

Testing a spatially distributed tracer-aided runoff model in a snow-influenced catchment: effects of multi-criteria calibration on streamwater ages

Journal:	<i>Hydrological Processes</i>
Manuscript ID	HYP-18-0196.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	26-Jun-2018
Complete List of Authors:	Piovano, Thea; University of Aberdeen, Northern Rivers Institute, School of Geosciences Tetzlaff, Doerthe; University of Aberdeen, Northern Rivers Institute, School of Geosciences; Leibniz-Institut für Gewässerökologie und Binnenfischerei, Ecohydrology; Humboldt-Universität zu Berlin Mathematisch Naturwissenschaftliche Fakultät, Department of Geography Ala-Aho, Pertti; University of Aberdeen, Northern Rivers Institute, Geography & Environment; Oulun Yliopisto Teknillinen Tiedekunta, Water Resources and Environmental Engineering Buttle, James; Trent University, Geography Mitchell, Carl; University of Toronto at Mississauga, Department of Geography Soulsby, Chris; University of Aberdeen, School of Geosciences; Leibniz-Institut für Gewässerökologie und Binnenfischerei, Ecohydrology
Keywords:	Tracer-aided model, Streamwater ages, STARR (Spatially distributed Tracer-Aided Rainfall-Runoff), Multi-criteria calibration, Stable isotopes, Snow-influenced catchments

SCHOLARONE™
Manuscripts

Testing a spatially distributed tracer-aided runoff model in a snow-influenced catchment: effects of multi-criteria calibration on streamwater ages

Thea I. Piovano¹, Doerthe Tetzlaff^{1,2,3}, Pertti Ala-aho^{1,4}, Jim Buttle⁵, Carl P.J. Mitchell⁶, Chris Soulsby^{1,2}

¹Northern Rivers Institute, University of Aberdeen, Aberdeen, AB24 3UF, United Kingdom.

²Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, 12587, Germany.

³Department of Geography, Humboldt University Berlin, Berlin, 10099, Germany.

⁴Department of Water Resources and Environmental Engineering, University of Oulu, Oulu, Finland.⁵ Department of Geography, Trent University, Peterborough, Ontario, Canada.

⁶Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada, M1C 1A4.

Abstract

Integrating stable isotope tracers into rainfall-runoff models allows investigation of water partitioning and direct estimation of travel times and water ages. Tracer data have valuable information content that can be used to constrain models and, in integration with hydrometric observations, test the conceptualisation of catchment processes in model structure and parameterisation. There is great potential in using tracer-aided modelling in snow-influenced catchments to improve understanding of these catchments' dynamics and sensitivity to environmental change. We used the spatially-distributed, tracer aided STARR (Spatially distributed Tracer-Aided Rainfall-Runoff) model to simulate the interactions between water storage, flux and isotope dynamics in a snow-influenced, long-term monitored catchment in Ontario, Canada. Multiple realisations of the model were achieved using a combination of single and multiple objectives as calibration targets. Whilst good simulations of hydrometric targets such as discharge and SWE could be achieved by local calibration alone, adequate capture of the stream isotope dynamics was predicated on the inclusion of isotope data in the calibration. Parameter sensitivity was highest, and most local, for single calibration targets. With multiple calibration targets key sensitive parameters were still identifiable in snow and runoff generation routines. Water ages derived from flux tracking subroutines in the model indicated a catchment where runoff is dominated by younger waters, particularly during spring snowmelt. However, resulting water ages were most sensitive to the partitioning of runoff sources from soil and groundwater sources, which was most realistically achieved when isotopes were included in the calibration. Given the paucity of studies where hydrological models explicitly incorporate tracers in snow-influenced regions this

1
2
3 study using STARR is an important contribution to satisfactorily simulating snowpack dynamics and
4 runoff generation processes, whilst simultaneously capturing stable isotope variability in snow-
5 influenced catchments.
6
7
8
9

10
11 **Keywords:** Tracer-aided model, streamwater ages, STARR (Spatially distributed Tracer-Aided
12 Rainfall-Runoff), multi-criteria calibration, stable isotopes, snow-influenced catchments
13
14
15
16

17 **1. Introduction**

18
19
20 Understanding water sources, flow paths and storage dynamics is crucial for the sustainable
21 management of water resources (Benettin et al., 2015). Conservative tracers, including stable
22 isotopes, have provided invaluable insights into catchment-scale water fluxes and storage dynamics.
23 An intrinsic value of tracers is that they help distinguish between the velocity and celerity of the
24 hydrological response; the first represents the velocity of water molecules, while the second is the
25 speed at which a rainfall or snowmelt perturbation is propagated via prevailing hydraulic gradients
26 (McDonnell & Beven, 2014). Integrating tracers into rainfall-runoff models allows investigation of
27 water partitioning and direct estimation of travel times (i.e. time spent by water travelling through a
28 catchment from input to output) and water ages (i.e. time spent by water within a control volume
29 from entry to the system to the present); and thus provides a basis for the examination of the
30 internal consistencies of models. Essentially, tracer-aided models combine the computation of water
31 fluxes and solute transport, the estimation of tracer concentrations both in storage elements and
32 discharge, and consequently, the identification of the storage dynamics controlling the catchment
33 discharge and solute flux responses (Birkel & Soulsby, 2015). Tracers have been integrated into
34 several lumped conceptual and analytical hydrological models (Barnes & Bonell, 1996; Hooper et al.,
35 1988; Neal et al., 1988; Weiler et al., 2003; Sayama & McDonnell, 2009; Birkel et al., 2015; McGuire
36 & McDonnell, 2015). Recently, there has been an increasing focus on statistical approaches aimed at
37 representing the probability distribution of transit times (Botter et al., 2011; Rinaldo et al., 2015).
38 Nevertheless, few spatially-distributed conceptualisations of tracers-aided models have been
39 developed (Stadnyk et al., 2013; van Huijgevoort et al., 2016a; Delavau et al., 2017; Uhlenbrook et
40 al., 2004; Kuppel et al., 2018).
41
42
43
44
45
46
47
48
49
50
51
52

53
54 Tracer data also have a valuable information content that can be used to constrain models and, in
55 integration with hydrometric observations, test the conceptualisation of catchment processes in
56
57
58
59
60

1
2
3 model structure and parameterisation (Uhlenbrook and Sieber, 2005; Fenicia et al., 2008).
4 Calibration of models based only on the hydrograph captures the celerity of the runoff response, but
5 does not necessarily further understanding of internal catchment processes, leading to equifinality in
6 model solutions (Beven & Binley, 1992; Beven, 1993; Kuczera & Mroczkowski, 1998; Kirchner, 2006;
7 Fenicia et al., 2008; Rinaldo et al., 2011). Although this problem has been widely examined, use of
8 stream discharge observations solely for calibration remains the most common practice. Using other
9 observed data to calibrate models (especially as parameterisation increases) can be useful in
10 constraining the model space and testing for internal consistencies (Kuppel et al., 2018). Moreover,
11 calibration based on different data often diversifies the choice of performance measures, as each
12 calibrated model can be related to the information content of observations (e.g. Birkel et al., 2014).
13 Accordingly, an intrinsic strength of tracer-aided models is the application of multi-criteria
14 calibration (McDonnell & Beven, 2014). Commonly, multi-criteria calibrations using “hard” or “soft”
15 data are based on hydrometric observations, e.g. snow depth, groundwater level, soil moisture data
16 (Seibert & McDonnell, 2002; Freer et al., 2004; Finger et al., 2011). Given that a multi-criteria
17 calibration in tracer-aided models is not aimed solely at improving the discharge prediction, such
18 calibrations can help identify model parameters and consequently improve the interpretation of
19 relative storages and fluxes for better understanding catchment processes (Gupta et al., 1998; Birkel
20 et al., 2011; Price et al., 2012; Soulsby et al., 2015).
21
22
23
24
25
26
27
28
29
30

31
32 In regions where the hydrology is dominated by snow, appropriate modelling of seasonal snowpack
33 accumulation and melt is essential for simulating the catchment hydrological response, both in
34 terms of timing and magnitude of spring freshet, soil-moisture content, soil thermal conductivity and
35 ground temperature (Kane et al. 1991; Hinzman et al. 1996; Liston & Elder, 2006). Processes of
36 snow accumulation and melt are extremely variable in space and time, driven by the interaction of
37 topography (e.g., slope and aspect), solar radiation intensity, precipitation inputs, wind fields and
38 patterns of drifting snow, vegetation distribution causing interception and different rates of
39 sublimation in canopy intercepted snow (Liston & Elder, 2006; Luce et al., 1998). For tracer-aided
40 modelling, in addition to the formidable challenges of incorporating such spatial and temporal
41 heterogeneity in energy-driven snowmelt models (e.g. DeBeer & Pomeroy, 2017), isotope mass
42 balances must also be tracked to simulate the isotopic composition of melt water (e.g. Ala-aho et al.,
43 2017b). Despite these challenges, there is great potential in using tracer-aided modelling in snow-
44 influenced catchments to improve understanding of their dynamics and sensitivity to environmental
45 change (Tetzlaff et al., 2015).
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The STARR (Spatially distributed Tracer-Aided Rainfall-Runoff) model was recently developed to
4 facilitate fully distributed simulations of hydrological storage dynamics and runoff processes, as well
5 as their associated isotopic compositions and age distributions (van Huijgevoort et al., 2016a,b). The
6 most recent advancement in the STARR model sought to improve modelling of spatially distributed
7 mass balances of snow accumulation and melt along with simulation of the isotopic composition of
8 melt water (Ala-aho et al., 2017a,b). A logical next step is to test the model at other catchments with
9 different intrinsic characteristics of snow accumulation and melt as a learning tool. The aim of this
10 paper is to test the STARR model at the Plastic Lake catchment near Dorset, Ontario in Canada. This
11 is a snow-influenced, long-term experimental site with a rich hydrological and isotopic tracer record.
12 In the model application to this new site, we also use a stepwise approach to multi-criteria
13 calibration. We specifically aim to: 1) evaluate the information content of different calibration
14 targets when simulating the snowpack and discharge, as well as the stable isotope dynamics in snow
15 melt and stream water; and 2) investigate the effects of multi-criteria calibration on the estimated
16 water age distributions in the catchment. This is the first time such work has been conducted in a
17 continental, southern boreal catchment. Plastic Lake is also part of a long-term study network where
18 effects of acid deposition and climate change have been investigated (Aherne et al., 2008). Thus, this
19 tracer-aided approach may assist in interpreting stream hydrochemistry in snow-influenced
20 catchments and help to evaluate catchment sensitivity to environmental change.
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 **2. Data and Methods**

36 **2.1 Study site**

37
38
39
40 The PC1 (Plastic-1) headwater catchment (0.23 km²) of Plastic Lake is located in south-central
41 Ontario, Canada (45°11'N, 78°50'W), 12 km south-east of Dorset (Fig. 1a). According to the Köppen
42 classification (Peel et al., 2007), the climate is humid-continental with long cool summers (Dfb).
43 Mean annual air temperature is 5.0 °C, with sub-zero monthly mean temperatures from December
44 to March. Mean January and June temperatures are -10.6 °C and 18.6 °C, respectively (Environment
45 Canada, 2010). Mean annual precipitation is 1058 mm, with >25% as snowfall (Environment Canada,
46 2010).
47
48
49
50

51
52 Elevation ranges between 351 and 385 m a.s.l. (Fig. 1b), and the catchment contains two vegetated
53 wetlands (Fig. 1c). The largest occupies a central bedrock depression, contains peat deposits > 1 m in
54 depth, and represents 7-10% of the total basin (Devito & Dillon, 1993; Devito et al., 1996). Bedrock is
55
56
57
58
59
60

1
2
3 composed of granitic biotite gneiss with crosscutting pegmatitic veins in the topographic highs and
4 amphibolite in the topographic lows (Kirkwood & Nesbitt, 1991; Dillon & LaZerte, 1992). The bedrock
5 is largely unfractured, and therefore groundwater is limited to the near surface. Relatively quick
6 infiltration and preferential lateral flow at the soil-bedrock interface are important runoff
7 mechanisms (Peters et al., 1995; Buttle & McDonald, 2002).
8
9

10
11 Soil thickness is highly variable, ranging from 0 to 1.5 m (Fig. 1d) and averaging about 0.5 m, with
12 bedrock outcrops covering 10% of total area. Dominant soils are weakly developed orthic humo-
13 ferric and ferro-humic podzols with humic mesisols under the largest wetland, formed upon thin
14 sandy basal tills (Dillon & LaZerte, 1992). Saturated hydraulic conductivities for the podzols are high:
15 vertical and horizontal conductivities, respectively, are approximately 10^{-4} m s⁻¹ and 10^{-3} m s⁻¹ near
16 the surface and decrease with depth to 10^{-5} m s⁻¹ above the bedrock (Peters et al., 1995). Major
17 macropores at the soil bedrock interface can lead to preferential lateral flow with greater rates than
18 the sole conductivity values suggest.
19
20
21
22
23

24
25 Average annual discharge is $4.3 \cdot 10^{-3}$ m³ s⁻¹ with peaks up to 0.24 m³ s⁻¹ during snowmelt and
26 occasionally no flow during summer droughts. The wetland has four intermittent channelized
27 inflows (Devito & Dillon, 1993) and drains to the gauged stream and subsequently into Plastic Lake.
28 Runoff in the wetland is generated by saturation-excess overland flow. The catchment is forested
29 and dominated by conifers: white pine (*Pinus strobus*) and hemlock (*Tsuga canadensis*) (Fig. 1c).
30 Upland areas contain some red oak trees, and the wetland is forested with birch (*Betula* spp.) and
31 black spruce (*Picea mariana*). The understorey is of *Alnus* spp., *Ilex verticillata*, and a well-developed
32 layer of *Sphagnum* in the wetland (Devito et al., 1996).
33
34
35
36
37
38

39 **2.2 The STARR model**

40
41 The STARR (Spatially distributed Tracer-Aided Rainfall-Runoff) model is a spatially explicit
42 hydrological model that simulates water fluxes, storage dynamics, isotope ratios and water ages.
43 Originally developed for a long-term experimental catchment in the Scottish Highlands (van
44 Huijgevoort et al., 2016a), recent advances include implementation of a snowmelt routine that
45 simulates the isotopic composition of the snowpack and melt water (Ala-aho et al., 2017a,b).
46
47
48
49

50
51 STARR is driven by standard meteorological variables and the isotope composition of precipitation
52 (full description in Section 2.3 and calibrated parameters are listed in Table 1). The model usually
53 runs at a daily time-step, but can be used at coarser or finer temporal resolution. The model builds
54 on a HBV-type conceptual model structure (Lindström et al., 1997), comprising routines for
55
56
57
58
59
60

1
2
3 interception (of rain and snow), snow accumulation and melt, soil water storage and flow,
4 groundwater storage and flow, and surface water routing. All routines use uniformly sized cells as
5 primary hydrological units (Fig. S1).
6
7

8
9 The snow routine is conceptualized by an energy balance based on heat advection from
10 precipitation, heat storage in the snowpack, net radiation, and latent and sensible heat exchange
11 (see Ala-aho et al., 2017b for details). Thermal processes are coupled with water mass balance
12 modelling to solve the snow accumulation, melt and sublimation fluxes. Estimation of the amount of
13 water retained in the soil uses physically-based parameters of volumetric field capacity and soil
14 depth; the latter is conceptualised as the depth to either water table or bedrock. Both recharge from
15 soil to groundwater storage and the soil outflows to discharge are conceptualized as power laws
16 parameterized by different recession coefficients (see Ala-aho et al., 2017b for details). Actual
17 evaporation is derived from potential evaporation as a function of soil water storage. The
18 groundwater module is a linear reservoir, with discharge to the stream network computed as a
19 linear function of water storage, while lateral groundwater flow between cells is based on DEM-
20 derived slopes, and a hydraulic conductivity parameter. Runoff fluxes from all cells in the different
21 modules are routed through the catchment according to the local drainage direction to simulate the
22 stream hydrograph.
23
24
25
26
27
28
29
30
31

32
33 Concurrent with the hydrological process modelling, isotope tracers are stored and mixed in each
34 conceptualized compartment (snow, soil and groundwater), similar to an earlier lumped tracer-aided
35 model (Birkel et al., 2011): for each model compartment, isotope ratios are estimated from a mixing
36 equation, assuming complete and instantaneous mixing within each cell. These mixing equations
37 enable estimation of water ages, where water age is included in the model as an “artificial tracer”.
38 New water from rainfall events or snowmelt is assigned an age of one day. During each time step,
39 water that is already stored in the model becomes a day older, and through mixing and water
40 exchange between model compartments and model cells, the water age evolves dynamically in the
41 landscape analogous to the water isotope composition. An additional passive storage in isotopic
42 mixing in the soil conceptualises the stored water that contributes to the isotope and water age
43 mixing volume, but does not contribute to runoff generation (Birkel et al., 2015). The snow module
44 accounts for sublimation fractionation (through the E_{frac} parameter) that would enrich heavy
45 isotopes in the remaining snow storage (both canopy intercepted snow and ground snowpack).
46 Another parameter (M_{frac}) is used in the estimation of the temporal dynamics of isotope
47 fractionation in meltwater and snowpack over the total melt period.
48
49
50
51
52
53
54
55
56
57
58
59
60

2.3 Data

PC1 has been monitored by the Dorset Environmental Science Centre (DESC) since 1987; however, the period with availability of all required data (including stable isotopes in precipitation and stream flow) was 2010-2016. Meteorological data used as model inputs are daily precipitation, air temperature, shortwave radiation, relative humidity and wind speed: they were all measured using standard instrumentation at a nearby automatic meteorological station (PCP, Fig. 1b).

Precipitation isotope composition is also required as a model input: liquid precipitation samples were collected at station PT1 (at Paint Lake, located around 10 km NW from Plastic Lake, Fig. 1a) from June 2010 to July 2014, while from May 2014 to May 2016 samples were collected at Plastic Lake (station PCP). The overlapping period in 2014 showed good correspondence of isotope ratios confirming that the two datasets could be merged. As only liquid precipitation was sampled (Fig. 2), to estimate isotopes in winter precipitation we infilled data gaps with the average monthly isotope compositions estimated by the Online Isotope Precipitation Calculator (Bowen, 2017). The estimation from OIPC is based on catchment latitude, longitude and mean elevation. According to (Bowen & Revenaugh, 2003) at our location the 95% confidence interval is $<4\text{‰}$ for $\delta^2\text{H}$ mainly due to the close proximity of PC1 to several stations. The method we used for filling the isotope record has been shown to be effective in reproducing the depleted winter signal in precipitation and allowing the snowpack isotope composition to be tracked in order to simulate the melt water depletion (Ala-aho et al., 2017a).

The digital elevation map had a spatial resolution of 25 m; this was used for model grid size (Fig. 1b). In order to have a spatially varied parameterization for vegetation, we derived a raster map from the Canadian Land Cover circa 2000 vector file available on GeoBase from classified Landsat 5 and Landsat 7 ortho-images (Fig. 1c). The catchment is classified as vegetated: almost 90% of the area is classed as mixed wood dense (as defined by the forest cover classification of the Earth Observation for Sustainable Development (EOSD) maps; greater than 60% crown closure and neither coniferous nor broadleaf tree account for 75% or more of total basal area). Some zones are coniferous dense cover (conifers $\geq 75\%$ of total basal area) and a small area, roughly overlapping with a wetland, is classified as mixed wood sparse (10-25% crown closure). Some LAI (Leaf Area Index) measurements were available for three vegetation species (red oak, white pine and hemlock) from upward hemispherical images analysed using the CAN EYE V6.1 software (Weiss & Baret, 2010). We averaged LAI values of white pine and hemlock to produce a conifers class value (LAI=4.3) and combined their LAI values with that of red oak to produce a mixed wood value (LAI=4.1). To model the hydrological characteristics of the wetland, we digitised the areas using empirical depth to bedrock

1
2
3 measurements from a previous study (a seismic refraction survey). From these we derived a raster
4 map of soil thickness (Fig. 1d), assuming an average value of 0.5 m (i.e. depth to water table) for the
5 wetland (Dillon and LaZerte, 1992). In absence of a specific model routine able to specifically
6 simulate the behaviour of lakes or wetlands, we incorporated the wetland conceptualisation in the
7 soil routine parametrization. Soil parameters that allow storage of different water amounts (i.e. f_{cap}
8 and K_{sat}) were separately set for wetland or non-wetland cells.
9
10

11
12 Streamwater and snowmelt isotope composition, snow water equivalent (SWE) and stream
13 discharge were all used in the multi-criteria calibration process. Streamwater isotope samples were
14 available at the catchment outlet (PC1) starting in June 2010. Average sampling frequency was 12
15 days, though this was more frequent during snowmelt and rainfall events and there are some longer
16 gaps due to both frozen and completely dry conditions, particularly during a dry period in summer
17 2012 (Fig. 2). The isotopic composition of snowmelt was sampled with snowmelt lysimeters between
18 2011 and 2016 at Paint Lake (PT1, Fig. 2) and utilized in the calibration process to constrain the
19 model routines simulating snow isotope evolution. We chose not to use the measured snowmelt
20 data as model input, because of (1) frequent temporal mismatches of snowmelt samples and model-
21 simulated snowmelt and (2) the problem of assigning a spatially distributed value from a melt
22 sample in a single location (Ala-aho et al., 2017a). Fig. 3 shows the seasonal variability of the isotopic
23 composition of precipitation (occurring in March-November) and snowmelt (occurring in November-
24 April) over the analysed period and the OIPC monthly average values that we coupled to liquid
25 precipitation data to produce a continuous time series that captured the depleted winter signal in
26 the absence of snowfall data.
27
28
29
30
31
32
33
34
35
36

37 SWE data were available from snow surveys conducted by DESC during 2011-2016 at Paint Lake
38 (PT1), with three replicates at 3 locations (South, Middle and North) and an average data frequency
39 of two days during the winter period. Ten replicates on eight sampling days between January and
40 April 2011 from PCP2 (located at Plastic Lake) were used to verify that the Paint Lake snow data
41 could be assumed to be representative for Plastic Lake (Pearson r correlation coefficient = 0.78,
42 slope of linear regression = 0.81). Daily discharge was recorded with a gauging station at the
43 catchment outlet (PC1, Fig. 1b), with some occasional gaps in 2015.
44
45
46
47
48

49 All isotopic samples were analysed by a Los Gatos DTL-100 laser analyser that has an instrument
50 precision of $\pm 0.4\text{‰}$ for $\delta^2\text{H}$ $\pm 0.1\text{‰}$ for $\delta^{18}\text{O}$. We chose to use $\delta^2\text{H}$ in the modelling as it was
51 characterised by slightly less uncertainty than $\delta^{18}\text{O}$.
52
53
54
55
56
57
58
59
60

2.4 Step-wise approach to multi-criteria calibrations

A Monte Carlo approach was used to calibrate the model. We ran 10,000 simulations, each characterized by a parameter set randomly drawn from a uniformly constrained range (Table 1). The calibrated parameters affect snow, soil and groundwater compartments and isotope modelling. Previous work demonstrated that these parameters are the most sensitive in the model (van Huijgevoort et al., 2016a; Ala-aho et al., 2017a,b). Values of other fixed parameters were selected according to available data, literature values, and preceding model experiences and trial runs. Initial conditions were based on trial simulations; however, each Monte Carlo simulation included a spin up period of two years to minimize the influence of initial conditions on the model performance. The calibration period for each of 10000 simulations in the Monte Carlo approach was selected as 1st January 2010 - 16th June 2016, according to the availability of all the required data.

We conducted a stepwise calibration approach. The goal was to retain 100 parameter sets from the 10,000 run ensemble according to different optimisation criteria. Firstly, we selected the 100 runs for each single variable calibration: 100 best runs with the highest efficiencies of simulated discharge (Q calibration), 100 best runs according to the efficiency of simulated streamwater $\delta^2\text{H}$ (H_s calibration), and the 100 best runs according to the snowpack simulation, estimated as SWE (S calibration) and snowmelt $\delta^2\text{H}$ (H_m calibration) to test the performance of the snow modelling in terms of both snowpack and isotopic composition. After these single variable optimisations, we applied a multi-criteria calibration, i.e. including simultaneously the efficiency of more than one variable. We conducted a calibration for the best efficiency for all the pairs of discharge plus one of the other variables (i.e. $Q+H_s$, $Q+S$, $Q+H_m$), then for discharge + streamwater isotopes + SWE ($Q+H_s+S$ calibration). Finally, we included all four available variables (discharge, streamwater isotopes, SWE and snowmelt isotopes, $Q+H_s+S+H_m$ calibration). Abbreviations are summarized in Table 2.

Calibration efficiencies for discharge were calculated from daily time step simulations. Both streamwater isotopes and SWE were compared on observation days. We compared snowmelt lysimeter values with spatially weighted values of simulated snowmelt water flux and isotope composition for days preceding the lysimeter sampling days, assuming that lysimeters integrate the isotopic snowmelt signal between the sampling days (Ala-aho et al., 2017a).

We used three different objective functions for each analysed variable: the Nash- Sutcliffe efficiency (NSE, Eq. 1) (Nash & Sutcliffe, 1970), the Kling-Gupta efficiency (KGE, Eq. 2) (Gupta et al., 2009) and the mean absolute error (MAE, Eq. 3).

$$NSE = 1 - \frac{\sum_{i=1}^n (o_i - s_i)^2}{\sum_{i=1}^n (o_i - \mu_o)^2} \quad (1)$$

$$KGE = 1 - \sqrt{(r - 1)^2 + (\mu_s / \mu_o - 1)^2 + (\sigma_s / \sigma_o - 1)^2} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^n |s_i - o_i|}{n} \quad (3)$$

where r is the Pearson correlation coefficient, μ is the mean, σ is the standard deviation, s refers to simulated values while o to observed, i to time step and n to the total number of time steps in simulations and observations. The NSE is commonly used in hydrology and emphasises the importance of discharge peaks. The KGE is based on equal weighting of linear correlation, bias ratio and variability, and simultaneously measures the Euclidian distance of the three components from the optimum. Thus, it is a more balanced approach than the NSE, with less biasing to peak runoff (Gupta et al., 2009). MAE assumes the same weight for all errors, does not square the errors and aims to minimize the bias ratio while ignoring correlation and variability between observations and simulations. Thus, the MAE gives an estimate of the error range in the same scale as observation variability, providing advantages over the root mean square error (Willmott & Matsuura, 2005). In calibration, we maximized either KGE or NSE while minimizing MAE. Unless stated, all the results refer to KGE for hydrometric variables (KGE_Q for discharge and KGE_{SWE} for SWE) and MAE for isotopes (both the streamwater MAE_{HS} and snowmelt composition MAE_{Hm}), although we also analysed the other objective function combinations.

We identified the 100 runs in the multi-criteria calibrations that best satisfied each target, based on the maximisation or minimisation of empirical cumulative distribution functions (eCDF) of the efficiencies. Details on this eCDF algorithm are given in Ala-aho et al. (2017b). An iterative process selects the quantile threshold common to all eCDFs that gives 100 model runs satisfying calibration targets. When applied to single variable calibration, the algorithm picks the 100 runs with the highest efficiency of the single target. This simple method, also applied by Kuppel et al. (2018), has the advantage of not combining the different objective functions in a single value.

To visualise the global parameter sensitivity of the calibrated model with different targets, we evaluated the maximum distance (MD) between the theoretical CDF (uniform or log-normal distribution) of each parameter randomised in the Monte Carlo approach and the eCDFs of that parameter after optimisation according to different calibration targets. Essentially the MD is the largest absolute Kolmogorov-Smirnov distance between the two curves (Kolmogorov, 1933). This method has been used by Spear & Hornberger (1980) and Whitehead & Young (1979), and

1
2
3 formalised by Pianosi & Wagener (2015) and gives a value that is easily compared between the
4 calibrations: the higher MD, the higher the model sensitivity to that parameter.
5
6

7 **3. Results**

8 **3.1 Hydroclimate and isotope hydrology at the PC1 catchment**

9
10
11 The study period of >6 years covered a wide range in hydroclimatic conditions affecting the PC1
12 catchment. Annual precipitation ranged from 855 mm in 2012 to 1221 mm in 2014 (Fig. 2a). On
13 average, the months with highest precipitation were usually October and June, while the driest were
14 usually February, March and July. Discharge was usually lowest in August ($0.0009 \text{ m}^3\text{s}^{-1}$ on average),
15 generally increasing in October ($0.0037 \text{ m}^3\text{s}^{-1}$), and dropping over winter to $0.0019 \text{ m}^3\text{s}^{-1}$ in February,
16 and then increasing during the snowmelt, peaking in April (mean over 6 years of $0.015 \text{ m}^3\text{s}^{-1}$, Fig. 2a).
17 The driest months in 2012 and 2015 occurred in early summer and there were two months of no
18 discharge in 2012 (from 08th July to 07th September).
19
20
21
22
23
24

25 Daily temperatures ranged between $-29 \text{ }^\circ\text{C}$ and $26 \text{ }^\circ\text{C}$ (Fig. 2b). On average, 110 days per year had
26 mean temperatures below $0 \text{ }^\circ\text{C}$. Winter 2013-2014 had the highest SWE values (Fig. 2b), with
27 monthly means of 172 mm in both February ($\pm 7 \text{ mm}$, standard deviation) and March ($\pm 12 \text{ mm}$) 2014.
28 SWE values in other winters had similar patterns of snow pack accumulation with peaks in March
29 (mean of $135 \text{ mm} \pm 16 \text{ mm}$).
30
31
32
33

34 Time series of $\delta^2\text{H}$ in precipitation (Fig. 2c and Figure 3) showed high variability and strong
35 seasonality, with depleted values in winter and relatively enriched values in summer (-135.8 to -3.9
36 ‰). Snowmelt samples were also depleted and showed high variability, ranging from -154.2 ‰ to $-$
37 40.7 ‰, usually increasing as snowmelt progressed. Compared to the variability in precipitation,
38 streamwater had a damped response (ranging from -91.1 ‰ to -45.8 ‰), with clear seasonal
39 depletion during snowmelt and more enriched values in summer. Median streamwater composition
40 of $\delta^2\text{H}$ (-69.1 ‰) was closer to summer precipitation than snowmelt.
41
42
43
44
45

46 Samples were plotted in dual isotope space (Fig. 4) showing the large range of precipitation and
47 more restricted variation in streamwater. The local meteorological water line (LMWL) was
48 $\delta^2\text{H} = 7.82 (\pm 0.09) \delta^{18}\text{O} + 10.72 (\pm 0.88)$ ($R^2 = 0.97$). Its slope and intercept were similar to
49 those of the global line (GMWL): 8 and 10, respectively. Few snowmelt and streamwater samples
50 plotted below the LMWL; thus, the effect of evaporative fractionation was small.
51
52
53
54
55
56
57
58
59
60

3.2 Information content of different calibration targets

The model was able to reproduce the hydrograph quite well (Fig. 5) with good capture of discharge peak timing and high KGE_Q efficiencies where discharge was included as a calibration target (Table 3). In general, summer low flows were slightly overestimated, with no days of zero flow being simulated. Simulated recession limbs were slightly steeper than measured. The temporal dynamics were similar across all calibrations. The best calibration used discharge as the sole criterion. That calibration also exhibited the smallest variability between the retained models.

Simulations of streamwater δ^2H (Fig. 6) generally captured the observed gradual winter depletion, marked snowmelt pulse and summer enrichment. Simulations slightly overestimated or underestimated observations in some years, notably in spring 2014 and spring 2016, respectively. Unlike the discharge simulations, uncertainty in the retained 100 simulations could be large: during dry periods (most evident in the summers of 2012 and 2015) the lower bounds of simulations were more negative than measured. However, the median and upper simulation bounds captured most of the observed variability, and calibrations based only on isotope composition showed the lowest uncertainty in simulations. Uncertainty increased in some of the multi-criteria calibrations. The worst was for $Q+Hm$, indicating that including snowmelt composition could not help constrain the model for a good simulation of streamwater isotopes (Table 3).

Snowpack SWE was also simulated reasonably well with maximum KGE_{SWE} up to 0.71 (Table 3, Fig. 7). Snowpack temporal dynamics were reasonably reproduced with most of the discrepancies in under- and over-estimating peaks in 2011 and 2013, respectively. Among the different multi-criteria calibrations, the higher variability in KGE_{SWE} values occurred when Hs was included in the calibration, while the best case was for the S calibration as expected.

The snowmelt isotope simulations, even if they did not always fit the data in terms of timing, were still insightful and worth including in the calibration (Fig. 8). The average isotope composition and the trend of gradual depletion during snowmelt episodes through early winter and gradual enrichment during spring melting were simulated reasonably well. The model tended to overestimate the most depleted measurements in January-February and the timing of snowmelt was not always well captured. The most obvious cases were in 2012 and 2014 which had more midwinter melt events than were simulated. Differences among the multi-criteria calibrations were small in terms of uncertainty and model efficiencies (Table 3).

1
2
3 Efficiency ranges of the 100 best simulations for each calibration target and each simulated variable
4 are summarised in Fig. 9. Discharge simulations were generally reasonable, so long as discharge data
5 were included in the set of calibration targets. For all multi-criteria calibrations that included
6 discharge, the minimum KGE_Q efficiency of retained simulations was >0.56 and the maximum was
7 ≤ 0.69 . For streamwater isotope composition, the best efficiency ($MAE_{H_s} = 5.66\%$, range $<1\%$) was
8 obtained using the single variable H_s calibration target, though multi-criteria calibration results
9 where H_s was included were also reasonable. Other single calibration targets, or flow coupled with S
10 or H_m , had higher MAE_{H_s} values. For the SWE simulations, KGE_{SWE} efficiencies for the best runs were
11 >0.52 (when the calibration targets included SWE). In single variable calibrations when SWE was not
12 included, the minimum KGE_{SWE} efficiency dropped to <0.3 , but could exceed 0.68. For snowmelt
13 isotope simulations, the range of efficiencies in terms of MAE_{H_m} was quite poor and similar among all
14 the calibration targets.

15
16 Unsurprisingly, single variable calibrations (Q , H_s , S , H_m) gave the best efficiencies (KGE or MAE
17 according to the variable) for the variable in question and large ranges in efficiency for all the other
18 variables. Adding one variable to discharge (e.g. $Q+H_s$, $Q+S$, $Q+H_m$) constrained simulations for both
19 discharge and the other calibrated variable, but large ranges persisted for the uncalibrated variables.
20 Both $Q+H_s+S$ and $Q+H_s+S+H_m$ calibrations produced more balanced models in terms of overall high
21 efficiencies and small ranges for all the other variables. The only difference was that the calibrations
22 that did not include snowmelt isotopes (H_m) gave slightly higher efficiencies for all other variables in
23 spite of the greater range in snowmelt δ^2H MAE_{H_m} . Similar relative trends in the efficiencies resulted
24 if NSE was used as the objective function to calibrate discharge and SWE (Table S1).

25
26 The maximum distance (MD) of parameter eCDFs prior to and after each calibration was used to
27 assess the overall model identifiability and to compare different selected parameter sets (Fig. 10). It
28 is particularly notable that the model was shown to be sensitive (i.e. high value of MD and greater
29 difference between eCDFs) for many contrasting parameters in the case of single or dual calibration
30 targets (whether hydrometric or isotopic), whilst it was relatively insensitive to changes in the other
31 parameter values. However, for many of the more sensitive parameters (e.g. a_{pow} , $B_{seepage}$, C_{corr} , C_{flux} ,
32 K_{pow} , SM_{pas}), the sensitivities in the calibration targets that included at least three variables (i.e.
33 $Q+H_s+S$, and $Q+H_s+S+H_m$) were just slightly lower than the corresponding highest sensitivities with
34 one of the single calibration targets, but higher than the other calibrations cases. Also, the MD for
35 the K_G parameter in the case of calibrations with three or all variables was even higher than
36 calibrations based solely on discharge or isotopes. Parameters related to the snow module (TT_{high} ,
37 TT_{low} and M_{frac}) had high MD values where calibration included SWE or snowmelt isotope

1
2
3 composition, as expected. However, the sensitivity decreased in calibrations including streamwater
4 isotopes ($Q+Hs+S$ and/or $Q+Hs+S+Hm$) and other calibration targets. E_{frac} (the parameter controlling
5 the snow sublimation fractionation) was identifiable in the isotope calibrations (including both
6 streamwater and snowmelt) and in the $Q+Hs+S+Hm$ calibration, but overall was one of the least
7 sensitive model parameters.
8
9

14 3.3 Influence of multi-criteria calibration on stream water age distributions

15
16
17 The model allowed us to derive and compare time-variant stream water ages using the different
18 calibration targets (Fig. 11). This resulted in different estimated water ages, with median ages over
19 the entire period, individual years and different calibrations generally being less than 1 year,
20 indicating a catchment with short transit times. Except for the Q calibration, which simulated the
21 oldest median age for 2015, the oldest median ages in all calibrations were during 2012 (Table S3).
22 The youngest median ages occurred in 2010 or 2013 according to different calibrations (2010 for Q ,
23 Hs , $Q+Hs$, $Q+S$, $Q+Hm$ calibrations; 2013 for S , Hm , $Q+Hs+S$, $Q+Hs+S+Hm$).
24
25
26

27
28 To compare the resulting ages obtained from the different calibrations, we analysed the cumulative
29 distribution functions (eCDF) of the median ages from each set of 100 retained simulations (Fig. 12).
30 The resulting eCDFs were all statistically significantly different ($p < 0.05$) according to the Kolmogorov-
31 Smirnov test, except for the $Q+Hs$ and $Q+Hs+S+Hm$ calibrations that had $p = 0.051$. Calibration based
32 only on discharge gave the highest probability of older water with both 50th and 95th percentiles
33 being highest at 330 and 1566 days, respectively (Table 4). Overall, adding data in the calibration
34 decreased the modelled proportion of old water: the 95th percentile decreased to 480 days (Hs
35 calibration), to 452 in $Q+Hs$ calibration, to 437 days both in $Q+Hs+S$ and $Q+Hs+S+Hm$. Calibrations
36 based only on SWE or only on snowmelt isotopes gave similar water age eCDFs. Likewise, combining
37 discharge with SWE or snowmelt isotopes resulted in a similar water age eCDF. The age distributions
38 of S and Hm calibrations had a median value similar to the average 50th percentile among the
39 calibrations, while the age distributions of $Q+S$ and $Q+Hm$ had third quantiles similar to the average
40 95th percentile among calibrations.
41
42
43
44
45
46
47
48
49

50 The choice of KGE or NSE as the objective function to calibrate discharge and SWE had little effect on
51 the water age estimations, with similar resulting eCDFs (Fig. S2, Table S2). The only notable
52 difference in the calibrations based on the NSE_Q and NSE_{SWE} was in the 95th percentile of the Q
53 calibration and of the $Q+S$ calibration that showed ages respectively 6 and 8 months younger than
54 the ages derived using KGE_Q and KGE_{SWE} .
55
56
57
58
59
60

4. Discussion

4.1 Importance of the information content of different calibration targets when simulating discharge, snowpack and stable isotope dynamics

The study confirmed the general skill of the STARR model in satisfactorily simulating snowpack dynamics and runoff generation processes, whilst simultaneously capturing stable isotope variability in snow-influenced catchments (Ala-aho et al., 2017b). This is significant given the paucity of studies where hydrological models explicitly incorporate tracers in snow-influenced regions (Tetzlaff et al., 2015). The model was able to produce concurrently good fits for discharge, snow water equivalent, and the stable isotope composition of streamwater and snowmelt. However, simulations of snowmelt isotopes were more problematic in matching point scale data. Nevertheless, given the long-term dataset, the model-data fits of the different variables showed good temporal consistency and this increases confidence that the estimated water ages are reasonable (McDonnell & Beven, 2014).

One of the main findings is that although the performance for a given variable was maximised in the local calibration, integrating other objectives in global solutions (e.g $Q+Hs+S$ and $Q+Hs+S+Hm$ cases) degraded the performances only marginally (though less so for snowmelt isotope composition). Furthermore, the performance range for a certain parameter in the global optimisation was usually significantly reduced compared to the performance of that parameter in a local optimisation (Fig. 9). These insights seem to be in line with the conclusion presented in Seibert & McDonnell (2002), i.e. incorporating diverse data-streams in the calibration conditions moved the model towards considering different dominant processes in the model simulations, giving a more balanced representation of catchment functioning. This conclusion is similar to Birkel et al. (2014), who showed that a coupled flow-tracers model with eight parameters and five calibrated variables (both hydrometric and isotopic: i.e. discharge, groundwater level, soil moisture, streamwater isotopes and soil water isotopes) exhibited better identifiability than model configurations with fewer parameters and calibration targets. Similarly, Hrachowitz et al. (2013) showed how a multiple objective calibration increased model truthfulness. One of the benefits of combining streamflow observations with other related data streams is that the spatio-temporal footprint of the information contained in these observations is more specific and more restricted behaviourally, as recently shown by Kuppel et al. (2018). In particular, based on the definition of an overall performance objective function, Kuppel et al. (2018) were also able to conclude that while model performance obtained with

1
2
3 calibration on an individual dataset was good for single objectives, using multiple datasets in the
4 calibration improved the overall model performance and yielded smaller overall uncertainty. A
5 similar comparison is not possible here as we did not define an overall measure of performance, in
6 favour of a compared analysis of single objective functions. However, the advantages of using
7 calibration targets in addition to discharge are clear, with a substantial improvement in overall
8 model-data fit and a general decrease of predictive uncertainty.
9
10
11

12
13 The good model-observation agreement for discharge ($KGE_Q=0.69/0.75$) based on discharge-specific
14 calibration provided a starting point for the model. However, such a simple analysis supplies an
15 incomplete understanding of catchment function with poorer simulations of stream isotopes and
16 SWE, indicating some likely equifinality (i.e. that the right answer may not have been reflecting the
17 right reasons in terms of catchment function, Kirchner, 2006). Despite the simplified
18 conceptualisation of the suite of snow processes, model fitting of SWE was quite good. This probably
19 reflects the limited influence of factors such as the redistribution of snow by wind in the forested
20 environment of PC1. However, Buttle (2009) showed the distinction between rain, snowmelt and
21 rain-on-snow to be important in applying the temperature-based WINTER snowmelt model for
22 estimating SWE at PC1. Whilst the temperature-based approach to snowmelt simulations in STARR
23 would capture the broad transitions between rainfall and snow, the simple conceptualisation and
24 the daily model time steps may have compromised the details of SWE simulation even though the
25 overall pattern was captured.
26
27
28
29
30
31
32
33

34
35 Overall, the models performed well in estimating streamwater isotope composition. The misfit
36 during dry periods (particularly in summer 2012 and 2015) gives an insight into other model
37 limitations (Fig. 6). The exaggerated isotopic depletion predicted by the model widens the
38 uncertainty bounds when flows become very low. Low discharge and flow cessation are captured
39 quite well, but the simulations conceptualise the dominant runoff source shifting to the
40 groundwater source where modelled stored water tends to be isotopically depleted as it is mainly
41 recharged by spring snowmelt. Previous empirical work in PC1 inferred that deep groundwater
42 fluxes are negligible (Devito & Dillon, 1993) and the ephemeral groundwater contributions are only
43 strong during autumn and spring (Devito et al., 1996). Moreover, Devito et al. (1996) emphasised the
44 dominant role of the wetlands in PC1, which become hydrologically disconnected under dry
45 conditions and play a dominant role in mediating the relationship between water storage and runoff
46 generation. In the STARR model, this is partially conceptualised by the parametrisation that gives
47 higher water storage in the wetland cells, which are also replenished by summer rainfall and become
48 less isotopically depleted as the summer progresses. Hence, when the model simulates cessation of
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 flows from this source, the shift to deeper storage dominance explains the high instability of
4 streamwater isotope estimation during dry periods. Similarly, poor models fits in stream isotope
5 simulations were already highlighted in a previous application of STARR (Ala-aho et al., 2017b)
6 where winter misfits at a site in Sweden (Krycklan) were explained by the model failing to capture
7 the switch from soil sources to a more isotopically depleted groundwater source during winter.
8
9

10
11 Although simulated snowmelt isotope compositions did not match the point observations in detail,
12 other benefits of including the snowmelt isotope routine in the analysis became apparent. The fact
13 that the depletion of stream isotope composition in spring was well-captured suggests the snowmelt
14 isotope routine helps identify the overall isotope dynamics at the catchment scale. There are
15 numerous issues that make the simulation of snowmelt isotopes extremely challenging (Ala-aho et
16 al., 2017a). First, the isotopic composition of snowmelt is highly variable both spatially and
17 temporally (Unnikrishna et al., 2002; Evans et al., 2016). This reflects the spatial distribution of the
18 snowpack and the layered structure that evolves. However, the conceptualisation in STARR assumes
19 complete isotopic mixing of the snowpack. Moreover, the effects of possible rain-on-snow events
20 render the simulation of midwinter snowmelt difficult, as shown by others (Eiriksson et al., 2013;
21 Evans et al., 2016; Juras et al., 2017). In addition, in order to compare simulation results with
22 observations, it was necessary: (1) to assume that lysimeters integrate the snowmelt isotopic signal
23 between the sampling days; and (2) to average the simulated snowmelt flux and its isotope
24 composition both spatially (across the catchment) and temporally (i.e. between sampling days since
25 the output signal of snowmelt flux is discontinuous). This method was originally applied by Ala-aho
26 et al. (2017a) and these pragmatic assumptions provide a successful basis for catchment scale
27 snowmelt isotope simulations, but the approximation gives crude results at the point scale. Despite
28 the discussed limitations of the snowpack and snowmelt process conceptualisation, the important
29 outcome of including both is the model's ability to decouple the isotopic composition of incident
30 precipitation from the actual liquid water inputs to the hillslope-stream system. In a snow-driven
31 catchment such as the one studied, this conceptualisation is valuable for the interpretation of water
32 composition and gives insights with respect to water fluxes, mixing and consequently, water age.
33
34
35
36
37
38
39
40
41
42
43
44
45
46

47 Contrary to previous work (Ala-aho et al., 2017b) which showed that some parameters related to the
48 snow module (TT_{high} , TT_{low} and M_{frac}) were relatively insensitive, here they exhibited greater
49 sensitivity (measured by MD) in the case of the calibration including SWE or snowmelt isotope
50 composition (S and/or H_s calibration). Nevertheless, as in the previous work, the overall parameter
51 sensitivity was low: in particular, E_{frac} (which governs the depletion of water that sublimates from the
52 snowpack) was less sensitive than shown in previous work (e.g. Ala-aho et al., 2017a,b). This was
53
54
55
56
57
58
59
60

1
2
3 unexpected at PC1 as we anticipated a strong influence of snowpack fractionation and isotopic
4 enrichment of the snowpack due to high canopy coverage and interception. However, as already
5 noted, the model reproduced the spring stream isotopes well, independent of a more enriched (high
6 E_{frac}) or less enriched (low E_{frac}) snowmelt signal. A further reason for the lower sensitivity of the E_{frac}
7 parameter compared to previous findings might be the approximated snowfall signal that could in
8 reality be more depleted than the OIPC monthly average values used in the absence of a consistent
9 winter dataset. As shown in Fig. 3, modelled OIPC values for the depleted signal of snowfall in
10 January and December were lower than the snowmelt range in the same period. However, in the
11 absence of data, we have no basis for assessing whether that the snowfall signal is sufficiently
12 depleted, although the low sensitivity of E_{frac} suggests that this is not the case.
13
14
15
16
17
18

19 **4.2 Effects of multi-criteria calibration and choice of objective function on the estimated water age** 20 **distributions in the catchment** 21

22
23 As well as their value in constraining conceptual models like STARR, integrating isotopes or other
24 tracers provides a basis for tracking fluxes and water ages, as has been successfully achieved in other
25 studies (e.g. Vaché & McDonnell, 2006; McMillan et al., 2012; Birkel et al., 2014). However, the
26 methodology that STARR uses to track water ages is spatially distributed and is intrinsically based on
27 spatial storage and flux distribution. Potentially, STARR also provides a detailed visualisation of how
28 fluxes, storage and water ages interact in time and space to underpin catchment functioning (van
29 Huijgevoort et al., 2016a,b). Other approaches to estimate water age, such as use of transit time
30 distributions (McGuire and McDonnell, 2006), lumped conceptual models (Birkel et al., 2014;
31 Soulsby et al., 2015) or storage selection functions (Rinaldo et al., 2015), do not necessarily bring
32 these insights. The method used by the STARR model for estimating water ages gives a mean stream
33 water age value between 7 and 11 months for Dorset depending on the calibration target. This
34 indicates that the catchment runoff is dominated by young water, especially in the spring snow melt,
35 similar to other boreal catchments such as Krycklan in Sweden (Ala-aho et al., 2017b). In addition,
36 the water age is generally consistent with mean transit times estimated in previous studies at the
37 PC1 where convolutional integral models yielded MTTs estimated to be between 7 and 9 months
38 (Lane-Coplen, 2015). Peralta-Tapia et al. (2016) also showed that convolution integral models with
39 long term data also gave comparable results to those derived by STARR.
40
41
42
43
44
45
46
47
48
49
50

51 Previous work also showed how tracking spatially distributed water ages with the STARR model
52 could give insights into a catchment's dominant flow processes (Ala-aho, 2017b). Estimated stream
53 water ages are generated by integrating when, where and which runoff generation processes are
54 activated. The benefit of this approach was shown in a comparison between different catchments,
55
56
57
58
59
60

1
2
3 with different characteristics and dominant runoff generation processes (Ala-aho, 2017b). Here, the
4 analysis of the time series of water ages revealed not only the dominant runoff processes and the
5 interactions between water storages, but also the impact of the different calibration targets on their
6 realisation in the model (Fig. 11). The time series reflected marked seasonality in temporally-variant
7 runoff processes. There were generally two periods each year with older water dominance, one
8 during winter baseflows and one during the dry summer season. The stream water rejuvenation
9 observed during the snowmelt period is consistent with the large contribution of younger water to
10 streamflow at that time (Wels et al., 1991), while rewetting of the catchment in the fall likely
11 accounts for the decrease in water ages following summer events. The ability of event water to
12 make a rapid contribution to catchment streamflow was also shown by Peters et al. (1995) and
13 explained by a dominance of preferential vertical movement of event water and its subsequent
14 lateral movement at the soil-bedrock interface.
15
16
17
18
19
20
21

22
23 Based on the multi-criteria calibration approach, different information content of the estimated
24 water ages was derived from different calibration targets (Fig. 11 and Fig. 12). This has shown how
25 the model can be constrained to appropriately reflect the different aspects of catchment function
26 (Birkel & Soulsby, 2015). A likely reason for the different eCDFs of water ages which we found for the
27 different calibration targets is how the calibration affects the main storages within PC1. The best
28 simulations calibrated to discharge constrain the model to simulate higher amounts of water held in
29 the deeper storage. This is likely related to the fact that there are no restrictions in the model on the
30 storage volume, and the discharge calibration selects the simulations with higher recession
31 constants for the soil box and lower for the groundwater box. As the calibration target resolves to
32 the best possible fitting of flow at the outlet, this resulted in a higher fraction of old water. Including
33 the streamwater isotopes in the calibration forced the model to route less water through deep flow
34 paths, thus better capturing the simulations of tracer dynamics and producing a more responsive
35 system dominated by younger water that fits better our conceptual understanding of the hydrology
36 of the catchment. It has been shown previously that a critical issue in water age estimation is related
37 to constraining the passive storage, i.e. the additional mixing volume that damps the tracer signal to
38 conceptualise the slower velocity of the water particles relative to the celerity of the runoff response
39 (Benettin et al., 2015; Birkel & Soulsby, 2016). The critical problem is the adequate representation of
40 relative storage, conceptualised in STARR explicitly in the SM_{pass} parameter and implicitly in the f_{cap}
41 parameter. The former is sensitive only in models where additional observations - e.g. volumetric
42 water content or groundwater levels - are included in the multi-criteria calibration. In these cases
43 the water age distribution can be more adequately simulated as the storage dynamics are more
44 appropriately captured (Kuppel et al., 2018). Explicitly including groundwater data in multi-objective
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 calibration could help address this issue (Fenicia et al., 2008). However, simply using the plausibility
4 of water ages to compare between multi-criteria calibration results can provide a basis for rejecting
5 models, if results are inconsistent with independent estimates of water age (Birkel & Soulsby, 2016).
6
7

8 **5. Conclusions**

9
10 We used the spatially-distributed, tracer aided STARR model to simulate the interactions between
11 water storage, flux and isotope dynamics in a snow-influenced, long-term monitored catchment
12 near Dorset, Ontario, Canada. Multiple realisations of the model were achieved using a combination
13 of single and multiple objectives as calibration targets. Whilst good simulations of hydrometric
14 targets such as discharge and SWE could be achieved by local calibration alone, adequate capture of
15 the stream isotope dynamics was predicated on the inclusion of isotope data in the calibration.
16 Parameter sensitivity was highest, and most local, for single calibration targets. With multiple
17 calibration targets the sensitivity of individual parameters was reduced, but key sensitive parameters
18 were still apparent in more balanced model solutions. Water ages derived from flux tracking
19 subroutines in the model indicated a catchment where runoff is dominated by younger waters,
20 particularly during spring snowmelt. The resulting water ages were most sensitive to the partitioning
21 of runoff sources from soil and groundwater sources. The stream water age became progressively
22 younger when adding tracer data in the model calibration. The isotope tracer data forced runoff
23 generation to take place in model compartments producing rapid flow response, and thereby
24 produced a more thorough understanding of the catchment runoff generation processes. The more
25 refined model behaviour demonstrates the usefulness of stream isotope composition in constraining
26 the stream water age through numerical simulations.
27
28
29
30
31
32
33
34
35
36
37
38
39

40 **Acknowledgments**

41
42 This work was funded by the European Research Council (project GA 335910 VeWa). Dr. Huaxia Yao,
43 Chris McConnell and Tim Field were instrumental to the collection of isotope samples and for many
44 aspects of hydrological data collection.
45
46
47
48
49

50 **References**

51
52
53 Aherne, J., Futter, M. N., & Dillon, P. J. (2008). The impacts of future climate change and sulphur
54 emission reductions on acidification recovery at Plastic Lake, Ontario. *Hydrology and Earth
55 System Sciences*, 12, 383–392.
56
57
58
59
60

- 1
2
3 Ala-aho, P., Tetzlaff, D., McNamara, J. P., Laudon, H., Kormos, P., & Soulsby, C. (2017a). Modeling the
4 isotopic evolution of snowpack and snowmelt: Testing a spatially distributed parsimonious
5 approach. *Water Resources Research*, 53(7), 5813–5830.
6 <https://doi.org/10.1002/2017WR020650>
7
8 Ala-aho, P., Tetzlaff, D., McNamara, J. P., Laudon, H., & Soulsby, C. (2017b). Using isotopes to
9 constrain water flux and age estimates in snow-influenced catchments using the STARR
10 (Spatially distributed Tracer-Aided Rainfall-Runoff) model. *Hydrology and Earth System
11 Sciences*, 21(10), 5089–5110. <https://doi.org/10.5194/hess-21-5089-2017>
12
13 Barnes, C. J., & Bonell, M. (1996). Application of Unit Hydrograph Techniques To Solute Transport in
14 Catchments. *Hydrological Processes*, 10(6), 793–802. [https://doi.org/10.1002/\(SICI\)1099-
15 1085\(199606\)10:6<793::AID-HYP372>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-1085(199606)10:6<793::AID-HYP372>3.0.CO;2-K)
16
17 Benettin, P., Bailey, S. W., Campbell, J. L., Green, M. B., Rinaldo, A., Likens, G. E., ... Botter, G. (2015).
18 Linking water age and solute dynamics in streamflow at the Hubbard Brook Experimental
19 Forest, NH, USA. *Water Resources Research*, 51(11), 9256–9272.
20 <https://doi.org/10.1002/2015WR017552>
21
22 Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty
23 prediction. *Hydrological Processes*, 6(3), 279–298. <https://doi.org/10.1002/hyp.3360060305>
24
25 Beven, K. J. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances
26 in Water Resources*, 16, 41–51.
27
28 Birkel, C., & Soulsby, C. (2015). Advancing tracer-aided rainfall-runoff modelling: A review of
29 progress, problems and unrealised potential. *Hydrological Processes*, 29(25), 5227–5240.
30 <https://doi.org/10.1002/hyp.10594>
31
32 Birkel, C., & Soulsby, C. (2016). Linking tracers, water age and conceptual models to identify
33 dominant runoff processes in a sparsely monitored humid tropical catchment. *Hydrological
34 Processes*, 30(24), 4477–4493. <https://doi.org/10.1002/hyp.10941>
35
36 Birkel, C., Soulsby, C., & Tetzlaff, D. (2014). Developing a consistent process-based conceptualization
37 of catchment functioning using measurements of internal state variables. *Water Resources
38 Research*, 50(4), 3481–3501. <https://doi.org/10.1002/2013WR014925>
39
40 Birkel, C., Soulsby, C., & Tetzlaff, D. (2015). Conceptual modelling to assess how the interplay of
41 hydrological connectivity, catchment storage and tracer dynamics controls nonstationary water
42 age estimates. *Hydrological Processes*, 29(13), 2956–2969. <https://doi.org/10.1002/hyp.10414>
43
44 Birkel, C., Tetzlaff, D., Dunn, S. M., & Soulsby, C. (2011). Using time domain and geographic source
45 tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models.
46 *Water Resources Research*, 47(2), 1–15. <https://doi.org/10.1029/2010WR009547>
47
48 Botter, G., Bertuzzo, E., & Rinaldo, A. (2011). Catchment residence and travel time distributions: The
49 master equation. *Geophysical Research Letters*, 38(11), 1–6.
50 <https://doi.org/10.1029/2011GL047666>
51
52 Bowen, G. J. (2017). The Online Isotopes in Precipitation Calculator, version 3.1. Retrieved from
53 <http://wateriso.utah.edu/waterisotopes>
54
55 Bowen, G. J., & Revenaugh, J. (2003). Interpolating the isotopic composition of modern meteoric
56 precipitation. *Water Resources Research*, 39(10), 1–13.
57 <https://doi.org/10.1029/2003WR002086>
58
59 Buttle, J. M. (2009). Using a temperature-based model of snow accumulation and melt to assess the
60

- 1
2
3 long-term hydrologic behaviour of forested headwater basins in south-central Ontario. In *66th*
4 *Eastern Snow Conference*. Niagara-on-the-Lake, Ontario, Canada.
- 5
6 Buttle, J. M., & McDonald, D. J. (2002). Coupled vertical and lateral preferential flow on a forested
7 slope. *Water Resources Research*, *38*(5), 18-1-18–16. <https://doi.org/10.1029/2001WR000773>
- 8
9 DeBeer, C. M., & Pomeroy, J. W. (2017). Influence of snowpack and melt energy heterogeneity on
10 snow cover depletion and snowmelt runoff simulation in a cold mountain environment. *Journal*
11 *of Hydrology*, *553*, 199–213. <https://doi.org/10.1016/j.jhydrol.2017.07.051>
- 12
13 Delavau, C. J., Stadnyk, T., & Holmes, T. (2017). Examining the impacts of precipitation isotope input
14 $\delta^{18}\text{O}_{\text{ppt}}$ on distributed, tracer-aided hydrological modelling. *Hydrology and Earth System*
15 *Sciences*, *21*(5), 2595–2614. <https://doi.org/10.5194/hess-21-2595-2017>
- 16
17 Devito, K. J., & Dillon, P. J. (1993). The influence of hydrologic fluctuations and peat oxia on the
18 phosphorus and nitrogen dynamics of a conifer swamp. *Water Resources Research*, *29*(8), 64.
19 Retrieved from <http://www.biodiversitylibrary.org/item/98791>
- 20
21 Devito, K. J., Hill, A. R., & Roulet, N. (1996). Groundwater-surface water interactions in headwater
22 forested wetlands of the Canadian Shield. *Journal of Hydrology*, *181*(1–4), 127–147.
23 [https://doi.org/10.1016/0022-1694\(95\)02912-5](https://doi.org/10.1016/0022-1694(95)02912-5)
- 24
25 Dillon, P. J., & LaZerte, B. D. (1992). Response of the Plastic Lake catchment, Ontario, to reduced
26 sulphur deposition. *Environmental Pollution*, *77*(2–3), 211–217. [https://doi.org/10.1016/0269-7491\(92\)90079-P](https://doi.org/10.1016/0269-7491(92)90079-P)
- 27
28 Eiriksson, D., Whitson, M., Luce, C. H., Marshall, H. P., Bradford, J., Benner, S. G., ... Mcnamara, J. P.
29 (2013). An evaluation of the hydrologic relevance of lateral flow in snow at hillslope and
30 catchment scales. *Hydrological Processes*, *27*(5), 640–654. <https://doi.org/10.1002/hyp.9666>
- 31
32 Environment Canada. (2010). Canadian climate Normals 1981-2010 (DORSET MOE). Retrieved from
33 http://climate.weather.gc.ca/climate_normals/
- 34
35 Evans, S. L., Flores, A. N., Heilig, A., Kohn, M. J., Marshall, H. P., & McNamara, J. P. (2016). Isotopic
36 evidence for lateral flow and diffusive transport, but not sublimation, in a sloped seasonal
37 snowpack, Idaho, USA. *Geophysical Research Letters*, *43*(7), 3298–3306.
38 <https://doi.org/10.1002/2015GL067605>
- 39
40 Fenicia, F., McDonnell, J. J., & Savenije, H. H. G. (2008). Learning from model improvement: On the
41 contribution of complementary data to process understanding. *Water Resources Research*,
42 *44*(6), 1–13. <https://doi.org/10.1029/2007WR006386>
- 43
44 Finger, D., Pellicciotti, F., Konz, M., Rimkus, S., & Burlando, P. (2011). The value of glacier mass
45 balance, satellite snow cover images, and hourly discharge for improving the performance of a
46 physically based distributed hydrological model. *Water Resources Research*, *47*(7), 1–14.
47 <https://doi.org/10.1029/2010WR009824>
- 48
49 Freer, J. E., McMillan, H., McDonnell, J. J., & Beven, K. J. (2004). Constraining dynamic TOPMODEL
50 responses for imprecise water table information using fuzzy rule based performance measures.
51 *Journal of Hydrology*, *291*(3–4), 254–277. <https://doi.org/10.1016/j.jhydrol.2003.12.037>
- 52
53 Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models:
54 Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4),
55 751–763. <https://doi.org/10.1029/97WR03495>
- 56
57 Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared
58 error and NSE performance criteria: Implications for improving hydrological modelling. *Journal*
59

- 1
2
3 *of Hydrology*, 377(1–2), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- 4
5 Hinzman, L. D., Kane, D. L., Benson, C. S., & Everett, K. R. (1996). Energy balance and hydrological
6 processes in an Arctic watershed. In J. F. Reynolds & J. D. Tenhunen (Eds.), *Landscape Function:
7 Implications for Ecosystem Response to Disturbance. A Case Study in Arctic Tundra* (pp. 131–
8 154).
- 9
10 Hooper, R. P., Stone, A., Christophersen, N., De Grosbois, E., & Seip, H. M. (1988). Assessing the
11 Birkenes Model of stream acidification using a multisignal calibration methodology. *Water
12 Resources Research*, 24(8), 1308–1316. <https://doi.org/10.1029/WR024i008p01308>
- 13
14 Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D., & Soulsby, C. (2013). What can flux tracking
15 teach us about water age distribution patterns and their temporal dynamics? *Hydrology and
16 Earth System Sciences*, 17(2), 533–564. <https://doi.org/10.5194/hess-17-533-2013>
- 17
18 Juras, R., Würzler, S., Pavlásek, J., Vitvar, T., & Jonas, T. (2017). Rainwater propagation through snow
19 pack during rain-on-snow events under different snow condition. *Hydrology and Earth System
20 Sciences*, 21, 4973–4987. <https://doi.org/10.5194/hess-2016-612>
- 21
22 Kane, D. L., Hinzman, L. D., Benson, C. S., & Liston, G. E. (1991). Snow hydrology of a headwater
23 Arctic basin: 1. Physical measurements and process studies. *Water Resources Research*, 27(6),
24 1099–1109. <https://doi.org/10.1029/91WR00262>
- 25
26 Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements,
27 analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(3), 1–
28 5. <https://doi.org/10.1029/2005WR004362>
- 29
30 Kirkwood, D. E., & Nesbitt, H. W. (1991). Formation and evolution of soils from an acidified
31 watershed: Plastic Lake, Ontario, Canada. *Geochimica et Cosmochimica Acta*, 55(5), 1295–1308.
32 [https://doi.org/10.1016/0016-7037\(91\)90308-R](https://doi.org/10.1016/0016-7037(91)90308-R)
- 33
34 Kolmogorov, A. N. (1933). Sulla Determinazione Empirica di Una Legge di Distribuzione. *Giornale
35 dell'Istituto Italiano Degli Attuari*, 4, 83–91.
- 36
37 Kuczera, G., & Mroczkowski, M. (1998). Assessment of hydrologic parameter uncertainty and the
38 worth of multiresponse data. *Water Resources Research*, 34(6), 1481–1489.
39 <https://doi.org/10.1029/98WR00496>
- 40
41 Kuppel, S., Tetzlaff, D., Maneta, M. P., & Soulsby, C. (2018). What can we learn from multi-data
42 calibration of a process-based ecohydrological model? *Environmental Modelling and Software*,
43 101, 301–316. <https://doi.org/10.1016/j.envsoft.2018.01.001>
- 44
45 Lane-Coplen, D. (2015). *Landscape Influences on Hydrological Transit Times in Precambrian Shield
46 Catchments*. University of Toronto.
- 47
48 Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and
49 test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201(1–4), 272–288.
50 [https://doi.org/10.1016/S0022-1694\(97\)00041-3](https://doi.org/10.1016/S0022-1694(97)00041-3)
- 51
52 Liston, G. E., & Elder, K. (2006). A Distributed Snow-Evolution Modeling System (SnowModel).
53 *Journal of Hydrometeorology*, 7(6), 1259–1276. <https://doi.org/10.1175/JHM548.1>
- 54
55 Luce, C. H., Tarboton, D. G., & Cooley, K. R. (1998). The influence of the spatial distribution of snow
56 on basin-averaged snowmelt. *Hydrological Processes*, 12(10–11), 1671–1683.
57 [https://doi.org/10.1002/\(SICI\)1099-1085\(199808/09\)12:10/11<1671::AID-HYP688>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1671::AID-HYP688>3.0.CO;2-N)
- 58
59 Mcdonnell, J. J., & Beven, K. (2014). Debates—The future of hydrological sciences: A (common) path
60 forward? A call to action aimed at understanding velocities, celerities and residence time

- distributions of the headwater hydrograph. *Water Resources Research*, 50(6), 5342–5350.
<https://doi.org/10.1002/2013WR015141>.Received
- McGuire, K. J., & McDonnell, J. J. (2015). Tracer advances in catchment hydrology. *Hydrological Processes*, 29(25), 5135–5138. <https://doi.org/10.1002/hyp.10740>
- 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(5).
<https://doi.org/10.1029/2011WR011688>
- Nash, J. E., & Sutcliffe, J. V. (1970). River Flow Forecasting Through Conceptual Models Part I - A Discussion of Principles. *Journal of Hydrology*, 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Neal, C., Christophersen, N., Neale, R., Smith, C. J., Whitehead, P. G., & 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(2), 155–165.
<https://doi.org/10.1002/hyp.3360020206>
- Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, 11(5), 1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>
- Peters, D. L., Buttle, J. M., Taylor, C. H., & LaZerte, B. D. (1995). Runoff Production in a Forested, Shallow Soil, Canadian Shield Basin. *Water Resources Research*, 31(5), 1291–1304.
<https://doi.org/10.1029/94WR03286>
- Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67, 1–11.
<https://doi.org/10.1016/j.envsoft.2015.01.004>
- Price, K., Purucker, S. T., Kraemer, S. R., & Babendreier, J. E. (2012). Tradeoffs among watershed model calibration targets for parameter estimation. *Water Resources Research*, 48(10), 1–16.
<https://doi.org/10.1029/2012WR012005>
- Rinaldo, A., Benettin, P., Harman, C. J., Hrachowitz, M., McGuire, K. J., van der Velde, Y., 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.
<https://doi.org/10.1002/2015WR017273>.Received
- Rinaldo, A., Beven, K. J., Bertuzzo, E., Nicotina, L., Davies, J., Fiori, A., ... Botter, G. (2011). Catchment travel time distributions and water flow in soils. *Water Resources Research*, 47(7), 1–13.
<https://doi.org/10.1029/2011WR010478>
- Sayama, T., & McDonnell, J. J. (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(7), 1–14. <https://doi.org/10.1029/2008WR007549>
- Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, 38(11), 23-1-23–14. <https://doi.org/10.1029/2001WR000978>
- 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(9), 7759–7776.
<https://doi.org/10.1002/2015WR017888>

- 1
2
3 Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in peel inlet-II. Identification of critical
4 uncertainties via generalized sensitivity analysis. *Water Research*, *14*(1), 43–49.
5 [https://doi.org/10.1016/0043-1354\(80\)90040-8](https://doi.org/10.1016/0043-1354(80)90040-8)
6
7 Stadnyk, T. A., Delavau, C., Kouwen, N., & Edwards, T. W. D. (2013). Towards hydrological model
8 calibration and validation: Simulation of stable water isotopes using the isoWATFLOOD model.
9 *Hydrological Processes*, *27*(25), 3791–3810. <https://doi.org/10.1002/hyp.9695>
10
11 Tetzlaff, D., Buttle, J., Carey, S. K., McGuire, K., Laudon, H., & Soulsby, C. (2015). Tracer-based
12 assessment of flow paths, storage and runoff generation in northern catchments: A review.
13 *Hydrological Processes*, *29*(16), 3475–3490. <https://doi.org/10.1002/hyp.10412>
14
15 Uhlenbrook, S., Roser, S., & Tilch, N. (2004). Hydrological process representation at the meso-scale:
16 The potential of a distributed, conceptual catchment model. *Journal of Hydrology*, *291*(3–4),
17 278–296. <https://doi.org/10.1016/j.jhydrol.2003.12.038>
18
19 Uhlenbrook, S., & Sieber, A. (2005). On the value of experimental data to reduce the prediction
20 uncertainty of a process-oriented catchment model. *Environmental Modelling & Software*, *20*,
21 19–32.
22
23 Unnikrishna, P. V., McDonnell, J. J., & Kendall, C. (2002). Isotope variations in a Sierra Nevada
24 snowpack and their relation to meltwater. *Journal of Hydrology*, *260*(1–4), 38–57.
25 [https://doi.org/10.1016/S0022-1694\(01\)00596-0](https://doi.org/10.1016/S0022-1694(01)00596-0)
26
27 Vaché, K. B., & McDonnell, J. J. (2006). A process-based rejectionist framework for evaluating
28 catchment runoff model structure. *Water Resources Research*, *42*(2), 1–15.
29 <https://doi.org/10.1029/2005WR004247>
30
31 van Huijgevoort, M. H. J., Tetzlaff, D., Sutanudjaja, E. H., & Soulsby, C. (2016a). Using high resolution
32 tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-
33 runoff model. *Hydrological Processes*, *30*(25), 4761–4778. <https://doi.org/10.1002/hyp.10902>
34
35 van Huijgevoort, M. H. J., Tetzlaff, D., Sutanudjaja, E. H., & Soulsby, C. (2016b). Visualization of
36 spatial patterns of connectivity and runoff ages derived from a tracer-aided model.
37 *Hydrological Processes*, *30*(25), 4893–4895. <https://doi.org/10.1002/hyp.10961>
38
39 Weiler, M., McGlynn, B. L., McGuire, K. J., & McDonnell, J. J. (2003). How does rainfall become
40 runoff? A combined tracer and runoff transfer function approach. *Water Resources Research*,
41 *39*(11), 1–13. <https://doi.org/10.1029/2003WR002331>
42
43 Wels, C., Taylor, C. H., Cornett, R. J., & LaZerte, B. D. (1991). Streamflow generation in a headwater
44 basin on the Precambrian shield. *Hydrological Processes*, *5*, 185–199.
45
46 Whitehead, P. G., & Young, P. (1979). Water quality in river systems: Monte Carlo analysis. *Water*
47 *Resources Research*, *15*(2), 451–459.
48
49 Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root
50 mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1),
51 79–82. <https://doi.org/10.3354/cr030079>
52
53
54
55
56
57
58
59
60

Table 1 Description and initial ranges of calibrated parameters in the model

Parameter	Description	Min	Max
Snow			
a_{pow} [-]	Coefficient to reduce snow albedo	1	3
c_{corr} [-]	Correction factor for snowfall in addition to wind correction	0	0.3
TT_{low} [°C]	Threshold temperature below which all precipitation is liquid	-2	0
TT_{high} [°C]	Threshold temperature above which all precipitation is liquid	0	2
Soil and groundwater			
$B_{seepage}$ (log)	Recession coefficient to determine soil recharge into groundwater	10^{-1}	10
C_{flux} (log)	Parameter for maximum capillary flux	10^{-2}	1
f_{cap} [-]	Field capacity ($f_{cap,w}$ for wetland, otherwise $f_{cap,nw}$)	0.2	0.8
K_G (log)	Recession coefficient baseflow	10^{-5}	10^{-3}
K_{pow} [-]	Power coefficient	1	3
K_s [day^{-1}]	Recession coefficient to determine outflow from soil storage ($K_{s,w}$ for wetland, otherwise $K_{s,nw}$)	5	50
K_{sat} (log $m s^{-1}$)	Saturated conductivity of lateral flow	10^{-1}	10
LP (log)	Fraction of limiting actual evaporation	10^{-2}	1
Isotopes			
E_{frac} [‰]	Fractionation parameter for δ^2H	-133	0
M_{frac} [‰]	Offset parameter, equilibrium ice-liquid for δ^2H	-19	0
SM_{pas} [mm]	Mixing volume soil	100	300

Table 2 Calibration target abbreviations and calibrated variables. Each calibrated variable that is included in each calibration target is marked with a cross. Q refers to discharge, Hs to streamwater isotopes, S to the Snow Water Equivalent (SWE) and Hm to the snowmelt isotopes. Colour legend is the same for all plots in Fig. 5-11.

	Calibration target								
	Q	Hs	S	Hm	Q+Hs	Q+S	Q+Hm	Q+Hs+S	Q+Hs+S+Hm
Discharge Q	X				X	X	X	X	X
Stream $\delta^2\text{H}$		X			X			X	X
SWE			X			X		X	X
SnowMelt $\delta^2\text{H}$				X			X		X

For Peer Review

Table 3 Comparison between ranges of objective functions (KGE of discharge and SWE, MAE of both isotope streamwater and snowmelt composition) for different calibration targets over the calibrated period (1st January 2010 – 16th June 2016).

Calibration target	KGE _Q [-]		MAE _{Hs} (stream $\delta^2\text{H}$) [‰]		KGE _{SWE} [-]		MAE _{Hm} (snowmelt $\delta^2\text{H}$) [‰]	
	min	max	max	min	min	max	max	min
Q	0.69	0.75	16.64	6.75	0.16	0.69	24.50	21.79
Hs	0.05	0.65	6.50	5.66	0.07	0.70	24.17	21.74
S	0.10	0.72	18.92	6.43	0.67	0.71	24.37	21.64
Hm	0.12	0.72	18.26	5.91	0.19	0.68	21.79	21.50
Q+Hs	0.61	0.70	7.47	6.17	0.17	0.69	24.05	21.52
Q+S	0.64	0.74	18.92	6.88	0.63	0.71	24.50	21.64
Q+Hm	0.62	0.73	17.33	6.66	0.25	0.68	22.07	21.52
Q+Hs+S	0.59	0.70	7.81	6.27	0.56	0.69	24.05	21.70
Q+Hs+S+Hm	0.56	0.69	8.28	6.54	0.52	0.67	22.35	21.70

Table 4 5th, 50th and 95th percentiles of eCDFs of median ages [in days] for the different calibration targets, with KGE as objective function for Q and SWE and MAE as objective function for isotopes.

Percentile	Q	Hs	S	Hm	Q+Hs	Q+S	Q+Hm	Q+Hs+S	Q+Hs+S+Hm
5 th	121	158	149	141	133	128	132	125	125
50 th	330	269	324	308	224	275	291	213	213
95 th	1566	480	841	785	452	1321	1182	437	437

For Peer Review

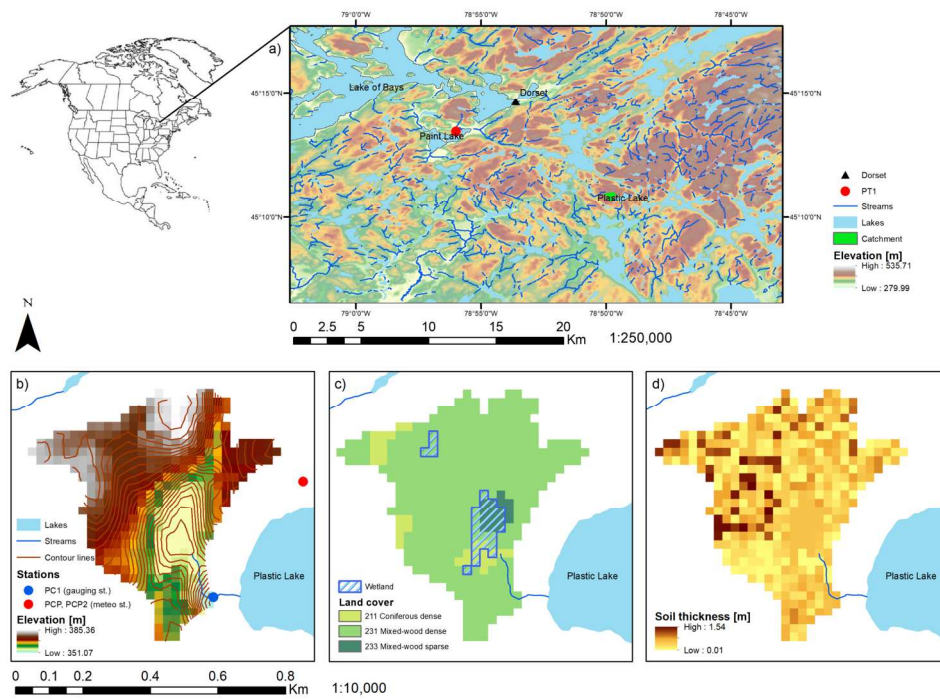


Figure 1: Location of the Plastic Lake catchment, the closest town (Dorset) and Paint Lake, b) map of topography (elevation in m a.s.l. and rough contour lines every 1 m, location of gauging and meteorological stations, c) land cover map and wetland areas, d) soil depth raster map.

177x127mm (300 x 300 DPI)

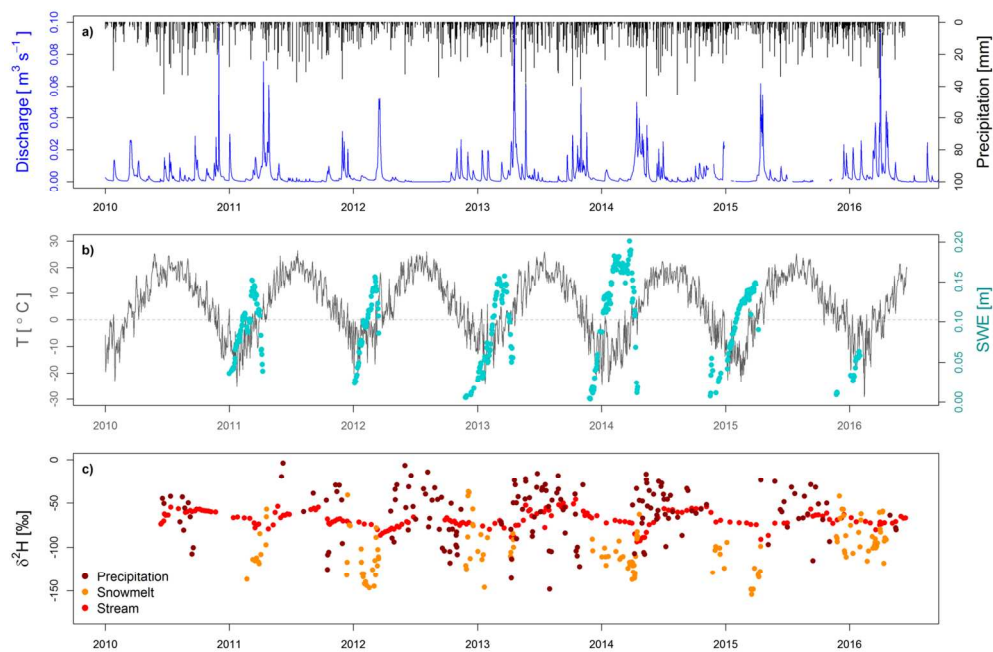


Figure 2: Daily data of precipitation and discharge (a), temperature and Snow Water Equivalent (SWE) (b) and stable isotopes of precipitation, snowmelt and stream water (c).

172x115mm (300 x 300 DPI)

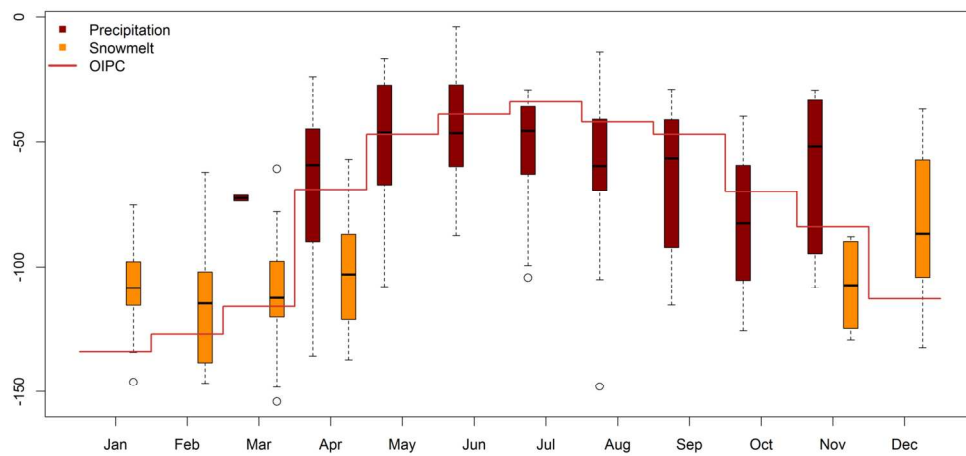


Figure 3: Boxplots showing the seasonality of precipitation and snowmelt isotope composition over the analysed period. The line shows the monthly average value estimated by OIPC.

162x78mm (300 x 300 DPI)

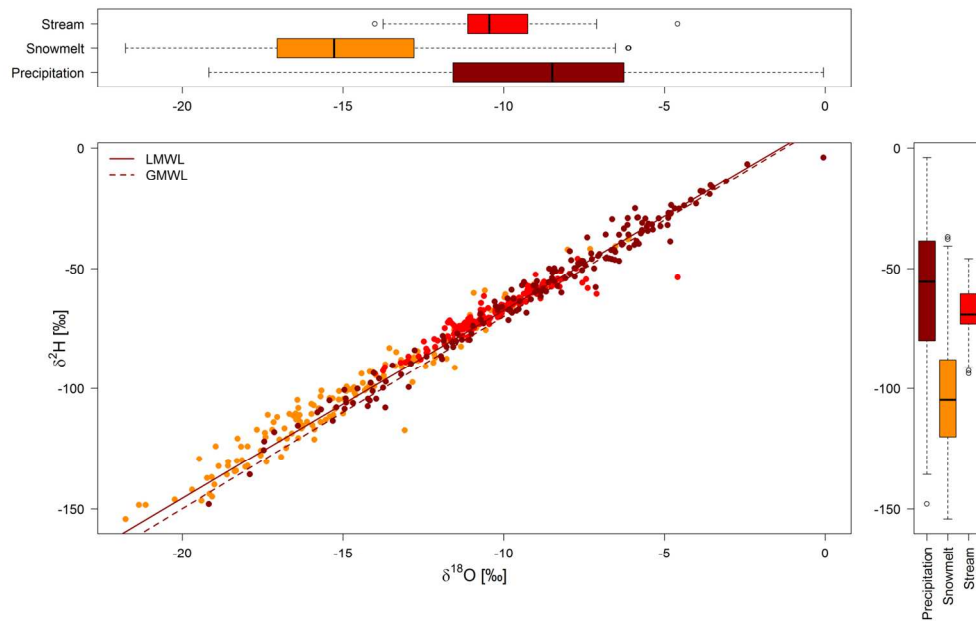


Figure 4: Dual isotope plot of precipitation, snowmelt and streamwater isotope samples with marginal boxplots showing ranges in $\delta^2\text{H}$ and $\delta^{18}\text{O}$. Solid line is the Local Meteorological Water Line (LMWL) and dotted line is the Global Meteorological Water Line GMWL.

169x110mm (300 x 300 DPI)

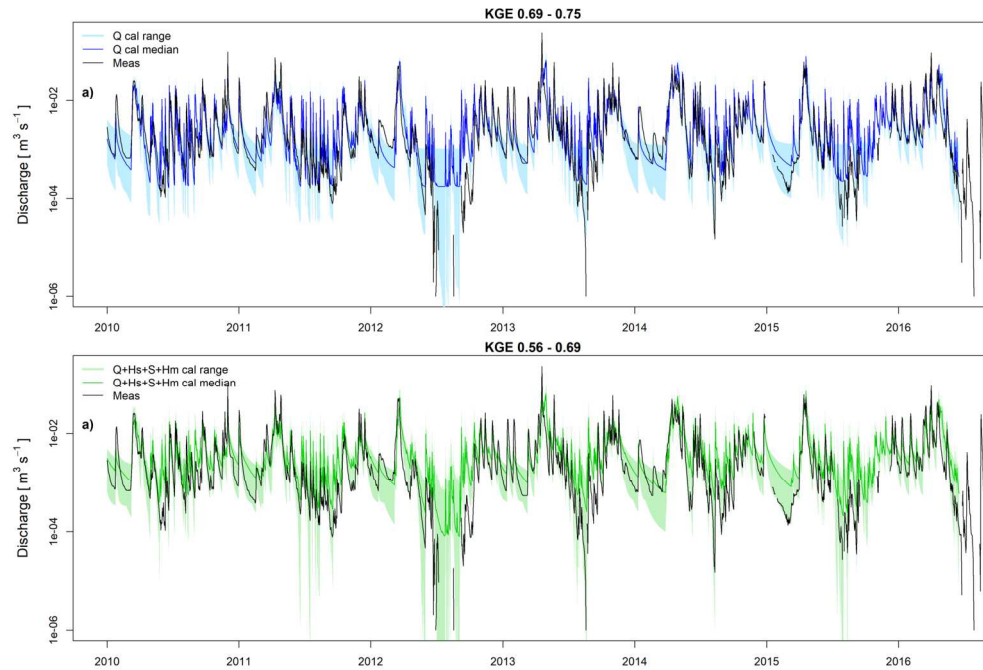


Figure 5: Discharge simulations based on the different calibration targets listed in Table 2 in semi-logarithmic plots: solid lines are simulated medians and colours refer to calibration target. Shaded areas encompass the highest and lowest 100 simulations selected according to each calibration. Measured discharge is drawn as black line.

177x118mm (300 x 300 DPI)

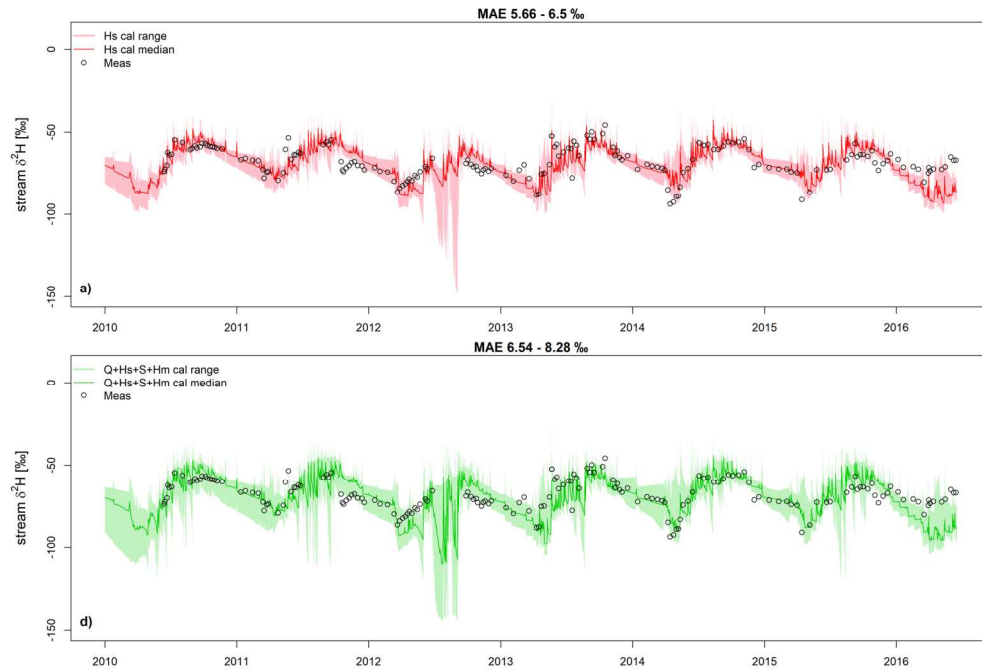


Figure 6: Streamwater $\delta^2\text{H}$ simulations based on the different calibration targets listed in Table 2: solid lines are simulated medians and colours refer to calibration target. Shaded areas encompass the highest and lowest 100 simulations selected according to each calibration. Black dots are the streamwater isotope samples.

177x118mm (300 x 300 DPI)

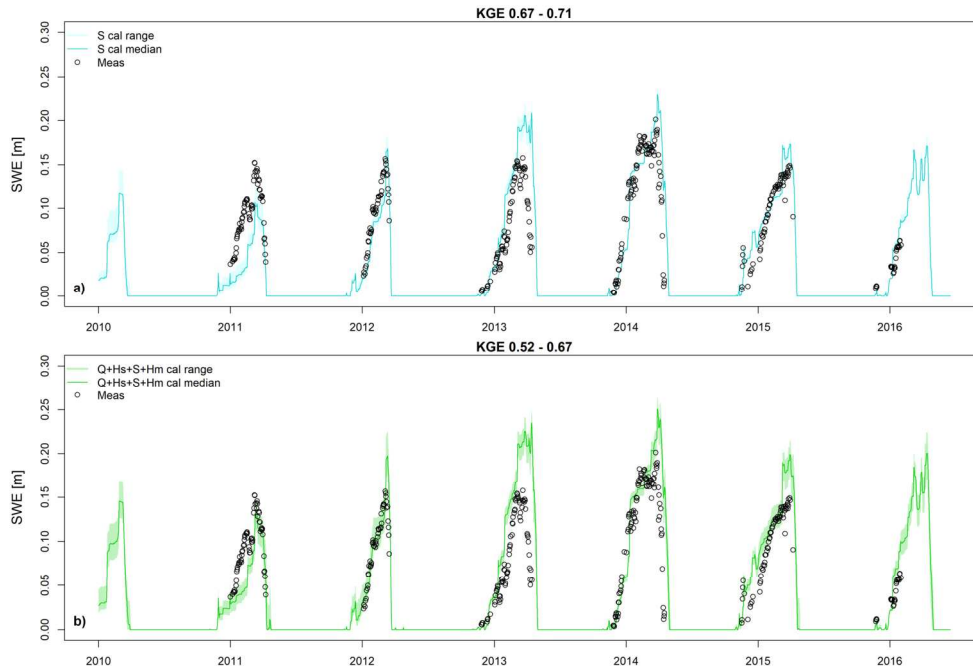


Figure 7: SWE simulations based on the different calibration targets listed in Table 2: solid lines are simulated medians and colours refer to calibration target. Shaded areas encompass the highest and lowest 100 simulations selected according to each calibration. Black dots are the SWE measurements.

177x118mm (300 x 300 DPI)

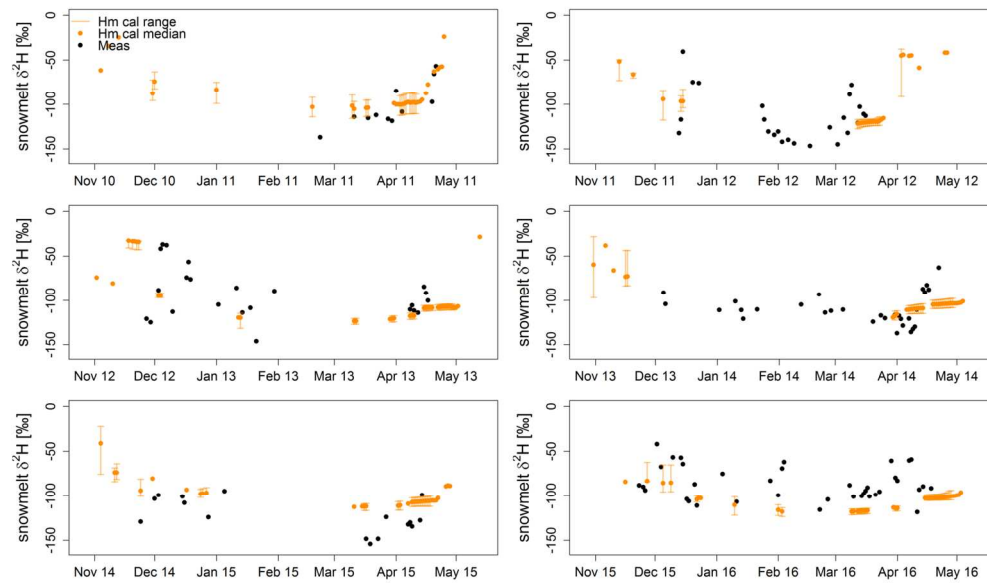


Figure 8: Snowmelt $\delta^2\text{H}$ based on the different calibration targets listed in Table 2, split by winter from 2010 to 2016: in each subplot, orange dots are simulated medians and colours refer to calibration target. Shaded error bars encompass the highest and lowest 100 simulations selected according to each calibration. Black dots are the snowmelt isotope samples.

177x107mm (300 x 300 DPI)

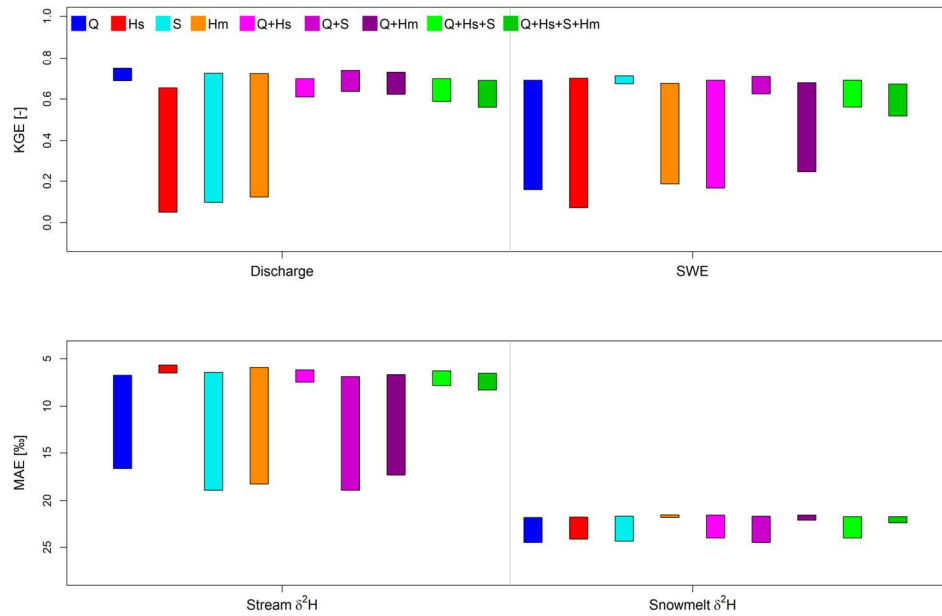


Figure 9: Comparison between maximum and minimum efficiency for each variable on the x-axis and each calibration target (by colour). The compared objective functions are KGE for discharge and SWE and MAE for streamwater and snowmelt isotope compositions. The MAE is reversed for improved readability, as the aim is to minimize MAE, while maximizing KGE.

177x118mm (300 x 300 DPI)

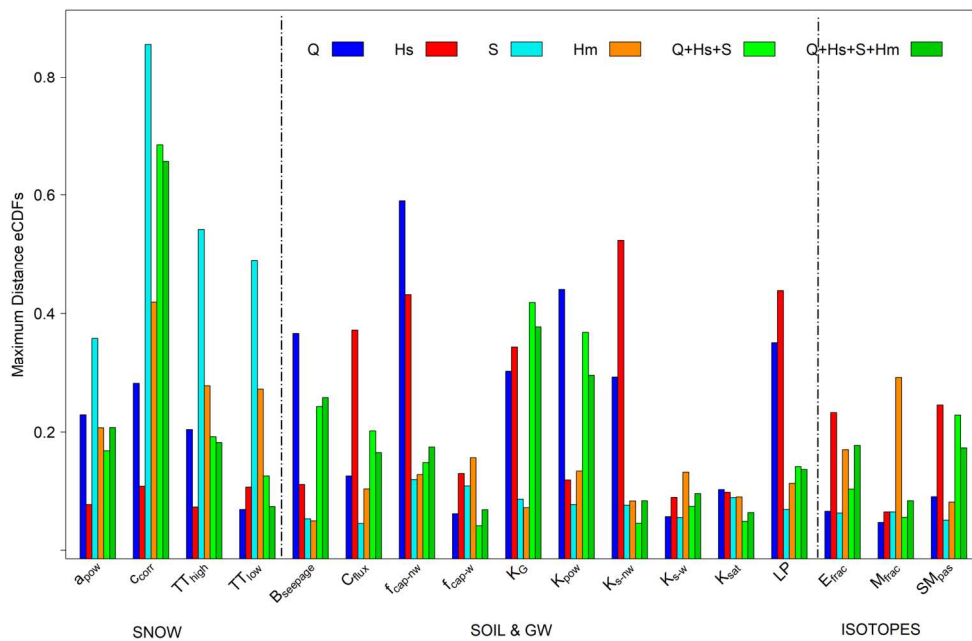


Figure 10: Comparison of parameter sensitivities based on the different calibration targets differentiated by colour (single variables, calibration of three variables and calibration of all variables). Ranking is based on the maximum distance of eCDFs before and after each type of calibration. The greater the maximum distance of eCDFs, the greater the model sensitivity to that parameter.

177x118mm (300 x 300 DPI)

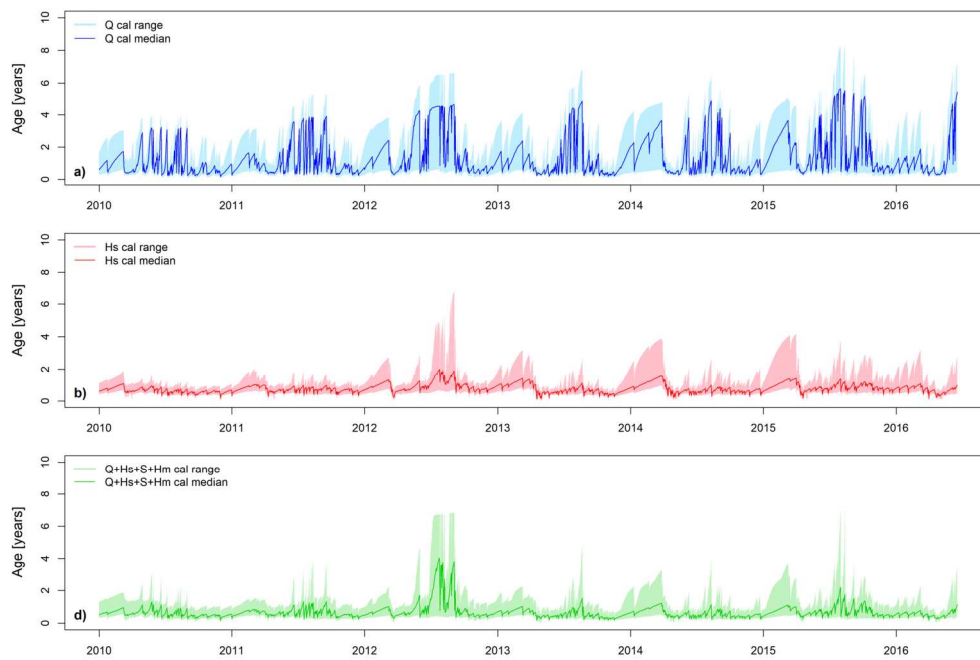


Figure 11: Estimated stream water ages based on different calibration targets. Solid lines are the simulated medians, colour refer to calibration target and shaded areas encompass the highest and lowest simulated values in 100 runs selected according to calibration.

177x118mm (300 x 300 DPI)

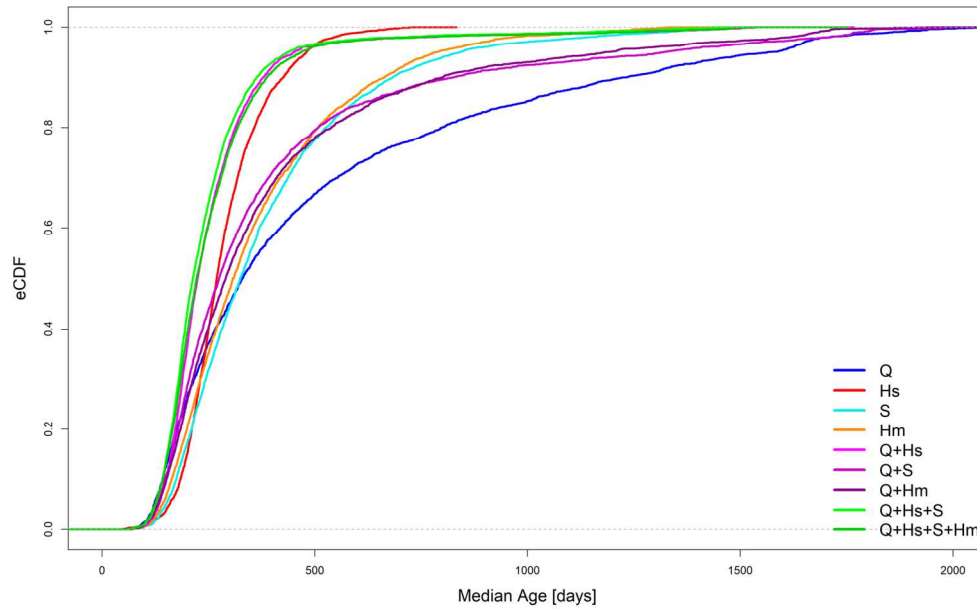


Figure 12: Empirical cumulative distribution functions (eCDFs) of the median water age (days) from selected 100 runs using the different calibration targets and KGE as objective function for Q (KGE_Q) and SWE (KGE_{SWE}), MAE of both streamwater and snowmelt isotope compositions (MAE_{Hs} and MAE_{Hm}).

177x118mm (300 x 300 DPI)