Environmental Modelling & Software 101 (2018) 301–316

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

What can we learn from multi-data calibration of a process-based ecohydrological model?

Sylvain Kuppel [a, *], Doerthe Tetzlaff [a, b, c], Marco P. Maneta [d], Chris Soulsby [a, b]

[a] Northern Rivers Institute, School of Geosciences, University of Aberdeen, AB24 3UF, United Kingdom
[b] Leibniz Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany
[c] Department of Geography, Humboldt University Berlin, Unter den Linden 6, 10099 Berlin, Germany
[d] Geosciences Department, University of Montana, 32 Campus Dr., Missoula, MT 59812-1296, USA

ARTICLE INFO

Article history:
Received 31 May 2017
Received in revised form
20 November 2017
Accepted 4 January 2018
Available online 12 January 2018

Keywords:
Catchment hydrology
Ecohydrology
Process-based modelling
Multi-objective calibration
Information content
EcH_2O

ABSTRACT

We assessed whether a complex, process-based ecohydrological model can be appropriately parameterized to reproduce the key water flux and storage dynamics at a long-term research catchment in the Scottish Highlands. We used the fully-distributed ecohydrological model EcH_2O, calibrated against long-term datasets that encompass hydrologic and energy exchanges, and ecological measurements. Applying diverse combinations of these constraints revealed that calibration against virtually all datasets enabled the model to reproduce streamflow reasonably well. However, parameterizing the model to adequately capture local flux and storage dynamics, such as soil moisture or transpiration, required calibration with specific observations. This indicates that the footprint of the information contained in observations varies for each type of dataset, and that a diverse database informing about the different compartments of the domain, is critical to identify consistent model parameterizations. These results foster confidence in using EcH_2O to contribute to understanding current and future ecohydrological couplings in Northern catchments.

1. Introduction

Numerical models are crucially important in the environmental sciences: models can complement and integrate theory and empirical data by incorporating testable hypotheses and by extending knowledge at spatial and/or temporal scales inaccessible to current observation methods. In particular, process-based models seek to explicitly represent the "*state variables and fluxes that are theoretically observable and can be used in the closure of assumed forms of the laws of conversation of mass, energy, and momentum at temporal scales characterizing the underlying physical processes*" (adapted from Fatichi et al., 2016). In contrast to conceptual and empirical approaches, physically-based models facilitate investigation of specific variables at local, process-specific scales (e.g., Endrizzi et al., 2014; Manoli et al., 2017; Niu and Phanikumar, 2015; Pierini et al., 2014). Additionally, a fully-distributed description of the simulation domain opens the possibility for tracking intra-system patterns and dynamics (e.g. Maxwell and Condon, 2016; Pierini et al., 2014), a task much less accessible to coarser spatial representations (i.e., lumped or semi-distributed models). Combining these two methodological choices with physically-based, fully-distributed models is thus a way to disentangle feedbacks and non-linear dynamics across fundamentally different processes (e.g. Drewry et al., 2010; Tague, 2009), and better predict system behaviour outside recorded environmental conditions (Seibert, 2003; Uhlenbrook et al., 1999). These tools are of particular relevance for the emerging field of critical zone science (National Research Council, 2012), which seeks integrated understanding of ecological, geological, geomorphological and pedological processes within a framework of hydrological partitioning (Brooks et al., 2015).

Within the field of hydrology the issue of appropriate model complexity is a focus of ongoing discussion. The corollary of expanding process-based approaches towards an "universal model" is an inevitable increase in complexity as explicit descriptions of additional system characteristics are added (e.g. topography, soil texture, tree height, canopy density etc.) (Band et al., 2001; Maxwell and Condon, 2016). Arguing that many of these numerous parameters cannot be appropriately measured,

* Corresponding author.
E-mail addresses: sylvain.kuppel@abdn.ac.uk (S. Kuppel), d.tetzlaff@igb-berlin.de (D. Tetzlaff), marco.maneta@mso.umt.edu (M.P. Maneta), c.soulsby@abdn.ac.uk (C. Soulsby).

https://doi.org/10.1016/j.envsoft.2018.01.001
1364-8152/

some fear that evolution of complex multi-disciplinary models only layer up unavoidable uncertainty and are prone to equifinality, whereby several combinations of parameter values –realistic or not– yield comparable performance (e.g. Beven and Binley, 1992; Beven and Freer, 2001; McDonnell et al., 2007).

The utility of measurements to help constrain the model solution space and identify feasible model configurations has been an increasingly central issue in hydrological model calibration. Sufficiently informative observations are necessary to ensure that the goodness of model-data fit attained effectively translates into physically-sound information for the internal model parameters; i.e., getting the right answers for the right reasons (Beven and Binley, 1992; Kirchner, 2006). The problem of equifinality—a particular case of *underdetermination* (Duhem, 1954)—is apparent when stream discharge is the only monitored variable available for calibration. Unfortunately, this remains the most common situation. The widespread use of streamflow time series to calibrate and validate models has spurred the development of elaborate single and multiple-criteria goodness-of-fit metrics (Kling et al., 2012; Krause et al., 2005; Legates and McCabe, 1999; Madsen, 2003; van Werkhoven et al., 2009) and calibration algorithms (Duan et al., 1992; Gupta et al., 1998; Sorooshian and Dracup, 1980; Tang et al., 2007; Tolson and Shoemaker, 2007) directed toward extracting a maximum of information content from this type of data (He et al., 2015; Rouhani et al., 2007; Shafii et al., 2017).

However, the information contained in streamflow time series is often insufficient to inform the parameterization of physically based models. Parameter values that represent physical properties of the catchment are usually poorly identified and become very sensitive to boundary conditions (Maneta et al., 2007). The situation deteriorates as more complex models incorporate increasingly detailed descriptions of catchment functioning. To constrain parameters of components associated with different subdomains of the model (ecological, surface, subsurface, etc.) it is desirable—but often impractical—to diversify data sources (Fang et al., 2013; Larsen et al., 2016; Rajib et al., 2016; Thorstensen et al., 2015). Combining different types of observations reduces information redundancy and provides direct insights into the different groups of physical processes represented in the model (Clark et al., 2011; Fatichi et al., 2016). A data-extensive approach to model calibration makes the choice of performance metrics easier because the information contained in observations is more directly related to the model compartment being calibrated (e.g. Birkel et al., 2014). Information diversity, however, brings other issues related to the assimilation of observations with diverse characteristics during calibration: some are technical e.g. combining spatio-temporal scales and associated uncertainties, while others are more fundamental to modelling, e.g. parameters compensating for model imperfections (Clark and Vrugt, 2006), or overlapping constraints and thus, possibly "pulling" the model in different directions (Efstratiadis and Koutsoyiannis, 2010). In other research fields, this approach is exemplified by the current efforts and associated challenges in assimilating multiple types of carbon cycle data to optimise Earth system models (Kaminski et al., 2013; Peylin et al., 2016).

The ecohydrology of high-latitude, energy-limited landscapes has traditionally been understudied despite the global ecological importance of this region. Since studies of plant-water couplings across disciplines gained momentum in the late 90s (Bonell, 2002), research efforts in ecohydrology have been primarily conducted in environments where water scarcity (Newman et al., 2006) or permanent presence (e.g., wetlands (Rodriguez-Iturbe et al., 2007)) makes hydrology an obvious, critical control upon how plants distribute and compete. Only recently, efforts have been directed towards understanding the specific ecohydrological processes of boreal, energy-limited regions (e.g. Cable et al., 2014) While there have been process-based model developments dedicated to the hydrology of high-latitude environments (e.g. Endrizzi et al., 2014; Kuchment et al., 2000; Lindström et al., 1997; Pomeroy et al., 2007), most model applications in these regions lack an explicit implementation of vegetation dynamics (e.g. Ala-aho et al., 2017a), and thus, cannot finely capture ecosystem imprints on water partitioning at the catchment scale.

High-latitude regions comprise mixed temperate forests, boreal forests and tundra, covering nearly 20% of the continental land mass (Tetzlaff et al., 2015a). These regions are subject to rapid climate change, with significant regional to global-scale implications (Hinzman et al., 2013), including shifts in precipitation regime and snow-mediated water balance (Bintanja and Andry, 2017; Jiménez Cisneros et al., 2014) and associated implications for runoff generation (Peterson et al., 2002; Zhang et al., 2014). While such environmental change has been observed to alter water pathways and flow regimes (Dye and Tucker, 2003; McClelland et al., 2006; Tetzlaff et al., 2013) and ecosystem dynamics (Naito and Cairns, 2015; Piao et al., 2008), further work is needed to identify the underlying mechanisms. Reasons for the limited understanding so far lie in the fine-scale landscape heterogeneity and the implications for spatial variation in energy inputs, as well as the logistical difficulties of collecting data in comparatively remote areas (Pomeroy et al., 2013; Tetzlaff et al., 2013), and the alarming recent decline in long-term monitoring of northern catchments (Laudon et al., 2017). However, we need to understand such processes and the related uncertainties of water cycling in these regions, while ongoing/projected biome shifts (e.g., (Beck et al., 2011; Williams et al., 2007)) call for particular scrutiny of ecosystem influence on water availability (Law, 1956) and vice-versa.

In this study, our main aim was to investigate to what extent a data-extensive approach to calibration can constrain the range of behavioural configurations of a highly-parameterized, physically-based model, such that the achieved parameter sets can be used as falsifiable hypotheses of the internal functioning of the catchment. For this, we used a distributed ecohydrologic model (EcH_2O, see (Maneta and Silverman, 2013) that integrates a kinematic hydrologic and energy balance model, with a vegetation dynamics model. The model is calibrated using several combinations of data types covering a range of ecohydrological variables collected at a long-term experimental northern montane catchment. We ask the following questions through our modelling experiments, *1)* what are the physical insights gained across ecohydrological processes? *2)* how valuable are the information contents brought by the different constraining datasets? Addressing these questions will help building a robust ecohydrological modelling framework dedicated to critical zone functioning in high-latitude environments.

2. Material and methods

2.1. Study site

The Bruntland Burn (Fig. 1) is a small catchment (3.2 km^2) located in the eastern Scottish Highlands (57°8′N 3°20′W). It is a headwater of the River Dee, which provides drinking water for the city of Aberdeen (250,000 people), ecosystem services such as an Atlantic salmon fishery, and has EU conservation designations. The region receives around 1100 mm of average annual precipitation (P), distributed quite evenly throughout the year, although November–February and June–August are usually wettest and driest periods, respectively. Less than 5% of P occurs as snowfall. The climatic water balance is energy-limited, with 400 mm of annual Potential Evapotranspiration (PET). The mean annual

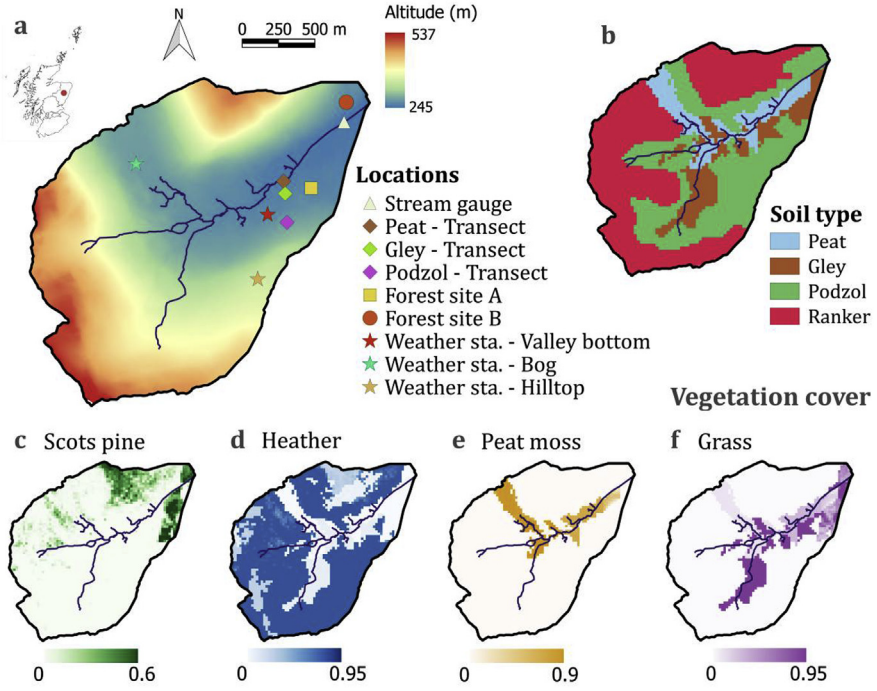

Fig. 1. Bruntland Burn catchment, showing **(a)** topography, stream network, and the monitoring locations and associated instrumentation: stream gauge (triangle), soil moisture sensors (diamonds and circle), square for sapflow transpiration (square and circle), and micrometeorology (including net radiation, star). The conceptualization used for simulations (b-f, $30 \times 30\ m^2$ resolution) comprises **(b)** pedology, aggregated from the Hydrology of Soil Types (HOST) classes, and **(c–f)** the pixel fraction covered by the four considered vegetation types (in addition to scree/bare soil, not shown).

temperature (T) is 7 °C with no monthly-averaged T below 0 °C in a transitional temperate/boreal oceanic climate.

The local topography reflects glacier retreat, with a wide valley bottom (~220 m.a.s.l.) surrounded by steeper slopes reaching up to 560 m.a.s.l. (Fig. 1a). This slope gradient is reflected by widespread glacial drift deposits (60% of the catchment) with depths ranging from ~40 m in the valley bottom to ~5 m on steeper slopes. These deposits are mostly saturated and form significant groundwater reservoirs that sustain stream base flow and maintain wet conditions in the valley bottom (Soulsby et al., 2016). The pedology comprises deep (0.5–4 m), organic-rich soils (histosols: peat and gley) in the riparian area bordering the stream channel network (Fig. 1b). These soils are persistently saturated and rapid overland flow is the dominant runoff generation mechanism following rainfall events (Tetzlaff et al., 2014). Hillslopes are characterized by shallower, freely-draining podzols (spodosols) overlying moraines and marginal ice deposits, while thin regosols (rankers) dominate above 400 m.a.s.l. Where the drift is thin or absent (Fig. 1b). These hydropedological units are somewhat reflected in the vegetation cover (Fig. 1c–f). Podzols and rankers predominantly support heather shrublands (*Calluna vulgaris* and *Erica* spp.), though this land cover is the result of overgrazing by red deer (*Cervus elaphus*) and sheep. Scots pine trees (*Pinus sylvestris*), the naturally-occurring vegetation, is confined to the northern steep hillslopes and to plantation stands near the catchment outlet. Riparian gley soils are characterized by herbaceous cover (*Molinia caerulea*), the latter being also found as secondary species in the peat where bog mosses (*Sphagnum spp.*) dominate the land cover.

2.2. The ecohydrological model EcH$_2$O

We used a new formulation of the spatially-distributed, process-based model EcH$_2$O (Maneta and Silverman, 2013). Here, EcH$_2$O couples a two-layer (canopy and understory) vertical energy balance scheme (Fig. 2a), a kinematic hydrologic module solving vertical and lateral water transfers (Fig. 2b), and a transpiration-based simulator of carbon uptake and allocation for plant growth (Lozano-Parra et al., 2014; Maneta and Silverman, 2013). The reasons for choosing EcH$_2$O lie in its original development aimed at filling a research gap between hydrology-focused catchment models and land surface models (LSMs) simulating biophysical and biogeochemical cycles in the critical zone. While catchment models provide a fit-for-purpose conceptualization of water pathways, in most cases they lack a process-based representation of energy balance and plant-water interactions. On the other hand, most state-of-art LSMs have historically been developed as surface components of climate models, to be run over large regions or continents. Despite the recent advances in representing land processes such as vegetation phenology and carbon/nutrient cycles, hydrology remained simplistic in most LSMs; overland flow routing, channel routing, and lateral subsurface flow are typically neglected or highly simplified (see Fan (2015) for a further discussion). Without these components, it is not possible to study the ecohydrological effects of upstream water subsidies, and the spatial organization of catchments imprinted by the water redistribution network. This is especially critical for studies in small catchments at high spatial resolution, like the one studied here. Finally, the parsimonious implementation of EcH$_2$O was preferred over other recent, potentially more sophisticated, ecohydrological models (e.g., Fatichi et al., 2012; Maxwell and Condon, 2016).

EcH$_2$O's calculation of energy and water fluxes, water storage dynamics and vegetation states is made at time steps subordinated to that of the meteorological forcing. The atmospheric boundary conditions for each time step are: Precipitation (P), incoming shortwave radiation (R_{SW}), downwelling longwave radiation (R_{LW}), T (maximum $T_{a,max}$, minimum $T_{a,min}$, and average $T_{a,mean}$ during the time step period), relative humidity (h_r) and wind speed (W_{xy}). New developments that we implemented in EcH$_2$O are

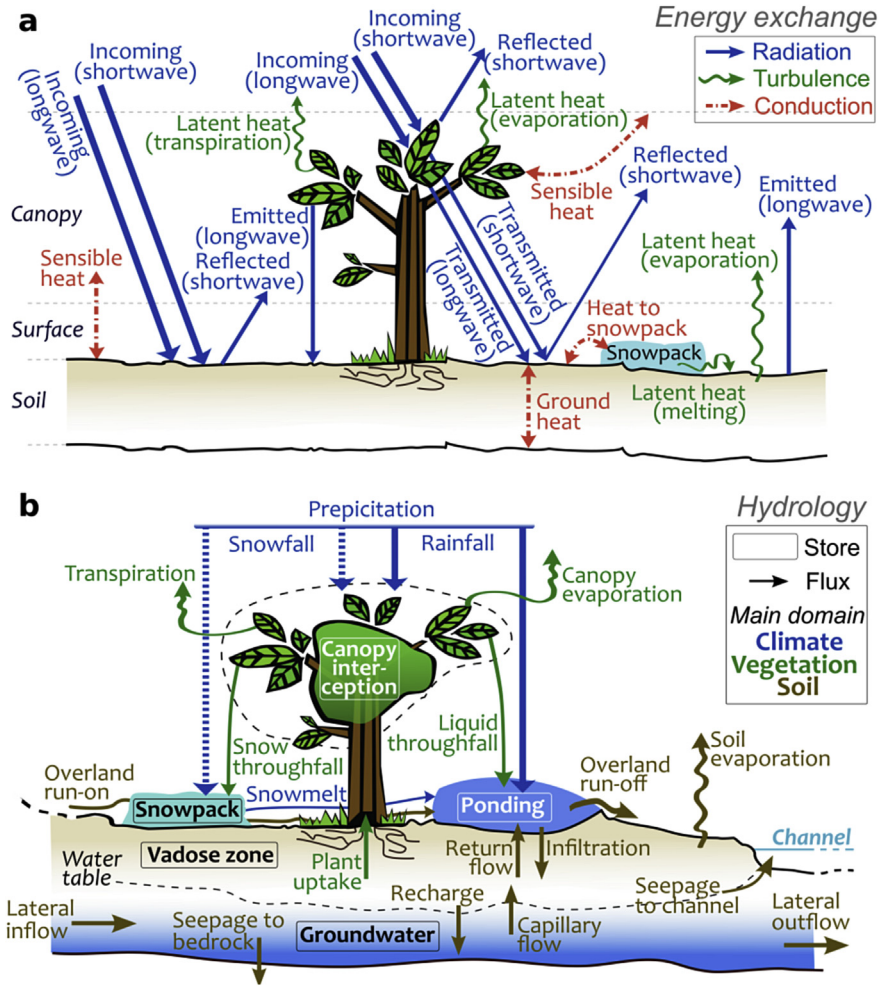

Fig. 2. Schematics of the processes taken into account in **a**) the energy balance and **b**) the hydrologic modules of the EcH$_2$O model.

documented in the Appendices, to complement other model details described elsewhere (Lozano-Parra et al., 2014; Maneta and Silverman, 2013). Here, we provide a brief summary of the model philosophy and main features.

The spatial domain of EcH$_2$O is mapped on a regular grid defined by that of the input digital elevation data. Each cell of the domain can have multiple vegetation covers (including bare soil). The energy and the water balance is solved for each cover and integrated over the cell area according to the fraction of the cell they occupy. The energy balance equations are solved at the top of the canopy (see Appendix A1 for details) and then at the soil or snowpack surface. Turbulent fluxes (momentum, heat, and vapour) are resolved using a first-order, local closure approximation under gradient similarity theory, valid for small eddies and neutral stratification conditions.

Canopy interception is simulated using a linear bucket approach. The partition of P and throughfall between the solid (snow) and liquid components during a time step is done according to the minimum and maximum air T during the time step, and to a snow-rain transition T threshold. Snowpack melt and liquid throughfall feed surface ponding which infiltrates following a Green and Ampt approximation of the Richard's equation (Mays, 2010). All ponded water at the end of each time step becomes overland flow, i.e., run-on to the downstream cell where it can reinfiltrate or, in turn, generate further overland flow. This calculation cascade follows the local drainage direction until the remaining surface water reaches the outlet or a channel cell. Flow in a channel is routed using a 1D solution for a kinematic wave (see Appendix A2 for details).

The soil is divided into three hydraulic layers: the shallow topsoil where soil evaporation takes place, the intermediate layer which typically shares the bulk of the roots with the topsoil, and the bottom layer where groundwater can be transferred laterally to the downstream cell or seep into the stream. Vertical water redistribution is based on the theory that only soil moisture in excess of field capacity (gravitational water) can move under gravity to deeper soil layers or laterally to the next cell downstream. Diffusive effects driven by local pressure gradients are therefore assumed to be negligible, and water below field capacity is retained by the soil and only removed by evapotranspiration.

Simulation of vegetation dynamics are adapted from the 3-PG and TREEDYN3 models (Bossel, 1996; Landsberg and Waring, 1997; Peng et al., 2002), with differentiated carbon allocation and growth schemes for ligneous and herbaceous species (Lozano-Parra et al., 2014). A Jarvis-type model is used to simulate the response of canopy conductance to environmental drivers (Cox et al., 1998; Jarvis, 1976).

2.3. Model setup and landscape characterization

All simulations were performed at a $30 \times 30\,m^2$ resolution. A LiDAR-derived $1 \times 1\,m^2$ DEM (Lessels et al., 2016) was used to delineate the catchment boundary, and further processed with the

PCRaster tool suite (http://pcraster.geo.uu.nl/) to obtain local slopes and local drainage direction; the latter was determined for each cell using the steepest descent among the eight adjacent cells (D8 algorithm, Fairfield and Leymarie, 1991). The model was run at daily time steps. Simulations covered a 64-month-long time period from June 2011 to September 2016, with the period from June 2011 to October 2012 used for model spin-up and therefore, discarded from the analysis (see Sect. 2.5).

P, $T_{a,mean}$, $T_{a,min}$, $T_{a,max}$, h_r, and W_{xy} data were collected at three meteorological stations installed in the catchment in different landscape positions: valley bottom, bog, and hilltop (Fig. 1a) and used from July 2014. Prior to that period, P was interpolated using a square elevation inverse distance-weighted algorithm applied to five Scottish Environment Protection Agency (SEPA) rain gauges located around the Bruntland Burn catchment within 10 km, similarly to (Birkel et al., 2011), while $T_{a,mean}$, h_r and W_{xy} fields were taken from the Balmoral station (~5 km NW) as available from the Centre for Environmental Data Analysis (CEDA) (Met Office, 2017). $T_{a,min}$ and $T_{a,max}$ (prior to July 2014), R_{SW} and R_{LW} (whole study period) were retrieved from the ERA-Interim climate reanalysis (Dee et al., 2011). Finally, we took into account altitudinal effects on P and T by respectively assuming a 5.5% increase of P every 100 m.a.s.l. as measured along a hillslope covering 200 m elevation difference (Ala-aho et al., 2017b), and a decrease of 0.6 °C/100 m.a.s.l based on the moist adiabatic temperature lapse rate (Goody and Yung, 1995).

Soil hydrological properties were mapped by splitting the model domain into four hydropedological units aggregated from the soil classes defined by the Hydrology of Soil Types (HOST) (Fig. 1b) (Tetzlaff et al., 2007). Energy-related soil characteristics were considered as spatially uniform. Five land cover types were considered: Scots pine stands, heather moorland, peat moss, grassland and bare rock/scree. Vegetation fraction of Scots pine in each cell (Fig. 1c) was estimated by aggregating a $1 \times 1\ m^2$-resolution LiDAR canopy cover measurements conducted over the catchment to the $30 \times 30\ m^2$ grid used for simulations. For the other vegetation fractions, we additionally used the soil classification, extensive land use mapping and aerial imagery (Tetzlaff et al., 2007). Heather was assumed to occupy 95% of the treeless surface of podzols and rankers, except for the steep northern rocky hillslopes (40% cover in treeless areas), a few sparsely vegetated moorlands on the west hillslopes (20% cover in treeless areas) and the gley (5% cover) (Fig. 1d). Peat moss was assumed to occupy 90% of an extensive raised peat bog in the NW parts of the catchment (see Fig. 1 in (Sprenger et al., 2017), and 70% of the remaining peat soils (Fig. 1e). In the latter areas, *Molinia* grasses compete (30% cover) with peat moss, while grasslands are dominant on gley soils (95% of the tree-and-shrub-free surface) and in patches of managed lands near the catchment outlet (Fig. 1f).

The list of model parameters calibrated in this study is given in Table 1. We selected 16 types of parameters, based on a preliminary sensitivity analysis following (Morris, 1991) performed with the calibration dataset described in section 2.4. As shown in Table 1, 10 parameters are soil-dependent and 6 are vegetation-specific. Since we considered 4 soil types and 4 vegetation types, the total number of calibrated variables was 64 ($10 \times 4 + 6 \times 4$). The uncalibrated model parameterization, including initial conditions, was prescribed based on literature values and expert knowledge (Supplementary Table S1).

2.4. Calibration and evaluation datasets

A specific advantage of the Bruntland Burn experimental catchment is the length and diversity of data records across eco-hydrological processes which is unusual for northern regions. We made use of 10 datasets for the calibration as summarized in Fig. 3. Daily discharge at the catchment outlet (Fig. 1a) was derived from 15-min stage height records (Odyssey capacitance probe, Christchurch, New Zealand) using a rating curve calibrated for a stable stream section.

Soil moisture data was collected at 15-min intervals at four locations: three of them along a transect representative of the main hydropedological units —podzol, gley and peat— on heather (Tetzlaff et al., 2014), and one plot in a Scots pine forest (Wang et al.,

Table 1
Calibrated parameter types and their allowed ranges of variation during calibration, for each of their four components: soil units or vegetation types (‡: logarithmic scale used for sampling).

Parameter		Calibration range			
Name	Description				
Soil-distributed					
		Peat	Gley	Podzol	Ranker
D_{soil}	Total soil depth (m)	0.8–8	0.8–8	0.65–4	0.65–4
D_{L1}	Depth of the 1st hydrological layer (m)	0.02–0.25	0.02–0.25	0.01–0.2	0.01–0.2
D_{L2}	Depth of the 2nd hydrological layer (m)	0.02–0.25	0.02–0.25	0.01–0.2	0.01–0.2
H	Porosity ($m^3.m^{-3}$)	0.7–0.98	0.4–0.9	0.3–0.6	0.3–0.6
K_{hx}	Saturated horizontal hydraulic conductivity ($m.s^{-1}$) ‡	10^{-7}–0.01	10^{-5}–10^{-3}	10^{-4}–0.1	10^{-4}–0.1
K_{hratio}	Ratio of vertical-to-horizontal hydraulic conductivity (–) ‡	10^{-3}–0.4	10^{-3}–0.6	10^{-3}–0.6	10^{-3}–0.6
λ_{BC}	Brooks-Corey exponent parameter (–)	3–12	3–12	3–12	3–12
ψ_{ae}	Air-entry pressure head (m)	0.05–0.8	0.05–0.8	0.05–0.8	0.05–0.8
θ_r	Residual soil moisture	0.02–0.2	0.02–0.2	0.02–0.2	0.02–0.2
k_{root}	Exponential root profile (m^{-1})	5–40	5–40	5–40	5–40
Vegetation-distributed					
		Pine	Heather	Peat moss	Grass
gs_{max}	Maximal stomatal conductance ($m.s^{-1}$) ‡	0.003–0.05	0.003–0.05	0.003–0.05	0.003–0.05
CWS_{max}	Maximum interception storage per unit LAI (m) ‡	10^{-4}–0.005	10^{-4}–0.005	10^{-4}–0.005	10^{-4}–0.005
T_{opt}	Optimal photosynthesis temperature (°C)	6–24	6–24	6–24	6–24
ψ_d	Soil water potential halving stomatal conductance (-m)	0.5–8	0.5–8	0.5–8	0.5–8
c	Sensitivity of stomatal conductance to soil water potential (–)	0.5–5	0.5–5	0.5–5	0.5–5
K_{beer}	Light attenuation coefficient (–)	0.3–0.6	0.3–0.6	0.3–0.6	0.3–0.6

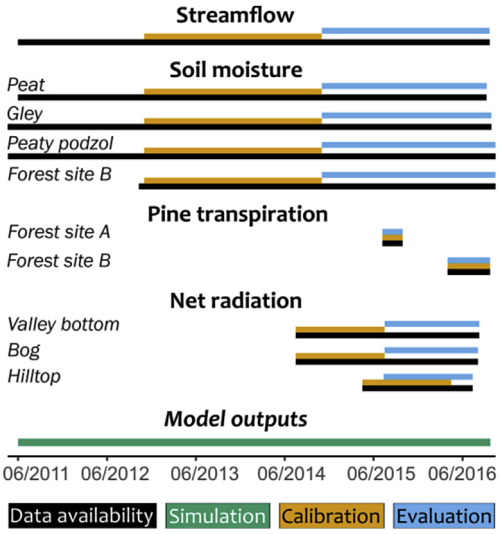

Fig. 3. Temporal window –at daily resolution– covered by each of the considered data sets (black), used to constrain the model during the calibration period (orange) after leaving at least a full year of spinup for simulations (green), and then to evaluate the calibrated model (blue). Note the shortness of some data sets (e.g. transpiration) explains the (partial or total) overlap between some calibration and evaluation periods. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2017a) (Forest site B, Fig. 1a). We used time domain reflectometry (TDR) soil moisture probes (model CS616, Campbell Scientific, Inc. USA) located 0.1, 0.2 and 0.4 m beneath the surface –corresponding to the main soil horizons (Geris et al., 2015)–except in the peat where only two probes were present, at depths of 0.1 and 0.2 m. Additionally, each group of probes was replicated at the same depths but ~2 m apart. These TDR sensors were calibrated using laboratory analyses of gravimetric soil water content and bulk density, from samples collected at each horizon (Geris et al., 2015). Finally, a single daily, vertically-averaged volumetric water content value was used for calibration and evaluation of the EcH_2O model.

Scots pine transpiration (*Tp*) was measured between July and September 2015 at Forest site A and between April and September 2016 at Forest site B (Fig. 1a), using 32 sets of Granier-type thermal dissipation sap flow sensors (Dynamax Inc., Houston, USA) installed on 10 and 14 trees in the Forest sites A and B, respectively, with 2-to-4 sensors per tree depending on the stem size (10–32 cm in diameter). Average stand-scale pine transpiration was derived using a sapwood-area-to-tree-diameter relationship estimated from incremental wood cores sampling in surrounding trees at the end of the study period (see (Wang et al., 2017a) for more details), and then daily averaged. Net radiation (*Rn*) was measured every 15 min at the three meteorological stations (Fig. 1a), and then daily averaged.

Finally, to provide a novel independent verification of the model's ability to represent seasonal storage dynamics, we compared modelled and empirical-based estimates of catchment-scale saturation area from June 2011 to September 2014. A conceptual rainfall-runoff model that linked antecedent wetness and soil moisture to saturation area was used to estimate the extent of saturation in the catchment. This model was previously calibrated against maps of measured saturation area extent and isotopic tracers measurements (Ali et al., 2014; Birkel et al., 2010). These estimates were compared with saturation extent in EcH_2O, which was defined as the proportion of cells in the domain where the volumetric water content in the top hydraulic layer exceeded 99% of soil porosity.

2.5. Model-data fusion method

We sampled the parameter space by conducting 100,000 Monte Carlo simulations using uniform parameter distributions with prescribed bounds based on literature values and prior experience (Table 1). Each run spanned the entire 64-month simulation period, but the first 16 months of each run were used to spin-up the model and stabilize water storages and flux simulations (notably soil moisture and stream discharge). After discarding the spin-up period, the calibration dataset was split whenever possible into two non-overlapping periods, one for calibration and one for evaluation (Fig. 3). However, this was not possible for the transpiration dataset (3 and 6 months long) or for the net radiation at the hilltop weather station (15 months long), and no split-sample evaluation was performed for these variables. For each simulation, the goodness-of-fit (GOF) for the calibration and for the evaluation datasets was quantified. We used the mean absolute error (MAE) for stream discharge, because other metrics based on squared model-data difference, such as root mean square error (RMSE) and Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970), are known to overemphasize the constraint brought by high-flow measurements and neglects low-flow portions of the dataset (Krause et al., 2005; Legates and McCabe, 1999). In addition, high flow measurements typically carry more uncertainty and in our system flow conditions vary over several orders of magnitude (see Sect. 3). Conversely, all other observables (volumetric water content, pine transpiration and net radiation) comparatively display a much more compact and symmetric distribution, with median values close to mean values. RMSE has been recommended when no information is given on model error distribution–the latter is then conservatively assumed as being Gaussian–(Chai and Draxler, 2014), which is why we chose this metric in this case. For cross-variable evaluations of model performance after calibration, the fit between the respective dispersion of model and data time series was quantified using Pearson's correlation coefficient *r*, complemented by the dataset-normalized RMSE extended to all variables, where model-data biases are notably taken into account.

To investigate the extent to which each type of measurement is capable of informing the calibration of a wide range of model parameters, we assessed multiple calibration scenarios each using a different subset of the 10 datasets for a multi-variable, multi-site calibration run as described in (Ala-aho et al., 2017b). In this method the GOF functions are used as informal measures of the likelihood. The GOF for each dataset was calculated for 100,000 Monte Carlo runs (MAE for streamflow, RMSE for the other outputs), and the dataset-specific empirical cumulative distribution function (CDF) of these was determined. Next, these CDFs were used to identify the 30 "best" model runs. The method iteratively identifies the quantile threshold common to all corresponding GOF's CDFs, below which the GOF of exactly 30 model runs *simultaneously* meets the calibration target for the objectives. In the case of using only one dataset as a constraint, this is equivalent to finding the 30 smallest values for MAE or RMSE. Although it remains an aggregative approach to the multi-objective problem (Cohon, 1978), this simple method advantageously avoids having to combine the different GOFs into a single numerical objective function. Note also that likelihood estimates are not used to guide the exploration of the model parameter space, which eliminates potential search biases if the characteristics of the model error residuals are incorrectly specified in the likelihood function. Overall, all this has five important advantages: 1) no need to choose pooling weights to combine the individual objective into a single function, 2) results are less sensitive to the choice of factors used to scale observations in dimensional GOF functions, 3) no compensatory effects between well-performing and poorly-fitted runs can occur,

4) less risk that potentially 'good' sections of the parameters space may be left unexplored if the specification of model errors are incorrect, and 5) independence of runs make the process trivially parallelizable. The first three of these advantages address some of the classical shortcomings in aggregated objective functions (Efstratiadis and Koutsoyiannis, 2010). The last two relax some of the disadvantages associated with more formal Markov Chain-based search methods.

Lastly, predictive uncertainty (PU) was taken as the 90%-spread for each simulated daily values across the 30 best runs (thus, avoiding making assumptions about the output distribution), then averaged over the whole simulation period (excluding the initial spin-up). Depending on the analysis being carried out (see Sect. 3 and 4), this uncertainty is kept dimensional (Eq. (1a)) or normalized (Eq. (1b)):

$$PU = \frac{1}{N_{eval}} \sum_{i=1}^{N_{eval}} M_{95}(t_i) - M_5(t_i), \tag{1a}$$

$$PU^* = \frac{1}{N_{eval}} \sum_{i=1}^{N_{eval}} \frac{M_{95}(t_i) - M_5(t_i)}{\langle |M(t_i)| \rangle_{\text{best runs}}}, \tag{1b}$$

where $M_{95}(t_i)$, $M_5(t_i)$, and $\langle |M(t_i)| \rangle$ are respectively the 95$^{\text{th}}$-percentile, the 5$^{\text{th}}$-percentile and average absolute value for the i-th time step in the evaluation period.

3. Results

3.1. Simulation of multiple data time series

The model captures well the main characteristics of the stream hydrograph for the 45 months period (Oct 2012–June 2016) shown in Fig. 4. Moderate and high flow conditions are well reproduced, with a slight underestimation of low flows during summers, especially in 2013 and 2015. The simulation of the hydrograph shows minor differences when the model was calibrated using the entire suite of observations versus using only streamflow data, with modified Kling-Gupta Efficiency (KGE) scores (Kling et al., 2012) over the evaluation period (Fig. 3) ranging from 0.60 to 0.95 across these best runs (not shown). In both calibration cases, the 90%-spread interval shows that the dispersal among the 30 best runs remains similarly small with respective PU values of 0.023 and 0.041 m^3 s^{-1}.

Fig. 5 shows the time series of volumetric soil water content (θ) in the shallow subsurface at the four monitored sites (shown in Fig. 1a). At each location, the depth-averaged measured data (see sect. 2.4) is compared to thickness-weighted averages of simulated θ in the two upper layers of EcH$_2$O in three calibration scenarios: using the local soil moisture dataset, using all four soil moisture datasets, and constraining against all datasets. The model generally provides consistent results in the peat in terms of timing and amplitude of θ dynamics (Fig. 5a). However, simulated soil moisture is often too prompt in rewetting the peat in autumn, while it displays an unrealistic drying event in the summer of 2015. Similar observations can be made about the gley (Fig. 5b) where the model tends to underestimate the annual amplitude because simulated soil saturation is reached at lower volumetric water contents than observed. In both peat and gley locations, the different calibration scenarios display a very similar average behaviour across best runs, but the 90%-spread interval among time series grows significantly when adding more constraints, with predictive uncertainties of 0.04, 0.11, and 0.22 m^3 m^{-3} (peat) and 0.02, 0.08, and 0.25 m^3 m^{-3} (gley) when respectively calibrating using θ measurements in peat, all θ data, and all datasets. At the podzol at the upslope end of the transect, the model satisfactorily captures soil moisture dynamics at daily-to-seasonal time scales (Fig. 5c). High-frequency peaks of θ tend, however, to be underestimated by EcH$_2$O, this discrepancy becoming more marked when adding more constraints in the calibration. Similar behaviour occurs in the podzolic soil at Forest site B (Fig. 5d): while the simulated high-frequency dynamics

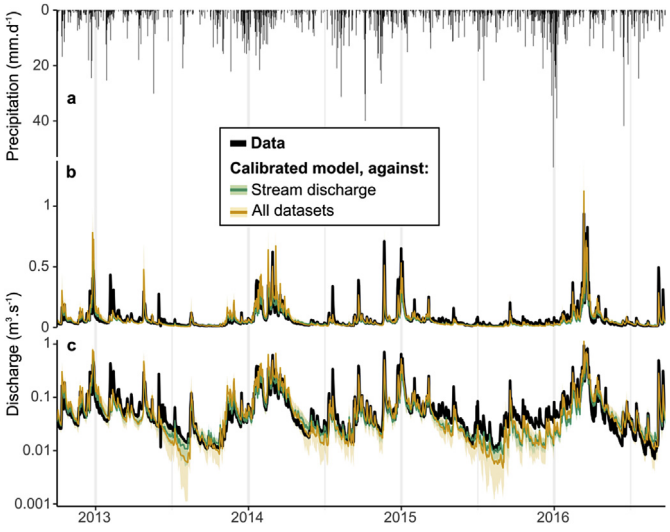

Fig. 4. Time series of **(a)** measured precipitation, and stream discharge, with **(b)** normal and **(c)** logarithmic scale the data in black and in colour the average of the 30 best and 90%-spread interval for two types of calibration: (green) using only stream discharge and (yellow) when simultaneously constraining the model against all datasets. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

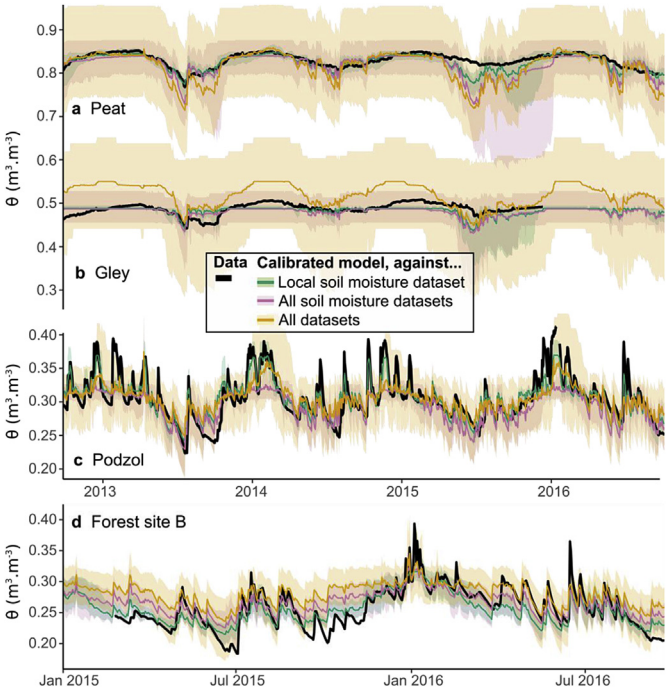

Fig. 5. Time series of profile-averaged soil volumetric water (**a**–**c**) in the transect–(**a**) peat plot, **b**) gley plot and **c**) podzol plot–and **(d)** at Forest site B, showing data in black and in colour the average of the 30 best runs and 90%-spread interval for three types of calibration: (green) using only the plotted quantities, (pink) using all soil moisture datasets and (yellow) using all collected datasets. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

are consistent with the measurements, the model markedly underestimates long-term θ variations. In both podzol locations, the increase in predictive uncertainty as constraints are diversified is less marked than in the valley bottom (peat and gley).

Summer pine transpiration is well simulated by EcH$_2$O at both Forest sites A and B (Fig. 6). This is particularly true when calibrating the model against local and all-transpiration datasets. Adding all other constraints leads EcH$_2$O to underestimate some peak values at Forest site B (Fig. 6b), while in this configuration the baseline transpiration becomes underestimated at Forest site A (Fig. 6a).

Fig. 7 compares measured net radiation to the simulated top-of-canopy value—averaged over vegetation and bare soil fractions—at the three weather stations. In all shown cases, the temporal dynamics of the seasonal signal are well reproduced by the model, as

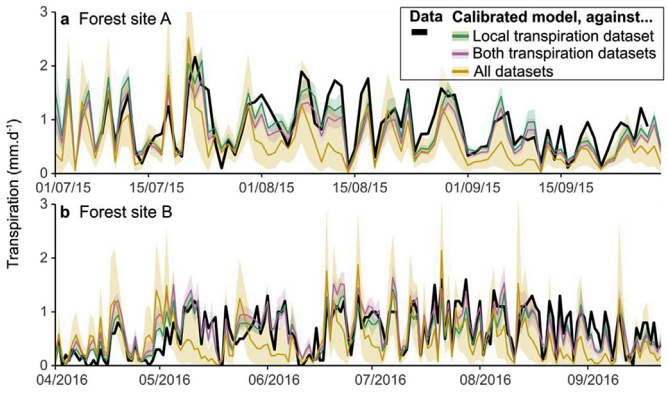

Fig. 6. Time series of measured and simulated pine stand transpiration at **(a)** Forest site A and **b)** Forest site B, showing data in black and in colour the average of the 30 best runs and 90%-spread interval for three types of calibration: (green) using only the plotted quantities, (pink) using all soil moisture datasets and (yellow) using all collected datasets. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

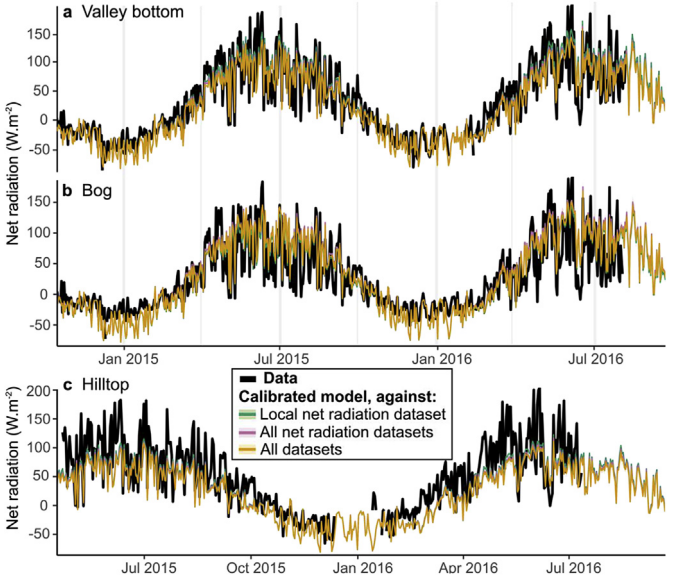

Fig. 7. Time series of measured and simulated net radiation at the 3 weather stations: **a)** Valley bottom, **b)** Bog and **c)** Hilltop, showing data in black and in colour the average of the 30 best runs and 90%-spread interval for three types of calibration: (green) using only the plotted quantities, (pink) using all soil moisture datasets and (yellow) using all collected datasets. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

are the day-to-day fluctuations, with a very small dispersal among best runs (PU < 5.5 W m^{-2}). However, EcH$_2$O tends to underestimate net radiation at all three sites, which may indicate an overestimation of soil temperatures to compensate for potentially low evaporative losses (see Eq. (A1) and discussion in sect. 4.1). This feature is especially marked at the heather-dominated hillslope location (Fig. 7c) where the simulated summer net radiation is only half of the observed values.

3.2. Overall performance and uncertainty reduction

The model performances are summarized using heat maps in the dual space of calibration scenarios and simulated variables (Fig. 8), quantifying model-data correlation ($r_{m,o}$, Fig. 8a) and data-average-normalized RMSE across the evaluation period ($RMSE^*_{m,o}$, Fig. 8b). In these plots, the columns θ_{all}, Tp_{all}, Rn_{all} and *All* show the metrics averaged over one type of output or all outputs. The most notable feature in these plots is that stream discharge is well reproduced regardless of the datasets used for calibration, with $r_{m,o} > 0.89$ ($p < .001$) in all cases and $RMSE^*_{m,o}$ values between 0.53 and 0.68. Secondly, the temporal dynamics of volumetric water content is generally best captured in the podzolic soils (podzol at transect and Forest site B, Fig. 8a), while the lowest $RMSE^*_{m,o}$ for soil moisture (θ) are found in the peat (Fig. 8b). Soil moisture is reasonably simulated in the gley only when the corresponding dataset is included in the calibration, although model-data

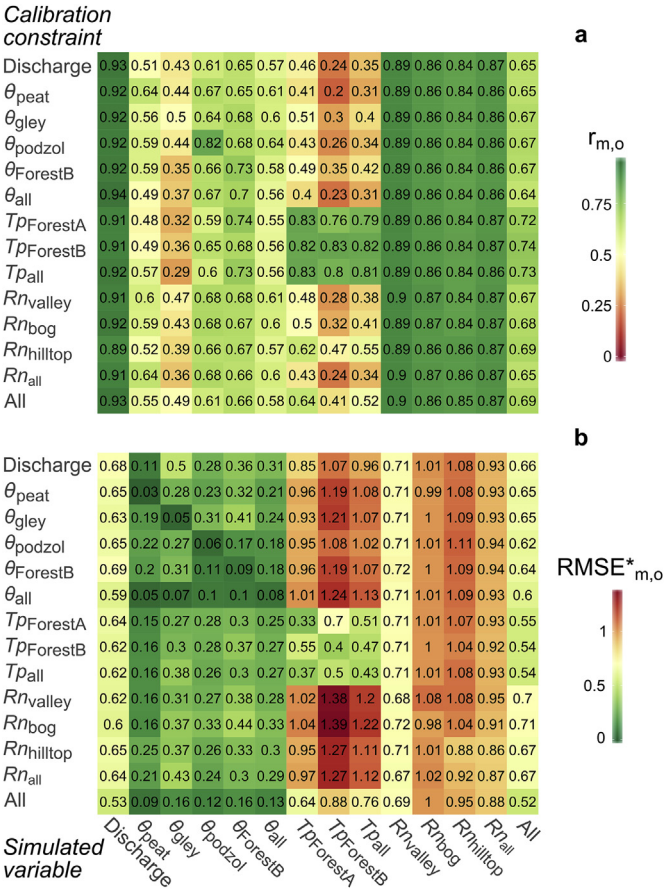

Fig. 8. Heat map of model-data fit over the evaluation period, as measured by **(a)** the Pearson's correlation coefficient and **(b)** normalized root mean square error averaged over the 30 best runs. The horizontal axis gives the variable or group or variables evaluated, depending on the dataset or combination of datasets used as a constraint over the calibration period (vertical axis).

correlation there remains low in most cases. Similarly, it is found that accurate simulation of transpiration is achieved only when the observations of transpiration are included in the calibration (rows Tp_{ForestA}, Tp_{ForestB}, Tp_{all} and *All*). Fourth, the scores for net radiation are somewhat insensitive to the calibration scenarios, with good temporal dynamics but high $RMSE^{*}_{m.o}$ due to the recurrent underestimation by the model mentioned in 3.1.

For most individual observables (column 1 to 5, 7 to 8, and 10 to 12) the all-datasets calibration scenario (last row) yields the highest scores, surpassed only by the scores of each simulated state calibrated against its direct observation (diagonals of Fig. 8). The improved model-data fit with more constraints becomes clearer when the average scores over observable types are considered (column 1, 6, 9, and 13). Finally, using all datasets as a simultaneous constraint yields the lowest overall model-data misfit across observables (bottom right square in Fig. 8b, as compared to the rest of the last column).

The impact of different calibration scenarios on individual and overall predictive uncertainty of simulations, i.e., the average 90%-spread interval among best runs (PU*, normalized as defined in Eq. 1b), is shown in Fig. 9 –following the same layout of Fig. 8. When the model is calibrated against individual datasets, PU × values for corresponding observables remain below 0.5 (diagonals in Fig. 9) and generally below 1 when simulating other variables (off-diagonal squares), even if the latter is not included in the calibration (e.g., uncertainty of θ_{peat} when calibrating against streamflow only–1st row and 2nd column). Notable exceptions are streamflow where PU × remains above 1 whenever discharge is not included in the calibration, and simulated pine transpiration, for which this feature is even more marked. A smaller overall predictive uncertainty is found for the all-dataset calibration scenario (PU* = 0.65), as it is the only case where large individual reductions in simulation dispersal are simultaneously achieved for streamflow, soil moisture, net radiation, and to a smaller extent, pine transpiration (Fig. 9, last row).

3.3. Parameter values

Fig. 10 shows the selected parameter values across the best runs within the prescribed sampling interval, displaying only 5 calibration scenarios; grouping all datasets of a same type and the "all-datasets constraint" case. Calibrated soil depths show consistent results, with deep (4–6 m) soils in the valley bottom and shallower podzol and rankers (~2 m), somewhat reflected in the depth of the two upper hydraulic layers (top row). Porosity takes markedly different values across soil types, from 0.85 in the peat, 0.5–0.7 in the gley down to 0.35–0.5 on the hillslope. This spatial variability is somewhat mirrored by an increasing saturated horizontal hydraulic conductivity K_{hx} over several orders of magnitude, from 10^{-5} m s^{-1} in the peat to nearly 0.01 m s^{-1} in the rankers. Other hydrological parameters such as air entry pressure (ψ_{ae}), residual soil moisture (θ_r) and anisotropy K_{ratio}) mostly displays similar values across soil types and calibration scenarios, centred in the sampling intervals. The same applies for vegetation parameters such as optimal photosynthesis temperature (T_{opt}), soil water potential control on stomatal closure ψ_d, and maximum stomatal conductance (gs_{max}). The light extinction coefficient (K_{Beer}) of heather and peat moss displays distinctively higher, well-constrained values when pine transpiration is used as a constraint. For pine trees, the sensitivity of stomatal conductance to soil water content (c) is much lower in the all-datasets calibration case, while canopy interception capacity (CWS_{max}) becomes much higher than in other cases.

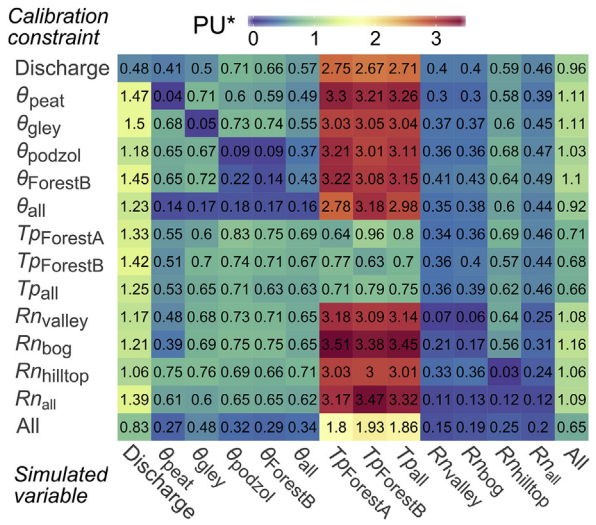

Fig. 9. Relative predictive uncertainty PU*, computed for each observable as the daily 90%-spread interval across the 30 best runs divided by the inter-run mean, then averaged over the evaluation period (Eq. (1b)). The horizontal axis gives the variable or group or variables evaluated, depending on the dataset or combination of datasets used as a constraint over the calibration period (vertical axis).

3.4. Independent evaluation at catchment scale

The empirically-based estimate of catchment-wide saturated area fraction ($Area_{\mathrm{sat}}$) is compared to the simulations provided by the 30 best parameters sets in five calibration scenarios (using each data type, plus the full suite of measurements) in Fig. 11. In all cases, the model broadly reproduces the observed temporal dynamics reasonably well, but apart from the peaks, the simulated saturation extent is generally overestimated. Slight differences appear between scenarios, with higher $Area_{\mathrm{sat}}$ values associated with the highest predictive uncertainty (24%) when only soil moisture is used as a constraint ($VWSC_{\mathrm{all}}$), while the stream discharge constraints brings the narrowest range of simulated $Area_{\mathrm{sat}}$ between best runs (PU = 17%).

4. Discussion

4.1. Insights into ecohydrological processes

This study shows the ability of a process-based model to consistently simulate not only water storages and fluxes in the critical zone at local-to-catchment scales, but also energy balance and ecohydrological couplings in a comprehensive model evaluation exercise. These are very encouraging results for the prospect of explicitly incorporating vegetation dynamics into a mechanistic description of catchment water partitioning, and towards improved prediction of the functional changes that catchments in northern latitude are likely to experience in the coming decades.

The most robustly simulated observable was stream discharge, with 90% of all optimisation-selected runs (30 for each of the 14 calibrations scenarios, 12 of the latter excluding discharge measurements) showing modified KGE scores between 0.67 and 0.87 across the simulation period (11/2012–06/2016, excluding spin-up, not shown). In particular, the model was able to capture well extreme events such as the 200-year return period flood during the winter 2015–2016. Note that the kinematic approximation for groundwater and stream routing in EcH$_2$O neglects diffusive water redistribution through pressure gradients. Therefore, this model-data consistency points at a reasonable adequacy of using a gravity-driven conceptualization of the Bruntland Burn catchment at the spatio-temporal scales considered. This hypothesis is

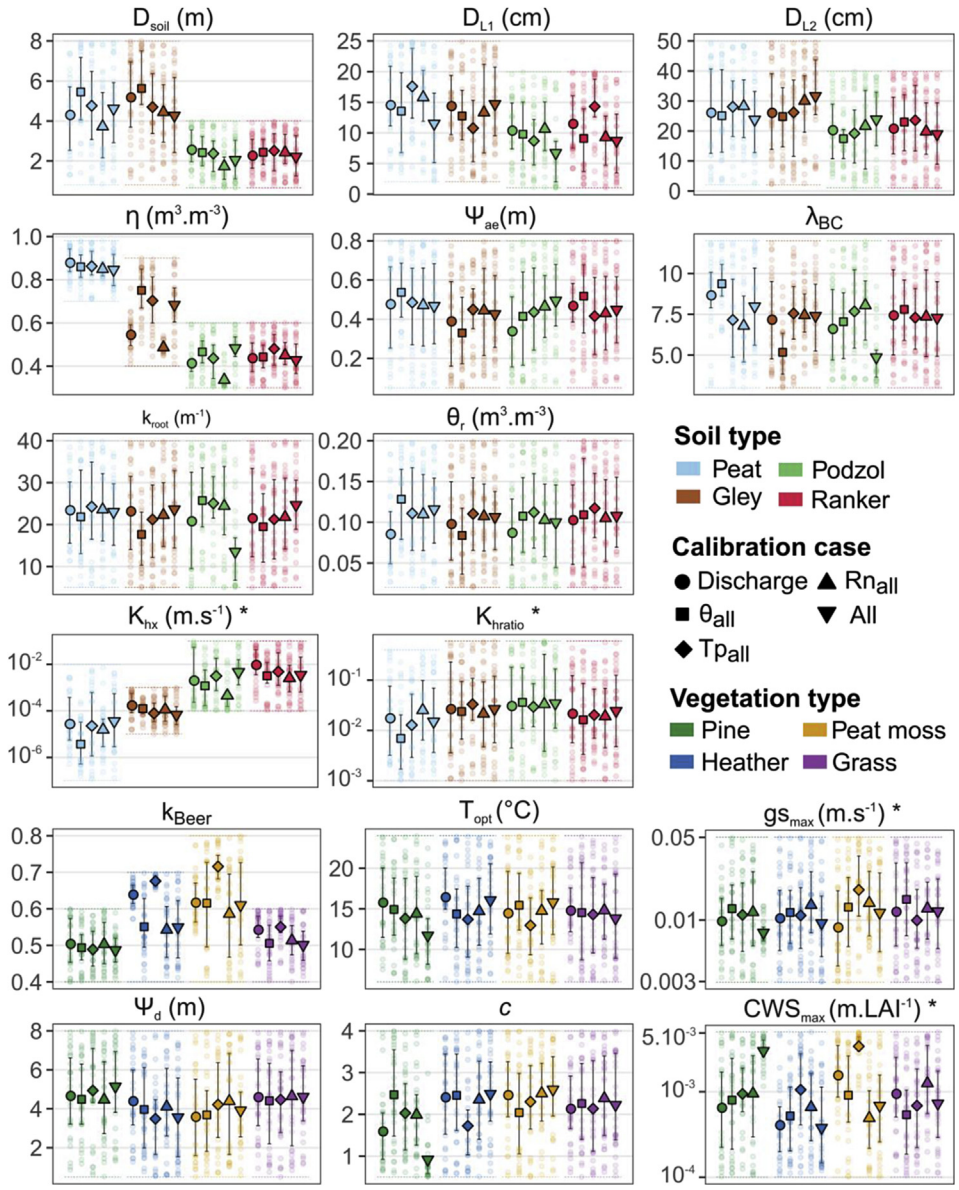

Fig. 10. Calibrated value of the constrained parameters, showing the individual and average value, and the interquartile range across the 30 best runs for each parameter component, as indicated by the colour code, and five different calibration case (symbols): against each data type (including all corresponding datasets) and all datasets together. Horizontal dashed lines indicate the allowed range of variation, and the asterisk that a logarithmic scale was used for sampling. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

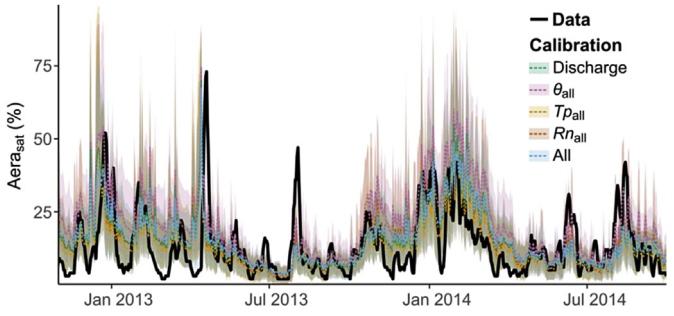

Fig. 11. Time series of saturated area at the catchment scale, comparing an independent data-driven estimate (Ali et al., 2014) (black) to simulation outputs using parameters sets from different calibration cases (average and 90%-spread interval).

consistent with strong topographic gradients in the catchment and the wet, low energy hydroclimate, which both sustain a quasi-permanently saturated valley bottom (Tetzlaff et al., 2014) and generally high water tables even on the steeper hillslopes (Blumstock et al., 2015). These settings result in a flashy response of the stream network to run-off events (Soulsby et al., 2015), generally driven by saturation overland flow from the peat and gleys, but in larger storm events the podzolic soils also connect to the saturated areas (Tetzlaff et al., 2014).

The general hydrological behaviour, which is broadly representative of other northern/boreal catchments (Tetzlaff et al., 2015b), contrasts with semiarid regions, which are characterized by a more transient hydraulic connection and disconnection within hillslopes and between hillslopes and the channel during dry

periods. Although bedrock topography remains critical to understand shallow subsurface flows in water-limited environments (e.g. Jobbágy et al., 2011; Maneta et al., 2008), diffusive effects by local pressure gradients are also highly relevant, reducing the spatial extent for which local measurements are representative, and limiting the propagation of information from the location where fluxes are measured to the location where model parameters need to be identified (Maneta and Wallender, 2013).

More detailed insights into storage dynamics were provided by including volumetric water content for the upper half-meter of soil profile into the calibration, at the four monitored locations across the main hydropedological units. A good model-data fit was achieved in most cases (Fig. 5); EcH_2O overall managed to capture the very different dynamics between locations across almost 13 data-years, including the variability within the same podzolic soil unit (i.e., with a common set of parameters) at two contrasting sites (Fig. 5c and d). Further, the calibration yields a depth of about 2 m of a "hydrologically-active" profile on the hillslope, which is much deeper than typical soil depth estimates from geophysical methods (Soulsby et al., 2016) but consistent with the hypothesis that groundwater recharge on the hillslope and downhill movement actively contributes to saturation overland flow through exfiltration in the valley bottom (Ala-aho et al., 2017a; Birkel et al., 2011). We note, nonetheless, that the riparian areas are the most challenging locations when attempting to capture soil moisture dynamics. In part, this reflects the very small variability measured, though the modelled θ remained too "reactive" as compared to the damped variability in the measurements (Fig. 5a and b). In addition, the gley porosity remained underestimated, leading to unrealistic saturated conditions outside the summer. It might also explain why the simulated gley soil seemed slightly too deep while it should be shallower than in the peat areas, as the model likely compensated in order to close the water balance in the valley bottom.

The other notable model-observation mismatch was the underestimation of the seasonal amplitude of net radiation at the hilltop site, a location mostly covered by heather shrubs (Fig. 1). Using independent estimates of transpiration and total evapotranspiration (ET) on a heather plot near the podzol transect in the same catchment (Wang et al., 2017b), a preliminary analysis hinted at an underestimation of transpiration and overestimation of soil evaporation in the EcH_2O model (not shown). By contrast, net radiation was well simulated in the riparian areas, suggesting an accurate estimation of energy conversion from radiative to turbulent fluxes such as soil evaporation and sensible heat. Further modelling-measurement comparison studies focusing on evaporative processes will help better constraining the energy balance on the catchment hillslopes (Gong et al., 2016; Larsen et al., 2016).

It is uncommon in catchment-scale calibration studies to use direct measurements of plot-scale tree transpiration (Du et al., 2014; Wei et al., 2016). This is not only because this type of measurements is rarely available, but also because most current hydrologic models cannot single out the transpiration fraction of evapotranspiration (Méndez-Barroso et al., 2014; Paniconi and Putti, 2015). The inclusion of an ecohydrological observable such as plot-scale transpiration of Scots pine, introduced direct knowledge of the exchanges between the physical and ecological components of the catchment, and reduced the number of possible internal model configurations that were consistent with observed soil moisture and streamflow. The calibrated model reproduced the major features of the transpiration time series surprisingly well, which increases our confidence that the internal water and energy exchanges at Bruntland Burn were adequately captured. At Forest site B, the simulated transpiration remained very similar across calibration scenarios over the summer 2016 (Fig. 6b), while soil moisture was more sensitive to the data used to calibrate the model (Fig. 5d). This is consistent with results reported using a data-oriented approach (Wang et al., 2017a), which showed weak controls of soil moisture on pine transpiration outside infrequent dry periods in this humid catchment.

4.2. Information content brought by the different observations

The multiplicity of datasets of this study, used in different combinations for model calibration and evaluation, brings novel insights in how informative and representative these measured quantities are for improving our modelling approach. This may help with the design of more efficient data collection campaigns. In the following we first discuss the spatio-temporal footprint, related to how time- and/or location-specific the measured signal is. The issue subsequently discussed is the behavioural footprint, i.e. how specific to some processes the retrieved information content is.

Streamflow was simulated reasonably well in all reported cases (Fig. 8), but with a substantially higher predictive uncertainty whenever discharge data was not included in the calibration (Fig. 9). Streamflow is well-known to integrate information of many catchment-scale processes (Beven and Binley, 1992), but this knowledge is too ambiguous to determine the exact catchment configuration that produces the observed signal. This is because streamflows integrate downstream following a convergent network towards a unique outlet, but the divergent nature of an upstream network makes it impossible to uniquely backtrack the locations where the flow was generated (Kirchner et al., 2001). This has two consequences: streamflow can be well simulated with numerous alternative model parameterizations (physically-consistent or not) (Kirchner, 2006), and the spatio-temporal and behavioural footprints are large and therefore less informative of individual processes happening at specific locations in the catchment (Guse et al., 2016). This was illustrated in the predictive uncertainties of the simulated catchment states: using only discharge as the calibration constraint yielded the most variable results for simulated soil moisture in the gley and net radiation (Fig. 9).

Some variables were only well-simulated when the model was calibrated against observations of that type, which indicates a more restricted behavioural and spatio-temporal footprint of the information. For instance, soil moisture in the gley displayed a significantly higher (up to ten-fold) model-observation mismatch whenever θ_{gley} was not part of the calibration constraints, even if other θ dataset were included in the calibration (Fig. 8b, 3rd column), therefore exhibiting a very localized spatial footprint. Moisture in podzols, on the other hand, displayed slightly more homogeneous performances across calibration scenarios (Fig. 8b, 4th and 5th columns), indicating that having two different podzolic soil moisture locations in this study additionally increased the spatial footprint of the associated calibration constraints. Transpiration in Scots pine stands (*Tp*) was also characterized by poor model-data fits unless the calibration scenarios involved a transpiration dataset (Fig. 8). Performances remained consistent when using data from Forest site B to calibrate *Tp* at Forest site A and vice-versa, even when the two sites cover different growing seasons (2015 and 2016). It indicates that in this catchment the *Tp* datasets has a narrow *behavioural* footprint but a more extensive spatial and temporal footprint. Conversely, this footprint of *Tp* datasets made them ill-suited as sole constraints to calibrate the model across processes, as seen for example from soil moisture goodness-of-fit (Fig. 8) and predictive uncertainty (Fig. 9).

The above considerations highlight the benefits of combining streamflow observations with other types of information that have a more specific footprint, such as measurements of volumetric water content and transpiration (Du et al., 2014). Compared to using streamflow alone, adding the two latter types of variables

improved model calibration by increasing model-data fit scores (Fig. 8) and by reducing the dispersion among best runs (Fig. 9). These enhanced performances were, moreover, generalizable to using a diversified combination of observations in the calibration, as seen from the improved overall model-data fit and low predictive uncertainty across model outputs. It supports a mitigation of the equifinality problem as the increased number of scale- and/or process-specific "diagnostics" (Clark et al., 2011) helped discarding unfeasible model configurations that may otherwise have given high performance scores.

However, we have also observed that the predictive uncertainty of some outputs (the case for peat and gley soil moisture) can increase substantially when the model was calibrated with increasing amounts of information. This may be an indication that with the catchment functioning hypothesis embedded in the model, the datasets have overlapping footprints that inform the calibration process with conflicting or inconsistent information. Note that such conflicts only occurred for individual outputs, as the overall uncertainty across all outputs was indeed lowest when using the full of datasets in the calibration (Fig. 9, bottom-right square).

The Monte Carlo approach adopted in this study uses the GOF as an informal measure of the likelihood of each of these parameters. After the best parameters are selected, the likelihood measure is not further used and the spread across best runs shown in Fig. 10 cannot be interpreted as a probability distribution. This avoids having to make assumptions about the structure of model residual errors, which in more formal statistical frameworks determine how the parameter space is sampled. If these assumptions are incorrect, some 'good' sections of the parameter space may end up being excluded from the search. Note that parameters with good performance have been found in the entire range of "permitted" parameter values. Also as a result, in our methodology all the selected parameter values contribute the same to the predictive spread, and their average does not necessarily represent better the hydrologic behaviour of a given catchment unit. In fact, from Fig. 10 no single combination of parameters can be picked to represent better the average behaviour of the predictive ensemble. For this, interpreting parameter means in terms of the hydrologic behaviour of the catchment behaviour can easily be misleading. However, the fact that the mean value of the parameters (Fig. 10) is often not at the centre of the feasible search space (or of the interquartile range) indicates that some values in the allowed range are more preferred than others, and that the information contained in the alternative calibration datasets informs these preferences differently. Understanding the mechanisms by which the parameters are nudged in a specific direction when calibrated with a specific dataset is desirable, but also difficult and complicates any meaningful interpretation of the differences between mean parameter values.

5. Conclusions

In the growing field of critical zone modelling, a process-based description of energy-plant-water relationships is a promising basis for a mechanistic understanding of vegetation influence on water pathways and stores, and projecting their responses to environmental change. More generally, these types of interdisciplinary models are increasingly needed in critical zone studies, where water is a fundamental medium for energy and material cycles in a wide range of processes at multiple time scales (White et al., 2015) likely to be altered over time (Goddéris and Brantley, 2013). Although the problem of equifinality is exacerbated with the increasing complexity of models, using multiple measurements informative of the range of processes implemented in the models can assist in constraining models to a limited subset of feasible configurations. When this is achieved, newer more integrated models offer an opportunity for deeper process insight. We demonstrated this by applying the fully-distributed model EcH$_2$O to a small northern headwater catchment, using different combinations of 10 datasets relative to 4 types of ecohydrological processes (discharge, soil moisture, pine stand transpiration and above-canopy net radiation). While EcH$_2$O was able to perform well for single objectives when calibrated against individual datasets, constraining its overall behaviour with multiple datasets in a rigorous multi-objective calibration experiment yielded an improved cross-output average model performance and smaller overall predictive uncertainty. The resulting model configuration also reproduced the main features of the temporal dynamics of an independent estimate of catchment-scale saturated area fraction. Successful comparison against this independent dataset indicated that the internal transient storage dynamics were generally captured correctly by the model. We also discussed the informational footprint resulting from each dataset across scales (*spatio-temporal* footprint) and processes (*behavioural* footprint). This modelling experiment increases our confidence that a data-intensive calibration approach constrains the set behavioural model configurations in an effective way, i.e., that our approach allows for broadly getting "the right answers for the right reasons" (Kirchner, 2006). Among other approaches, ongoing model extensions to include tracking of stable water isotopes (^{2}H and ^{18}O) fluxes and water age across ecohydrological compartments will provide means to test this further. The experiments to date provide a foundation for using EcH$_2$O to project the impact of climate variability on catchment functioning in sensitive high-latitude systems. Acknowledging that we have applied EcH$_2$O in a location where snowfall is quite modest, a critical next step will be to conduct simulations in snowier catchments. More generally, we intend to assess the reciprocal links between ecosystem functioning, land cover change and the mediating role of vegetation in buffering atmospheric impacts on water fluxes and storage.

Software availability

The source code of the EcH$_2$O model (c++ programming language) is available at https://bitbucket.org/maneta/ech2o, while the associated documentation, compiled binaries and case study files can be found at http://hs.umt.edu/RegionalHydrologyLab/software/default.php. The Python routine and dataset used for calibration are available upon request to the authors.

Acknowledgements

This work was funded by the European Research Council (project GA 335910 VeWa). M. Maneta acknowledges support from the U.S National Science Foundation (project GSS 1461576) and U.S National Science Foundation EPSCoR Cooperative Agreement #EPS-1101342. All model runs were performed using the High Performance Computing (HPC) cluster of the University of Aberdeen, and the IT Service is thanked for its help in installing PCRaster and other libraries necessary to run EcH$_2$O and post-processing Python routines on the HPC cluster. Finally, the authors are grateful to the many people who have been involved in establishing and continuing data collection at the Bruntland Burn, particularly Christian Birkel, Maria Blumstock, Jon Dick, Josie Geris, Konrad Piegat, Claire Tunaley, and Hailong Wang.

AppendixRecent developments in EcH$_2$O model

In the following equations, the parameters calibrated in this study (Table 1) are highlighted in bold font.

A1. Canopy processes

Canopy-level processes link the radiation budget (solar radiation, incoming longwave, and outgoing longwave radiation) to conductive energy transfers (sensible heat), evaporative losses from the canopy (evaporation of intercepted water and plant transpiration), soil water availability, and soil water potential energy. The core of the canopy processes is a set of 3 equations (energy balance, soil water balance, and soil water potential energy) and 3 unknowns (canopy temperature T_c, plant-available soil moisture S_{t+1}, and soil matric potential, ψ_{soil}), solved for each vegetation type present in the grid cell:

$$\begin{aligned} \frac{(S_{t+1}-S)(\eta-\boldsymbol{\theta_r})Z_{r95}}{\Delta t} - \frac{LET(T_c,\psi_{soil})}{\rho_w \lambda_v} &= 0 \\ \frac{\boldsymbol{\psi_{ae}}}{(S_{t+1})^{\boldsymbol{\lambda_{BC}}}} - \psi_{soil} &= 0 \\ NR(T_c) + H(T_c) + LET(T_c,\psi_{soil}) + LE &= 0 \end{aligned} \quad (A.1)$$

The topmost equation in (A.1) is the water balance in the root zone after infiltration has been accounted for in the initial soil moisture, where S_t is the weighted average degree of soil saturation in the soil over Z_{r95} at the beginning of the time step, S_{t+1}, is the weighted average degree of soil saturation over Z_{r95} at the end of the time step. In both cases, the weights for the saturation averages are given by the fraction of roots in each layer of the soil. Additionally, η is soil porosity, θ_r is residual moisture content, Z_{r95} is the total soil depth of the layers containing 95% of roots, Δt is the size of the time step, ρ_w is density of liquid water, λ_v is latent heat of vaporization, and $LET(T_c,\psi_{soil})$ is a function calculating latent heat consumed for transpiration, which is dependent on the temperature of the canopy and the soil water potential. The second equation in (A.1) is the Brooks and Corey prognostic equation, where ψ_{ae} is the soil air entry pressure, ψ_{soil} is soil matric potential, and λ_{BC} is the pore size index (Brooks and Corey exponent parameter). The lower equation in (A.1) is the energy balance in the canopy as described in Maneta and Silverman (2013), but reproduced here for completeness. This equation assumes that the available radiative energy (NR) will be consumed as sensible heat (H), as latent heat from transpiration (LET), or as latent heat from evaporation of intercepted water (LE):

$$NR(T_c) = R_{SW}(1-\alpha_c)(1-\exp(-\boldsymbol{K_{beer}} \times LAI)) + \varepsilon_c R_{LW} - \varepsilon_c \sigma T_c^4, \quad (A.2)$$

$$H(T_c) = \frac{\rho_a c_a (T_a - T_c)}{r_a}, \quad (A.3)$$

$$LET(T_c,\psi_{soil}) = \frac{\rho_a c_a (e_a - e_s^* h_r)}{\gamma(r_a + r_s(\psi_{soil}))}, \quad (A.4)$$

$$LE = \frac{\rho_a c_a \left[e_a - e_c^* \left(h_r + \frac{(1-h_r)\times CWS}{\boldsymbol{CMS_{max}}}\right)\right]}{\gamma \times r_a}, \quad (A.5)$$

where R_{SW} is incoming solar radiation, α_c is the effective canopy albedo, K_{beer} is a light extinction coefficient as per Beer's law, LAI is leaf area index, ε_c is the canopy emissivity, R_{LW} is downwelling long wave radiation, σ is the Stefan-Boltzmann constant, T_c is the effective canopy temperature, ρ_a is density of air at air temperature T_a, c_a is the heat capacity of air, e_a is air vapour pressure at T_a, e_c^* is the canopy saturation vapour pressure at T_c, h_r is air relative humidity, γ is the psychrometric constant, r_a is aerodynamic resistance, r_s is stomatal resistance, CWS is the current amount of canopy water storage, and CWS_{max} is the maximum canopy storage allowed by unit LAI. See Maneta and Silverman (2013) for additional details on the calculation of these quantities.

The current calculation of stomatal resistance differs from the original formulation in that the efficiency factor that provided the stomatal dependency on soil moisture has been changed to a dependency on soil water potential. The new stomatal resistance formulation and the soil water potential efficiency factor are:

$$\begin{aligned} r_s &= [\boldsymbol{gs_{max}} \cdot LAI \cdot f(R_{SW}) \cdot f(T_a) \cdot f(vpd) \cdot f(\psi_{soil})]^{-1} \\ f(\psi_{soil}) &= \frac{1}{1 + \left(\frac{\psi_{soil}}{\boldsymbol{\psi_d}}\right)^{\boldsymbol{c}}} \end{aligned} \quad (A.6)$$

in which ψ_d is the soil water potential at which 0.5 efficiency for soil water potential is achieved, and c is a function shape parameter. Other efficiency factors are calculated as in Maneta and Silverman (2013).

A2. Soil hydrology

The soil hydrology component has been improved over the original formulation in Maneta and Silverman (2013) by including a vertical soil water redistribution model with three hydraulic layers. The topmost layer receives infiltration from the surface, its soil moisture content controls infiltration rates in the Green and Ampt infiltration equation, and is also the only layer from which evaporation occurs. The middle layer typically contains most of the root system and therefore its moisture controls the hydrologic limitation to transpiration. The deepest layer transfers gravitational water laterally using the original kinematic wave formulation. The condition at the bottom of the third layer can now be impervious (no flow) or can leak at a rate given by a bedrock leakance parameter.

As in the original formulation, only water in excess of field capacity can move by gravity. When soil water exceeds field capacity in the topmost or second layer, the water excess (gravitational water) moves downward to the next layer at a rate determined by the linearized unsaturated hydraulic conductivity function $K(\theta) = \boldsymbol{K_{hx}} \boldsymbol{K_{hratio}} L_b \left(\frac{\theta - \boldsymbol{\theta_r}}{\theta_{fc} - \boldsymbol{\theta_r}}\right)$, where K_{hx} is the saturated horizontal hydraulic conductivity, K_{hratio} is the anisotropy ratio, while L_b is a bedrock leakance parameter set to 1 for layers 1 and 2, and that can vary between 0 (no flow) and 1 (free gravitational drainage) for soil hydraulic layer 3. Water leaking through the bedrock leaves the domain.

Horizontal water transfers to the downstream cell only occur in the third soil hydraulic layer following the linearized kinematic wave formulation described in Maneta and Silverman (2013). When the storage capacity of the bottom layer is exceeded, saturation excess water is transferred to the middle layer, increasing the middle layer moisture content. If the middle layer saturates, saturation excess is transferred to the topmost layer, increasing the topmost layer moisture content. If the topmost layer saturates, saturation excess produces return flow to the surface (seepage face). Return flow is added to the pool of ponded water that generates overland flow the following time step.

A3. Channel routing

Channel flow is simulated using a 1D solution of the kinematic wave equation. The equation is solved for stream discharge, using a

power function ($A = \alpha Q^{\beta}$) to relate cross-section flow area to stream discharge:

$$\frac{\partial Q}{\partial x} + \alpha\beta Q^{\beta-1}\frac{\partial Q}{\partial t} - q_{bf} - q_{ovl} = 0, \quad \text{(A.7)}$$

where Q [L^3T^{-1}] is stream discharge, q_{bf} [L^2T^{-1}] are groundwater contributions to streamflow per unit length of channel, q_{ovf} [L^2T^{-1}] are overland flow contributions to streamflow per unit length of channel, and x [L] and t [T] are distance in the flow direction and time, respectively.

Using Manning's equation to approximate flow velocity, and assuming rectangular channel cross-sections, parameters α and β can be determined to be $\left[\frac{nP^{2/3}}{\sqrt{S}}\right]^{3/5}$ and $\beta = \frac{3}{5}$, in which n [$TL^{-1/3}$] is Manning's roughness coefficient, P [L] is the channel wetted perimeter, approximated by channel width, and S is the streambed slope. Equation (A.7) is solved using a first-order implicit finite-difference scheme and is unconditionally stable.

A4. Exponential root profile

The new model formulation requires the fraction of roots in each soil layer, $f_{root}^{L1,2,3}$, to be specified. In this paper, we have assumed an exponential root profile modulated by a single parameter k_{root}:

$$\begin{aligned} f_{root}^{L1} &= \frac{1-\exp(\boldsymbol{k_{root}} \times \boldsymbol{D_{L1}})}{1-\exp(\boldsymbol{k_{root}} \times \boldsymbol{D_{soil}})} \\ f_{root}^{L2} &= \frac{\exp(\boldsymbol{k_{root}} \times \boldsymbol{D_{L1}}) - \exp(\boldsymbol{k_{root}} \times (\boldsymbol{D_{L1}} + \boldsymbol{D_{L2}}))}{1-\exp(\boldsymbol{k_{root}} \times \boldsymbol{D_{soil}})}, \quad \text{(A.8)} \\ f_{root}^{L3} &= 1 - f_{root}^{L1} - f_{root}^{L2} \end{aligned}$$

where D_{soil} is the total depth of the soil hydrologically active layer, and $D_{L1,2}$ are the depth of layers 1 or 2.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envsoft.2018.01.001.

References

Ala-aho, P., Soulsby, C., Wang, H., Tetzlaff, D., 2017a. Integrated surface-subsurface model to investigate the role of groundwater in headwater catchment runoff generation: a minimalist approach to parameterisation. J. Hydrol. 547, 664–677.

Ala-aho, P., Tetzlaff, D., McNamara, J.P., Laudon, H., Soulsby, C., 2017b. Using isotopes to constrain water flux and age estimates in snow-influenced catchments using the STARR (Spatially distributed Tracer-Aided Rainfall–Runoff) model. Hydrol. Earth Syst. Sci. 21, 5089.

Ali, G., Birkel, C., Tetzlaff, D., Soulsby, C., McDonnell, J.J., Tarolli, P., 2014. A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions. Earth Surf. Process. Landforms 39, 399–413.

Band, L.E., Tague, C.L., Groffman, P., Belt, K., 2001. Forest ecosystem processes at the watershed scale: hydrological and ecological controls of nitrogen export. Hydrol. Process. 15, 2013–2028. https://doi.org/10.1002/hyp.253.

Beck, P.S., Juday, G.P., Alix, C., Barber, V.A., Winslow, S.E., Sousa, E.E., Heiser, P., Herriges, J.D., Goetz, S.J., 2011. Changes in forest productivity across Alaska consistent with biome shift. Ecol. Lett. 14, 373–379.

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298. https://doi.org/10.1002/hyp.3360060305.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249, 11–29.

Bintanja, R., Andry, O., 2017. Towards a rain-dominated Arctic. Nat. Clim. Change 7, 263–267. https://doi.org/10.1038/nclimate3240.

Birkel, C., Tetzlaff, D., Dunn, S.M., Soulsby, C., 2010. Towards a simple dynamic process conceptualization in rainfall–runoff models using multi-criteria calibration and tracers in temperate, upland catchments. Hydrol. Process. 24, 260–275. https://doi.org/10.1002/hyp.7478.

Birkel, C., Tetzlaff, D., Dunn, S.M., Soulsby, C., 2011. Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models. Water Resour. Res. 47, W02515. https://doi.org/10.1029/2010WR009547.

Birkel, C., Soulsby, C., Tetzlaff, D., 2014. Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. Water Resour. Res. 50, 3481–3501. https://doi.org/10.1002/2013WR014925.

Blumstock, M., Tetzlaff, D., Malcolm, I.A., Nuetzmann, G., Soulsby, C., 2015. Baseflow dynamics: multi-tracer surveys to assess variable groundwater contributions to montane streams under low flows. J. Hydrol. 527, 1021–1033.

Bonell, M., 2002. Ecohydrology—a completely new idea? Hydrol. Sci. J. 47, 809–810.

Bossel, H., 1996. TREEDYN3 forest simulation model. Ecol. Model. 90, 187–227.

Brooks, P.D., Chorover, J., Fan, Y., Godsey, S.E., Maxwell, R.M., McNamara, J.P., Tague, C., 2015. Hydrological partitioning in the critical zone: recent advances and opportunities for developing transferable understanding of water cycle dynamics. Water Resour. Res. 51, 6973–6987.

Cable, J.M., Ogle, K., Bolton, W.R., Bentley, L.P., Romanovsky, V., Iwata, H., Harazono, Y., Welker, J., 2014. Permafrost thaw affects boreal deciduous plant transpiration through increased soil water, deeper thaw, and warmer soils. Ecohydrology 7, 982–997.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250.

Clark, M.P., Vrugt, J.A., 2006. Unraveling uncertainties in hydrologic model calibration: addressing the problem of compensatory parameters. Geophys. Res. Lett. 33.

Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resour. Res. 47, W09301. https://doi.org/10.1029/2010WR009827.

Cohon, J.L., 1978. Multiobjective Programming and Planning. Academic Press, New York, USA.

Cox, P.M., Huntingford, C., Harding, R.J., 1998. A canopy conductance and photosynthesis model for use in a GCM land surface scheme. J. Hydrol. 212, 79–94.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828.

Drewry, D.T., Kumar, P., Long, S., Bernacchi, C., Liang, X.-Z., Sivapalan, M., 2010. Ecohydrological responses of dense canopies to environmental variability: 1. Interplay between vertical structure and photosynthetic pathway. J. Geophys. Res. Biogeosci. 115, G04022. https://doi.org/10.1029/2010JG001340.

Du, E., Link, T.E., Gravelle, J.A., Hubbart, J.A., 2014. Validation and sensitivity test of the distributed hydrology soil-vegetation model (DHSVM) in a forested mountain watershed. Hydrol. Process. 28, 6196–6210. https://doi.org/10.1002/hyp.10110.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031.

Duhem, P., 1954. The aim and structure of physical theory. Am. J. Phys. 22, 503–503.

Dye, D.G., Tucker, C.J., 2003. Seasonality and trends of snow-cover, vegetation index, and temperature in northern Eurasia. Geophys. Res. Lett. 30.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrol. Sci. J. 55, 58–78. https://doi.org/10.1080/02626660903526292.

Endrizzi, S., Gruber, S., Dall'Amico, M., Rigon, R., 2014. GEOtop 2.0: simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. Geosci Model Dev 7, 2831–2857. https://doi.org/10.5194/gmd-7-2831-2014.

Fairfield, J., Leymarie, P., 1991. Drainage networks from grid digital elevation models. Water Resour. Res. 27, 709–717.

Fan, Y., 2015. Groundwater in the Earth's critical zone: relevance to large-scale patterns and processes. Water Resour. Res. 51. https://doi.org/10.1002/2015WR017037.

Fang, X., Pomeroy, J.W., Ellis, C.R., MacDonald, M.K., DeBeer, C.M., Brown, T., 2013. Multi-variable evaluation of hydrological model predictions for a headwater basin in the Canadian Rocky Mountains. Hydrol. Earth Syst. Sci. 17, 1635.

Fatichi, S., Ivanov, V.Y., Caporali, E., 2012. A mechanistic ecohydrological model to investigate complex interactions in cold and warm water-controlled environments: 1. Theoretical framework and plot-scale analysis. J. Adv. Model. Earth Syst. 4.

Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B., others, 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J. Hydrol. 537, 45–60.

Geris, J., Tetzlaff, D., McDonnell, J., Soulsby, C., 2015. The relative role of soil type and tree cover on water storage and transmission in northern headwater catchments. Hydrol. Process. 29, 1844–1860. https://doi.org/10.1002/hyp.10289.

Goddéris, Y., Brantley, S.L., 2013. Earthcasting the future critical zone. Elem. Sci.

Anthr 1, 19.
Gong, J., Jia, X., Zha, T., Wang, B., Kellomäki, S., Peltola, H., 2016. Modeling the effects of plant-interspace heterogeneity on water-energy balances in a semiarid ecosystem. Agric. For. Meteorol. 221, 189–206. https://doi.org/10.1016/j.agrformet.2016.01.144.
Goody, R.M., Yung, Y.L., 1995. Atmospheric Radiation: Theoretical Basis. Oxford university press.
Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resour. Res. 34, 751–763.
Guse, B., Pfannerstill, M., Gafurov, A., Fohrer, N., Gupta, H., 2016. Demasking the integrated information of discharge: advancing sensitivity analysis to consider different hydrological components and their rates of change. Water Resour. Res. 52, 8724–8743. https://doi.org/10.1002/2016WR018894.
He, Z.H., Tian, F.Q., Gupta, H.V., Hu, H.C., Hu, H.P., 2015. Diagnostic calibration of a hydrological model in a mountain area by hydrograph partitioning. Hydrol. Earth Syst. Sci. 19, 1807–1826.
Hinzman, L.D., Deal, C.J., McGuire, A.D., Mernild, S.H., Polyakov, I.V., Walsh, J.E., 2013. Trajectory of the Arctic as an integrated system. Ecol. Appl. 23, 1837–1868.
Jarvis, P.G., 1976. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. Philos. Trans. R. Soc. Lond. B Biol. Sci. 273, 593–610.
Jiménez Cisneros, B.E., Oki, T., Arnell, N.W., Benito, G., Cogley, J.G., Doll, P., Jiang, T., Mwakalila, S.S., 2014. Freshwater Resources.
Jobbágy, E.G., Nosetto, M.D., Villagra, P.E., Jackson, R.B., 2011. Water subsidies from mountains to deserts: their role in sustaining groundwater-fed oases in a sandy landscape. Ecol. Appl. 21, 678–694. https://doi.org/10.1890/09-1427.1.
Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P.J., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., others, 2013. The BETHY/JSBACH carbon cycle data assimilation system: experiences and challenges. J. Geophys. Res. Biogeosciences 118, 1414–1426.
Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42, W03S04. https://doi.org/10.1029/2005WR004362.
Kirchner, J.W., Feng, X., Neal, C., 2001. Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations. J. Hydrol. 254, 82–101. https://doi.org/10.1016/S0022-1694(01)00487-5.
Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrol. 424, 264–277.
Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97.
Kuchment, L.S., Gelfan, A.N., Demidov, V.N., 2000. A distributed model of runoff generation in the permafrost regions. J. Hydrol. 240, 1–22.
Landsberg, J.J., Waring, R.H., 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. For. Ecol. Manag 95, 209–228.
Larsen, M.A.D., Refsgaard, J.C., Jensen, K.H., Butts, M.B., Stisen, S., Mollerup, M., 2016. Calibration of a distributed hydrology and land surface model using energy flux measurements. Agric. For. Meteorol. 217, 74–88. https://doi.org/10.1016/j.agrformet.2015.11.012.
Laudon, H., Spence, C., Buttle, J., Carey, S.K., McDonnell, J.J., McNamara, J.P., Soulsby, C., Tetzlaff, D., 2017. Save northern high-latitude catchments. Nat. Geosci. 10.
Law, F., 1956. The effect of afforestation upon the yield of water catchment areas. J. Br. Waterworks Assoc. 38, 489–494.
Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241.
Lessels, J.S., Tetzlaff, D., Birkel, C., Dick, J., Soulsby, C., 2016. Water sources and mixing in riparian wetlands revealed by tracers and geospatial analysis. Water Resour. Res. 52, 456–470. https://doi.org/10.1002/2015WR017519.
Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201, 272–288.
Lozano-Parra, J., Maneta, M.P., Schnabel, S., 2014. Climate and topographic controls on simulated pasture production in a semiarid Mediterranean watershed with scattered tree cover. Hydrol. Earth Syst. Sci. 18, 1439.
Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Adv. Water Resour. 26, 205–216. https://doi.org/10.1016/S0309-1708(02)00092-1.
Maneta, M.P., Silverman, N.L., 2013. A spatially distributed model to simulate water, energy, and vegetation dynamics using information from regional climate models. Earth Interact. 17, 1–44.
Maneta, M.P., Wallender, W.W., 2013. Pilot-point based multi-objective calibration in a surface–subsurface distributed hydrological model. Hydrol. Sci. J. 58, 390–407.
Maneta, M.P., Pasternack, G.B., Wallender, W.W., Jetten, V., Schnabel, S., 2007. Temporal instability of parameters in an event-based distributed hydrologic model applied to a small semiarid catchment. J. Hydrol. 341, 207–221.
Maneta, M., Schnabel, S., Jetten, V., 2008. Continuous spatially distributed simulation of surface and subsurface hydrological processes in a small semiarid catchment. Hydrol. Process. 22, 2196–2214.
Manoli, G., Huang, C.-W., Bonetti, S., Domec, J.-C., Marani, M., Katul, G., 2017. Competition for light and water in a coupled soil-plant system. Adv. Water Resour. 108, 216–230. https://doi.org/10.1016/j.advwatres.2017.08.004.
Maxwell, R.M., Condon, L.E., 2016. Connections between groundwater flow and transpiration partitioning. Science 353, 377–380.
Mays, L.W., 2010. Water Resources Engineering. John Wiley & Sons.
McClelland, J.W., Déry, S.J., Peterson, B.J., Holmes, R.M., Wood, E.F., 2006. A pan-arctic evaluation of changes in river discharge during the latter half of the 20th century. Geophys. Res. Lett. 33.
McDonnell, J.J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M.L., Selker, J., Weiler, M., 2007. Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. Water Resour. Res. 43, W07301. https://doi.org/10.1029/2006WR005467.
Méndez-Barroso, L.A., Vivoni, E.R., Robles-Morua, A., Mascaro, G., Yépez, E.A., Rodríguez, J.C., Watts, C.J., Garatuza-Payán, J., Saíz-Hernández, J.A., 2014. A modeling approach reveals differences in evapotranspiration and its partitioning in two semiarid ecosystems in Northwest Mexico. Water Resour. Res. 50, 3229–3252. https://doi.org/10.1002/2013WR014838.
Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33, 161–174.
Naito, A.T., Cairns, D.M., 2015. Patterns of shrub expansion in Alaskan arctic river corridors suggest phase transition. Ecol. Evol. 5, 87.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — a discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
National Research Council, 2012. New Research Opportunities in the Earth Sciences. https://doi.org/10.17226/13236.
Newman, B.D., Wilcox, B.P., Archer, S.R., Breshears, D.D., Dahm, C.N., Duffy, C.J., McDowell, N.G., Phillips, F.M., Scanlon, B.R., Vivoni, E.R., 2006. Ecohydrology of water-limited environments: a scientific vision. Water Resour. Res. 42, W06302. https://doi.org/10.1029/2005WR004141.
Niu, J., Phanikumar, M.S., 2015. Modeling watershed-scale solute transport using an integrated, process-based hydrologic model with applications to bacterial fate and transport. J. Hydrol. 529 (Part 1), 35–48. https://doi.org/10.1016/j.jhydrol.2015.07.013.
Met Office, 2017. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). NCAS Br. Atmospheric Data Cent.
Paniconi, C., Putti, M., 2015. Physically based modeling in catchment hydrology at 50: survey and outlook. Water Resour. Res. 51, 7090–7129.
Peng, C., Liu, J., Dang, Q., Apps, M.J., Jiang, H., 2002. TRIPLEX: a generic hybrid model for predicting forest growth and carbon and nitrogen dynamics. Ecol. Model. 153, 109–130. https://doi.org/10.1016/S0304-3800(01)00505-1.
Peterson, B.J., Holmes, R.M., McClelland, J.W., Vörösmarty, C.J., Lammers, R.B., Shiklomanov, A.I., Shiklomanov, I.A., Rahmstorf, S., 2002. Increasing river discharge to the arctic ocean. Science 298, 2171–2173.
Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maignan, F., Chevallier, F., others, 2016. A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. Geosci. Model Dev. 9, 3321.
Piao, S., Ciais, P., Friedlingstein, P., Peylin, P., Reichstein, M., Luyssaert, S., Margolis, H., Fang, J., Barr, A., Chen, A., others, 2008. Net carbon dioxide losses of northern ecosystems in response to autumn warming. Nature 451, 49–52.
Pierini, N.A., Vivoni, E.R., Robles-Morua, A., Scott, R.L., Nearing, M.A., 2014. Using observations and a distributed hydrologic model to explore runoff thresholds linked with mesquite encroachment in the Sonoran Desert. Water Resour. Res. 50, 8191–8215. https://doi.org/10.1002/2014WR015781.
Pomeroy, J.W., Gray, D.M., Brown, T., Hedstrom, N.R., Quinton, W.L., Granger, R.J., Carey, S.K., 2007. The cold regions hydrological model: a platform for basing process representation and model structure on physical evidence. Hydrol. Process. 21, 2650–2667.
Pomeroy, J.W., Fang, X., Shook, K., Whitfield, P.H., 2013. Predicting in ungauged basins using physical principles obtained using the deductive, inductive, and abductive reasoning approach. Putt. Predict. Ungauged Basins Pract 43–63.
Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. J. Hydrol. 536, 192–207. https://doi.org/10.1016/j.jhydrol.2016.02.037.
Rodriguez-Iturbe, I., D'Odorico, P., Laio, F., Ridolfi, L., Tamea, S., 2007. Challenges in humid land ecohydrology: interactions of water table and unsaturated zone with climate, soil, and vegetation. Water Resour. Res. 43, W09301. https://doi.org/10.1029/2007WR006073.
Rouhani, H., Willems, P., Wyseure, G., Feyen, J., 2007. Parameter estimation in semi-distributed hydrological catchment modelling using a multi-criteria objective function. Hydrol. Process. 21, 2998–3008. https://doi.org/10.1002/hyp.6527.
Seibert, J., 2003. Reliability of model predictions outside calibration conditions. Nord. Hydrol 34, 477–492.
Shafii, M., Basu, N., Craig, J.R., Schiff, S.L., Van Cappellen, P., 2017. A diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework. Water Resour. Res. 53. https://doi.org/10.1002/2016WR019736.
Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: correlated and heteroscedastic error cases. Water Resour. Res. 16, 430–442.
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 Resour. Res. 51, 7759–7776.
Soulsby, C., Bradford, J., Dick, J., McNamara, J.P., Geris, J., Lessels, J., Blumstock, M., Tetzlaff, D., 2016. Using geophysical surveys to test tracer-based storage estimates in headwater catchments. Hydrol. Process. 30, 4434–4445.

Sprenger, M., Tetzlaff, D., Tunaley, C., Dick, J., Soulsby, C., 2017. Evaporation fractionation in a peatland drainage network affects stream water isotope composition. Water Resour. Res. 53, 851–866. https://doi.org/10.1002/2016WR019258.

Tague, C.L., 2009. Assessing climate change impacts on alpine stream-flow and vegetation water use: mining the linkages with subsurface hydrologic processes. Hydrol. Process. 23, 1815–1819. https://doi.org/10.1002/hyp.7288.

Tang, Y., Reed, P.M., Kollat, J.B., 2007. Parallelization strategies for rapid and robust evolutionary multiobjective optimization in water resources applications. Adv. Water Resour. 30, 335–353. https://doi.org/10.1016/j.advwatres.2006.06.006.

Tetzlaff, D., Soulsby, C., Waldron, S., Malcolm, I.A., Bacon, P.J., Dunn, S.M., Lilly, A., Youngson, A.F., 2007. Conceptualization of runoff processes using a geographical information system and tracers in a nested mesoscale catchment. Hydrol. Process. 21, 1289–1307.

Tetzlaff, D., Soulsby, C., Buttle, J., Capell, R., Carey, S.K., Laudon, H., McDonnell, J., McGuire, K., Seibert, J., Shanley, J., 2013. Catchments on the cusp? Structural and functional change in northern ecohydrology. Hydrol. Process. 27, 766–774.

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 Resour. Res. 50, 969–985.

Tetzlaff, D., Buttle, J., Carey, S.K., Huijgevoort, M.H., Laudon, H., McNamara, J.P., Mitchell, C.P., Spence, C., Gabor, R.S., Soulsby, C., 2015a. A preliminary assessment of water partitioning and ecohydrological coupling in northern headwaters using stable isotopes and conceptual runoff models. Hydrol. Process. 29, 5153–5173.

Tetzlaff, D., Buttle, J., Carey, S.K., McGuire, K., Laudon, H., Soulsby, C., 2015b. Tracer-based assessment of flow paths, storage and runoff generation in northern catchments: a review. Hydrol. Process. 29, 3475–3490.

Thorstensen, A., Nguyen, P., Hsu, K., Sorooshian, S., 2015. Using densely distributed soil moisture observations for calibration of a hydrologic model. J. Hydrometeorol. 17, 571–590. https://doi.org/10.1175/JHM-D-15-0071.1.

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res. 43.

Uhlenbrook, S., Seibert, J., Leibundgut, C., Rodhe, A., 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. Hydrol. Sci. J. 44, 779–797. https://doi.org/10.1080/02626669909492273.

Wang, H., Tetzlaff, D., Dick, J.J., Soulsby, C., 2017a. Assessing the environmental controls on Scots pine transpiration and the implications for water partitioning in a boreal headwater catchment. Agric. For. Meteorol. 240, 58–66.

Wang, H., Tetzlaff, D., Soulsby, C., 2017b. Testing the maximum entropy production approach for estimating evapotranspiration from closed canopy shrubland in a low-energy humid environment. Hydrol. Process. n/a-n/a https://doi.org/10.1002/hyp.11363.

Wei, L., Link, T.E., Hudak, A.T., Marshall, J.D., Kavanagh, K.L., Abatzoglou, J.T., Zhou, H., Pangle, R.E., Flerchinger, G.N., 2016. Simulated water budget of a small forested watershed in the continental/maritime hydroclimatic region of the United States. Hydrol. Process. 30, 2000–2013. https://doi.org/10.1002/hyp.10769.

van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Adv. Water Resour. 32, 1154–1169. https://doi.org/10.1016/j.advwatres.2009.03.002.

White, T., Brantley, S., Banwart, S., Chorover, J., Dietrich, W., Derry, L., Lohse, K., Anderson, S., Aufdendkampe, A., Bales, R., others, 2015. The role of critical zone observatories in critical zone science. Dev. Earth Surf. Process 19, 15–78.

Williams, J.W., Jackson, S.T., Kutzbach, J.E., 2007. Projected distributions of novel and disappearing climates by 2100 AD. Proc. Natl. Acad. Sci. 104, 5738–5742. https://doi.org/10.1073/pnas.0606292104.

Zhang, X., Tang, Q., Zhang, X., Lettenmaier, D.P., 2014. Runoff sensitivity to global mean temperature change in the CMIP5 Models. Geophys. Res. Lett. 41, 2014GL060382. https://doi.org/10.1002/2014GL060382.