). Since tracer transport was directly coupled to water transport (Equation 3) the corresponding mixing assumption is that of "instantaneous" and "well-mixed" reservoirs to produce the observed tracer damping in stream. This has been shown to be an overly simplistic concept (e.g. Fenicia *et al*., 2010; Botter *et al*., 2012) with advective-dispersion processes most likely controlling mixing in real world situations (Kirchner and Neal, 2013). That said, Tetzlaff *et al*. (2014) showed that virtually all precipitation damping in the catchment occurs in the upper 0.3 m of the catchments organic soils. Nevertheless, in this study we relaxed this assumption using the model's dynamic, non-linear connectivity to convert additional storage parameters into different mixing volumes according to the wetness state of the catchment. This resulted in a type of partial mixing mechanism whilst maintaining our key goal of keeping model parameters to a minimum. However, the effect of such a partial mixing mechanism could really only be discerned for climatically extreme periods (Figure 7) similar to the results of Hrachowitz *et al*. (2013). The incorporation of additional parameters could improve this, but this is statistically difficult to justify. Work from Benettin *et al*. (2013) might help in this regard to pre-define mixing assumptions based on advection-dispersion models. Other types of integrated flow-tracer models such as the multiple interactive pathways approach by Davies *et al*. (2011) also allows the direct incorporation of partial mixing with minimal parameterization. However, in wet environments like the Scottish Highlands, the relatively even distribution of P throughout the year, in combination with high water storage in saturated organic soils (e.g. histosols) fringing the riparian zone seems to integrate partial mixing in other geographic source areas resulting in a more complete mixing at the catchment scale. Therefore, Tetzlaff *et al*. (2014) used the term "isostat" to describe this characteristic mixing behaviour of the saturation areas.

5.2 Long-term storage and water age's response to hydroclimatic variability

Storage simulations showed that much larger mixing volumes (i.e. an order of magnitude) determine *TT*s rather than dynamic storage changes as shown by e.g. Birkel *et al*. (2011a). Building on previous work, this longer-term data analysis clearly demonstrated the non-stationary character of mean annual *TT*s (Figure 8) oscillating between 1.4 and 2.4 years (Table 3). It is clear that the interpretation of *TT*s is relative due to large uncertainties as expressed in the error bars in Figure 8d and must therefore be cautious. However, despite not being exactly the same (e.g. Botter *et al*., 2011), the latter *TT* values were consistent with previously determined mean transit time estimates in the catchment applying lumped convolution integral methods and spectral analysis (Hrachowitz *et al*. 2010, Kirchner *et al*., 2010). Furthermore, the flux water ages derived from daily tracking of model states and fluxes (Hrachowitz *et al*., 2013) to assess intra-annual variability if used to derive mean annual water age were shorter than the *TT* (e.g. 1.1 against 1.4 yr for 06/07). They do, however, also fall well within the estimated errors of *TT*s. Both water age and *TT* corresponded to climate extremes (specifically wetter periods decrease *TT* and flux water age) as previously reported by Hrachowitz *et al*. (2011).

The non-stationary character of mean annual *TT*s could be related to the variability of saturation area dynamics (Figure 8) or – in other words – hydrological connectivity, which is based on antecedent wetness rather than topographic controls (Ali *et al*., 2014). This finding also holds for intra-annual water ages and is consistent with work by Roa-Garcia and Weiler (2010) in that high antecedent wetness increases flow path connectivity which results in shorter water transit times and reduced recharge and deeper mixing. In the Girnock, this is reflected in the dominance of near-surface runoff generation processes mostly in form of saturation overland flow which contribute increasingly younger waters to streamflow under high-frequency and large saturation area connectivity (Figure 9e – flux water age estimates and derived CDF).

Despite these tentative conclusions, we also recognize the limitations of stable water isotopes in determining such mixing volumes (Stewart *et al*., 2010) due to dependence on input variability and the degree of damping in output (Kirchner *et al*., 2010). This was reflected in better simulations where input variability was greater and damping was lower (Figure 8, Table 1). It might be that the lowest flows are dominated by storage levels with a older water age than can be detected by simple isotope input-output variability (which in this model is represented by calibrated additional storage parameters to be able to reproduce the degree of observed damping). Recently, Peters *et al*. (2013) showed how long-term Tritium data could be used to estimate total catchment storage and fluctuations at smaller scales and relatively impermeable bedrock substrate.

The quantification of total catchment storage, however, remains a challenge (hence we preferred to use the term catchment storage rather than total catchment storage to emphasize this), but we hope to have presented a step in the right direction using coupled tracer and hydrology models. Certainly recent geophysical surveys in the Girnock (J. Bradford, Boise State University, Pers. Comm.) have indicated up to 40m of saturated drift in valley bottom areas. However, this drift has a low porosity (~10-20%) and low hydraulic conductivity (~10mm d^{-1}) so catchment scale storage is limited to about 1000-2000mm. These figures are of the same order of magnitude as those derived in the current study, and other tracer-based studies using more conventional convolution integral models (Tetzlaff *et al.*, 2014).

VI Conclusions

In this study we showed how the interplay of tracer dynamics, hydrological connectivity and catchment storage affects non-stationary water ages using longer-term hydrometric and isotopic data sets. We also have shown how hydroclimatic variability relates to catchment storage dynamics and resulting time-variable water age. A coupled tracer and runoff model showed that relatively modest dynamic storage fluctuations (<100mm) account for seasonal flow variability in most years including the driest year 2003 (on a 42 year data record). However, storage volumes an order of magnitude greater (>1000mm) need to be invoked to account for the mixing needed to simulate isotope dynamics. This clearly demonstrated the non-stationarity of catchment water age at daily and annual time scales, but in the Scottish environment these only vary by a factor of two in the more extreme wet and dry periods. Such extreme conditions have a major influence on hydrological connectivity and contributing flow pathways which in turn affect the proportion of younger near-surface waters (~ 10 to 50 days old) or older groundwaters (~ 1 to 10 years old) reaching the stream. However, in most years, the riparian wetlands represent a large water storage available for mixing which acts as an “isostat” moderating isotope variability and limiting the time variance of water age. Even though, the “isostat” effect is most likely more prominent in northern upland catchments with histosols fringing the stream, the methods and model concepts could easily be transferred and tested at other sites. Despite the limitations associated with using isotope tracers and inevitable model errors, we present a step towards characterizing catchment storage and its associated dynamics. We conclude that much can be learnt from long-term analysis of hydrometric and isotopic data and their integration in hydrological models to understand the potential impact of extreme climatic events on hydrological systems. Specifically, the approach helps assess the degree to which such extremes are mediated by internal dynamic storage changes and additional storage available for mixing.

Acknowledgements

We would like to gratefully acknowledge data provided by SEPA, Iain Malcolm (MSS) and BADC. Mark Speed, Susan Waldron and many MSS staff helped with sample collection and lab analysis. We thank the European Research Council ERC (project GA 335910 VEWA) for funding and are grateful for constructive comments provided by three anonymous reviewers

References

Ali G, Birkel C, Tetzlaff D, Soulsby C, McDonnell JJ, Tarolli P. 2014. A comparison of wetness indices for the prediction of connected saturated areas under contrasted conditions. *Earth Surface Processes and Landforms*. doi: 10.1002/esp.3506.

Andrews F, Croke B, Jakeman A. 2011. An open software environment for hydrological model assessment and development. *Environmental Modelling & Software* **26**: 1171-1185. doi:10.1016/j.envsoft.2011.04.006.

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

Beven K. 2012. *Rainfall-runoff modelling: The Primer*. Second edition. Wiley-Blackwell.

Benettin P, Rinaldo A, Botter G. 2013. Kinematics of age mixing in advection-dispersion models. *Water Resour. Res*. **49**: 8539–8551. doi:10.1002/2013WR014708.

Birkel C, Tetzlaff D, Dunn SM, Soulsby C. 2010. Towards simple dynamic process conceptualization in rainfall-runoff models using multi-criteria calibration and tracers in temperate, upland catchments. *Hydrol. Process.* **24**: 260-275.

Birkel C, Tetzlaff D, Dunn SM, Soulsby C. 2011a. Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models. *Water Resources Research.* ***47***. DOI: 10.1029/2010WR009547

Birkel C, Soulsby C, Tetzlaff D. 2011b. Modelling catchment-scale water storage dynamics: reconciling dynamic storage with tracer-inferred passive storage. *Hydrol. Process.* ***25***: 3924–3936. doi: 10.1002/hyp.8201.

Birkel C, Soulsby C, Tetzlaff D. 2014. Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. *Water Resources Research*. Doi: 10.1002/2013WR014925.

Bishop K, Seibert J, Nyberg L, Rodhe A. 2011. Water storage in a till catchment. II: Implications of transmissivity feedback for flow paths and turnover times. *Hydrol. Process*. **25**: 3950–3959. doi: 10.1002/hyp.8355

Botter G, Bertuzzo E, Rinaldo A. 2011. Catchment residence and travel time distributions: The master equation. *Geophys. Res. Lett*. **38**: L11403. doi:10.1029/2011GL047666.

Botter G. 2012. Catchment mixing processes and travel time distributions. *Water Resour. Res*. **48**: W05545. doi:10.1029/2011WR011160.

Brooks JR, Barnard HR, Coulombe R, McDonnell JJ. 2009. Ecohydrologic separation of water between trees and streams in a Mediterranean climate. *Nature Geoscience* **3**: 100 – 104.

Brauer C., Teuling AJ, Torfs PJJF, Uijlenhoet R. 2013. Investigating storage-discharge relations in a lowland catchment using hydrograph fitting, recession analysis, and soil moisture data. *Water Resour. Res.* **49**: 4257–4264. doi:10.1002/wrcr.20320.

Capell R, Tetzlaff D, Soulsby C. 2013. Will catchment characteristics moderate the projected effects of climate change on flow regimes in the Scottish Highlands?, *Hydrol. Process.* **27**: 687–699. doi: 10.1002/hyp.9626

Davies J, Beven K, Rodhe A, Nyberg L, Bishop K. 2013. Integrated modelling of flow and residence times at the catchment scale with multiple interacting pathways. *Water Resour. Res*. **49**: 4738–4750, doi:10.1002/wrcr.20377.

Deb K, Pratap A, Agarwal S, Meyarivan T. 2002. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transaction on Evolutionary Computation* **6**(2): 181-197.

Dunn SM, Birkel C, Tetzlaff D, Soulsby C. 2010. Transit time distributions of a conceptual model: their characteristics and sensitivities. *Hydrol. Process.* **24**: 1719–1729. doi: 10.1002/hyp.7560.

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. *Hydrol. Process.* **24**(12): 1730–1741. doi: 10.1002/hyp.7595.

Heidbüchel I, Troch PA, Lyon SW, Weiler M. 2012. The master transit time distribution of variable flow systems. *Water Resour. Res.* **48**: W06520. doi: 10.1029/2011WR011293.

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 timevariant transit time assessment. *Water Resour. Res*. **46**: W10536. doi:10.1029/2010WR009148.

Hrachowitz M, Soulsby C, Tetzlaff D, Malcolm IA. 2011. Sensitivity of mean transit time estimates to model conditioning and data availability. *Hydrol. Processes* **25**(6): 980–990. doi:10.1002/hyp.7922.

Hrachowitz M, Savenije H, Bogaard TA, Tetzlaff, Soulsby C. 2013. What can flux tracking teach us about water age distribution patterns and their temporal dynamics? *Hydrol. Earth Syst. Sci.* **17**: 533-564. doi: 10.5194/hess-17-533-2013.

Kirchner JW, Feng X, Neal C. 2000. Fractal stream chemistry and its implications for contaminant transport in catchments. *Nature* **403**: 524–527. doi:10.1038/35000537.

Kirchner JW. 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res*. **42**: W03S04. doi: 10.1029/2005WR004362.

Kirchner JW. 2009. Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resour. Res*. **45**: W02429. doi: 10.1029/2008WR006912.

Kirchner JW, Tetzlaff D, Soulsby C. 2010. Comparing chloride and water isotopes as hydrological tracers in two Scottish catchments. *Hydrol. Process.* **24**: 1631–1645. DOI: 10.1002/hyp.7676.

Kirchner JW, Neal C. 2013. Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection. *PNAS* **110**: 12213–12218. doi/10.1073/pnas.1304328110.

Kling H, Fuchs M, Paulin M. 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol*. **424–425**: 264–277. doi:10.1016/j.jhydrol.2012.01.001.

Maloszewski P, Zuber A. 1982. Determining the turnover time of groundwater systems with the aid of environmental tracers, 1. Models and their applicability. *J. Hydrol*. **57**: 207-231.

McDonnell JJ, Beven K. 2014. Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities, and residence time distributions of the headwater hydrograph. *Water Resour. Res*., **50**: doi:10.1002/2013WR015141.

McMillan H, Tetzlaff D, Clark M, Soulsby C. 2012. Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach. *Water Resour. Res.* **48**: W05501. doi:10.1029/2011WR 011688.

McNamara JP, Tetzlaff D, Bishop K, Soulsby C, Seyfried M, Peters NE, Aulenbach BT, Hooper R. 2011. Storage as a Metric of Catchment Comparison. *Hydrol. Process.* **25**. DOI: 10.1002/hyp.8113

Peters NE, Burns D, Aulenbach BT. 2013. Evaluation of High-Frequency Mean Streamwater Transit-Time Estimates Using Groundwater Age and Dissolved Silica Concentrations in a Small Forested Watershed. *Aquat Geochem*, DOI: 10.1007/s10498-013-9207-6

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 Resour. Res*. **47**: W07537. doi:10.1029/2011WR010478.

Roa-García MC, Weiler M. 2010. Integrated response and transit time distributions of watersheds by combining hydrograph separation and long-term transit time modeling. *Hydrology and Earth System Sciences* **14**: 1537-1549, DOI: 10.5194/hess-14-1537-2010, 2010.

Rodell M, Velicogna I, Famiglietti JS. 2009. Satellite-based estimates of groundwater depletion in India. *Nature* **460**: 999-1002. DOI: 10.1038/nature08238.

Salve R, Rempe DM, Dietrich WE. 2012. Rain, rock moisture dynamics, and the rapid response of perched groundwater in weathered, fractured argillite underlying a steep hillslope. *Water Resour. Res*. **48**: W11528. doi: 10.1029/2012WR012583.

Schaefli B, Gupta H. 2007. Do Nash values have value? *Hydrological Processes* **21**: 2075-2080. DOI: 10.1002/hyp.6825.

Seibert J, McDonnell J.J. 2010. Land-cover impacts on streamflow: A change-detection modeling approach that incorporates parameter uncertainty. *Hydrological Sciences Journal* **55**(3): 316-332.

Shaw S, Harpold AA, Taylor JC, Walter MT. 2008. Investigating a high resolution, stream chloride time series from the Biscuit Brook catchment, Catskills, NY. *J. Hydrol*. **348**: 245–256.

Sivakumar B. 2004. Dominant processes concept in hydrology: moving forward. *Hydrological Processes* **18**: 2349–2353.

Soulsby C, Tetzlaff D, van den Bedem N, Malcolm IA, Bacon PJ, Youngson AF. 2007. Inferring groundwater influences on surface water in montane catchments from hydrochemical surveys of springs and streamwaters. *J. Hydrol*. **333**: 199– 213.

Soulsby C, Neal C, Laudon H, Burns DA, Merot P, Bonell M, Dunn SM, Tetzlaff D. 2008. Catchment data for process conceptualization: Simply not enough? *Hydrol. Processes* **22**: 2057– 2061.

Soulsby C, Tetzlaff D, Hrachowitz M. 2009. Tracers and transit times: windows for viewing catchment scale storage? *Hydrol. Process*. **23**: 3503–3507. DOI: 10.1002/hyp.7501

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, Soulsby C, Waldron S, Malcolm IA, Bacon PJ, Dunn SM, Lilly A. 2007. Conceptualisation of runoff processes using GIS and tracers in a nested mesoscale catchment. *Hydrological Processes* **21**: 1289–1307.

Tetzlaff D, Uhlenbrook S, Eppert S, Soulsby C. 2008. Does the incorporation of process conceptualisation and tracer data improve the structure and performance of a simple rainfall-runoff model in a Scottish mesoscale catchment? *Hydrological Processes* **22**: 2461-2474.

Tetzlaff D, Birkel C, Dick J, Soulsby C. 2014. Storage dynamics in hydropedological units control hillslope connectivity, runoff generation and the evolution of catchment transit time distributions. *Water Resour. Res*., doi: 10.1002/2013WR014147.

van der Velde Y, de Rooij GH, Rozemeijer JC, van Geer FC, Broers HP. 2010. Nitrate response of a lowland catchment: On the relation between stream concentration and travel time distribution dynamics. *Water Resour. Res.* **46**: W11534. doi:10.1029/2010WR009105.

679 Westra S, Thyer M, Leonard M, Kavetski D, Lambert M. 2014. A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resour. Res.*, **50**: 5090–5113, doi:10.1002/2013WR014719.

682 Zreda M, Shuttleworth WJ, Zeng X, Zweck C, Desilets D, Franz TE, Rosolem R. 2012. COSMOS: the COsmic-ray Soil Moisture Observing System. *Hydrol. Earth Syst. Sci.* **16**: 4079– 4099. doi:10.5194/hess-16-4079-2012, 2012.

685 Zuber A. 1986. On the interpretation of tracer data in variable flow systems. *J. Hydrol*. **86**: 45–57.

TABLES

Table 1: Isotopic signatures for input ($P\_\delta^{18}O$) and output ($Q\_\delta^{18}O$) annual mean, minimum, maximum (in brackets), isotopic ranges (in per mil) and the coefficient of variation (CV) over the study period. The annual water balance and the CV of the % catchment saturation area extent (dSAT) is also given.

Table 2: Mean parameter values and ranges expressed as minimum and maximum values (in parenthesis) for each annual time window derived from the best parameter (500 parameter sets) population after multi-objective calibration. This indicates the accepted parameter variability and mean performance measures (KGE) derived from the Pareto fronts in Figure 4. The initial parameter ranges were constrained based on previous tracer-based studies (Birkel *et al.*, 2010).

Table 3: Storage totals (S) are given for the annual means of the individual calibration periods in the context of mean annual transit times (TT). Note that S_{up} is only given for values exceeding the baseline when lateral and vertical fluxes were generated from this reservoir.

FIGURES

Figure 1: Location and topography of the Girnock experimental catchment.

Figure 2: The model concept used to simulate discharge and oxygen-18. Three reservoirs (upper, lower and saturation area) conceptualize water and tracer fluxes with associated dynamic storage (S_{up}, S_{low} and S_{sat}) and additional storage available for mixing (upS_p, $lowS_p$ and $satS_p$). The latter calibrated additional storage parameters are converted into dynamic mixing volumes MV according to wetness state (*dSAT*) assuming that during expansion of the saturation areas (> wetness) more additional storage is available for mixing contrary to the hillslopes. Note that calibrated parameters are shown in red.

Figure 3: Mean annual water balance (P, Q and PET) and runoff coefficients (RC) emphasize the climate variability in form of the drought period 2003.

Figure 4: Discharge and isotope simulations are plotted using the range of the individual calibration periods. The simulation ranges were derived from the best parameter population (500 sets) of each calibration window (shown as Pareto fronts in Figure 5 with parameter ranges given in Table 2). Simulation bounds indicate parameter variability as a proxy for uncertainty. Note that the same colour code is used throughout and that discharge is plotted on a logarithmic scale.

Figure 5: Pareto fronts of annual time windows (water years) derived from multi-objective calibration (NSGA2) using discharge and $\delta^{18}O$ time series. Performance was assessed using the Kling-Gupta efficiency (KGE).

Figure 6: a) Daily mean rain (blue), discharge (black) and PET. b) Model derived mean daily dynamic storage (S_d - blue line), daily mean individual storages (dashed lines) and daily dynamic storage for selected years (2003, 2004 and 2007). Note that the scale is Julian days emphasizing the carry-over effect of the drought period 2003.

Figure 7: a) Active hillslope and b) saturation area storages are individually plotted against dynamic mixing volumes for the six complete water years spanning the study period 2003 to 2009.

Figure 8: Annual dynamic and additional storage available for mixing (sum of three reservoirs, respectively) are shown in relation to a) mean annual discharge Q and resulting mean transit time (TT) estimates. b) shows the dynamic storage against additional mixing volumes and c) show the relationship of precipitation isotope input variability (CV P_$\delta^{18}O$) to model performance (KGE) simulating stream $\delta^{18}O$ and d) the variability of catchment wetness (CV dSAT) to TT. Error bars reflect standard deviations across all simulations reflecting interannual variability.

Figure 9: The conceptual diagram shows the main controls in terms of a) tracer damping (measured oxygen-18 input-output time series), b) near-surface flow pathways (daily histogram), c) connectivity (daily histogram) and d) dynamic storage (simulated time series) on e) streamwater age exemplified using two contrasting (dry 2004/05 in red and wet 2006/07 in orange) water years. The degree of tracer damping is highly variable on an inter-annual basis, but reflects the additional storage needed for mixing and resulting streamwater age time series and CDF (see also Figure 8 and Table 2 for mean annual average values). Note that the time series are coloured according to the previously used scheme.

TABLES

Table 1: Isotopic signatures for input (P_δ^{18}O) and output (Q_δ^{18}O) annual mean, minimum, maximum (in brackets), isotopic ranges (in per mil) and the coefficient of variation (CV) over the study period. The annual water balance and the CV of the % catchment saturation area extent (dSAT) is also given.

Water year	P_ δ^{18}O	P_ δ^{18}O range	CV P_ δ^{18}O	Q_δ^{18}O	Q_δ^{18}O range	CV Q_δ^{18}O	P (mm yr^{-1})	Q (mm yr^{-1})	CV dSAT
03/04	-9.03 [-16.1, -5.2]	10.9	0.34	-8.88 [-11.7, -7.6]	4.1	0.09	845	435	0.69
04/05	-7.74 [-11.5, -3.2]	8.3	0.26	-8.21 [-9.1, -7.3]	1.9	0.04	766	461	0.68
05/06	-8.28 [-16.0, -3.7]	12.4	0.32	-8.50 [-10.5, -7.0]	3.5	0.07	821	466	0.71
06/07	-8.33 [-15.9, -5.6]	10.3	0.25	-8.54 [-10.9, -7.4]	3.5	0.06	1038	740	0.81
07/08	-8.49 [-17.5, -2.9]	14.6	0.38	-8.67 [-9.9, -6.9]	3.0	0.08	835	593	0.73
08/09	-8.49 [-18.5, -3.4]	14.9	0.34	-8.47 [-10.6, -7.2]	3.4	0.09	843	485	0.63

Table 2: Mean parameter values and ranges expressed as minimum and maximum values (in parenthesis) for each annual time window derived from the best parameter (500 parameter sets) population after multi-objective calibration. This indicates the accepted parameter variability and mean performance measures (KGE) derived from the Pareto fronts in Figure 4. The initial parameter ranges were constrained based on previous tracer-based studies (Birkel *et al.*, 2010).

	a	***b***	***Re***	***k***	***α***	***upS_p***	***$satS_p$***	***$lowS_p$***	**Mean KGE_Q**	**Mean KGE_$\delta^{18}O$**
Initial range	[0.2, 0.8]	[0.0001, 0.1]	[0.2, 0.9]	[0.001, 0.1]	[0.1, 0.9]	[0, 1000]	[0, 1000]	[0, 1000]		
Unit	d^{-1}	d^{-1}	d^{-1}	d^{-1}	(-)	mm	mm	mm	(-)	(-)
03/04	0.39 (0.2, 0.6)	0.002 (0.001, 0.004)	0.74 (0.35, 0.79)	0.005 (0.001, 0.016)	0.84 (0.66, 0.9)	125 (102, 386)	8 (4, 59)	878 (277, 899)	0.40	0.80
04/05	0.47 (0.36, 0.61)	0.006 (0.003, 0.007)	0.63 (0.47, 0.8)	0.070 (0.04, 0.1)	0.47 (0.33, 0.57)	255 (83, 583)	36 (2.1, 47)	931 (370, 990)	0.84	0.66
05/06	0.43 (0.28, 0.48)	0.009 (0.002, 0.03)	0.57 (0.44, 0.78)	0.019 (0.01, 0.04)	0.79 (0.54, 0.86)	641 (340, 902)	38 (3, 75)	742 (248, 931)	0.73	0.76
06/07	0.54 (0.26, 0.71)	0.011 (0.004, 0.06)	0.56 (0.29, 0.74)	0.069 (0.02, 0.1)	0.59 (0.3, 0.78)	139 (62, 215)	53 (32, 122)	768 (328, 945)	0.68	0.74
07/08	0.42 (0.2, 0.55)	0.026 (0.01, 0.06)	0.66 (0.31, 0.79)	0.078 (0.04, 0.09)	0.37 (0.13, 0.53)	908 (794, 989)	32 (9, 56)	620 (370, 926)	0.68	0.87
08/09	0.41 (0.27, 0.54)	0.087 (0.07, 0.09)	0.47 (0.21, 0.68)	0.096 (0.05, 0.1)	0.71 (0.39, 0.83)	859 (580, 978)	102 (39, 174)	976 (745, 999)	0.50	0.77

Table 3: Storage totals (S) are given for the annual means of the individual calibration periods in the context of mean annual transit times (TT). Note that S_{up} is only given for values exceeding the baseline when lateral and vertical fluxes were generated from this reservoir.

Calibration period	S_{up}	S_{sat}	S_{low}	MV_{up}	MV_{sat}	$lowS_p$	S	Q	TT
Units	mm	mm	mm	mm	mm	mm	(mm yr^{-1})	(mm yr^{-1})	yr
03/04	3.2	5.3	5.7	33.1	7.2	878	933	435	2.14
04/05	4.9	20.3	23.4	104.0	28.8	931	1112	461	2.42
05/06	5.7	14.2	72.4	93.7	46.2	742	974	466	2.09
06/07	10.3	26.9	133.0	24.5	33.5	768	996	740	1.35
07/08	4.8	18.7	64.6	447.1	32.2	620	1187	593	2.00
08/09	8.6	5.8	37.6	92.8	16.5	976	1137	485	2.35

35

4

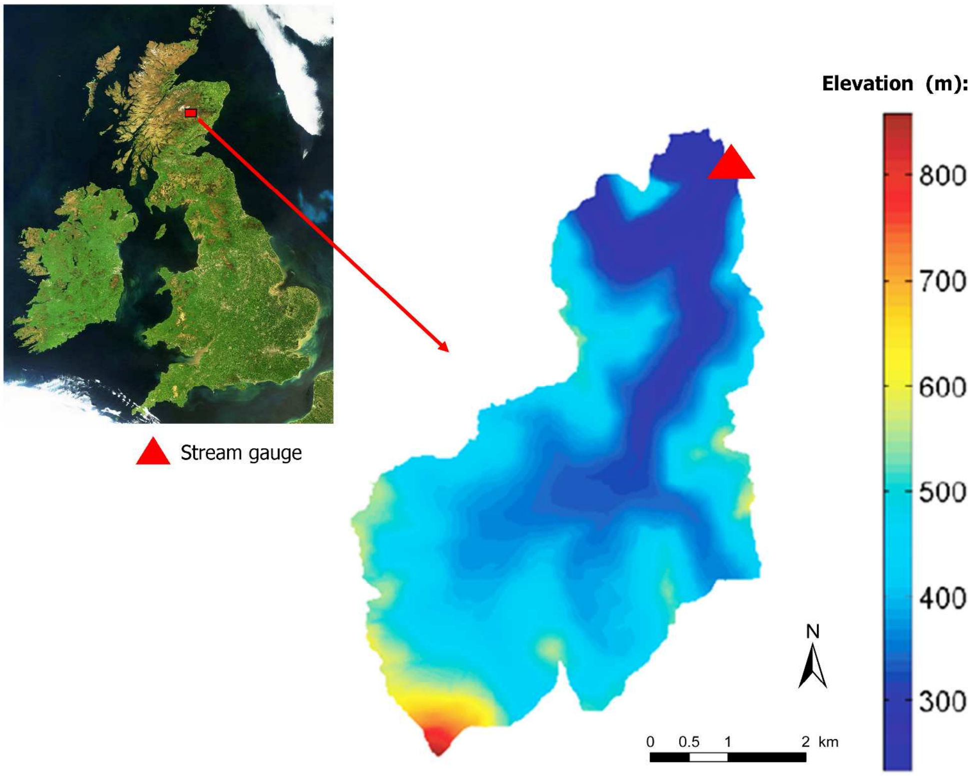

Figure 1: Location and topography of the Girnock experimental catchment.
287x226mm (300 x 300 DPI)

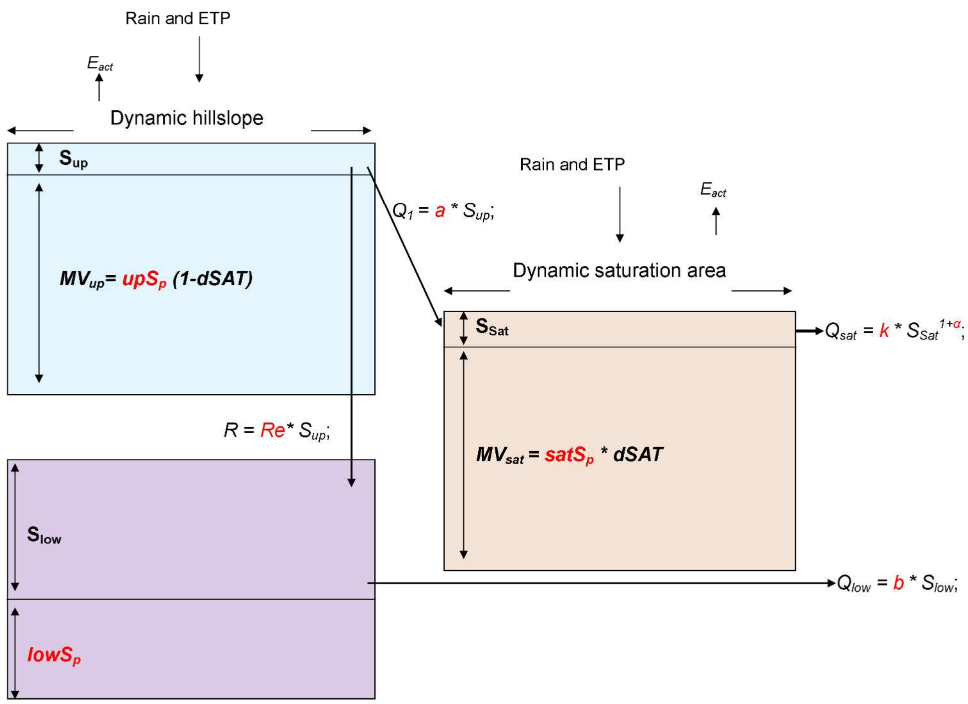

Figure 2: The model concept used to simulate discharge and oxygen-18. Three reservoirs (upper, lower and saturation area) conceptualize water and tracer fluxes with associated dynamic storage (Sup, Slow and Ssat) and additional storage available for mixing (upSp, lowSp and satSp). The latter calibrated additional storage parameters are converted into dynamic mixing volumes MV according to wetness state (dSAT) assuming that during expansion of the saturation areas (> wetness) more additional storage is available for mixing contrary to the hillslopes. Note that calibrated parameters are shown in red.
249x177mm (300 x 300 DPI)

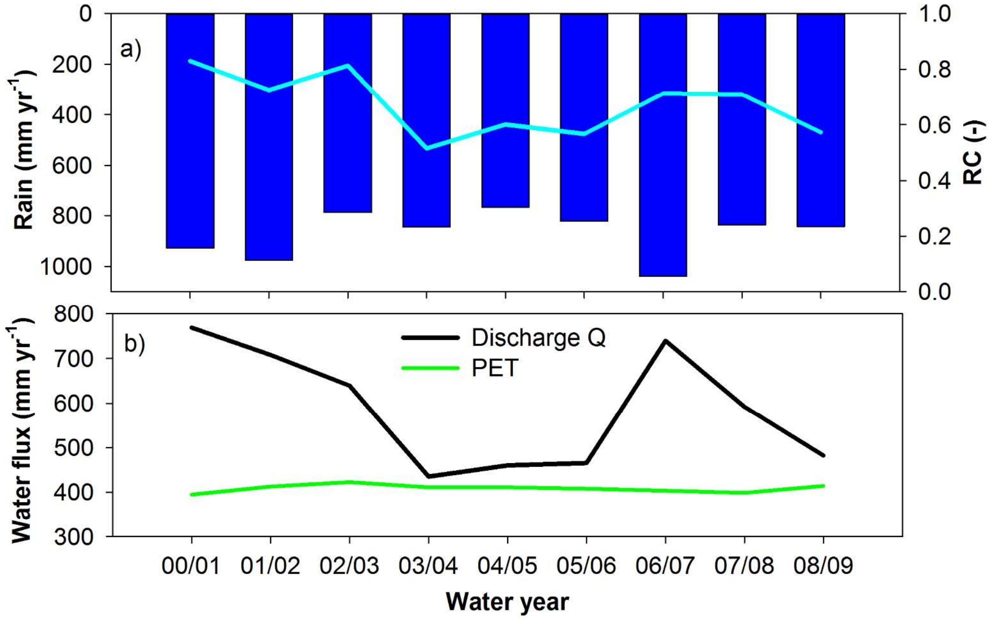

Figure 3: Mean annual water balance (P, Q and PET) and runoff coefficients (RC) emphasize the climate variability in form of the drought period 2003.
201x128mm (300 x 300 DPI)

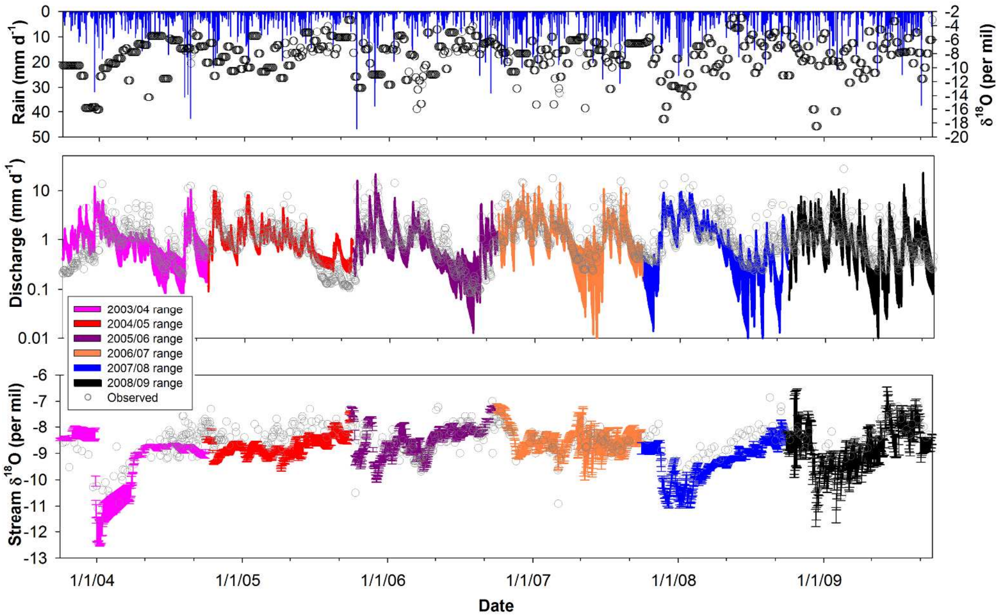

Figure 4: Discharge and isotope simulations are plotted using the range of the individual calibration periods. The simulation ranges were derived from the best parameter population (500 sets) of each calibration window (shown as Pareto fronts in Figure 5 with parameter ranges given in Table 2). Simulation bounds indicate parameter variability as a proxy for uncertainty. Note that the same colour code is used throughout and that discharge is plotted on a logarithmic scale.
290x180mm (300 x 300 DPI)

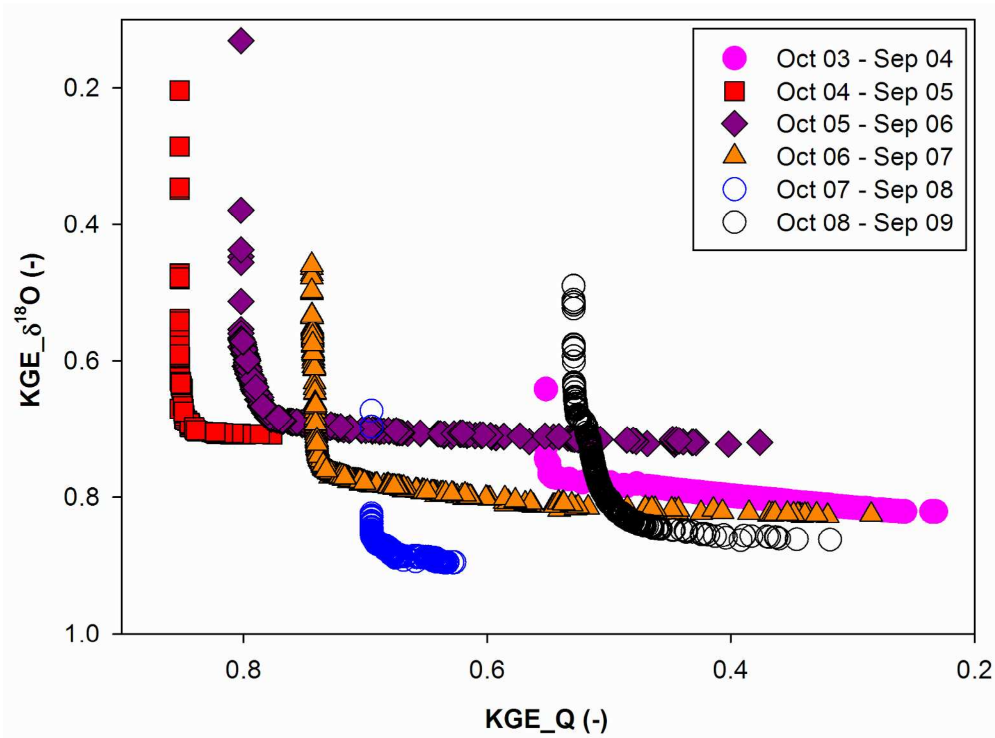

Figure 5: Pareto fronts of annual time windows (water years) derived from multi-objective calibration (NSGA2) using discharge and $\delta^{18}O$ time series. Performance was assessed using the Kling-Gupta efficiency (KGE).
188x140mm (300 x 300 DPI)

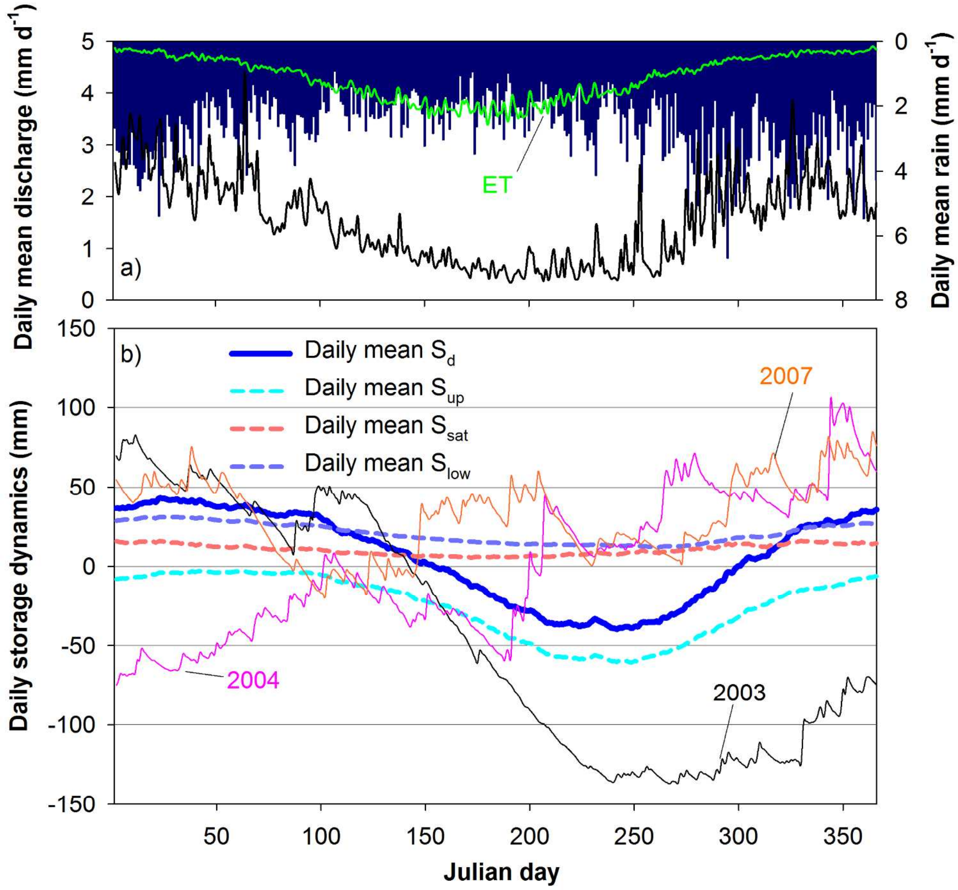

Figure 6: a) Daily mean rain (blue), discharge (black) and PET. b) Model derived mean daily dynamic storage (S^d - blue line), daily mean individual storages (dashed lines) and daily dynamic storage for selected years (2003, 2004 and 2007). Note that the scale is Julian days emphasizing the carry-over effect of the drought period 2003.
203x185mm (300 x 300 DPI)

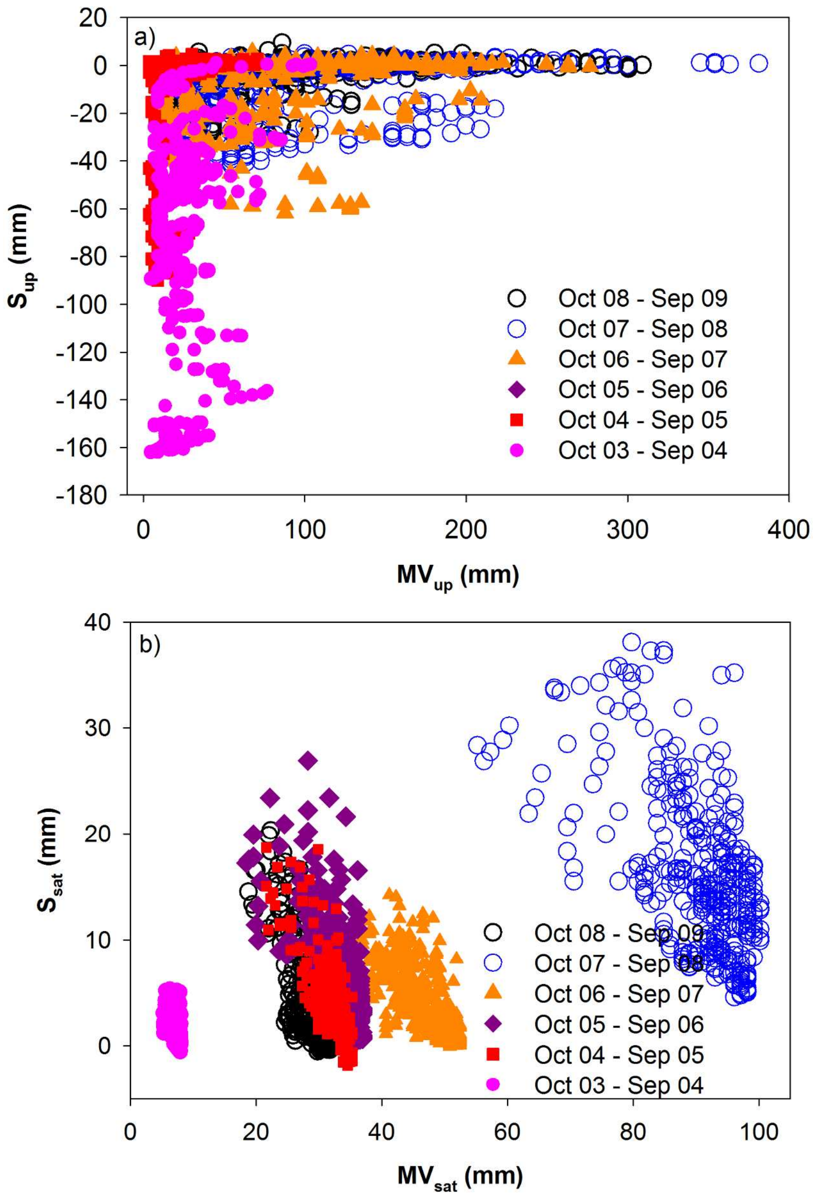

Figure 7: a) Active hillslope and b) saturation area storages are individually plotted against dynamic mixing volumes for the six complete water years spanning the study period 2003 to 2009.
164x240mm (300 x 300 DPI)

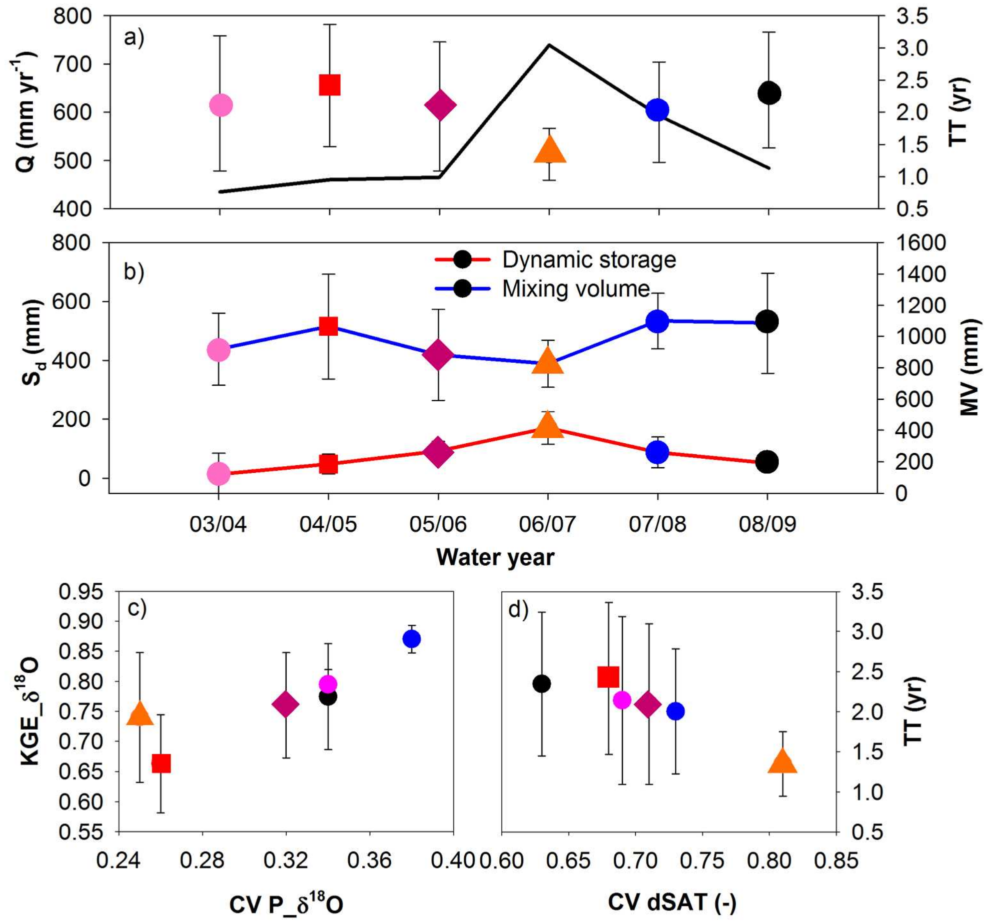

Figure 8: Annual dynamic and additional storage available for mixing (sum of three reservoirs, respectively) are shown in relation to a) mean annual discharge Q and resulting mean transit time (TT) estimates. b) shows the dynamic storage against additional mixing volumes and c) show the relationship of precipitation isotope input variability (CV P_δ^{18}O) to model performance (KGE) simulating stream δ^{18}O and d) the variability of catchment wetness (CV dSAT) to TT. Error bars reflect standard deviations across all simulations reflecting interannual variability.
209x195mm (300 x 300 DPI)

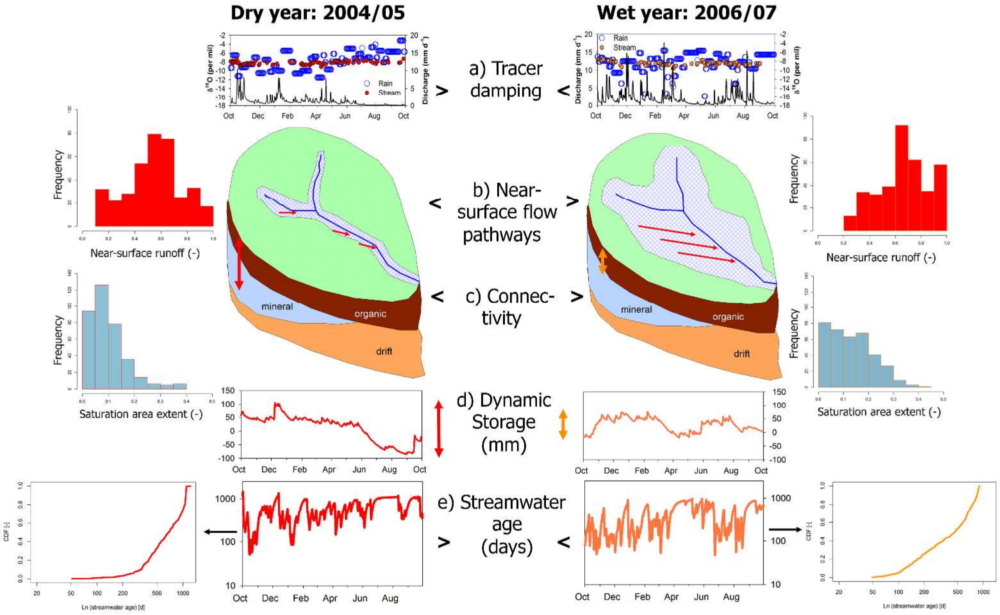

Figure 9: The conceptual diagram shows the main controls in terms of a) tracer damping (measured oxygen-18 input-output time series), b) near-surface flow pathways (daily histogram), c) connectivity (daily histogram) and d) dynamic storage (simulated time series) on e) streamwater age exemplified using two contrasting (dry 2004/05 in red and wet 2006/07 in orange) water years. The degree of tracer damping is highly variable on an inter-annual basis, but reflects the additional storage needed for mixing and resulting streamwater age time series and CDF (see also Figure 8 and Table 2 for mean annual average values). Note that the time series are coloured according to the previously used scheme.
451x281mm (300 x 300 DPI)