HYDROLOGICAL PROCESSES
Hydrol. Process. (2016)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.10902

Using high resolution tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-runoff model

M. H. J. van Huijgevoort,[1] D. Tetzlaff,[1]* E. H. Sutanudjaja[2] and C. Soulsby[1]

[1] *Northern Rivers Institute, School of Geosciences, University of Aberdeen, Aberdeen AB31 5XY, UK*
[2] *Faculty of Geosciences, Utrecht University, Utrecht 3508 TC, Netherlands*

Abstract:

Models simulating stream flow and conservative tracers can provide a representation of flow paths, storage distributions and mixing processes that is advantageous for many predictive purposes. Compared with models that only simulate stream flow, tracer data can be used to investigate the internal consistency of model behaviour and to gain insight into model performance. Here, we examine the strengths and weaknesses of a data-driven, spatially distributed tracer-aided rainfall-runoff model. The model structure allowed us to assess the influence of landscape characteristics on the routing and mixing of water and tracers. The model was applied to a site in the Scottish Highlands with a unique tracer data set; ~4 years of daily isotope ratios in stream water and precipitation were available, as well as 2 years of weekly soil and ground water isotopes. The model structure was based on an empirically based, lumped tracer-aided model previously developed for the catchment. The best model runs were selected from Monte Carlo simulations based on dual calibration criteria using objective functions for both stream isotopes and discharge at the outlet. Model performance for these criteria was reasonable (Nash–Sutcliffe efficiencies for discharge and isotope ratios were ~0.4–0.6). The model could generally reproduce the variable isotope signals in the soils of the steeper hill slopes where storage was low, and damped isotope responses in valley bottom cells with high storage. The model also allowed us to estimate the age distributions of internal stores, water fluxes and stream flow. Average stream water age was ~1.6 years, integrating older groundwater in the valley bottom and dynamic younger soil waters. By tracking water ages and simulating isotopes, the model captured the changes in connectivity driven by distributed storage dynamics. This has substantially improved the representation of spatio-temporal process dynamics and gives a more robust framework for projecting environmental change impacts. Copyright © 2016 The Authors Hydrological Processes Published by John Wiley & Sons Ltd.

KEY WORDS isotopes; tracer-aided modelling; runoff processes; hydrological modelling

Received 10 December 2015; Accepted 28 April 2016

INTRODUCTION

Conservative tracers, including stable isotopes, are increasingly used in hydrological models to explore the mixing relationships between fluxes and storages in catchment systems (see review by Birkel and Soulsby, 2015). Tracers can be used to directly infer travel time distributions in hydrological models and thus, to investigate the internal consistency of the model behaviour (Dunn *et al.,* 2010). Tracer data can also provide additional insight into model performance and can be used as extra criteria besides the more common hydrometric-based goodness of fit measures (Uhlenbrook and Sieber, 2005; Fenicia *et al.,* 2008). This can help in selecting a model that gives the right answer for the right reasons (Kirchner, 2006). The importance of using both hydrograph data and tracer data in catchment investigations and integrating tracer data in models was emphasised by McDonnell and Beven (2014). They argued that such a dual approach was essential to be able to distinguish and reconcile both the rapid celerity of the hydrological response and the slower movement of water particles inferred by conservative tracers. In many studies, tracers have been traditionally used simply to estimate the mean transit time of the water in a catchment (McGuire and McDonnell, 2006), but more recent empirical and theoretical work has stressed the need to characterise the transit time distribution and its time variance (Botter *et al.,* 2010; McDonnell *et al.,* 2010; Heidbüchel *et al.,* 2012; Hrachowitz *et al.,* 2013). This then allows us to understand the time-variance of stream water age in terms of the dynamics and integration of hydrological flow paths which source water of different ages to the stream (Rinaldo *et al.,* 2015).

*Correspondence to: Doerthe Tetzlaff, Northern Rivers Institute, School of Geosciences, University of Aberdeen, Aberdeen AB31 5XY, UK.
E-mail: d.tetzlaff@abdn.ac.uk

Within models, stable isotopes can be used to 'track' water fluxes, infer mixing relationships in internal stores and explore how the evolution of water ages occurs in relation to flow path dynamics (Sayama and McDonnell, 2009; McMillan *et al.,* 2012; Davies *et al.,* 2013; Hrachowitz *et al.,* 2013; Birkel *et al.,* 2015). Building on earlier work (e.g. Neal *et al.,* 1988; Barnes and Bonell, 1996), several conceptual models that link flow with tracers have been developed over recent years (e.g. Fenicia *et al.,* 2008; Birkel *et al.,* 2011b). Some of these models also include the simulation of water ages in a catchment that are spatially explicit (Birkel *et al.,* 2015). These models have been insightful with regards to how catchments function in terms of partitioning, storing and releasing water. However, such models have the tendency to be lumped and can give overly simplistic representations of complex, spatially distributed processes. Beyond using tracers to identify dominant flow paths, stable isotopes can be of added value in hydrological modelling, because they enable mixing relationships in internal stores to be characterised whilst the evolution of water ages is estimated (McMillan *et al.,* 2012; Davies *et al.,* 2013; Birkel *et al.,* 2015). Within such models, it is important to appropriately conceptualise the key spatial controls on the fluxes of water and tracers, such as how distributed storage dynamics drive non-linearities in connectivity, as well as the associated mixing relationships in order to estimate the age of stream flow. The non-linearities in connectivity refer to the non-linear increase in streamflow response to precipitation events because of the expansion of the saturated area, which connects runoff generating areas to the channel network, during wetter periods (Birkel *et al.,* 2010). Recent work suggested that more spatially distributed tracer-aided models could help capture these interactions more successfully and improve tracer simulations in particular (Birkel and Soulsby, 2015). Several spatially-distributed models including either water ages (e.g. Sayama and McDonnell, 2009) or tracer data (e.g. Uhlenbrook *et al.,* 2004) have been developed previously; however, they did not combine tracking water ages and simulating stable isotopes.

In previous model studies, relatively coarse (weekly) or short term (1–2 years) data sets have also usually been a constraint (Birkel *et al.,* 2010). Decreasing costs of stable isotopes analysis have increased the availability of high quality data sets that have a high spatial and temporal resolution. The data sets now available are usually of either daily or sub-daily resolution and extend over several years (Kirchner and Neal, 2013). This creates opportunities for more detailed modelling and increases the potential for using spatially-distributed models instead of lumped models.

Here, we develop such a spatially semi-distributed tracer-aided rainfall-runoff model for an experimental site in the Scottish Highlands with rich stable isotope data sets and empirically based perceptual models of catchment function. Previous tracer-based models have been developed for this catchment (e.g. Birkel *et al.,* 2011b), and existing models [e.g. Hydrologiska Byråns Vattenbalansavdelning (HBV), Seibert and Vis, 2012] have been used for the simulation of discharge (Tetzlaff *et al.,* 2015), but these showed limitations in capturing more subtle internal catchment processes which affect variations in the tracer response. With the model presented here, we improve the conceptualisation of these internal catchment processes by using a more spatially distributed approach. We have also changed the interception of precipitation in the model compared with previous model work to more accurately characterise the influence of vegetation on water partitioning (Tetzlaff *et al.,* 2015). At the same time, the aim was to keep the model relatively simple in order to derive a generic tool that can be applied across northern sites (Tetzlaff *et al.,* 2015). The main research aims of this study are therefore to: (1) develop a spatially explicit representation of water and tracer fluxes at the catchment scale using a conceptual spatially distributed, tracer-aided runoff model, (2) understand how landscape structure affects the routing and mixing of water, (3) characterise how these interactions affect the dynamics of water ages in different components of the system and how this affects the non-stationarity in stream water ages.

STUDY SITE AND DATA

Study site

The Bruntland Burn (3.2 km^2) is a tributary of the 30 km^2 Girnock experimental site in the northeast of Scotland (Figure 1). Detailed descriptions of the Bruntland Burn can be found in Tetzlaff *et al.* (2014) and Geris *et al.* (2015). Mean annual precipitation and mean annual discharge are ~1000 and ~600 mm, respectively. The catchment has been glaciated and has a wide flat valley bottom surrounded by steeper slopes (average 13°). The elevation ranges from 248 to 538 m asl. The catchment can be divided in distinct hydropedological units. In the riparian zone located in the valley bottom the main soil type is peat (~9% of the area), covering glacial drift deposits, which can be up to 30 m deep. The riparian zone includes a quasi-permanently saturated area around the stream that has a significant influence on the hydrology (Birkel *et al.,* 2011b). On the lower slopes, the soil type changes to a peaty gley (~12% of the area), whereas the steeper hill slopes are covered by freely draining podzols (~36%) that become thinner with higher elevation and grade to shallow regosols (leptosols, ~14%) and bedrock outcrops

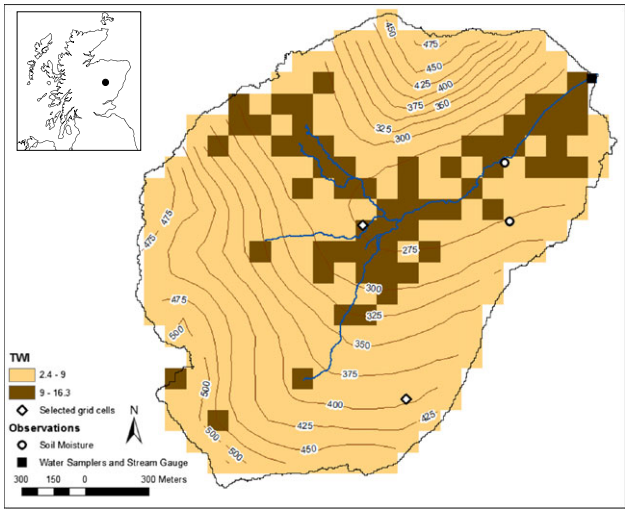

Figure 1. Location and topographical wetness index (TWI) for the Bruntland Burn catchment based on a 100 by 100 metres grid with an overview of the measurement locations and cells selected as examples of the response in the hill slope and valley bottom landscape units. Contour lines indicate the elevation (m). The two classes of the TWI are based on the threshold chosen to represent the difference between the freely draining soils at the hill slopes and the semi-permanently saturated soils in the valley bottom

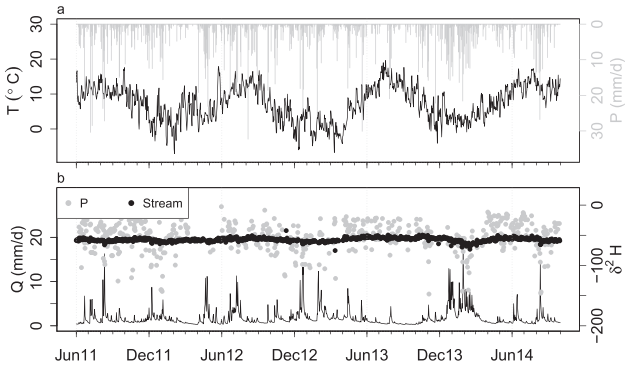

Figure 2. Observed data of the Bruntland Burn (June 2011 to October 2014), (a) Daily temperature and precipitation data, (b) Daily discharge data and isotope ratios (δ^2H) for stream water and precipitation

(~29%). The glacial drifts get thinner with elevation and are absent above 400 m asl. The dominant land cover is heather (*Calluna* and *Erica* species) moorland and around 10% of the catchment is forested, mainly with Scots Pine (*Pinus sylvestris*).

Previous work has shown that the hydrological regime of the stream has a flashy response to precipitation inputs, which is largely generated from saturation overland and/or shallow subsurface flow in the peaty soils in the riparian area (Birkel *et al.,* 2011b). In the largest runoff events (>10 mm d^{-1}) with large precipitation inputs, the steeper hill slopes also saturate due to wet antecedent conditions and connect to an extended saturated riparian area for transient periods lasting from a few days to a few weeks (Tetzlaff *et al.,* 2014). Deeper groundwater accounts for 25–35% of annual runoff. Geochemical and geophysical surveys in the catchment suggest that this is mainly sourced from the drift deposits in the valley bottom (Blumstock *et al.,* 2015; Soulsby *et al.*, 2016).

Data

Daily discharge from the outlet (Figure 2) was available from 1st June 2011 until 15th October 2014. Precipitation measurements were available from within the catchment from December 2013 till October 2014, but precipitation data from several neighbouring stations (max. 20 km away) were available for the whole period. The data from nearby stations were compared with the precipitation data from within the catchment. The station that corresponded best to these values (distance ± 5 km) was selected as input for the modelling. Potential Evapotranspiration (PET) was estimated with the Penman–Monteith method (Allen *et al.,* 1998) using observations from a nearby automatic weather station (distance ±1 km) for wind speed, humidity and net radiation. The annual estimated value for PET was ~420 mm. Temperature data were available from the same weather station. The available data included several events that were relatively extreme for this region and which are described in detail by Geris *et al.* (2015). The spring in 2013 (January to April) was the coldest for over 50 years, whereas the summer 2013 (June to September) was the warmest and driest period for over 10 years. This was then followed by the wettest winter period on record (December 2013–January 2014) (Geris *et al.,* 2015).

Daily samples for analysis of stable isotopes were collected from precipitation and the stream at the outlet using ISCO 3700 autosamplers. Paraffin was added to each bottle to prevent sample evaporation and freezing in the field. Soil water samples were taken on a weekly basis from the hill slope and in the valley bottom (Figure 1) from June 2011 until November 2013. Soil water was sampled from a 30 cm depth using small suction lysimeters (see Geris *et al.,* 2015 for details). All samples were analysed for deuterium (δ^2H) and oxygen-18 (δ^{18}O) ratios using a Los Gatos DLT-100 laser spectrometer (Los Gatos Research, San Jose, California, USA) (precision ±0.4‰ for δ^2H and ±0.1‰ for δ^{18}O). Analytical results were expressed as parts per mille relative to Vienna Standard Mean Ocean Water . Because of the greater relative precision, we used the δ^2H data in the modelling. Occasional gaps in the precipitation isotope ratios were infilled with a multiple linear regression model based on climatic variables (Soulsby *et al.,* 2015).

Additionally, spatial data of the catchment was needed for the modelling. A digital elevation map (DEM) and Light Detection and Ranging (LiDAR) data were available for the Bruntland Burn at 1 m resolution. Coarser scale land use, soil and geology maps were also available (Tetzlaff

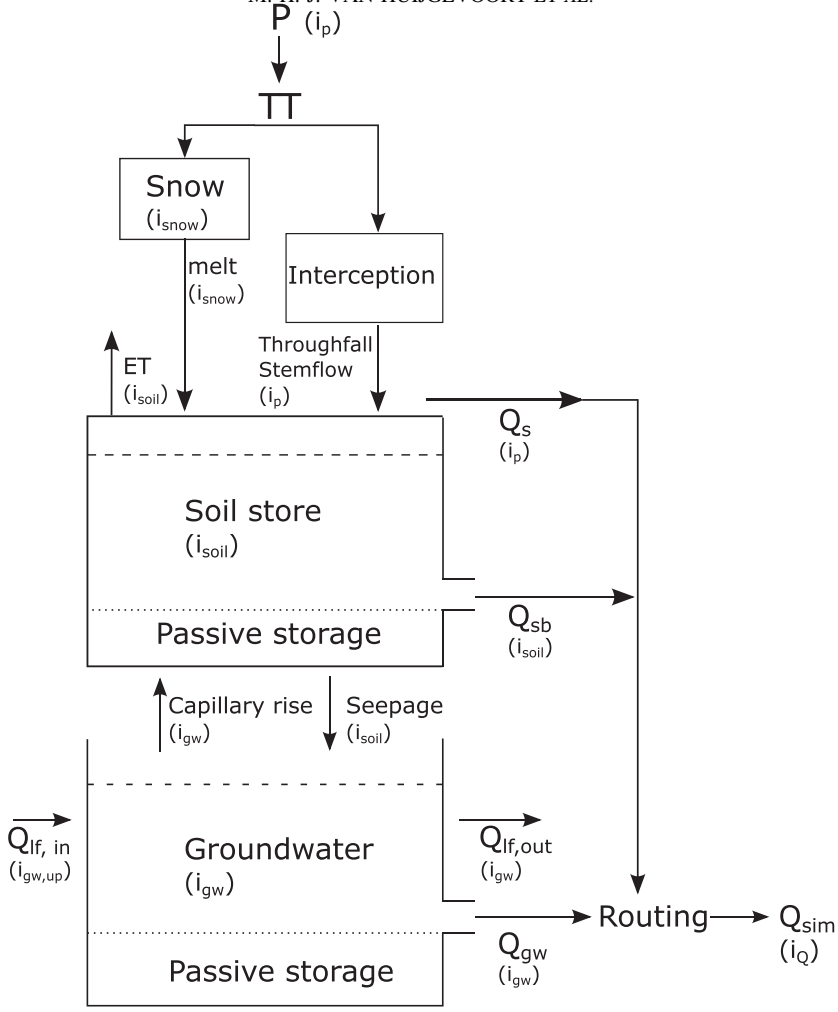

Figure 3. Schematic model structure of the STARR model, isotope ratios in the different stores and fluxes are indicated with i. This structure represents the conceptual model for each grid cell with P, precipitation; TT, threshold temperature to partition snow and rain; ET, evaporation; Q_s, surface runoff; Q_{sb}, subsurface runoff; Q_{lf}, lateral groundwater flow in and out of the grid cell; $i_{gw,up}$, isotope ratio in groundwater upstream from the cell; Q_{gw}, outflow from groundwater; and Q_{sim}, total discharge

et al., 2007). The DEM data were used to spatially distribute the temperature, PET and precipitation data, to estimate the local drainage direction for each cell and to determine the slope for each cell. The LiDAR data and spatial maps were used to determine the vegetation patterns and extent of the saturated area in the catchment.

MODEL DESCRIPTION

The Spatially distributed Tracer-Aided Rainfall-Runoff (STARR) mod18el integrated the general hydrological structure of the HBV-light runoff model (Lindström *et al.,* 1997; Seibert and Vis, 2012) and the parameterisation of tracer mixing and flux tracking developed as part of an earlier conceptual tracer-aided model for the catchment (Birkel *et al.,* 2011b). STARR was built in a PCRASTER PYTHON framework (Karssenberg *et al.,* 2010, 2007) consisting of several separate modules (Figure 3 and Appendix 1), which were conceptualised using a range of free and fixed parameters (Table I). Main balance equations are included in the model description section; a complete overview of the equations is given in Appendix 1. Input time series needed to drive the model were precipitation, temperature, potential evapotranspiration and isotopes ratios in precipitation. Observed discharge data and stream isotope ratios were used for dual calibration. To reduce computation time the model was run on a 100 by 100 m resolution grid (Figure 1) using a daily time step, although the framework could be used with a finer grid.

Model set-up

Snow module. The influence of snow in the Bruntland Burn catchment is usually limited (< 5% of annual precipitation). However, we incorporated a snow module to capture years when the snowfall contribution is higher

Table I. Overview of the parameters and their units in the STARR model

Parameter	Unit	Description	Initial range
Snow module			
TT**	°C	Threshold temperature snow and rain	0 (−1.5–2.5)
SFCF*	–	Snowfall correction factor	1
CFMAX**	mm day^{-1} °C	Degree day factor	5 (1–10)
CFR*	–	Refreezing coefficient	0.05
CWH*	–	Maximum snowmelt retained by snow pack	0.1
Interception module			
Canopy Gap Fraction*	–	Fraction of precipitation that does not hit canopy	0.3 (forest)$^{+}$ 0.6 (heather)$^{+}$
$\overline{R}$*	mm hr^{-1}	Mean rainfall ratec	1.38^{+}
$\overline{E}$*	mm hr^{-1}	Mean canopy evaporation rate	0.2 (forest)$^{++}$ 0.02 (heather)$^{++}$
Cmax*	mm	Canopy interception storage capacity	1.5 (forest)$^{\#}$ 0.7 (heather)$^{\#}$
Soil module			
FC	mm	Fieldcapacity, total water holding capacity of the soil	200–500 (valley) 0–100 (hill slope)
LP**	–	Fraction of fieldcapacity below which actual evaporation equals potential evaporation	0.8 (0.3–1)
BetaSeepage	–	Exponent to determine soil recharge into lower box	0.01–4
K_s	day^{-1}	Recession coefficient of discharge from soil store	0.001–0.5
Cflux*	mm day^{-1}	Maximum capillary flux	2 (valley) 0.5 (hill slope)
SMpas valley	mm	Passive soil store in valley bottom	0–300
Frac SMpas	–	Fraction passive soil store hill slopes	0–1
Groundwater module			
K_g	day^{-1}	Recession coefficient baseflow	0.001–0.1
K_{sat}	mm day^{-1}	Saturated conductivity to determine lateral flow	10^{-4}–200
GWpas	mm	Passive groundwater storage	0–1000

STARR, Spatially distributed Tracer-Aided Rainfall-Runoff.
*These parameters were fixed and not calibrated.
**Parameters fixed after initial STARR runs and sensitivity analysis, initial range given between brackets
$^{+}$ Values derived from observations
$^{++}$ Values derived using both observations and literature
$^{\#}$ Values derived from literature (Dunn and Mackay, 1995)

and to make the model more generic and transferrable to other northern snow-influenced catchments. The snow module was identical to that in the HBV model and used the degree-day method to divide precipitation into rain and snow (Table I).

Interception module. The interception module used the Gash interception model (Gash *et al.,* 1995). Parameters related to the interception module (Table I) were derived either from literature or from observed meteorological data. The catchment was divided in two major vegetation units based on LiDAR data to estimate the interception from vegetation higher than 1 m (forested parts) and lower than 1 m (heather dominated parts). Average canopy evapotranspiration for Scots pine and heather were calculated using the Penman–Montheith equation with values derived from literature (Dunn and Mackay, 1995) and observations from a nearby weather station (Section on Data). The values for canopy storage were derived from Dunn and Mackay (1995) as well. Mean rainfall rate was calculated from hourly observed precipitation data with a threshold rate of 0.5 mm/h to represent saturated canopy conditions (Gash, 1979).

Soil storage module. The soil storage module was based on that of HBV, except for an additional runoff component from the soil store (Q_{sb}, Figure 3). This was added to conceptualise the faster-responding flow paths in

the catchment, and because although an overland flow component (Q_s) was included, it rarely activates. The soil store was treated as a linear reservoir to determine this direct discharge (Q_{sb}). The change in storage in the soil store (SM) for each time step was given by

$$SM(t) = SM(t-1) + P_{eff} - ET - Q_s - Q_{sb} - Seepage + CapFlux$$

where P_{eff} is effective liquid precipitation (sum of throughfall, stemflow and snowmelt), *Seepage* is recharge to the groundwater store, and *CapFlux* is capillary flux from the groundwater store.

To distinguish between the valley bottom and the hill slopes, the FC and CFlux parameters related to the soil storage module (Table I) were divided over the catchment using the Topographic Wetness Index (TWI; Figure 1). The TWI was calculated using the System for Automated Geoscientific Analyses Wetness Index routine (Böhner and Selige, 2006). A threshold value of 9 was chosen to represent the difference between the freely draining soils at the hill slopes and the semi-permanently saturated soils in the valley bottom, based on previous field campaigns and model studies in the Bruntland Burn (Birkel *et al.,* 2011b). The TWI threshold gave a better representation of the difference between the hill slopes and the valley bottom at 100 m resolution than the soil map.

A passive storage component (SMpas) was added to the soil store to increase the mixing volume available in the catchment. This passive storage represents storage that did not contribute directly to the discharge, but increased the total mixing volume for isotope ratios (Birkel *et al.,* 2015). The passive storage in the soil store was distributed between the valley bottom (SMpas valley) and hill slope (fracSMpas*SMpas valley) using the TWI threshold, where the passive storage at the hill slopes was calibrated as a fraction of the passive storage in the valley bottom.

Groundwater storage module. The groundwater module represented the slower pathways in the catchment. The groundwater outflow (Q_{gw}, Figure 3) was derived from a linear reservoir. In addition to the outflow to the stream, lateral flow between the grid cells was simulated as well. This was calculated using the hill slope gradient and a calibrated saturated conductivity parameter (K_{sat}). A passive storage component (GWpas) was added to the groundwater to increase the mixing volume available in the catchment. The change in groundwater storage (*GW*) for each time step was given by

$$GW(t) = GW(t-1) + Seepage - Q_{GW} + \Delta Q_{lf} - CapFlux$$

where ΔQ_{lf} is net lateral groundwater flow.

Routing module. The outflow fluxes from the different modules, i.e. direct runoff (Q_s), subsurface runoff (Q_{sb}) and groundwater runoff (Q_{gw}) for each cell, were routed through the catchment based on the local drainage direction map derived from the DEM and a mean velocity value based on observed velocities at the catchment outlet. From the mean velocity, a hydraulic travel time was computed; this indicated the time it took the water to cross each grid cell.

Simulation of isotope ratios and estimation of water ages

The model simulated the isotope ratios for every cell separately; the daily precipitation isotope ratios were used as the input. Because the precipitation isotope ratios were only available at the outlet, the input values were similar for each grid cell. Previous work in the catchment established that the relatively small area and limited altitudinal range resulted in insignificant spatial variability in precipitation isotope inputs (Birkel *et al.,* 2011b). For each of the model stores, a complete and instantaneous mixing of the isotope ratios was assumed within each cell. For example, for the isotope ratios in the soil store (i_s), this led to the following mixing equation:

$$\frac{di_s(SM + SMpas)}{dt} = i_p P_{eff} - i_s Q_{sb} - i_s Q_s - i_s ET - i_s Seepage + i_{gw} CapFlux$$

where i_p is precipitation isotope ratio and i_{gw} is isotope ratio in groundwater store.

The values in the outflow fluxes for each store were then also mixed during routing, as were the values routed from upstream cells. Although the model assumed complete mixing at the scale of a single cell, the coupling of water and solute transport resulted in time-variant partial mixing at the catchment scale depending on spatial and temporal variation in the dynamics of fluxes and storage at this larger scale. Although it was recognised that the 100 m by 100 m cell size is relatively large, more detailed assessment of the spatial extent of mixing zones in the catchment showed that it was an appropriate initial scale for capturing much of the variation empirically observed (Lessels *et al.,* 2016). Fractionation was not parameterised in the model, although previous work has shown that this can affect summer stream flow isotope ratios (Birkel *et al.,* 2011b).

To further quantify how catchment functioning affected the partitioning, storage and mixing of water, the estimation of water ages was included in the model. For the water ages in the soil and groundwater stores, again a complete, instantaneous mixing of the inputs was assumed according to the dynamic and passive storage volumes, and the ages of the water stores were tracked on

a daily time step. For the age of the evaporated water from the soil, it was assumed that the youngest water would evaporate first and that the water in the soil was never older than 2000 days. This assumption was based on the previous estimates of the transit times in the different hydropedological units reported in Tetzlaff *et al.* (2014). For example, the water age in the soil (Age_{sm}) was determined by

$$\frac{dAge_{SM}(SM + SMpas)}{dt} = Age_P P_{eff} - Age_{SM} Q_{sb} - 1^* Q_s - Age_E ET - Age_{SM} Seepage + Age_{GW} CapFlux$$

where Age_P is age of precipitation, Age_E is age of evaporated water and Age_{gw} is age of groundwater store.

Model calibration

The STARR model was calibrated using Monte Carlo simulations for the period 1/4/12 until 31/3/14. In total, nine parameters were calibrated; some parameter values were based on observations or literature values, and other parameters were fixed based on previous model experience and sensitivity analysis of previous model runs (Table I). After preliminary testing, STARR was calibrated for the 9 parameters of which the initial range in the Monte Carlo simulations is given in Table II. From this initial range, the 100 best runs from a total of 30 000 were retained (Table II). These were selected based on a dual calibration criteria which took into account both the isotope ratios and discharge at the outlet of the catchment. Both the isotope ratios (NS_{iso}) and the discharge data (NS_Q) were compared using the Nash–Sutcliffe (NS) efficiency statistic (Nash and Sutcliffe, 1970). These criteria were combined in a single objective function (OF),

$$OF = \sqrt{(1 - NS_Q)^2 + (1 - NS_{iso})^2}$$

For the evaluation of the model, several other goodness-of-fit measures were calculated over the entire time series, the Kling–Gupta efficiency (KGE; Gupta *et al.,* 2009) for both the discharge and isotope ratios, the logarithmic Nash–Sutcliffe efficiency ($logNS_Q$) for the discharge and the correlation coefficient (R_{iso}) for isotope ratios. This way, the effect of the chosen dual calibration criteria could be estimated by comparing the of values with values of the single goodness of fit measures over a range of conditions.

The initial conditions were determined by looping the input data for the best performing parameter set derived from a preliminary model evaluation. After the selection of the 100 best runs, the input data were looped twice for each run for the whole period to make sure that the stores were filled and the isotopic composition stable.

RESULTS

Simulated discharge and tracer dynamics

In general, the retained parameters gave a reasonable performance in the simulation of the discharge and isotope ratios (Figure 4). The NS efficiencies for both variables were generally in the range of ~0.4–0.6 for the calibration period. For the 'best run' – according to the minimised dual objective function – the NS_Q and NS_{iso} were 0.52 and 0.56, respectively, reflecting the trade off in the goodness of fit in flow simulations needed to get reasonable isotope simulations. Over the whole period, NS_Q decreased slightly for the best simulations to 0.5, although other objective functions for the retained models remained high with a KGE_Q of ~0.6 and the $logNS_Q$ had values around 0.5. The NS_{iso} also decreased for the whole

Table II. Resulting range in parameter values after calibration for the 100 selected runs

Parameters	FC valley	FC hill slope	Beta Seepage	K_s	K_g	K_{sat}	GWpas	SMpas valley	Frac SMpas
	(mm)	(mm)	(-)	(day^{-1})	(day^{-1})	($mm\,day^{-1}$)	(mm)	(mm)	(-)
Initial range	200–500	0–100	0.01–4	0.001–0.5	0.001–0.1	10^{-4}-200	0–1000	0–300	0–1
Values best run based on calibration period	232	46	0.16	0.1	0.016	176	603	295	0.96
Min values	201	34	0.01	0.012	0.0011	0.05	33	154	0.37
Max values	499	100	3.22	0.23	0.0997	198	996	300	1

FC, Fieldcapacity, total water holding capacity of the soil; K_s, recession coefficient of discharge from soil store; K_g, recession coefficient baseflow; GWpas, passive groundwater storage; SMpas, passive soil store in valley bottom; Frac SMpas, fraction passive soil store hill slopes.

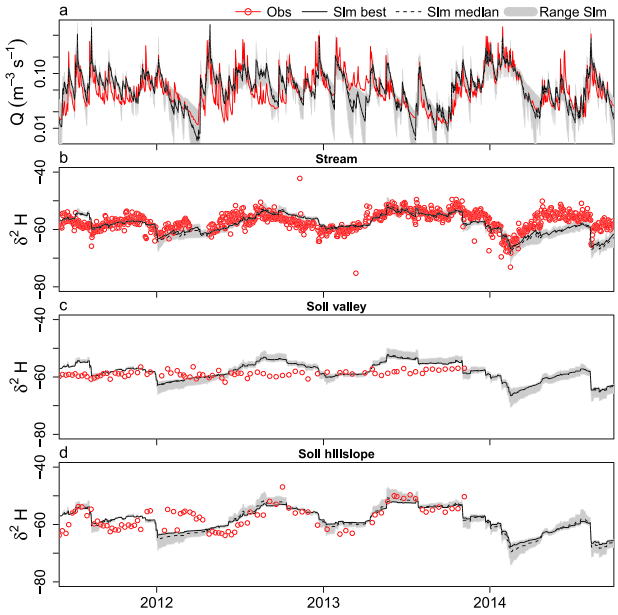

Figure 4. Model results for the Bruntland Burn, solid lines for the simulations give the results of the best run, dotted lines for the simulations give the median from the 100 runs, and the grey areas indicate the range for the 100 selected runs, (a) simulated and observed discharge data for the outlet of the Bruntland Burn, (b) simulated and measured isotope ratios in the stream at the outlet of the Bruntland Burn, (c) simulated isotope ratios for the soil store from a cell in the valley bottom corresponding to the measurement location and measured isotope ratios from soil water sampled in the valley bottom and (d) simulated isotope ratios for the soil store from a cell on the hill slope corresponding to the measurement location and measured isotope ratios from soil water sampled on the hill slope

period, although the KGE_{iso} for the best models varied between 0.42 and 0.64 and R_{iso} values ranged between 0.52 and 0.65 indicating a reasonable fit.

The simulated discharge captured the main dynamics of stream flow and generally; the uncertainty bounds from the retained parameter sets bracketed the range of measured discharge (Figure 4a). There was a tendency for the best simulations to underestimate the low flows that occurred, for example, in spring 2012 and spring 2013. Conversely, small peaks sometimes tended to be overestimated, for example, autumn 2012. However, the more extreme events that occurred during the simulated period, the dry period in summer 2013 followed by a very wet winter in 2014, were quite well captured by the model. In general, the observed response in the catchment seemed to be more dynamic than the response simulated by the best performing parameter set and the median of the ranges of the retained sets.

The simulated isotope ratios in the stream (Figure 4b) successfully captured the large damping of the signal observed in precipitation (Figure 2b). Similarly, the seasonal cycles in the isotope ratios were well reproduced by the model, as well as the range and dynamics in daily fluctuations although more extreme outliers (e.g. winter 2013) were not captured (Figure 4b). The model underestimated the isotope ratios in spring and summer 2014. This underestimation occurred in periods where fractionation of water stored in the riparian peatlands can cause enrichment of isotopes in stream water (Blumstock *et al.,* 2015). The very wet winter in 2014 was followed by a warm, dry spring; therefore, more fractionation could take place due to the very wet conditions in the riparian areas (Lessels *et al.,* 2016). The model also underestimated the isotope ratios after a single extreme event in August 2014, where observed ratios were restored to values similar to the pre-event conditions faster than the modelled ratios.

To provide a qualitative evaluation of the internal consistency of the model, the simulated isotope ratios in the soil stores were compared with the measured isotope ratios in the soil water (Section on Data) at different locations. Within the model a distinction was made between the soils in the saturated valley bottom and the hill slopes. The isotope ratios simulated at the locations of the observations showed the difference in dynamics between these landscape units (Figure 4c–d). Simulated values had a comparable mean and range with the observed isotope ratios, where the isotope signal in the valley was more damped than the signal on the hill slope. However, the simulations in the valley were more variable than the measured isotope ratios; this would suggest even more mixing occurs in the riparian area than estimated by the model or that the sample point was missing less well-mixed overland flow waters. The model underestimated the isotope ratios on the hill slope in the winter 2011-2012, but followed the observations before and after that period.

Influence of landscape structure on routing and mixing of water

The spatial distribution of water in the dynamic storage across the catchment was determined for the soil store and the groundwater store (Figures 5 and 6). Two randomly selected cells in the model were used to represent the valley and the hill slope (see their location in Figure 1). As an example of the spatial distribution, results from the 'best' model run (in terms of the OF) are given for a day in the dry period (31/8/13) and the wettest period (1/2/14). The mean values for the stores differed substantially between the 100 runs (Table III), but the spatial patterns and temporal dynamics were similar.

As dictated by the model set-up and calibration, the saturated area is parameterized differently than the hill slopes in the soil store, and the resulting dynamic storage showed large differences between these landscape units (Figures 5 and 6). The valley bottom cells had higher values in the soil store on both days reflecting the higher

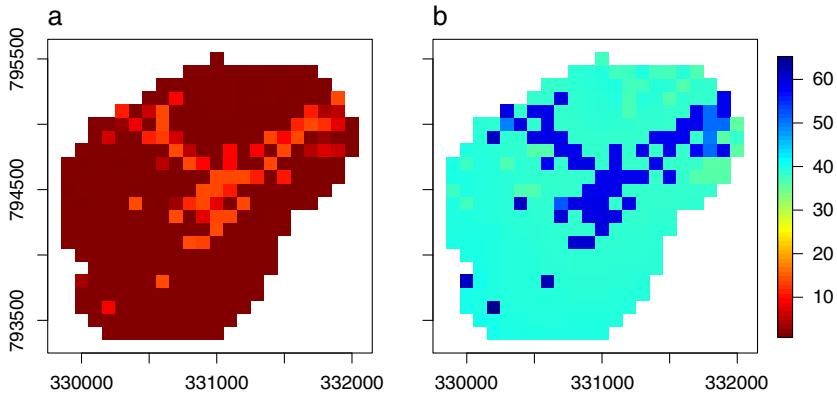

Figure 5. Spatial distribution of the simulated dynamic storage (mm) in the soil store derived from the best model run, (a) simulated storage during a dry period (31/8/13) and (b) simulated storage during a wet period (1/2/14)

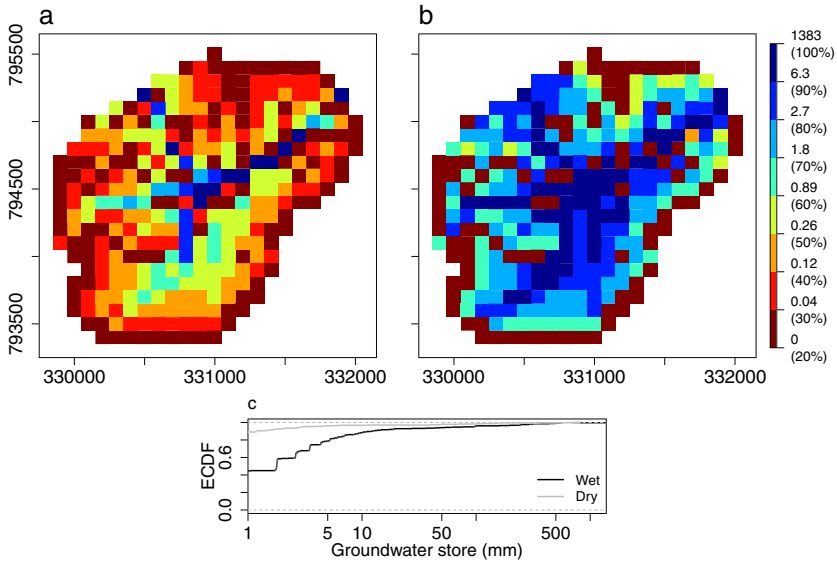

Figure 6. Simulated dynamic storage (mm) in the groundwater store derived from the best model run, (a) spatial distribution of the simulated storage during a dry period (31/8/13), the scale is based on the percentile values of the simulated storage values, (b) spatial distribution of the simulated storage during a wet period (1/2/14) and (c) cumulative distribution of the storage values over the catchment for 1 day in a wet and dry period. ECDF, empirical cumulative distribution function

Table III. Ranges in mean storage values over the 100 retained model runs for the dynamic storage in soil store and the dynamic storage in groundwater store

	Soil valley	Soil hill	GW valley	GW hill
Mean storage values (best run) (mm)	27	13	1022	1
Standard deviation (mm)	25	10	1461	72
Coefficient of variation (-)	62	53	186	885

GW, ground water.

calibrated FC parameter. The dynamic storage values in the soil store showed mean values over the 100 runs up to 137 and 51 mm in the valley and the hill slopes, respectively. On both selected days, the area with higher values is broadly consistent with the maximum saturation area in the valley bottom (Figure 5). But whilst maximum

storage is around 13 mm on the driest day, it is 65 mm on the wettest. Similarly, the hill slope storage is only 1–2 mm on the example dry day, but increases to 40 mm on the wettest. The effects of interception losses from forest cover were captured in some of the most northern and most eastern cells, which had lower storage values than the other valley bottom cells. The distribution of storage indicated that the main mixing zones for isotopes ratios were the valley bottom riparian soils rather than the hill slopes. These patterns corresponded with the time series from the soil isotope ratios (Figure 4c-d), where the signal in the valley was more damped than the signal on the hill slope.

In the ground water store, the dynamic storage values were also much higher in the valley bottom than on the hill slopes, with the mean values for the best run at 1022 and 1 mm, respectively (Table III). It should be emphasised that there is also an additional passive storage component involved in the tracer damping in each cell; this was 603 mm in the best run. The large differences in storage values were caused by the lateral flow of groundwater, which accumulated in certain cells in the valley where flow lines converged (Figure 6). The scaling in the plots is based on percentiles to emphasise the much higher values of dynamic storage in the valley bottom cells, with only 10% of the whole catchment exceeding 1.2 and 12.1 mm on the example dry and wet days, respectively. On both selected days, there were parts of the catchment, mainly the upper hill slope cells, where dynamic storage in groundwater was 0 mm; this ranged from 30% of the catchment during dry conditions to 20% during wetter conditions (Figure 6). The spatial patterns of dynamic storage were very similar on both days with upper hill slopes with no water and large storage values in the valley bottom (Figure 6a-b). As would be expected, the storage values on the wetter day were higher than on the dry day in the lower hill slope cells and the valley bottom. The cumulative probability distribution of the dynamic storage values are shown in Figure 6c, the probability of low storage values is much higher on the dry day.

Evolution of water ages in different components of the system and non-stationarity in stream water ages

The simulated water ages in the model were closely linked to the simulated storage values in the catchment. The cells with the higher storage values contained older water than cells with less storage. The age of the water therefore increased in fluxes from different components of the system from the evaporation < soil < groundwater (Figure 7). The median water age was taken from the time series of the 100 retained model runs in the selected cells or at the outlet (Age Q). The evaporation was assumed to always consist of the youngest water that entered the soil.

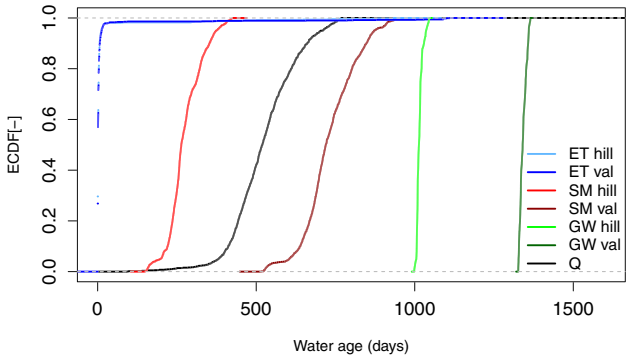

Figure 7. Cumulative distribution functions for the median of the simulated water ages from the best 100 runs for the different landscape units in the catchment (val, valley bottom landscape unit; hill, hill slope landscape unit). ECDF, empirical cumulative distribution function; ET, evaporation; SM, water in the soil store; GW, ground water

If no young water was available (e.g. no precipitation events in preceding days), the evaporation was sourced from the older water in the soil store or from capillary rise from the groundwater. This resulted in very young evaporation water for most of the period, but for a few days evaporation was very old because of this capillary rise (see difference mean and median age for evaporation in Table IV). The average age of hill slope soil water was generally younger (0.85 years) than that in the valley bottom (~2 years), and both were younger than hill slope groundwater (2.9 years) in most model realisations. The groundwater in the valley bottom was estimated as the oldest water in the catchment with a mean age of ~3.4 years for the best run (Table IV). However, the uncertainty in these estimates was large, with the largest coefficient of variation in the evaporation ages (Table IV). The water in the stream reflected the integration of runoff derived from both the groundwater and the soil stores, as the average stream water age at ~1.6 years was intermediate between that of groundwater and hill slope soil water (Table IV and Figure 7). A visualisation of the spatio-temporal dynamics of runoff generation from each cell and the associated water ages is discussed by van Huijgevoort *et al.* (in review).

In addition to the mean ages and age distributions, the variations in the water age in all components of the system showed marked contrasts (Figure 8). The ages in the soil store were dynamic in time and changed most rapidly in the hill slopes in response to most precipitation inputs, reflecting the lower storage. In contrast, the age of soil water in the valley bottom was older and exhibited a more moderated response given the larger storage. The age of the water in the groundwater store reacted much more slowly; the age was almost constant over the whole period in all model realisations. In the hill slopes where the dynamic storage is low (Table III), this was mainly caused by the constant value of the additional mixing

Table IV. Ranges in the simulated water ages over the 100 retained model runs for different landscape units

	Age ET valley	Age ET hill slope	Age SM valley	Age SM hill slope	Age GW valley	Age GW hill slope	Age Q
Mean water age (best run) (days)	24	25	727	311	1237	1056	578
Median water age (best run) (days)	1	1	718	302	1232	1056	551
Standard deviation (days)	13	27	294	79	667	503	183
Coefficient of variation (-)	64	101	40	29	49	48	35

GW, ground water; ET, evaporation;SM; water in the soil store.

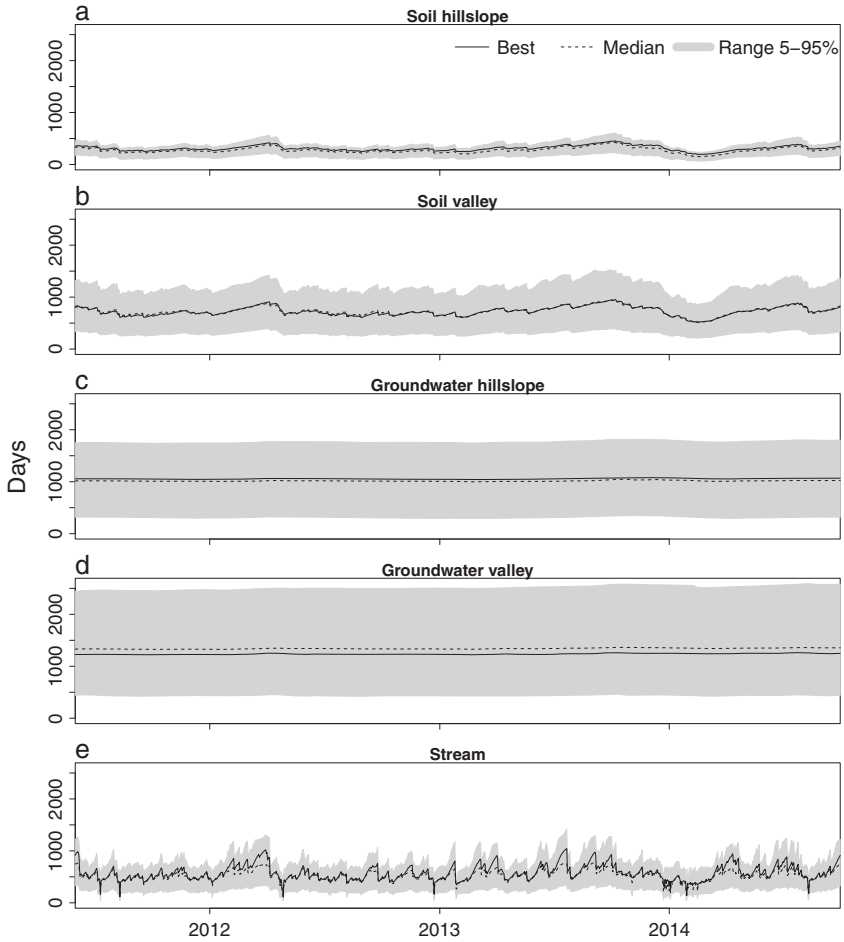

Figure 8. Simulated water ages within the catchment, solid line gives the ages derived from the best model run, dotted line gives the median of the 100 best runs, grey areas indicate the range between the 5th and 95th percentile of the 100 runs. (a) Simulated water ages for the soil store for a cell on the hill slope, (b) Simulated water ages for the soil store for a cell in the valley, (c) Simulated water ages for the groundwater store for a cell on the hill slope, (d) Simulated water ages for the groundwater store for a cell in the valley, (e) Simulated stream water ages for the discharge at the outlet of the catchment

volume, which provided a more stable age. The stream again integrated the differences in the contributing stores, with a higher average age than the hill slope soil store, but a more dynamic time series than the stable groundwater store.

Focusing on the direct relationship between water age, discharge and total storage (Figure 9), the role of storage-driven connectivity in the catchment and the non-stationarity of the water age could be investigated in more detail. The dynamics of the stream water age at the

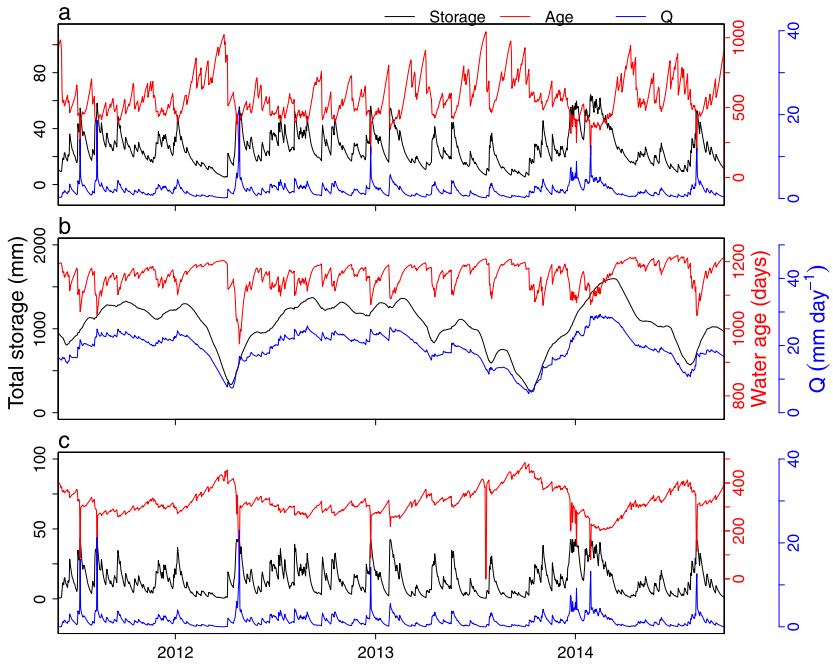

Figure 9. Relation between total dynamic storage, water age and discharge from best model run for (a) catchment mean total dynamic storage and simulated discharge and stream water age at the outlet, (b) total dynamic storage, total runoff and water age of runoff for 1 cell in the valley, and (c) total dynamic storage, total runoff and water age of runoff for 1 cell on the hill slope

outlet were related to the peak events in discharge, with youngest waters generally dominant during the largest events (Figure 9a). During wet periods and peak events, the discharge and mean total storage in the catchment increased and the water in the stream became younger. The uncertainty bound of the youngest stream water age in large events ranged from ~0.1 to 1.2 years, with a median of about 0.25 years (Figure 8). Because the age of the groundwater showed limited variability (~2.9–3.4 years), this reflected the increased dominance of contributions from faster reacting flow pathways in the hill slope soil zones and their much younger characteristics (~0.5 years) and surface runoff (Figure 9c). The groundwater influence increased again during dry periods, when the stream water became older. During the dry periods, the hill slope cell was not connected for a few days in summer 2013, indicated by the water age of zero days (Figure 9c). The other very young water ages that occurred in the soil hill slope indicate surface runoff.

The spatial distribution of the water age of total runoff from each cell for a day in a dry period and a day in a wetter period also showed younger water during the wet period across the catchment (Figure 10). In this realisation, the mean runoff age across the catchment on the dry day was 540 days, whereas for the day in the wetter period the mean age decreased to 306 days. Furthermore, the spatial distribution of the water ages indicated again that, driven by the available storage, the oldest water was found in the valley bottom and the youngest water on the hill slopes on both days. The influence of the increased flux of younger hill slope water under wet conditions in decreasing water ages in the valley bottom became apparent though (Figure 10).

DISCUSSION

Performance of a distributed, tracer-aided runoff model

Many hydrological models only focus on the simulation of the hydrograph; however, this often gives little information about the actual processes within a catchment, as different model structures and parameter sets can lead to the same result; the widely known equifinality problem (e.g. Beven, 1993). Besides equifinality, many studies have also described the importance of differentiating between velocity and celerity, or between long residence times of water particles and the short-term response of the stream hydrograph (e.g. McDonnell and Beven, 2014). According to McDonnell and Beven (2014), 'velocity indicates the mass flux of the water itself, whereas celerity relates to the speed with which a perturbation to the flow propagates'. By incorporating tracer data into models, the velocity or transit time can be estimated, while the hydrograph response is also simulated. Tracers can have an important role in constraining models; thus, tracer data have been increasingly integrated into models (e.g. Fenicia *et al.,* 2008; Birkel *et al.,* 2011b; Capell *et al.,*

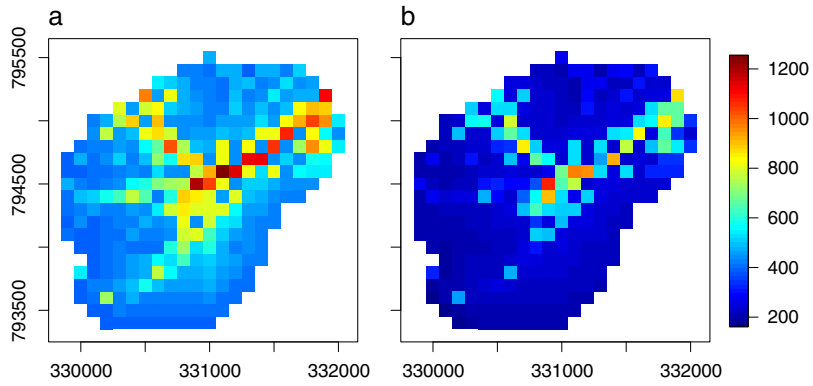

Figure 10. Spatial distribution of water age (days) of total runoff from each cell derived from the best model run, (a) Simulated runoff ages (days) during a dry period (31/8/13), (b) Simulated runoff ages (days) during a wet period (1/2/14)

2012). The downside is that models incorporating tracer data become more complex, and additional parameters are needed to estimate the mixing volumes (McDonnell and Beven, 2014). The STARR model presented here showed that with three additional parameters (GWpas, SMpas valley and fracSMpas); there was substantial information gain from the simulation of isotope ratios on the spatial distribution of mixing and partitioning of water. The model improved the simulation of the small scale dynamics in the isotope ratios compared with previous more lumped modelling approaches (Soulsby *et al.,* 2015), but this was a trade off in terms of higher uncertainty because of the increase in parameters. The simulated isotope ratios in the retained models provided realisations of the possible distribution of stores in the system and their temporal dynamics and connectivity. In turn, this identified the areas most important for mixing, in this case the riparian soils and drift in the valley bottom. By tracking water ages and simulating isotopes, STARR also captured the non-linearities in connectivity driven by distributed storage dynamics successfully without being highly parameterised.

Using tracer data in the hydrological model allowed us to use an objective function for dual calibration criteria based on discharge and isotope ratios. This has significant advantages over using single-objective optimalisation methods, which all have limitations (Wagener *et al.,* 2003). As indicated by McDonnell and Beven (2014), an acceptable model performance for both tracers and the hydrograph can help identify a model giving reasonable simulations for the right reasons. However, as shown in this study, using the dual calibration can lead to compromises on the performance of the model for the separate variables. The assessment of model performance is also highly dependent on the specific choice of the objective function used. In this study, we chose the NS efficiency for discharge and isotope ratios. The NS efficiency puts more emphasis on peak events, therefore identification of parameters linked to slow pathways is more difficult (e.g. Dunn, 1999; Wagener *et al.,* 2003). Different objective functions that focus less on peak events for discharge and isotope ratios, like the KGE (focus on variability in streamflow) and logNS (focus on low flows), were tested in the dual calibration. The combination of the NS efficiency for both variables gave the best overall calibration result for this specific catchment. In particular, it provided a better reproduction of event-based responses in terms of rapid depressions or peaks in δ^2H which corresponded to the flashy nature of the hydrograph.

The STARR modelling framework was developed as a generic approach and kept as simple as possible to facilitate transferability to other catchments with similar landscape units and hydroclimates where both hydrometric and tracer data are available. Time series of daily meteorological variables and daily precipitation isotope ratios are still needed as input data. This is one of the limitations of the model, because the opportunities for application in ungauged basins are constrained (Tetzlaff *et al.,* 2013). However, the continued progress in the development of analytical methods for isotopes will provide future opportunities for measuring isotopes at lower costs (Birkel and Soulsby, 2015), and identification of minimal requirements for stream isotope time series needed to usefully characterise the system response (Seibert and Beven, 2009; Hrachowitz *et al.,* 2011). Thus, exploring the potential to use such minimal tracer data with models may be a useful future framework for sparsely monitored and ungauged catchments (Tetzlaff *et al.,* 2013).

Influence of landscape structure on storage, routing and mixing of water

The model results showed much higher storage volumes in the valley bottom cells compared with those

on the hill slopes and more damped signals in isotope ratios as a result. This indicated that mixing of the water mainly occurred in the saturated area in the valley bottom. These findings are in agreement with previous model and field-related studies in the Bruntland Burn catchment that have indicated the quasi-permanently saturated riparian area plays a key role in determining the hydrological response and mixing processes in the catchment (Birkel *et al.,* 2011b; Tetzlaff *et al.,* 2014). The importance of the riparian zone in controlling runoff and chemistry is not catchment specific, but has been found in many other studies (Bishop *et al.,* 2004; Burt, 2005; Allan *et al.,* 2008; Seibert *et al.,* 2009). It is therefore important to account for the riparian zone in modelling. Our semi-distributed model makes it possible to map out the riparian zone in detail and assign different hydrological properties to that area to assess the critical role in a more quantitative way.

The higher storage values found in the valley bottom are also in agreement with field studies using geophysical measurements. These measurements showed that the depth of the drift deposits in the valley can be more than 30 m and equate to a catchment scale storage of >2000 mm (Soulsby *et al.*, 2016). The values for dynamic groundwater storage in the valley bottom found in this study ranged up to 8.7 m, with a spatial pattern that was consistent with independent geophysical measurements. The dynamic storage values for the catchment estimated by previous lumped model studies (e.g. Birkel *et al.,* 2011a) were lower than the storage values from the best run estimated here. It is however difficult to compare these storage estimates directly because of the different model conceptualizations.

Evolution of water ages in different components of the system and non-stationarity in stream water ages

The estimated mean stream water age from the best 100 parameter sets was 1.4 years and ranged between 0.5 and 2.4 years. Using a two parameter gamma distribution, Tetzlaff *et al.* (2014) estimated the mean transit time for the catchment at 2.8 years, but this was based on only 1 year of data. Other studies using the gamma distribution in the same catchment found values between 1.9 (Hrachowitz *et al.,* 2010) and 1.2 years (Soulsby *et al.,* 2015). Soulsby *et al.* (2015) also used a lumped tracer-aided conceptual model to track water fluxes and estimate their ages similar to model presented here. They found an average stream water age of 1.8 years using the same data set, varying between around 0.3 years at high flows and >3 years at low flows. These values corresponded well with the dynamics found in the stream water age in this study. The range of the estimated age of the groundwater in the valley (0.9–12 years) in the STARR model is large, but the mean age of the best runs (3.4 years) is comparable with the age found in the previous lumped modelling study (~4 years, Soulsby *et al.,* 2015).

Water ages in the model have been estimated by tracking the input and output of water following the approach of Hrachowitz *et al.* (2013). Other approaches to estimate water ages are the approach by Porporato and Calabrese (2015) and the use of storage age selection functions (SAS, Botter *et al.,* 2010; Harman, 2015; Rinaldo *et al.,* 2015). The complete mixing assumption in the presented model for the outflow from the soil store and the groundwater in individual cells coincides with specific cases of both these approaches. A loss function was included by Calabrese and Porporato (2015) with a specific parameter that can be varied to define the selection type (in this case uniform selection). The SAS functions are equal to the 'uniform selection' assumption, which has clear limitations in terms of oversimplifying the complexity of actual processes (Harman, 2015). Although complete mixing is not realistic usually, the use of two different reservoirs that are spatially distributed means that at the catchment scale partial mixing is achieved leading to non-stationary estimates of water ages. Compared with these approaches, the proposed model framework gives more insight into the spatial distribution of the evolution of water ages across the catchment. This comes at the expense of more parameters, but provides a basis for testing against empirical observations and gives reasonable results (Lessels *et al.,* 2016).

The stream water ages found in this study showed a decrease in water age with an increase in total storage but decrease in the storage involved in mixing the tracer signal. This is similar to the 'inverse storage effect' identified by Harman (2015) which was also shown by Birkel *et al.* (2015) and Soulsby *et al.* (2015). During wet periods or peak events, incoming water (and tracers) does not completely mix with groundwater, but partly bypasses this on a faster pathway through and over the soil leading to younger water in the stream and reducing the volumes of storage involved in damping the tracer signal. During dry periods, the influence from groundwater is larger, which increases the stream water age. This is consisted with results from previous studies in this catchment and a wider range of catchments in the Scottish Highlands (Hrachowitz *et al.,* 2013; Soulsby *et al.,* 2015).

CONCLUSIONS

This study presented the development and application of the conceptual STARR model for an experimental catchment in the Scottish Highlands. STARR simulates the hydrological

storages and fluxes, isotope ratios and water ages of different landscape components. The distributed nature of the model made it possible to explore the connectivity-driven relationships between different landscape units (e.g. hill slopes and valley bottom riparian zones) and their importance in the mixing and partitioning of water. Compared with previous lumped conceptual models used in the same catchment, more detailed, spatially nuanced descriptions of water partitioning and an improved simulation of stream water isotope dynamics were achieved.

For the selected catchment, the model performed reasonably for both discharge simulations and the isotope ratios. Given the dual calibration on both variables, some compromises had to be made regarding the performance of the model for the separate variables. The spatial distribution of the hydrological stores showed a clear distinction between the hill slopes and the valley bottom with much higher storage values in the latter. These different storage values also created a distinctive pattern in the isotope mixing where the isotope ratios were damped in the valley compared with the hill slopes, which was broadly consistent with observed soil isotope ratios. The simulated water ages were linked to the simulated storage values, and the oldest water was found in the groundwater stores in the valley bottom. The stream water ages reflected the time variant integration of the groundwater and soil stores. The influence of the groundwater led to mean stream water ages older than the water in the soil store, but the influence of the water from the hill slope soil store led to a non-stationary stream water age with a larger young water component during wet periods. This indicated fast lateral flow paths during wet events that bypassed the complete mixing within the groundwater.

The capability of the model to track storage dynamics and simulate hydrological fluxes, isotope ratios and water ages led to a more comprehensive description of the partitioning within the catchment. In that way, STARR can be a possible way forward for investigating the relationships within catchments between the different stores and the role of landscape structure in more detail. The effect of land use changes, for example reforestation, on the partitioning of the water can also be explored with this model. The model was developed as generic and as simple as possible to be applicable to other catchments. One of the next steps will be testing the model for inter-site comparison in other northern catchments (Tetzlaff *et al.,* 2015). Another next step will be the development of a more sophisticated calibration strategy. STARR has substantially improved the representation of the spatial and temporal dynamics of dominant hydrological processes and gives a more robust framework for projecting the effects of environmental change.

ACKNOWLEDGEMENTS

The authors would like to thank Jonathan Dick, Josie Geris, Jason Lessels and Claire Tunaley for data collection and Audrey Innes for lab sample preparation. We also thank Christian Birkel for discussions about the model structure and comments on an earlier draft of the paper. Climatic data were provided by Iain Malcolm and Marine Scotland Fisheries at the Freshwater Lab, Pitlochry. Additional precipitation data were provided by the UK Meteorological Office and the British Atmospheric Data Centre (BADC). We thank the European Research Council ERC (project GA 335910 VEWA) for funding the VeWa project.

REFERENCES

Allan CJ, Vidon P, Lowrance R. 2008. Frontiers in riparian zone research in the 21st century. *Hydrological Processes* **22**: 3221–3222. DOI:10.1002/hyp.7086

Allen, RG, Pereira, LS, Raes, D, 1998. Crop evapotranspiration : guidelines for computing crop water requirements, FAO irrigation and drainage papers 56. FAO, Rome.

Barnes CJ, Bonell M. 1996. Application of unit hydrograph techniques to solute transport in catchments. *Hydrological Processes* **10**: 793–802. DOI:10.1002/(SICI)1099-1085(199606)10:6<793::AID-HYP372>3.0.CO;2-K

Beven K. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* **16**: 41–51. DOI:10.1016/0309-1708(93)90028-E

Birkel C, Soulsby C. 2015. Advancing tracer-aided rainfall-runoff modelling: a review of progress, problems and unrealised potential. *Hydrological Processes* DOI:10.1002/hyp.10594

Birkel C, Soulsby C, Tetzlaff D. 2015. Conceptual modelling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls nonstationary water age estimates. *Hydrological Processes* **29**: 2956–2969. DOI:10.1002/hyp.10414

Birkel C, Soulsby C, Tetzlaff D. 2011a. Modelling catchment-scale water storage dynamics: reconciling dynamic storage with tracer-inferred passive storage. *Hydrological Processes* **25**: 3924–3936. DOI:10.1002/hyp.8201

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. DOI:10.1029/2010WR009547

Birkel C, Tetzlaff D, Dunn SM, Soulsby C. 2010. Towards a simple dynamic process conceptualization in rainfall–runoff models using multi-criteria calibration and tracers in temperate, upland catchments. *Hydrological Processes* **24**: 260–275. DOI:10.1002/hyp.7478

Bishop K, Seibert J, Köhler S, Laudon H. 2004. Resolving the Double Paradox of rapidly mobilized old water with highly variable responses in runoff chemistry. *Hydrological Processes* **18**: 185–189. DOI:10.1002/hyp.5209

Blumstock M, Tetzlaff D, Malcolm IA, Nuetzmann G, Soulsby C. 2015. Baseflow dynamics: multi-tracer surveys to assess variable groundwater contributions to montane streams under low flows. *Journal of Hydrology* **527**: 1021–1033. DOI:10.1016/j.jhydrol.2015.05.019

Böhner J, Selige T. 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. *Gottinger Geographische Abhandlungen* **115**: 13–28.

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. DOI:10.1029/2009WR008371

Burt TP. 2005. A third paradox in catchment hydrology and biogeochemistry: decoupling in the riparian zone. *Hydrological Processes* **19**: 2087–2089. DOI:10.1002/hyp.5904

Calabrese S, Porporato A. 2015. Linking age, survival, and transit time distributions. *Water Resources Research* **51**: 8316–8330. DOI:10.1002/2015WR017785

Capell R, Tetzlaff D, Soulsby C. 2012. Can time domain and source area tracers reduce uncertainty in rainfall-runoff models in larger heterogeneous catchments? *Water Resources Research* **48**: W09544. DOI:10.1029/2011WR011543

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

Dunn SM. 1999. Imposing constraints on parameter values of a conceptual hydrological model using baseflow response. *Hydrology and Earth System Sciences* **3**: 271–284. DOI:10.5194/hess-3-271-1999

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. DOI:10.1002/hyp.7560

Dunn SM, Mackay R. 1995. Spatial variation in evapotranspiration and the influence of land use on catchment hydrology. *Journal of Hydrology* **171**: 49–73. DOI:10.1016/0022-1694(95)02733-6

Fenicia, F, McDonnell, JJ, Savenije, HHG, 2008. Learning from model improvement: on the contribution of complementary data to process understanding. *Water Resources Research* **44**, W06419. doi:10.1029/2007WR006386

Gash J. 1979. An analytical model of rainfall interception by forests. *Royal Meteorological Society, Quarterly Journal* **105**: 43–55.

Gash JHC, Lloyd CR, Lachaud G. 1995. Estimating sparse forest rainfall interception with an analytical model. *Journal of Hydrology* **170**: 79–86. DOI:10.1016/0022-1694(95)02697-N

Geris J, Tetzlaff D, Soulsby C. 2015. Resistance and resilience to droughts: hydropedological controls on catchment storage and run-off response. *Hydrological Processes* **29**: 4579–4593. DOI:10.1002/hyp.10480

Gupta HV, Kling H, Yilmaz KK, Martinez GF. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* **377**: 80–91. DOI:10.1016/j.jhydrol.2009.08.003

Harman CJ. 2015. Time-variable transit time distributions and transport: theory and application to storage-dependent transport of chloride in a watershed. *Water Resources Research* **51**: 1–30. DOI:10.1002/2014WR015707

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

Hrachowitz M, Savenije H, Bogaard TA, Tetzlaff D, Soulsby C. 2013. What can flux tracking teach us about water age distribution patterns and their temporal dynamics? *Hydrology and Earth System Sciences* **17**: 533–564. DOI:10.5194/hess-17-533-2013

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. DOI:10.1002/hyp.7922

Hrachowitz M, Soulsby C, Tetzlaff D, Speed M. 2010. Catchment transit times and landscape controls-does scale matter? *Hydrological Processes* **24**: 117–125. DOI:10.1002/hyp.7510

Karssenberg D, de Jong K, van der Kwast J. 2007. Modelling landscape dynamics with Python. *International Journal of Geographical Information Science* **21**: 483–495. DOI:10.1080/13658810601063936

Karssenberg D, Schmitz O, Salamon P, de Jong K, Bierkens MFP. 2010. A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling & Software* **25**: 489–502. DOI:10.1016/j.envsoft.2009.10.004

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

Kirchner JW, Neal C. 2013. Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection. *Proceedings of the National Academy of Sciences* **110**: 12213–12218. DOI:10.1073/pnas.1304328110

Lessels JS, Tetzlaff D, Birkel C, Dick J, Soulsby C. 2016. Water sources and mixing in riparian wetlands revealed by tracers and geospatial analysis. *Water Resources Research* **52**: 456–470. DOI:10.1002/2015WR017519

Lindström G, Johansson B, Persson M, Gardelin M, Bergström S. 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology* **201**: 272–288. DOI:10.1016/S0022-1694(97)00041-3

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 Resources Research* **50**: 5342–5350. DOI:10.1002/2013WR015141

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, Wörman A, Wrede S. 2010. How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. *Hydrological Processes* **24**: 1745–1754. DOI:10.1002/hyp.7796

McGuire KJ, McDonnell JJ. 2006. A review and evaluation of catchment transit time modeling. *Journal of Hydrology* **330**: 543–563. DOI:10.1016/j.jhydrol.2006.04.020

McMillan, H, Tetzlaff, D, Clark, M, Soulsby, C, 2012. Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach. *Water Resources Research* **48**, W05501. doi:10.1029/2011WR011688

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology* **10**: 282–290.

Neal C, Christophersen N, Neale R, Smith CJ, Whitehead PG, Reynolds B. 1988. Chloride in precipitation and streamwater for the upland catchment of river Severn, Mid-Wales; some consequences for hydrochemical models. *Hydrological Processes* **2**: 155–165. DOI:10.1002/hyp.3360020206

Porporato A, Calabrese S. 2015. On the probabilistic structure of water age. *Water Resources Research* **51**: 3588–3600. DOI:10.1002/2015WR017027

Rinaldo A, Benettin P, Harman CJ, Hrachowitz M, McGuire KJ, van der Velde Y, Bertuzzo E, Botter G. 2015. Storage selection functions: a coherent framework for quantifying how catchments store and release water and solutes. *Water Resources Research* **51**: 4840–4847. DOI:10.1002/2015WR017273

Sayama T, McDonnell JJ. 2009. A new time-space accounting scheme to predict stream water residence time and hydrograph source components at the watershed scale. *Water Resources Research* **45**: W07401. DOI:10.1029/2008WR007549

Seibert J. 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology* **98-99**: 279–293. DOI:10.1016/S0168-1923(99)00105-7

Seibert J, Beven KJ. 2009. Gauging the ungauged basin: how many discharge measurements are needed? *Hydrology and Earth System Sciences* **13**: 883–892. DOI:10.5194/hess-13-883-2009

Seibert J, Grabs T, Köhler S, Laudon H, Winterdahl M, Bishop K. 2009. Linking soil- and stream-water chemistry based on a Riparian Flow-Concentration Integration Model. *Hydrology and Earth System Sciences* **13**: 2287–2297. DOI:10.5194/hess-13-2287-2009

Seibert J, Vis MJP. 2012. Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences* **16**: 3315–3325. DOI:10.5194/hess-16-3315-2012

Soulsby C, Birkel C, Geris J, Dick J, Tunaley C, Tetzlaff D. 2015. Stream water age distributions controlled by storage dynamics and nonlinear hydrologic connectivity: Modeling with high-resolution isotope data. *Water Resources Research* **51**: 7759–7776. DOI:10.1002/2015WR017888

Soulsby C, Bradford J, Dick J, McNamara JP, Geris J, Lessels J, Blumstock M, Tetzlaff D. 2016. Using geophysical surveys to test tracer-based storage estimates in headwater catchments. *Hydrological Processes* DOI: 10.1002/hyp.10889, accepted article

Tetzlaff D, Al-Rawas G, Blöschl G, Carey SK, Fan Y, Hrachowitz M, Kirnbauer R, Jewitt G, Laudon H, McGuire KJ, Sayama T, Soulsby C, Zehe E, Wagener T. 2013. Process realism: Flow paths and storage. In *Runoff Prediction in Ungauged Basins- Synthesis across Processes*, Blöschl G, Sivapalan M, Wagener T, Viglione A, Savenije H (eds). Cambridge University Press: Places and Scales; 53–70.

Tetzlaff D, Birkel C, Dick J, Geris J, Soulsby C. 2014. Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions. *Water Resources Research* **50**: 969–985. DOI:10.1002/2013WR014147

Tetzlaff D, Buttle J, Carey SK, van Huijgevoort MHJ, Laudon H, McNamara JP, Mitchell CPJ, Spence C, Gabor RS, Soulsby C. 2015. A preliminary assessment of water partitioning and ecohydrological coupling in northern headwaters using stable isotopes and conceptual runoff models. *Hydrological Processes* DOI:10.1002/hyp.10515

Tetzlaff D, Soulsby C, Waldron S, Malcolm IA, Bacon PJ, Dunn SM, Lilly A, Youngson AF. 2007. Conceptualization of runoff processes using a geographical information system and tracers in a nested mesoscale catchment. *Hydrological Processes* **21**: 1289–1307. DOI:10.1002/hyp.6309

Uhlenbrook S, Roser S, Tilch N. 2004. Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model. *Journal of Hydrology* **291**: 278–296. DOI:10.1016/j.jhydrol.2003.12.038

Uhlenbrook S, Sieber A. 2005. On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model. *Environmental Modelling & Software* **20**: 19–32. DOI:10.1016/j.envsoft.2003.12.006

van Huijgevoort, MHJ, Tetzlaff, D, Sutanudjaja, EH, Soulsby, C n/d Visualisation of spatial patterns of connectivity and runoff ages derived from a tracer-aided model. Hydrological Processes, in review.

Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV. 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes* **17**: 455–476. DOI:10.1002/hyp.1135

APPENDIX 1. ALGORITHMS IN STARR MODEL INVOLVING HYDROMETRIC AND ISOTOPE DATA

Snow module (degree day method, Seibert, 1999)		
Snow melt	$melt = CFMAX(T(t) - TT)$	TT = Threshold temperature snow and rain CFMAX = Degree day factor
Refreezing liquid water in snowpack	$refreezing = CFR * CFMAX(TT - T(t))$	CFR = Refreezing coefficient
Interception module (Gash *et al.,* 1995)		
P needed for saturation of canopy	$P_G^{`} = -\frac{\overline{R}S_c}{\overline{E}_c}\ln\left(1 - \frac{\overline{E}_c}{\overline{R}}\right)$	$\overline{R}$ = average rainfall rate c = canopy cover $\overline{E}_c$ = evaporation rate from canopy S_c = canopy capacity per unit area of cover
Interception small storms, insufficient to saturate the canopy	cP_G	P_G = rainfall
Wetting up the canopy	$cP_G^{`} - cS_c$	
Evaporation from saturation until rainfall ceases	$c\frac{E_c}{R} * \left(P_G - P_G^{`}\right)$	
Evaporation after rainfall ceases	cS_c	
Stemflow	$p_t P_G$	p_t = proportion of the rainfall diverted to stemflow
Soil module		
Evaporation soil	$ET = PET * \min\left(\frac{SM}{FC * LP}, 1\right)$	PET = Potential evaporation SM = Soil moisture FC = Field capacity LP = Fraction of fieldcapacity below which actual evaporation equals potential evaporation
Recharge from soil to the groundwater	$Seepage = \left(\frac{SM}{FC}\right)^{BetaSeepage}$	$BetaSeepage$ = Exponent to determine soil recharge into groundwater
Soil discharge	$Q_{sb} = k_s SM$	k_s = Recession coefficient discharge from soil store
Direct runoff	$Q_s = \max(SM - FC, 0)$	
CapFlux	$CapFlux = CFlux * \left(\frac{FC - SM}{FC}\right)$	$CFlux$ = Maximum capillary flux

Snow module (degree day method, Seibert, 1999)		
Soil moisture store	$SM(t) = SM(t-1) + P_{eff} - ET - Q_s - Q_{sb} - Seepage + CapFlux$	P_{eff} = effective precipitation (sum throughfall, stemflow and snowmelt)
Groundwater module		
Groundwater discharge	$Q_{gw} = k_g GW$	k_g = Recession coefficient baseflow
Lateral groundwater flow	$Q_{lf} = k_{sat} slope(DEM + GW)$	k_{sat} = Saturated conductivity DEM = elevation difference between cells
Groundwater store	$GW(t) = GW(t-1) + Seepage - Q_{gw} + \Delta Q_{lf} - CapFlux$	ΔQ_{lf} = net lateral flow
Routing		
Total discharge	$Q_{tot,cell} = Q_s + Q_{sb} + Q_{gw} Q_{tot} = accutraveltimeflux(ldd, Q_{tot}, velocity)$	ldd = map with local drainage direction accutraveltimeflux = routing function in PCRaster[1]
Isotopes ratios		
Isotopes ratio soil (i_s)	$\frac{di_s(SM+SMpas)}{dt} = i_p P_{eff} - i_s Q_{sb} - i_s Q_s - i_s ET - i_s Seepage + i_{gw} CapFlux$	i_p = isotopes ratios effective precipitation $SMpas$ = passive storage component
Isotopes ratio groundwater (i_{gw})	$\frac{di_{gw}(GW+GWpas)}{dt} = i_s Seepage - i_{gw} CapFlux - i_{gw} Q_{gw} - i_{gw} Q_{lf,out} + i_{gw,up} Q_{lf,in}$	$i_{gw,up}$ = isotopes ratios inflow lateral groundwater flow $GWpas$ = passive storage component
Water ages		
Water age soil store (Age_{SM})	$\frac{dAge_{SM}(SM+SMpas)}{dt} = Age_P P_{eff} - Age_{SM} Q_{sb} - 1{*}Q_s - Age_E ET - Age_{SM} Seepage + Age_{GW} CapFlux$	Age_p = age of the precipitation (for rain equal to 1) Age_E = age of evaporation
Water age groundwater (Age_{GW})	$\frac{dAge_{GW}(GW+GWpas)}{dt} = Age_{SM} Seepage - Age_{GW} CapFlux - Age_{GW} Q_{gw} - Age_{GW} Q_{lf,out} + Age_{GW,up} Q_{lf,in}$	

[1]For each cell the function computes a travel time which is the time in time steps it takes for material to cross the cell. This is calculated by dividing the distance from the centre of the cell to the centre of the next downstream cell on the local drainage direction by the velocity which is given in distance per time step. Then for each cell, the material in the cell is transported downstream over the local drainage direction while taking the sum of the travel times of the cells through which the material is routed. The transported material is then deposited in the cell for which the travel time reaches 1. (see http://pcraster.geo.uu.nl/pcraster/4.1.0/doc/manual/op_accutraveltime.html)