

Evaluating the effects of alternative model structures on dynamic storage simulation in heterogeneous boreal catchments

Shirin Karimi ^{a,*}, Jan Seibert ^{b,c} and Hjalmar Laudon ^a

^a Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, Sweden

^b Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden

^c Department of Geography, University of Zurich, Zurich, Switzerland

*Corresponding author. E-mail: Shirin.karimi@slu.se

 SK, 0000-0003-1009-0938; JS, 0000-0002-6314-2124; HL, 0000-0001-6058-1466

ABSTRACT

Estimating dynamic storage as a metric can be used to make an overall assessment of catchment resilience to extreme weather events such as droughts and floods. Because of the complexity of direct empirical measurements, bucket-type hydrological models can be a suitable tool to simulate the catchment storage across a broad range of scales as they require minimal input data. However, these models consist of one or more conceptual structures based on several linear or nonlinear reservoirs and connections between these reservoirs. Therefore, choosing the most appropriate model structure to represent storage-discharge functioning in catchments is difficult. To bridge this gap, this study evaluated the performance of three different HBV model structures on 14 heterogeneous boreal catchments classified into four distinct catchment categories. The results showed that the three-bucket structure performed better in larger catchments with deeper sediment soils. In contrast, a single reservoir structure is sufficient to predict the storage-discharge behavior for a lake-influenced catchment with lower elevation above the stream network. Moreover, our results indicate that while the estimates of mean catchment storage varied between the different model structures, the ranking between the catchments largely agreed for the different structures. Hence, our results suggest that instead of a single model structure, using an ensemble averaging approach would not only better address the structural uncertainty but also facilitate further storage comparison between different catchments. Finally, based on Spearman rank correlation results, we found that catchment size and sediment soil were positively correlated with dynamic storage estimation.

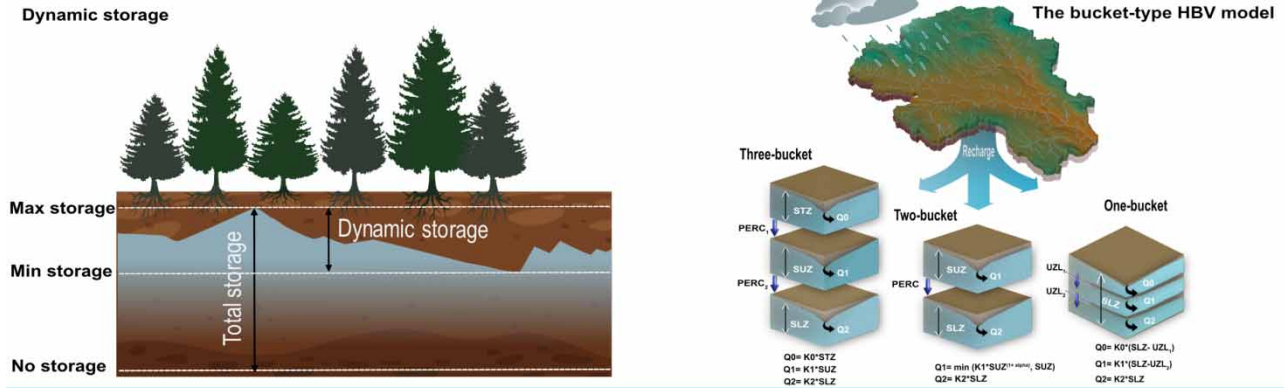
Key words: boreal, bucket-type, dynamic storage, ensemble, HBV, hydrological models, storage-discharge

HIGHLIGHTS

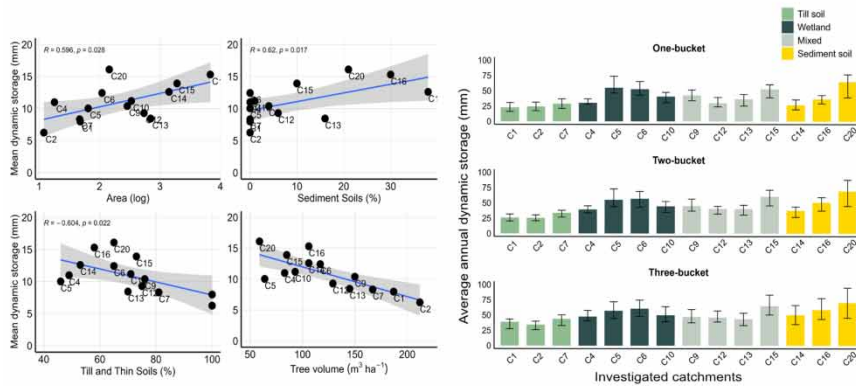
- Differences were found in storage estimates between the different model structures.
- Three-bucket structure performed well for all catchments.
- Lake-influenced catchments were well represented by one-bucket structure.
- One-bucket structure performed poorly in the large catchments with deep sediment soils.
- An ensemble averaging approach can be used as a point of comparison between hydrological functioning of catchments.

GRAPHICAL ABSTRACT

Methodology



Results and findings



1. INTRODUCTION

Storage and release of water are essential catchment functions, with large effects on modulating hydrological extremes such as drought and floods (Creutzfeldt *et al.* 2012). Water entering catchments is stored in different forms, including snow, ice, lakes, soil moisture, and groundwater (Riegger & Tourian 2014). Depending on soil physical properties, water can be retained in the unsaturated zone, move through the soil column and percolate to saturated layers as groundwater recharge. Accordingly, there are various pathways that water can take when it flows between unsaturated and saturated zones before leaving the catchment, depending on soil physical properties, catchment characteristics, and climate (Jutebring Sterte *et al.* 2021).

The temporal dynamics of stream discharge depend largely on catchment storage (Kirchner 2009; Creutzfeldt *et al.* 2012; Brauer *et al.* 2013). In periods when evapotranspiration and precipitation are negligible, it can be assumed that the only driver of discharge is the amount of water stored in the catchment (Moore 1997; Kirchner 2009). Water content freely available for drainage and flow is known as 'mobile storage' (Farrick & Branfireun 2014; Staudinger *et al.* 2017). However, not all catchment storage contributes to the runoff generation process, which is sometimes referred to as 'hydraulically decoupled storage' (Dralle *et al.* 2018). Changes in the amount of these hydrologically decoupled storages can have a negligible effect on runoff generation, at least in the short term, because of deep groundwater storage, or have no effect, such as canopy interception that leaves the basin through evaporation without entering the soil.

In many previous studies, catchment storage at different spatial scales has been quantified using a range of different methods: water balance approaches (Sayama *et al.* 2011; Wang & Alimohammadi 2012), hydrometric analysis (Kirchner 2009; McNamara *et al.* 2011; Brauer *et al.* 2013; Amvrosiadi *et al.* 2017a, 2017b), analysis of stable isotope tracers (Soulsby

et al. 2009; Sprenger *et al.* 2018), gravimeter measurements (Creutzfeldt *et al.* 2012; Kennedy *et al.* 2014), and application of hydrological models (Staudinger *et al.* 2017; Sterte *et al.* 2018; Karlsen *et al.* 2019). One of the most common methods for direct estimation of catchment water storage involves using a network of piezometer wells and soil moisture probes to provide detailed information on whether a sufficient number of sensors has been installed (Kalbus *et al.* 2006). However, due to the high spatiotemporal variability, collecting massive hydrological data across a broad range of scales is time-consuming and costly (Kirchner 2009). Hence, efforts to reduce field investigations have resulted in various modeling attempts to understand better the complexities of the hydrological behavior in heterogeneous catchments.

There are several definitions of catchment water storage (Staudinger *et al.* 2017). Here, we focus on the part of the storage that is relevant for catchment discharge. This leads to an operational definition where the storage is zero when streamflow ceases completely. Various terms have been proposed for this discharge-generating storage; these include ‘dynamic storage’ (Kirchner 2009; Sayama *et al.* 2011; Staudinger *et al.* 2017), ‘active storage’ (McNamara *et al.* 2011), and ‘direct storage’ (Dralle *et al.* 2018). Here we use the term dynamic storage as a critical metric to assess the sensitivity of a catchment to extreme weather events such as flooding or its resistance to drought by maintaining low flows (McNamara *et al.* 2011). Revealing the dynamic storage-discharge relationship can also help make more accurate predictions of streamflow changes in response to global warming.

Bucket-type models, also called conceptual models, are suitable tools for simulating runoff and water availability based on storage–discharge relationships and represent in general terms how precipitation results in groundwater recharge, evapotranspiration, and discharge (Hrachowitz & Clark 2017; Sitterson *et al.* 2018). These models consist of several reservoirs, which can be linear or nonlinear and be connected in serial or parallel (Stoelzle *et al.* 2015; Parra *et al.* 2019a). The most important advantage of these bucket-type models, compared to more detailed and complex fully physically-based models, is a lower number of free parameters, which means that the parameter values of the models can, in principle, be estimated by calibration only. However, even for such models, model parameter uncertainty has to be considered when deriving storage estimates by modeling. Furthermore, there is model structure uncertainty as there are usually various possible alternatives to arrange the different buckets. For example, the proper choice of linear and nonlinear storage-discharge relationships has been debated for a long time (Stoelzle *et al.* 2015). Some researchers have concluded that groundwater release is a linear process (Chapman 1999; Fenicia *et al.* 2006), while others have demonstrated that nonlinear relationships are more appropriate (Wittenberg 1999; Mishra *et al.* 2003; Botter *et al.* 2009; Maneta *et al.* 2018; Rezaei-Sadr 2019). These findings show that choosing appropriate model structures is challenging and uncertain. In other words, the mathematical formulation of hydrological functions that control the transformation of rainfall to runoff might vary by type of catchment and climate, and that a ‘one size fits all’ is not easily found (Hogue *et al.* 2006; Ajami *et al.* 2007).

Many studies have used bucket-type models to improve the understanding of hydrological functioning in catchments, such as estimation of water storage (Krasnostein & Oldham 2004; Mendoza-Sanchez *et al.* 2013; Staudinger *et al.* 2017) as well as streamflow signatures (Hailegeorgis & Alfredsen 2015; Ledesma & Futter 2017; Teutschbein *et al.* 2018). These studies have also demonstrated that bucket models can provide suitable representations of catchment hydrology and result in good runoff simulations for both calibration and validation periods. However, as pointed out in previous studies, achieving an acceptable value of model performance based on a good agreement between modeled and observed streamflow data does not necessarily correspond to a good simulation for other hydrological functioning (Gupta *et al.* 2012; Teutschbein *et al.* 2015; Lane *et al.* 2019; Seibert *et al.* 2019). Therefore, choosing a model structure with an appropriate degree of complexity is a crucial step in catchment modeling.

Hence, despite a good model performance for a certain calibration or validation period, model outputs can still be unreliable when extrapolated beyond calibration conditions (Seibert 1997). Therefore, a proper analysis of how the model parameters and the model structures are correlated and a detailed assessment of their effects on the model output can improve the model reliability. To our knowledge, few previous studies have comprehensively analyzed the most important parameters controlling the amount of storage in the catchment reservoirs (Tetzlaff *et al.* 2015; Teutschbein *et al.* 2015; Ledesma & Futter 2017). This is especially true in northern latitude catchments defined by long-lasting winters and snow-melt-dominated hydrological conditions.

The main purpose of the approach lies in the evaluation of the performance of different model structures, which only differ in the number of storage reservoirs, in estimating dynamic storage using a bucket-type model for a large number of catchments with contrasting landscape characteristics but similar climatic conditions. In addition, we aimed to better understand the relationship between dominant catchment characteristics and dynamic storage variability across a

heterogeneous boreal landscape. We addressed the following research questions: (1) How does the uncertainty of parameters in the response routine of a bucket-type model affect the dynamic storage estimation? (2) Is there a major difference between dynamic storage estimates obtained using different model structures? If yes, which model structure is most reliable at different catchment scales? We then used this approach to evaluate how simulated dynamic storage varies in both time and space for different subcatchments. Furthermore, we studied potential relationships between the different types of storage zones and dominant catchment characteristics.

2. MATERIALS AND METHODS

2.1. Study site

The study was conducted in the Krycklan catchment located in northern Sweden, about 50 km west of the Baltic Sea coast (64°25' N, 19°46' E), with an area of approximately 68 km². The catchment area comprises 14 nested subcatchments with elevation ranging from 130 to 370 meters above sea level, as shown in Figure 1. The higher elevations of the catchment are dominated by till (58% of total area) and peat soil, and lower elevations are covered by sediment soils (Laudon *et al.* 2021). Forests cover 87% of the entire catchment and are dominated by Scots pine (*Pinus sylvestris*) and Norway spruce (*Picea abies*).

The hydrological characteristics of the catchment are well studied and are monitored by 14 streamflow gauging stations (named C1-C20, with some of the original stations abandoned for various reasons) (Table 1). The climate in this area is characterized by relatively cold winters with consistent snow coverage for four to six months of the year, between November to April. The annual average precipitation was 614 mm for 1981–2010, and the average daily mean temperature was 1.8 °C. The mean temperature was –9.5 °C in January and +14.7 °C in July (Laudon *et al.* 2013).

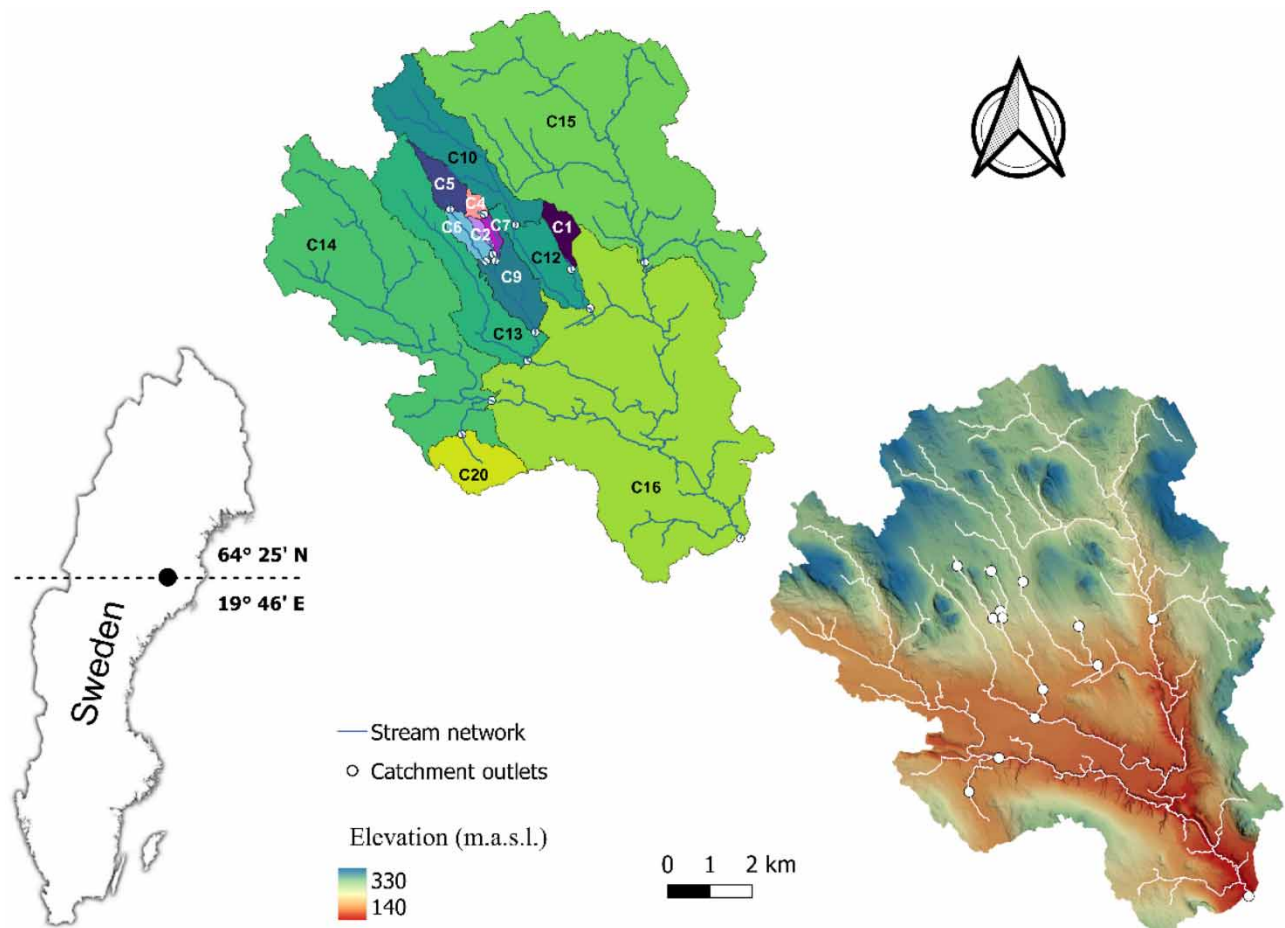


Figure 1 | Topography and subcatchments of the Krycklan catchment and its location in Sweden.

Table 1 | Catchment gauging set up and mean specific discharge values over the study period

Catchment	Gauge type	Mean specific discharge (mm d ⁻¹)	Water level logger
C1	90° V-notch weir	0.83	PT, TT
C2	90° V-notch weir ^a	0.63	PT, TT
C4	90° V-notch weir ^a	1.03	PT, TT
C5	120° V-notch weir, H-flume ^b	1.12	PT, TT
C6	Culvert	1.05	PT, TT
C7	90° V-notch weir ^c	0.84	PT, TT, Float
C9	Culvert	0.91	PT, TT
C10	Culvert	0.90	PT, TT
C12	Venturi flume	0.93	TT
C13	Trapezoidal flume	0.78	PT, TT
C14	Natural section	0.71	TT
C15	Natural section	1.03	PT, TT
C16	Natural section (bridge)	0.90	TT, RLS
C20	Culvert	0.97	TT

^aHeated weir since 2011.

^bHeated flume since 2012.

^cHeated weir since 1981.

PT, Pressure transducer (MJK 3400, with Campbell Scientific CR1000); TT, TruTrack capacitance rods (WTHR 1000); Float, OTT model X float strip chart recorder; RLS, OTT RLS Radar.

2.2. Input data

2.2.1. Hydrological data

Discharge observations from 14 subcatchments within the Krycklan catchment were used (Karlsen *et al.* 2019). Water levels, measured using automatic stage loggers, were possible year-round for five gauging stations in heated houses (C2 and C4 have been heated since 2011, C5 and C13 since 2012, and C7 since 1981). The monitoring program began in 1981 inside a heated hut (with an hourly resolution), and in 2003, gauging started in all 14 subcatchments. The lengths of all hydro-climatic data series in our study were adjusted to match the span of subcatchments with the least amount of data, covering a period of seven hydrologic years (1 October–30 September) from 2011 to 2017. Frequent manual water-level measurements (monthly during winter, minimum bi-weekly during the rest of the year) were made to calibrate automatic water level data, and stage-discharge relationships were defined using manual flow gauging (Karlsen *et al.* 2019). Specific discharge, defined as streamflow per unit drainage area, was calculated for each catchment.

2.2.2. Meteorological data

Air temperature, humidity, net radiation, and wind speed for estimating potential evapotranspiration (PET) via the Penman equation were measured in the central part of the Krycklan catchment at the Svartberget research station (Laudon *et al.* 2021). All climate data were assumed to be uniform for the whole catchment area. Precipitation from one centrally placed weather station was used (64°14' N, 19°46' E, 225 m a.s.l.) for all catchments as there is no significant elevation gradient observed for the region by the network of the Swedish Meteorological and Hydrological Institute (SMHI) (Karlsen *et al.* 2016).

2.2.3. Catchment characterization

A LiDAR DEM with 2 m resolution was created from a point cloud with a point density of 15–25 points/square meter. This DEM was hydrologically corrected by burning streams and culverts across roads as described in Lidberg *et al.* 2017. Catchment areas were delineated from the hydrologically correct DEM using Deterministic-8 (D8) (O'Callaghan & Mark 1984). The Swedish property map (1:12,500, Lantmäteriet Gävle, Sweden) was used to calculate forest, lake and, wetland coverage for each catchment. The proportion of soil type cover was calculated for sediment soils, till, and thin soils using the

Table 2 | Catchment characteristics of the entire Krycklan catchment (in bold) and its subcatchments (sorted from left to right by increasing catchment area)

Properties	Unit	C02	C04	C01	C07	C05	C06	C20	C09	C10	C12	C13	C14	C15	C16
TOPOGRAPHY															
Area	[km ²]	01	0.2	0.5	0.5	0.7	1.1	1.5	2.9	3.4	5.4	7.0	13.8	19.1	67.9
Median catchments area ^a	[km ²]	0.1	0.2	0.1	0.2	0.7	0.8	1.0	0.3	2.0	1.7	0.5	2.2	0.7	1.4
Elevation above sea level	[m]	275	287	279	275	293	282	211	252	297	277	251	229	278	239
Elevation above stream ^b	[m]	10.1	9.0	10.9	7.5	2.3	4.2	13.5	4.4	8.3	7.4	6.3	10.2	9.6	10.0
Slope	[%]	9	8	9	9	5	9	11	8	10	9	9	12	12	12
Aspect	[°]	139	146	177	149	168	164	164	161	161	164	156	171	180	172
GEOLOGY/SOIL															
Sediment	[%]	16	27	8	15	30	26	42	20	19	20	29	47	27	42
Till	[%]	84	22	92	65	16	39	45	62	52	61	57	43	56	47
LAND COVER															
Wetlands	[%]	0	44	2	18	40	25	10	14	26	17	10	5	15	9
Agricultural land	[%]	0	0	0	0	0	0	3	0	0	0	1	3	0	2
Lakes	[%]	0	0	0	0	6	4	0	1	0	0	1	1	2	1
Tree volume ^c	[m ³ ha ⁻¹]	212	83	187	167	64	117	59	150	93	129	145	106	85	106
Forest	[%]	100	56	98	82	54	71	88	84	74	83	88	90	82	87

^aMedian catchment area from 5 m LIDAR DEM, calculated similar to McGuire *et al.* (2005).

^bCatchment mean elevation above stream from 5 m LIDAR DEM, calculated similar to McGuire *et al.* (2005).

^cCalculated for the entire catchment using correlations between a forest inventory (from 110 plots) and LIDAR measurements (Laudon *et al.* 2013).

Note: Values in bold refer to the Krycklan catchment outlet (C16).

quaternary deposits map (1:100,000, Geological Survey of Sweden, Uppsala, Sweden). Landscape characteristics of all catchments, including the main outlet (C16, in bold), are summarized in Table 2.

2.3. HBV model description

A modified version of a bucket-type, semi-distributed hydrological model, namely the HBV model (Lindström *et al.* 1997) in the version HBV light (Seibert & Vis 2012) was used to simulate dynamic storage at a daily time step in the catchment. The HBV model is a widely used bucket-type model for simulating runoff. One advantage of HBV is the limited demand for input data. The input data for the HBV model were rainfall, air temperature data, and derived PET. The observed streamflow data were used for the calibration of the model. In general, the model consists of four commonly used routines which represent: (1) snow by a degree-day method; (2) soil moisture, and (3) groundwater by three linear reservoir equations; and (4) channel routing by a triangular weighting function. In the snow routine, a threshold temperature is used to distinguish between rainfall and snowfall, and snowmelt is considered by a degree-day approach.

A more detailed description of HBV's routines can be found in Seibert & Vis (2012). HBV-light includes three alternative model structures based on a varying number of conceptual reservoirs (one, two, or three) and different shapes of the storage-discharge relationship (linear versus nonlinear).

We implemented the three model structures of the response routine. This included a one-bucket structure with a single groundwater reservoir and three linear outflows (Q0, Q1, and Q2) at three different thresholds; a two-bucket structure with two reservoirs in parallel with a nonlinear outflow (Q1) in an upper bucket and one linear outflow (Q2) in a lower bucket (basic version); and a three-bucket structure with three parallel reservoirs (STZ, SUZ, and SLZ) and one linear outflow in each reservoir (Figure 2). It should be noted that these model structures differ only in the response routine, i.e., the same snow and soil routines were used in all model variants.

The built-in Genetic Algorithm followed by a Powell optimization (GAP, Seibert 2000) was used to calibrate the models for each catchment. For calibration, warm-up periods of at least one year were used for each catchment based on data availability. The initially chosen possible parameter ranges were adapted from Uhlenbrook *et al.* (1999) where preliminary simulations indicated that suitable parameter values were close to the limits.

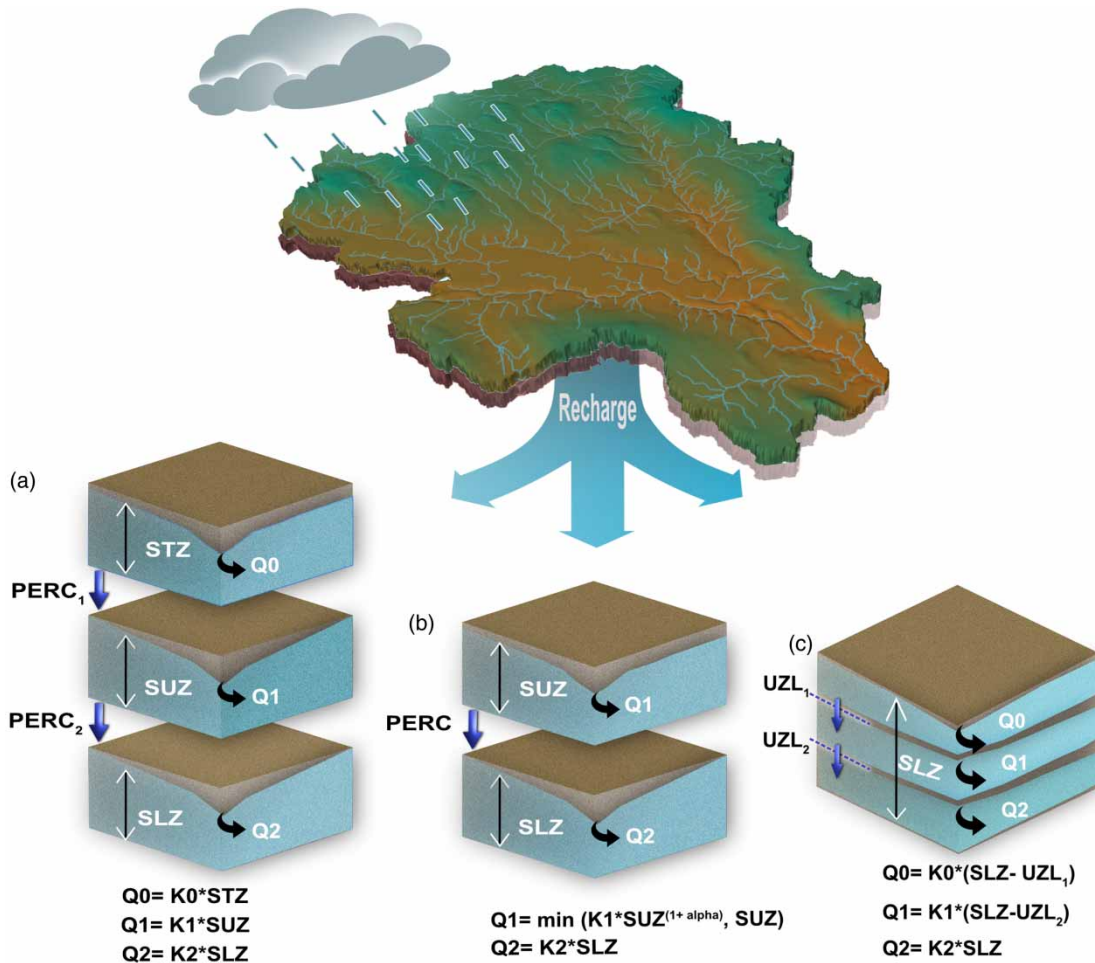


Figure 2 | Groundwater routines considered in this study. Three-bucket (a); two-bucket (b) and one-bucket (c). Recharge = Input from soil routine (mm/d); STZ = Storage in soil top zone (mm/d); SUZ = Storage in upper groundwater zone (mm/d); SLZ = Storage in lower groundwater zone (mm/d); UZL_i = Threshold parameter (mm/d); PERC_i = Maximum percolation to the lower zone (mm/d); Alpha = Non-linearity coefficient; K_i = Recession coefficient (d⁻¹); Q_i = Runoff component (mm/d).

To consider parameter uncertainty, each calibration trial was repeated 100 times (5000 iterations each) and the best 100 parameter sets were selected according to the Nash–Sutcliffe model efficiency. The Nash–Sutcliffe efficiency (NSE) coefficient, here called R_{eff} , is calculated with the following equation (Uhlenbrook *et al.* 1999):

$$R_{eff} = 1 - \frac{\sum (Q_{obs} - Q_{sim})^2}{\sum (Q_{obs} - \bar{Q}_{obs})^2} \quad (1)$$

where Q_{obs} and Q_{sim} are the measured (observed) and modeled (simulated) flows, respectively.

We considered parameterizations resulting in good runoff simulations in terms of Nash–Sutcliffe efficiencies (R_{eff}) as plausible. From the 100 calibration trials, we thus derived an ensemble of plausible simulations that result in a range of storage estimates for the upper (SUZ) and lower storage (SLZ) reservoirs. In the next step, the estimated storage from all reservoirs was combined and the difference between the minimum and maximum values was considered as dynamic storage.

A qualitative preparatory evaluation of parameter sensitivity indicated that automatic calibration could sometimes result in very small values for K₂ (baseflow recession coefficient) with tiny model performance improvements for these very small K₂ values (Figure 3 and Supplementary Material, Figure A). These very small K₂ values, in turn, resulted in extremely large estimates of the derived catchment storage for some of the catchments. This optimization issue, which greatly affected the storage

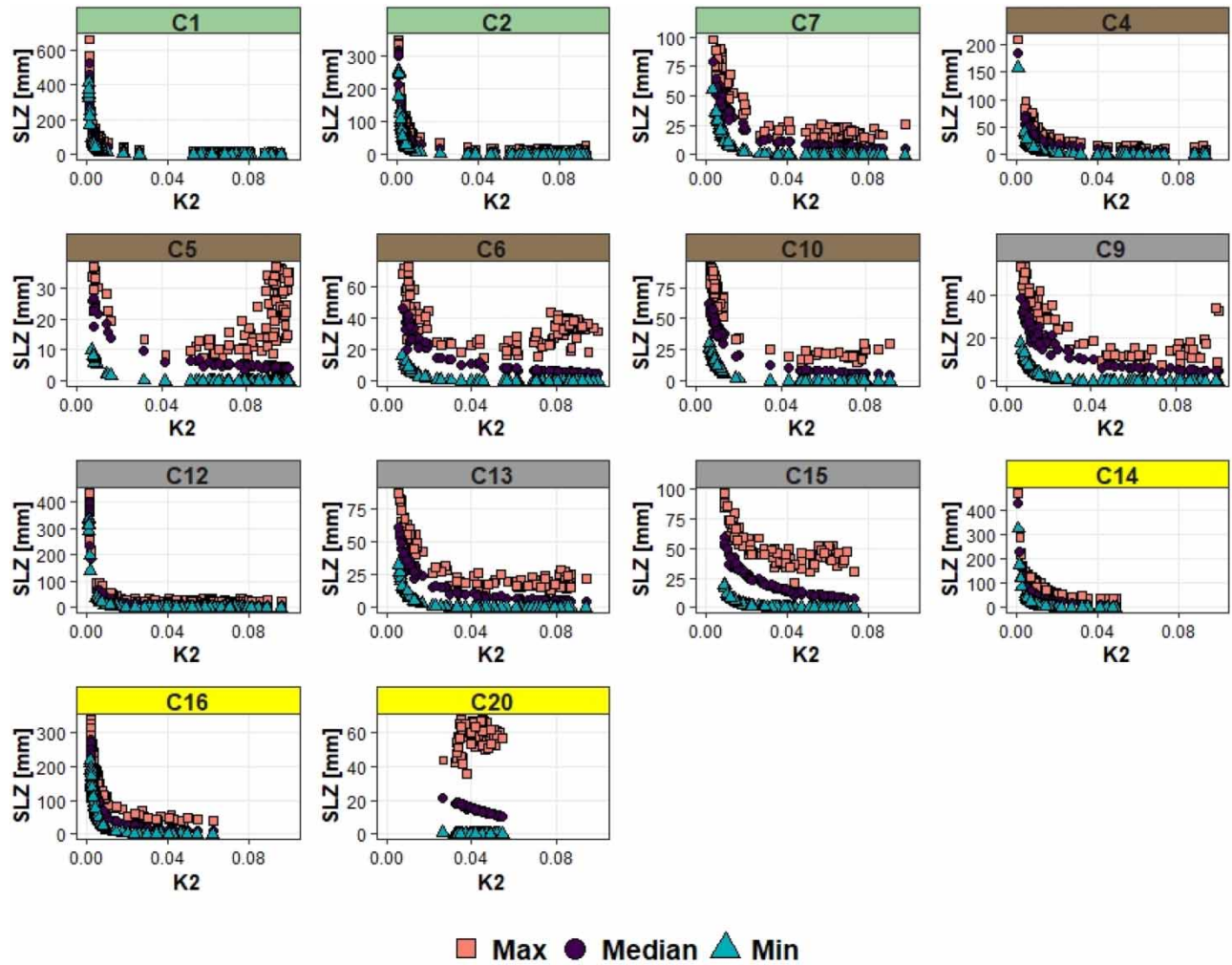


Figure 3 | Scatter plots of recession coefficient values for the lower groundwater zone (K2) versus the minimum, maximum, and median simulated lower groundwater storage (SLZ) derived from the first 100 calibration trials for each of the subcatchments. Plots are colored based on different dominant landscape types (green = forest on till, brown = wetland, gray = catchments with mixed characteristics, yellow = forest on sedimentary deposits). Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/nh.2022.121>.

simulation results, occurred mainly with the three-bucket model structure. All but one of the catchments (C20) showed a strongly nonlinear relationship between K2 and storage in the lower zone. To avoid artefacts like this, we limited the lower range of K2 to 0.02. The summary statistics (min, max, and mean) of calibrated parameter values with limited K2 parameter values are summarized in Table 3.

2.4. Water balance method

Based on simple water balance accounting (Equation (2)), time series of catchment water storage were obtained for each catchment.

$$V(t) = V_0 + \Delta t \sum_{i=1}^t (P_i - Q_i - E_i) \quad (2)$$

where P is the precipitation, E is the evapotranspiration, Q is the streamflow, and V(t) and V₀ are the storage at time step t and t = 0, respectively. We then calculated the total water storage, ΔV as the difference between the minimum and maximum of the estimated storage volumes over the observation period (Staudinger *et al.* 2017).

Table 3 | Calibration results with limited K2 parameter for the catchment outlet, C16

Parameter	Description	Calibration range	Min	Max	Median
One-bucket					
PERC	Maximum flow from upper to lower zone (mm/d)	0–4	3.07	3.99	3.94
UZL	Threshold parameter (mm)	0–70	20.31	60.47	24.17
K0	Recession coefficient of fast runoff (d^{-1})	0.1–0.5	0.10	0.37	0.18
K1	Recession coefficient of upper box (d^{-1})	0.01–0.2	0.01	0.10	0.04
K2	Recession coefficient of lower box (d^{-1})	0.02–0.1	0.02	0.07	0.03
Two-bucket					
PERC	Maximum flow from upper to lower zone (mm/d)	0–4	0.93	4.00	3.04
Alpha	Nonlinearity coefficient	0–1	0.14	0.60	0.31
K1	Recession coefficient of upper box (d^{-1})	0.01–0.2	0.02	0.19	0.11
K2	Recession coefficient of lower box (d^{-1})	0.02–0.1	0.02	0.07	0.05
Three-bucket					
PERC	Maximum flow from upper to lower zone (mm/d)	0–4	0.80	3.56	1.66
UZL	Threshold parameter (mm)	0–70	4.08	11.43	7.15
K0	Recession coefficient of fast runoff (d^{-1})	0.1–0.5	0.30	0.49	0.43
K1	Recession coefficient (d^{-1})	0.01–0.2	0.07	0.18	0.11
K2	Recession coefficient of lower box (d^{-1})	0.02–0.1	0.02	0.06	0.02

2.5. Correlation analysis between landscape characteristic and dynamic storage

To investigate the relationship between landscape characteristics and simulated dynamic storage, we used non-parametric Spearman rank correlations (Spearman 1904). The storage metrics used in this analysis were dynamic storage (the difference between the minimum and maximum storage), mean storage (mean value during the entire analysis period), storage in the upper and lower zone separately (SUZ and SLZ), and mean storage for each season. Instead of using a single model output, we used the average of all three model structures to calculate these storage metrics and then examined their relationship with landscape characteristics.

3. RESULTS

3.1. Model performance analysis

Based on the results, differences in model performance were found between catchments with different landscape characteristics. There are also performance differences between the different model structures. The range of the best 100 model performances (R_{eff}) of the three model structures for each subcatchment is shown in Figure 4. The plot suggests that model performance decreases as the mean elevation above the streams increases ($r = -0.85$, $p < 0.05$). Moreover, the three-bucket model gave good agreement between the simulated and observed streamflow in all the catchments (R_{eff} between 0.76 and 0.87). The one-bucket model performed well in the lake-influenced catchments, C5 and C6 ($0.80 < R_{\text{eff}} < 0.87$), while it performed worse for large catchments with deep soils, C14 and C16 ($0.68 < R_{\text{eff}} < 0.75$). This indicates that the one-bucket model could not represent the hydrological functioning as well as the three-bucket model in the large catchments with deep sediment soils. The one-bucket model resulted in the poorest simulations in C14 and the best simulation in C6 with maximum R_{eff} values of 0.74 and 0.87, respectively. Additionally, the two-bucket model with two storage reservoirs and only two outflows (intermediate and baseflow) performed similarly to the one-bucket model for C2, C4, C9, and C20. The highest model performances were obtained for C5, C6 and, C9, which have a high lake percentage and lower elevation above their stream network. The simulated discharge time series derived from the three model structures calibrated against observed discharge, for four representing catchments including C4, C5, C7, and C16, are shown in supporting information Figure B1–4. These catchments are also representing the highest (C4 and C5), average (C7), and lowest NS values (C16). We also provided a detailed analysis of snow water equivalent and soil moisture estimated using the three model structures for the above-

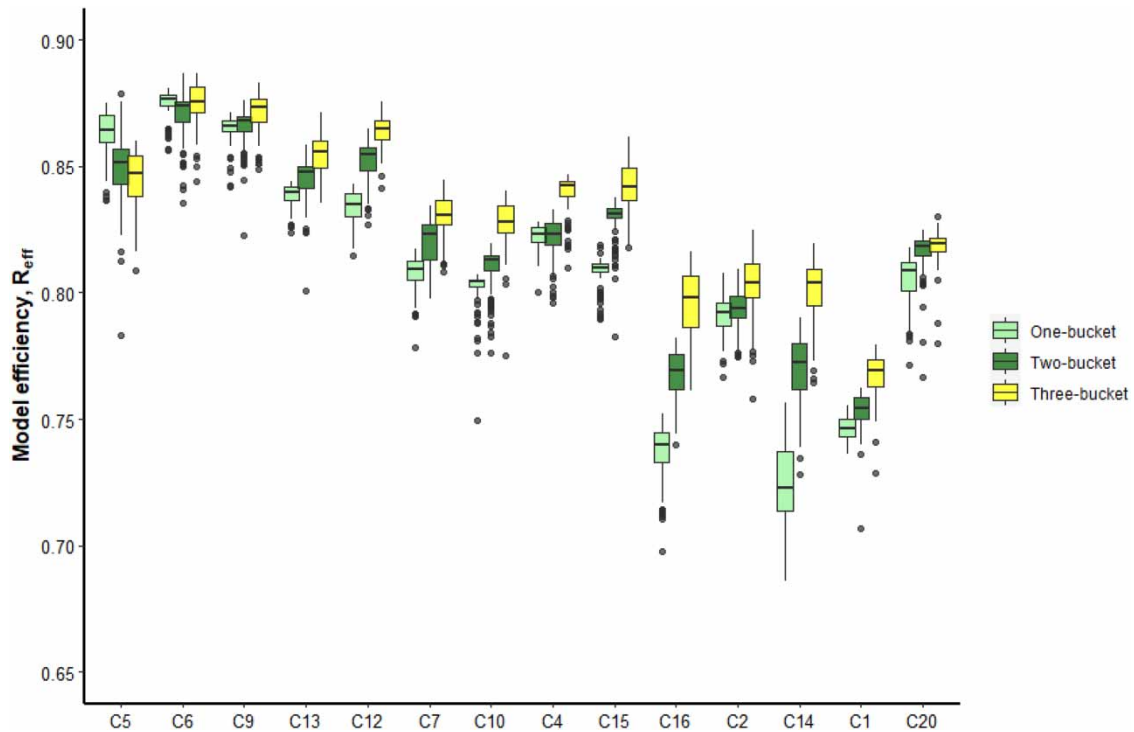


Figure 4 | Boxplots represent the variation of model performance during the calibration period. The catchments are sorted from left to right in ascending order of their mean elevation above the stream network. The lower and upper endpoints of the boxes represent the 25th and 75th percentiles of model performance over 100 parameterizations. The whiskers show the range of model performance between the minimum and maximum values. The horizontal bar and dots show the median value and outliers, respectively.

mentioned catchments. Results showed that all model structures simulated snow water equivalent relatively similar for the four representing catchments, though C4 and C5 had a slightly higher amount (supporting information Figure C1-4).

3.2. Flow characteristics

The contribution (as a fraction of the total discharge) of outflow from the surface (Q0), upper (Q1), and the lower (Q2) flow paths for all the model structures are shown in Figure 5 (note that the standard version of HBV has only two outflows, Q1 and Q2). The boxplots are based on 100 streamflow characteristic values derived from 100 different parameter sets. When comparing the runoff components calculated by all three model structures, remarkable differences were seen between different subcatchments and the three HBV model structures. For example, the difference in the contribution of different Q components between the sediment and till-dominated catchments was apparent. It can be seen in Figure 5 that the calibrated contribution of discharge from the lower groundwater bucket (baseflow) was higher in larger catchments with more sediment soils (e.g., C14, C15, and C16). In contrast, in catchments C5 and C4, the fast flow (Q1) showed a more significant contribution to total runoff generation.

3.3. Comparisons of dynamic storage within and between catchments

Simulation results from the three-bucket structure, which showed reasonably good performance across all catchments, demonstrate differences in storage components between catchments. The simulated maximum storage in the upper and lower groundwater bucket are plotted in Figure 6. It can first be noted that in wetland-dominated catchments, the maximum storage in the upper groundwater zone was higher than the other subcatchments. In contrast, catchments dominated by sediment soils retained a higher amount of water in their lower groundwater bucket. It should also be noted that the upper reservoir (SUZ) is nearly empty during long periods and the mean value is, therefore, small. The simulated daily storage in upper and lower reservoirs estimated by each model structure was then aggregated to compare changes in dynamic storage estimated by different model structures.

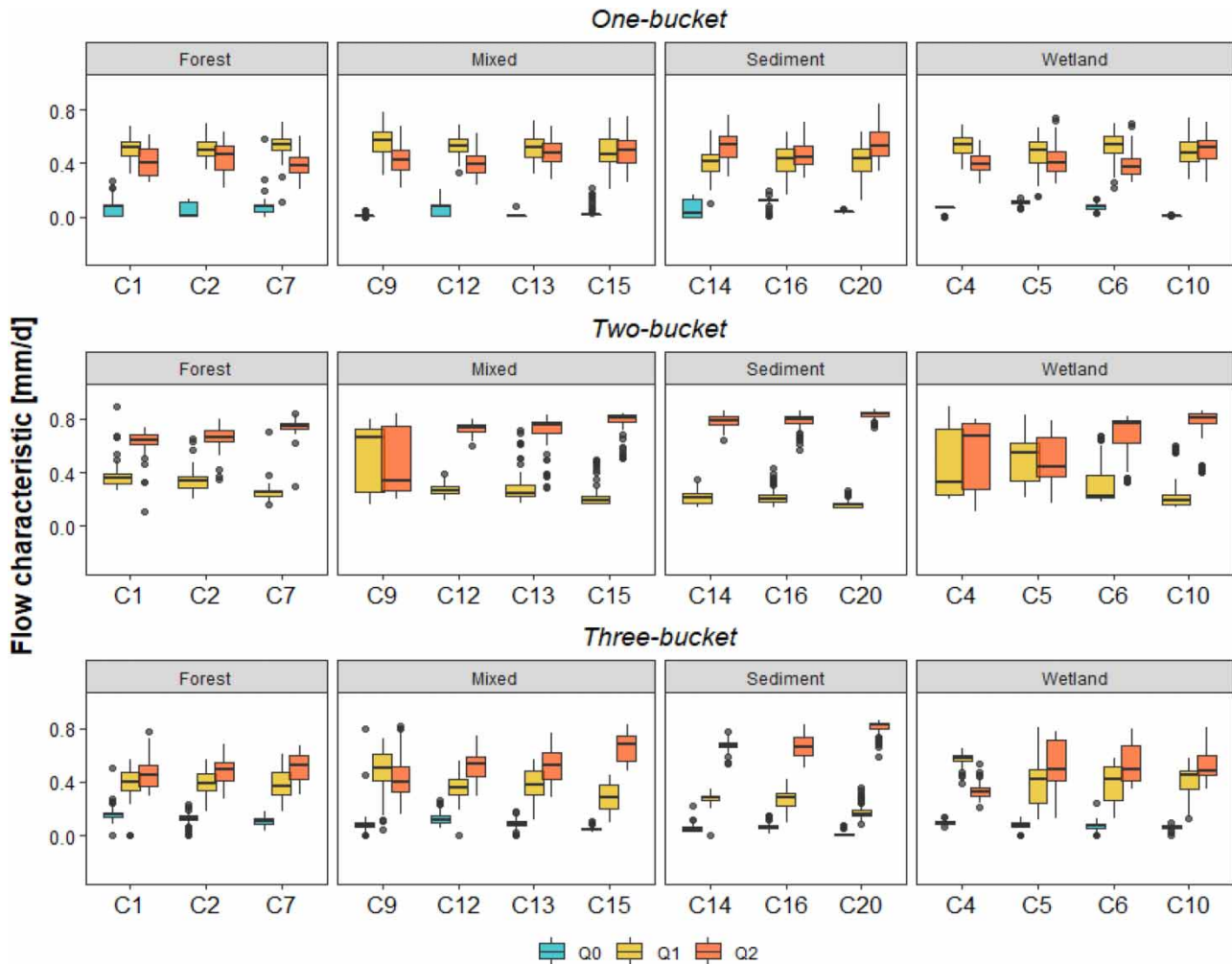


Figure 5 | Fraction of fast runoff (Q0), delayed runoff (Q1), and slow runoff (Q2) to catchment outflow derived from different model structures.

The dynamic storage estimates calculated as the difference between the maximum and minimum groundwater storages from the different model structures are presented in Figure 7. While the estimated storage volumes varied, a consistent overall pattern among the catchments could be observed. In general, the forested catchments had lower storage estimates than the catchments with a lake or large wetlands, and storage estimates were even larger for the sediment catchments. For some of the catchments (C5, C6, and C9), the estimated dynamic storages were similar for the different model structures. In contrast, the differences were higher for others (C14, C15, and C16).

In contrast to the overall range of dynamic storage, the mean storage value calculated by each model structure showed large variation among catchments (Figure 8, Table 4).

For most of the catchments there was only a slight increase in estimated dynamic storage from the one-bucket to the two-bucket structure, but a significant increase from the two-bucket to the three-bucket in most of the catchments (Figure 8). This pattern was opposite for the lake dominated-catchment (C5).

Simulated dynamic storage, however, exhibited more variation both within and among the catchments on shorter time-scales (Figure 9). The variation in dynamic storage between catchments was highest in spring with a minimum value of 9 mm for C1 and a maximum value of 30.9 mm for C20, and lowest in winter with a range of 5 mm. According to the three-bucket model, C5 had the largest seasonal variation with the mean storage approximately 5 times greater in spring compared to winter. Additionally, we quantified the similarity degree between the mean seasonal storage estimates by the three model structures using the Spearman correlation coefficient (Table 5). The comparison of similarity metrics revealed that the

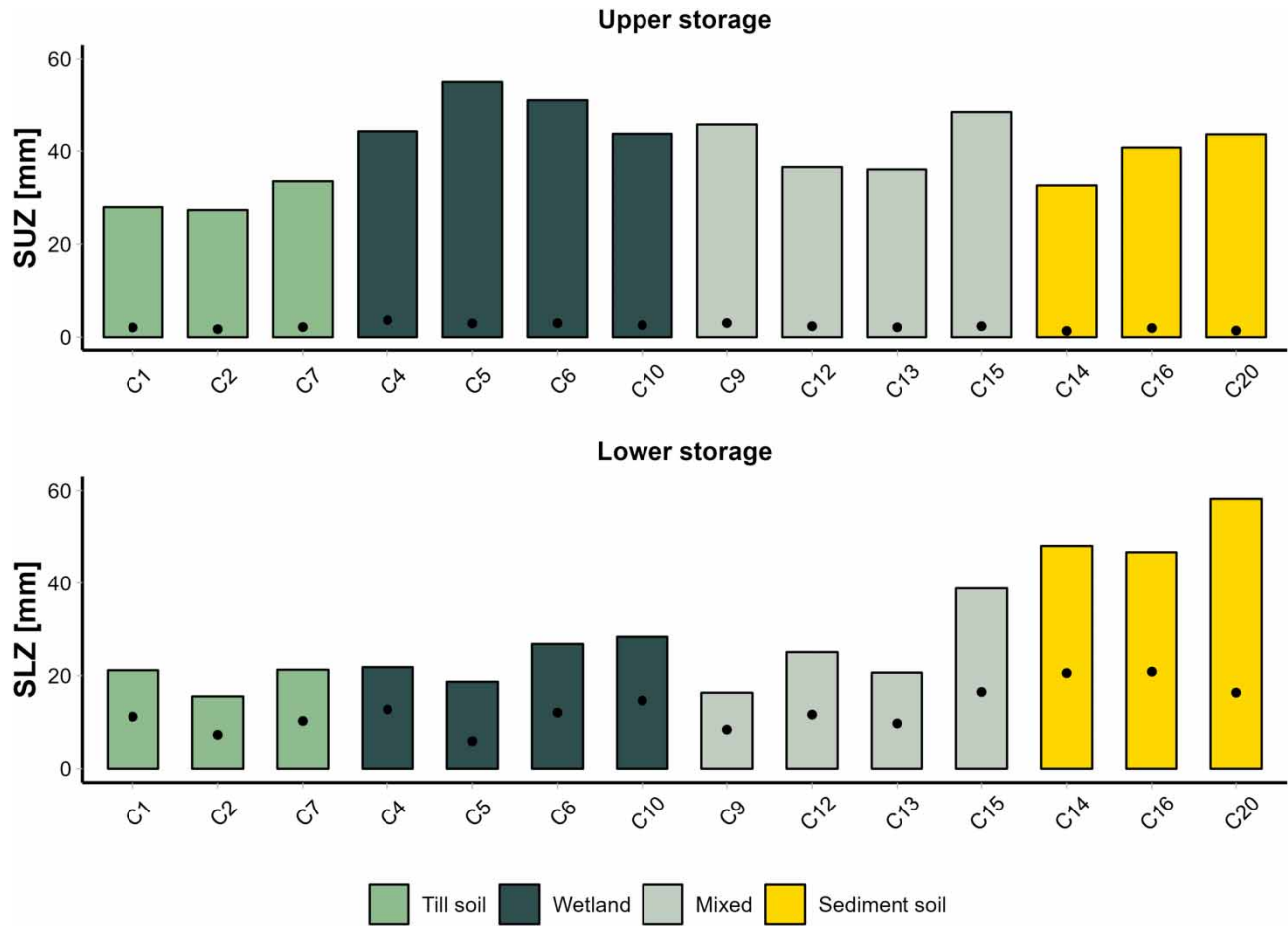


Figure 6 | The maximum amount of storage (mm) in the upper (SUZ) and lower (SLZ) groundwater zones estimated by the three-bucket model structure over the whole study period. The black dots show the mean storage over the entire period.

correlation between the mean storages estimated by the three model structures was highest for spring and lowest for winter. Moreover, the one-bucket and three-bucket structures had the largest differences in estimating dynamic storage during all seasons. These differences were highest for C14 and lowest for C5.

3.4. The role of landscape characteristics

The results of Spearman's correlation analysis between catchment storage and dominant landscape characteristics are summarized in Table 6. The test showed a positive relationship between catchment tree volume, till soils, and total water storage, ΔV . This indicates that as the amount of tree volume and till soils increase, the total water storage, ΔV also increases ($r = 0.60$, $p < 0.05$). In contrast, with increasing wetland percentage, the amount of total water storage, ΔV decreased ($r = -0.59$, $p < 0.05$). Dynamic storage (HBV), on the other hand, showed a negative correlation with tree volume ($r = -0.77$, $p < 0.05$) and till soil ($r = -0.60$, $p < 0.05$). It also showed a positive correlation with lake percentage. Positive correlations were also found between catchment area, mean storage during the entire period, and mean seasonal storage (winter, spring, and summer). Elevation above stream (EAS) and slope both showed a negative relationship with storage in the upper zone (SUZ) and a positive relationship with storage in the lower zone (SLZ). In contrast, there was no correlation between mean catchment elevation and any of the storage metrics.

Catchment area and sediment soil were positively correlated to mean SLZ, which means that catchments with a larger area, higher EAS, larger amounts of sediment soil, and steeper slope would have higher storage in the lower groundwater bucket. For the mean storage over the whole study period, a positive relationship was observed with the mean catchment area and the proportion of sediment soil. However, the mean storage was inversely affected by till soils and tree volume

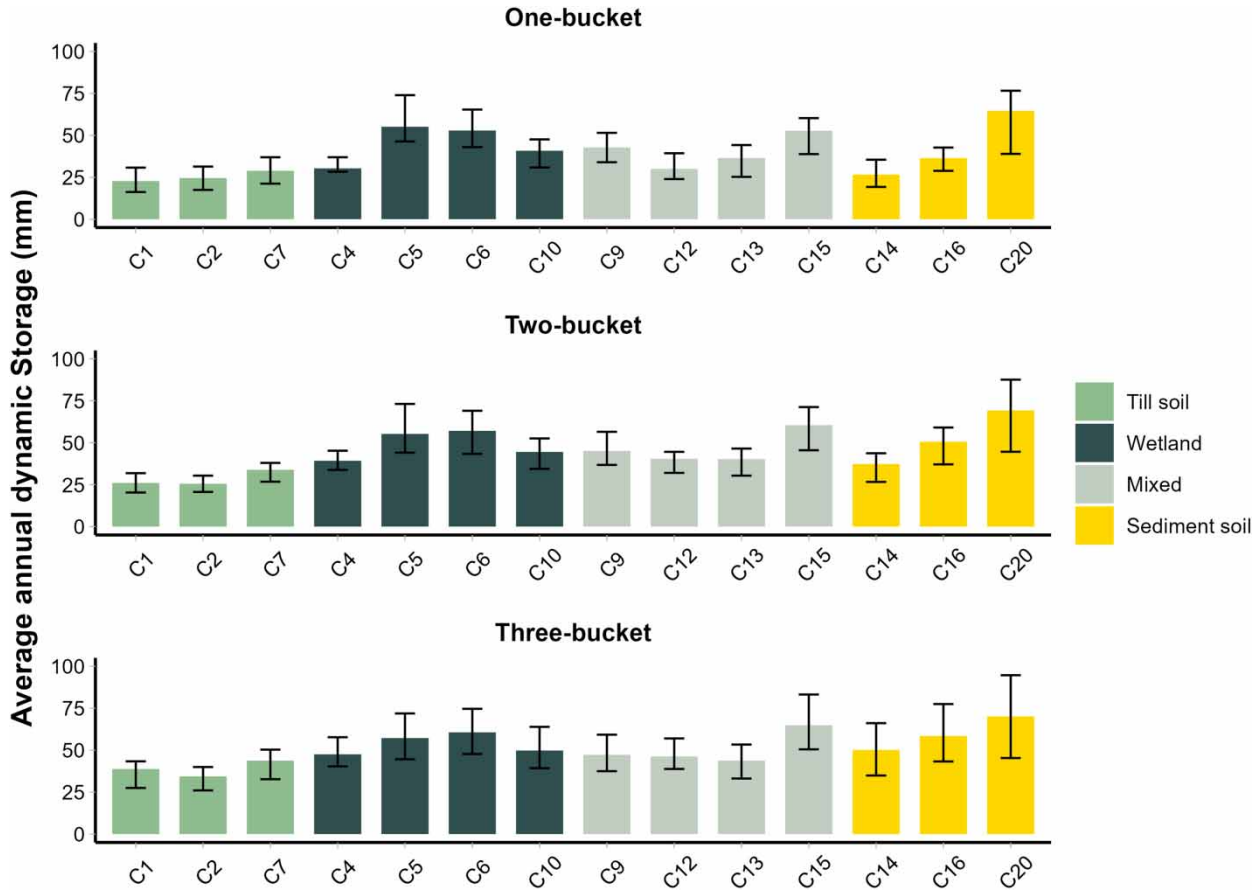


Figure 7 | Dynamic storage estimated by the three HBV model structures. The error bars represent the intra-annual variability of dynamic storage for the study period (2011–2017).

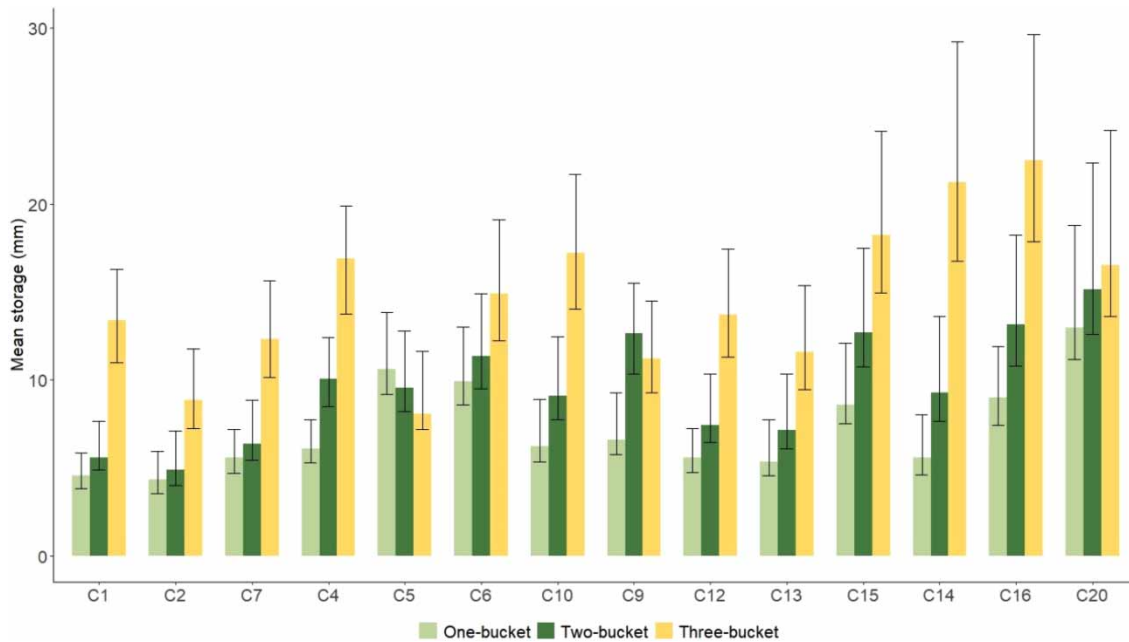


Figure 8 | Mean storage estimates by all model structures. The error bars represent the intra-annual variability of mean storage for the study period (2011–2017).

Table 4 | Dynamic storages as estimated by the three model structures

	C2	C4	C1	C7	C5	C6	C20	C9	C10	C12	C13	C14	C15	C16
One-bucket														
Min	0.01	0.00	0.01	0.02	0.10	0.13	0.19	0.02	0.01	0.01	0.00	0.03	0.05	0.09
Max	31.37	37.03	30.76	37.04	74.25	65.82	77.32	51.61	47.62	39.44	44.24	35.56	60.48	43.11
Mean	4.48	6.38	4.66	5.73	11.16	10.47	14.17	7.20	6.75	5.82	5.77	5.92	9.38	9.29
Two-bucket														
Min	0.01	0.35	0.03	0.01	0.10	0.27	0.04	0.67	0.22	0.00	0.08	0.13	0.14	0.23
Max	30.41	45.61	31.91	37.99	73.21	69.43	87.91	58.24	53.09	44.58	46.73	44.13	71.61	59.80
Mean	5.27	10.18	6.01	6.84	10.06	11.75	16.39	12.53	9.58	8.07	7.74	9.99	13.53	13.83
Three-bucket														
Min	0.29	0.98	0.73	0.46	0.04	0.50	0.13	0.38	0.89	0.41	0.43	1.22	0.52	1.22
Max	40.78	60.25	44.37	51.52	71.95	76.14	95.25	60.32	66.34	58.14	54.57	69.89	84.72	80.90
Mean	9.00	16.43	13.26	12.43	8.89	15.08	17.76	11.45	17.23	14.00	11.84	21.89	18.86	22.82

Catchments are sorted from left to right by increasing their area. Values with darker colors correspond to larger values.
 Note: The color scale shows the range of data in each row (e.g. the max value in the range has the darkest color).

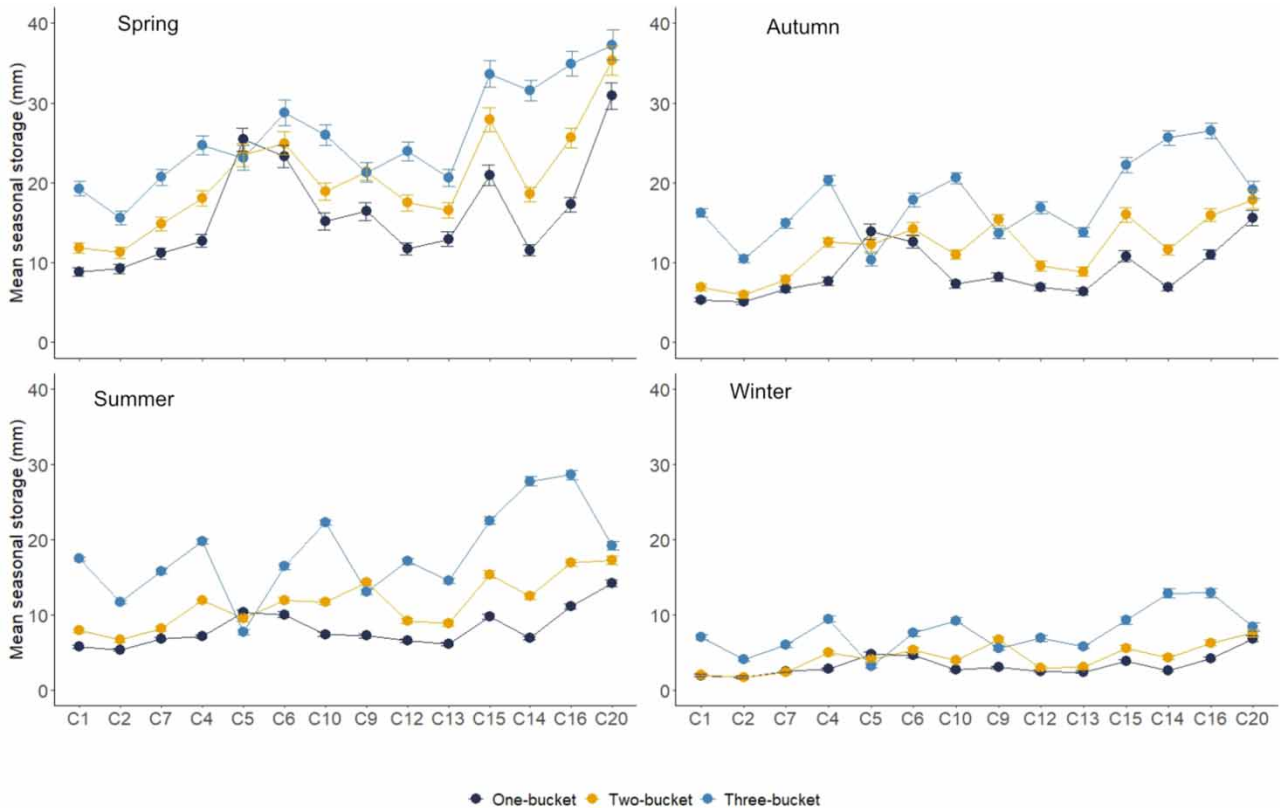


Figure 9 | Mean seasonal water storage estimated by the three model structures. Seasons were divided into four seasons: winter (NDJFM), spring (AM), summer (JJA) and autumn (SO).

Table 5 | Spearman proximity matrix between mean seasonal storages calculated using the three different model structures

	One-bucket	Two-bucket	Three-bucket
Spring			
One-bucket	1	0.916	0.657
Two-bucket	0.916	1	0.864
Three-bucket	0.657	0.864	1
Summer			
One-bucket	1	0.824	0.319
Two-bucket	0.824	1	0.565
Three-bucket	0.319	0.565	1
Autumn			
One-bucket	1	0.895	0.284
Two-bucket	0.895	1	0.486
Three-bucket	0.284	0.486	1
Winter			
One-bucket	1	0.846	0.266
Two-bucket	0.846	1	0.442
Three-bucket	0.266	0.442	1

Table 6 | Spearman rank correlation coefficients between physical catchment properties and storage metrics

Variables	ΔV [mm]	Dynamic storage	Mean storage	Summer	Autumn	Winter	Spring	SUZ	SLZ
Area (log)	0.042	0.437	0.596	0.596	0.534	0.618	0.543	-0.231	0.604
Elevation	-0.015	-0.044	-0.073	-0.112	-0.079	-0.073	-0.040	0.385	0.002
EAS	0.332	-0.187	0.213	0.429	0.165	0.262	0.029	-0.749	0.525
Slope	0.226	0.124	0.513	0.670	0.469	0.509	0.374	-0.670	0.783
Mire	-0.590	0.306	0.130	-0.002	0.187	0.057	0.222	0.816	-0.084
Lake %	-0.265	0.633	0.307	0.033	0.326	0.272	0.503	0.298	-0.085
Till	0.615	-0.601	-0.604	-0.388	-0.604	-0.617	-0.619	-0.148	-0.361
Sediment	0.032	0.371	0.624	0.636	0.531	0.657	0.499	-0.524	0.631
Tree volume (m ³ ha ⁻¹)	0.603	-0.770	-0.750	-0.576	-0.719	-0.689	-0.796	-0.158	-0.475
Soil depth (m)	0.011	0.383	0.432	0.383	0.374	0.471	0.407	-0.476	0.306
ΔV [mm]	1	-0.534	-0.393	-0.169	-0.433	-0.336	-0.490	-0.292	-0.125
Dynamic storage	1	0.793	0.538	0.767	0.697	0.938	0.138	0.332	
Mean storage		1	0.921	0.982	0.978	0.938	-0.081	0.811	
Summer				1	0.912	0.938	0.763	-0.156	0.934
Autumn					1	0.969	0.925	-0.002	0.802
Winter						1	0.877	-0.103	0.855
Spring							1	0.007	0.618
SUZ								1	-0.310
SLZ									1

Values in bold are different from 0 with a significance level $\alpha = 0.05$.

Note: Mean storage is the average of model-based storage for the whole period of 2011–2017. Dynamic storage is the difference between min and max model-based storage.

(Figure 10(c) and 10(d)). For the mean seasonal storage, an inverse correlation was found between till soils and the mean storage in autumn, spring, and winter, with a much weaker correlation in summer ($r = -0.388$, $p < 0.05$).

Tree volume also revealed a negative correlation with all seasonal storage, although this relationship was stronger in winter and summer. Furthermore, a positive correlation was observed between sediment soil and the mean storage in summer and winter. This highlighted the role of sediment soils in maintaining a relatively high flow, mainly during the dry periods. The mean slope showed a significant correlation only to the mean storage during summer. Contrary to what was expected, the mean catchment soil depth did not correlate with any storage metrics.

4. DISCUSSION

4.1. Model performance comparison

The analysis of model performances showed that the three-bucket model captures the hydrological behavior best in all catchments. This finding is similar to *Stoelzle et al. (2015)*, who suggested the simple structures of one or two reservoirs are less efficient in simulating catchment baseflow. *Fenicia et al. (2014)* also found that single-reservoir models are too simplistic to identify the different runoff-generating processes and that models with multiple parallel flow paths performed better. In addition, for all model structures, the best model performances were achieved in catchments with lower slope and mean elevation above streams (EAS) (Figure 4). A possible explanation is that with increasing elevations above the stream, the spatial heterogeneity of flow paths and processes within the respective catchments increases. Therefore, models that treat catchments as a single unit might have difficulties in representing the hydrologic functioning of these catchments.

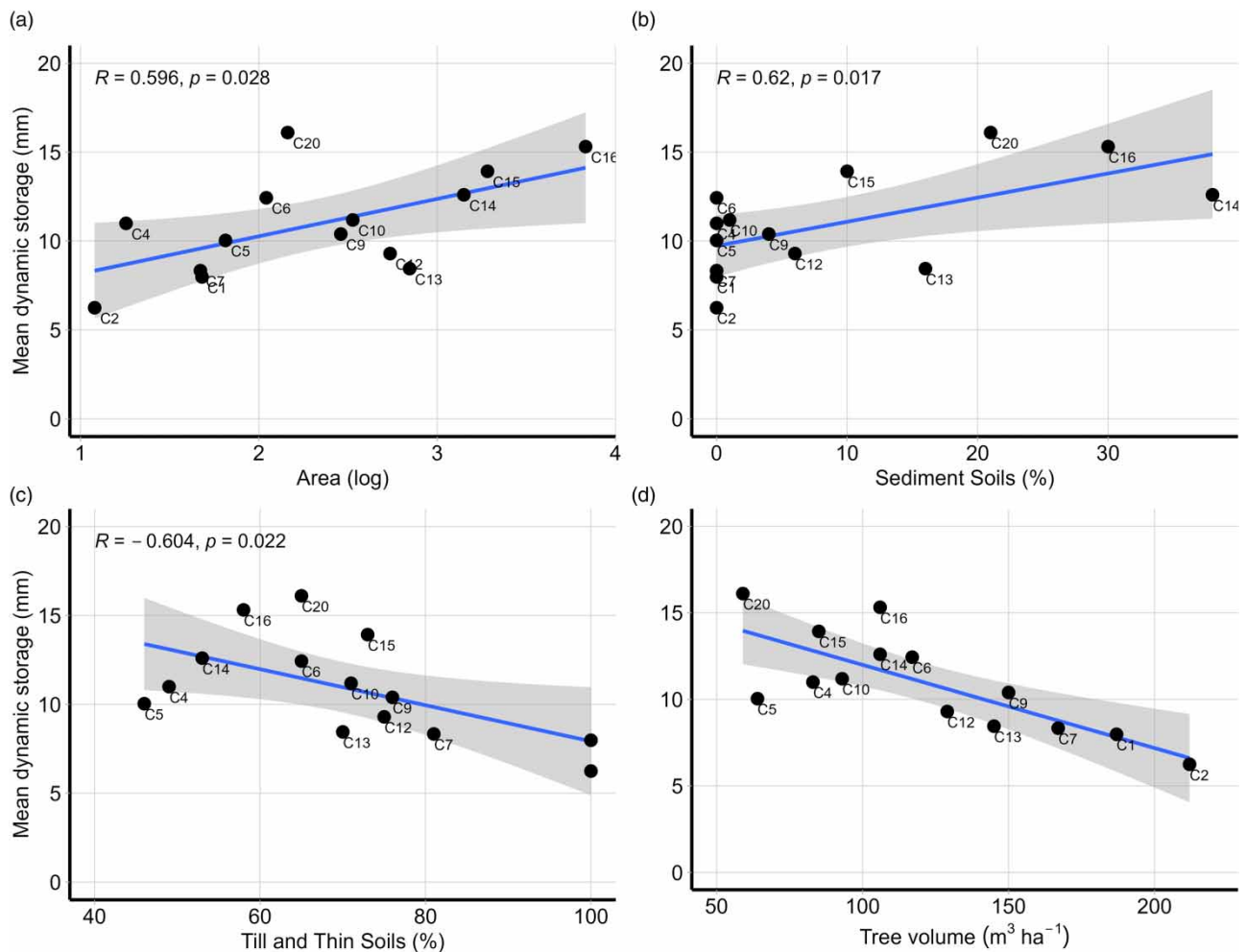


Figure 10 | Scatter plots illustrating the correlations between dynamic storage and dominant catchment characteristics.

Our results also suggest that in large catchments with higher slope and EAS, the one-bucket and two-bucket reservoirs were not adequate, and better performance was achieved by multiple reservoirs and flow pathways. The underlying reason could be that for the larger catchments there also is a higher percentage of sediment cover and deeper soils, which both can increase spatial variability within a catchment and, as a result, affect model performance. Furthermore, the relatively poor performance of all three model structures in large sediment catchments suggests that capturing the storage–discharge functioning is more challenging than the other catchments, likely because of more complex hydrological responses (Jutebring-Sterte *et al.* 2021). However, our results contradict the findings of other studies. For example, Broderick *et al.* (2016) used the NSE criterion and found higher model performance for catchments with greater storage capacity and baseflow as they are less sensitive to storm events, and therefore produce a less variable flow series. Additionally, Van Esse *et al.* (2013) tested 12 different conceptual model structures on 237 catchments in France and found that conceptual models have higher efficiency in larger catchments.

4.2. Runoff components differences

The analysis of simulated runoff components suggests various dominant hydrological functioning in different catchments that can potentially be attributed to the spatial heterogeneity of soil properties in the studied catchment. The results of 100 calibration trials showed the importance of baseflow in sustaining runoff generation, especially in sediment soil catchments. This is in line with findings from Karlsen *et al.* (2019), Teutschbein *et al.* (2015), and Peralta-Tapia *et al.* (2015) that also demonstrated that the amount of baseflow draining into the streams has a positive correlation with increasing catchment area and sediment cover.

Conversely, in wetland-dominated catchments (C4 and C5), overland flow from peat soils dominates storm runoff generation. Laudon *et al.* (2007) and Seibert *et al.* (2003) have also pointed out that in peatlands, groundwater levels reach the soil surface and, combined with overland flow from snowmelt, results in higher specific discharge compared to other catchments. In contrast, larger catchments with deep sediment soils have deeper flow paths and higher contribution to baseflow.

4.3. Partitioning of storage between upper and lower reservoirs

We also illustrated and compared the differences in the simulated storage components and flow paths among the catchment categories. We observe differences in dynamic storage at various locations (upper and lower storage zones) within the catchment between different landscape classes. Since the studied catchments are partially nested and have a similar climatological regime, the variability in storage behavior should be mainly due to different physical attributes such as soils, landform, and vegetation cover.

The large sediment catchments in the lower part of Krycklan had a higher infiltration rate and deeper subsurface flow paths. This suggests that water is stored mainly in the lower storage zone and thus can help maintain baseflow during dry periods. Our results are consistent with those obtained by Peralta-Tapia *et al.* (2015), who used stable isotopes to show that the contribution of deeper groundwater flow paths to catchment discharge increases with the catchment size. Conversely, the wetland-dominated catchments had more storage volume in the upper groundwater zone during high flows. This supports the finding of Laudon *et al.* (2007) that in wetland-dominated catchments, the proportion of new or event water was much higher than forest on till catchments. Using isotopic data, they also concluded that there is a shallow pathway in wetland catchments close to the surface, which might be formed by a concrete frost layer inhibiting the infiltration of rain and meltwater.

4.4. Influence of different model structures on dynamic storage

The different model structures had a varying number of storage reservoirs, which could partly explain differences in model performance as they capture hydrological catchment functioning differently. However, our main question was how much the estimated storages differed between these structures. As expected, using multiple storage reservoirs, even with the same number of parameters (one-bucket and three-bucket structures), makes a large difference in some catchments. Typically, for the most lake-influenced catchments (C5, C6, and C9), as well as the sediment dominated catchment C20, the modeling results were similar for all model structures, which implies that regardless of the high model performance value, model structures with only one reservoir can adequately represent the dynamic storage behavior. The explanation for C20, which was an outlier of sediment categories, could be due to its smaller drainage area compared to other sediment-dominated catchments (C14 and C16).

Although applying different model structures resulted in different estimates of dynamic storage for each catchment, all dynamic storage estimates calculated for different model structures ranked similar among the catchments. For instance, all model structures resulted in less dynamic storage for till catchments, while a higher amount of dynamic storage was obtained for those most influenced by lakes. These results are in line with [Karlsen *et al.* \(2019\)](#), who calculated dynamic storage, as the difference between the observed daily minimum and maximum specific discharge for each catchment and found the lowest dynamic storage for C2 (20 mm), and the highest for C6 (75 mm).

Moreover, we found that the spatial variability of dynamic storage is higher in spring, possibly due to differences in soil frost extent during snowmelt, depending on the vegetation cover and soil type. The difference between the two dominant soil groups (till and sediment) in transferring snowmelt water inflow within the catchment may also account for this difference. Catchments with sediment soil have deeper, or probably more groundwater paths, while catchments with till soils are less permeable so that most of the snowmelt leaves the catchment as overland flow and through the upper storage box.

As mentioned above, the boreal ecosystems are characterized by long winter and large snow accumulation. Therefore, spring snowmelt has a dominant contribution to the annual storage-discharge magnitude in northern regions ([Laudon & Ottosson Löfvenius 2016](#)). The simulation results determined that although the precipitation and temperature were similar across all catchments, the differences in catchment topography, soil, and land cover resulted in differences in the snow accumulation among catchments. Compared to other forested similar size catchments, the higher amount of discharge and dynamic storage of lake and wetland-dominated catchments (C5 and C4) during spring snowmelt can be explained by the larger snowpacks in these open canopy catchments ([Kozii *et al.* 2017](#)). In seasons when rainfall has the dominant effect on runoff, the lake catchment, C5, had the smallest amount of dynamic storage compared to other catchments. This agrees with previous findings suggesting that deep glacial till soils, wetlands, and forests on sediment soils have more storage during low flow conditions, while shallower till soils and open-water wetlands have been shown to sustain less water as they already contain a lot of water and their storage capacity (active storage) is small ([Meriö *et al.* 2019](#)).

The three different model structures showed that although the pattern of dynamic storage (the difference between the minimum and maximum annual storage values) estimates among catchments were similar, the catchments ranking for mean storage values differed greatly. However, for every model structure, catchments with higher tree volume and more till soils had a lower amount of dynamic storage compared to sediment catchments with less tree volume. This is likely due to transpiration and interception losses that reduce groundwater recharge ([Ilstedt *et al.* 2016](#); [Bonnesoeur *et al.* 2019](#)), or soil water uptake within the tree root zone ([Allen & Chapman 2001](#)).

4.5. Importance of using an ensemble average of different model outputs in the final prediction

To consider parameter uncertainty and to ensure that results are not affected by it, as a first step we calculated the conceptual dynamic storage using an ensemble average taken over 100 best simulations, to have more robust storage estimations. Second, the results revealed that although the three-bucket model performed relatively better for all catchments, for some catchments the one-bucket model with less complexity also yielded a high model performance. Additionally, our findings indicated that the three-bucket model, despite its high performance, had more uncertainty in simulating dynamic storage than the one-bucket model, which gained the lowest model performance. Most studies, on the other hand, have focused on the impacts of model structures on discharge simulation ([Uhlenbrook *et al.* 1999](#); [Van Esse *et al.* 2013](#); [Fencia *et al.* 2014](#); [Parra *et al.* 2019a](#)), and considered the performance criteria to determine the predictive power of a model. With our focus on estimating a fraction of total storage that is actively controlling discharge release, and using it as a comparison of hydrological functioning across contrasting boreal catchments, makes the validation more difficult.

Therefore, we argue that using a simple ensemble averaging method that combines the prediction of each model equally would decrease the weakness of every single model in representing the hydrological functioning. This approach will also reduce the biases resulting from modelers' personal preferences

4.6. Dependence of dynamic storage on catchment characteristics

The results from the Spearman rank correlation tests suggest that the largest mean storage estimates were found in larger catchments, and catchments characterized by a large proportion of sedimentary soil, especially during winter. This finding is in accordance with other studies in Krycklan. For example, [Karlsen *et al.* \(2016\)](#), who applied partial least square regression to quantify the linkage between recession characteristics and catchment properties found that with increasing catchment size, the relative contribution of groundwater, especially during winter baseflow, increased. This is also in agreement with [Tiwari](#)

et al. (2014) and Peralta-Tapia *et al.* (2016), who have shown that deep groundwater contribution in the catchment increases with catchment size.

When comparing our findings with Jutebring Sterte *et al.* (2021), a similar pattern was seen between travel time and dynamic storage among the catchments. For example, the modeled mean travel time (MTT) increased with increasing slope, sediment soil proportion, and catchment size. In their study, C20 was seen as an outlier in most of the correlation analyses, and the longest MTTs were found for that catchment despite its relatively small size. A direct association of sedimentary soil on prolonged groundwater storage-release processes has also been found in a study conducted in south-central Chile (Parra *et al.* 2019b), while Staudinger *et al.* (2017) found no correlation between catchment area and any of the storage metrics in alpine catchments.

We also found a strong correlation between catchment tree volume and dynamic storage, which is in line with findings of Barrientos & Iroumé (2018), who explored the impacts of forest management on dynamic storage in 15 forested catchments and concluded that catchments with lower forest cover (biomass volume and plantation density) have higher storage volumes. This might be due to the increasing evapotranspiration rate by deep-rooted trees with water uptake in sub-surface storage and therefore results in reducing the available amounts of dynamic storage. Furthermore, we assumed that the positive association between the total water storage (estimated by the water balance method) and till soils could be because densely forested catchments store significant amounts of input water, and this large quantity of immobile storage was included in the calculations of the water balance method.

5. CONCLUSION

In this study, we used the HBV model with three model structures, each with a different number of storage reservoirs, to evaluate differences in dynamic storage simulation. First, we found that there is a high variability among different parameterizations. This means that to achieve robust results for further analyses, it is important to consider parameter uncertainty, for instance, by an approach such as the 100 calibration trials used in our study. Second, we found high variability in dynamic storage estimates not only between the catchments but also between the different model structures.

Moreover, this study highlighted the importance of a three-bucket conceptual model that generates runoff using multiple storage reservoirs, especially in large groundwater-dominated catchments with deeper sediment soils. In contrast, for the lake-influenced catchments, a single-reservoir structure performed equally well, indicating that these simple structures can adequately represent the hydrological functioning in such catchments due to the lower storage capacity and shallow groundwater levels.

We also concluded that an ensemble average of simulations can be used as a point of hydrological functioning comparison between different catchments. Considering that the model validation using real data in these large-scale heterogeneous catchments is not possible in practice, this can reduce the time, cost, and risk associated with uncertainty of each single model structure.

Finally, as demonstrated by our results, the dynamic catchment storage could be explained by landscape features such as drainage area, slope, soil characteristics, and tree volume.

DATA AND CODE AVAILABILITY

The GIS and hydroclimate data are publicly available from (<https://www.slu.se/Krycklan>). The HBV-light software and its tutorial are freely available from (<https://www.geo.uzh.ch/en/units/h2k/Services/HBV-Model.html>). The source code of the different response routine model structures can be requested from the second author.

COMPETING INTERESTS

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The GIS and hydroclimate data are publicly available from <https://www.slu.se/Krycklan>. The HBV-light software and its tutorial are freely available from <https://www.geo.uzh.ch/en/units/h2k/Services/HBV-Model.html>. The source code of the different response routine model structures can be requested from the second author.

REFERENCES

- Ajami, N. K., Duan, Q. & Sorooshian, S. 2007 An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research* **43** (1), W01403.
- Allen, A. & Chapman, D. 2001 Impacts of afforestation on groundwater resources and quality. *Hydrogeology Journal* **9** (4), 390–400.
- Amvrosiadi, N., Bishop, K. & Seibert, J. 2017a Soil moisture storage estimation based on steady vertical fluxes under equilibrium. *Journal of Hydrology* **553**, 798–804.
- Amvrosiadi, N., Seibert, J., Grabs, T. & Bishop, K. 2017b Water storage dynamics in a till hillslope: the foundation for modeling flows and turnover times. *Hydrological Processes* **31** (1), 4–14.
- Barrientos, G. & Iroumé, A. 2018 The effects of topography and forest management on water storage in catchments in south-central Chile. *Hydrological Processes* **32** (21), 3225–3240.
- Bonnesoeur, V., Locatelli, B., Guariguata, M. R., Ochoa-Tocachi, B. F., Vanacker, V., Mao, Z., Stokes, A. & Mathez-Stiefel, S. L. 2019 Impacts of forests and forestation on hydrological services in the Andes: a systematic review. *Forest Ecology and Management* **433**, 569–584.
- Botter, G., Porporato, A., Rodriguez-Iturbe, I. & Rinaldo, A. 2009 Nonlinear storage-discharge relations and catchment streamflow regimes. *Water Resources Research* **45** (10), W10427.
- Brauer, C. C., Teuling, A. J., Torfs, P. J. J. F. & Uijlenhoet, R. 2013 Investigating storage-discharge relations in a lowland catchment using hydrograph fitting, recession analysis, and soil moisture data. *Water Resources Research* **49** (7), 4257–4264.
- Broderick, C., Matthews, T., Wilby, R. L., Bastola, S. & Murphy, C. 2016 Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. *Water Resources Research* **52** (10), 8343–8373.
- Chapman, T. 1999 A comparison of algorithms for stream flow recession and baseflow separation. *Hydrological Processes* **13** (5), 701–714.
- Creutzfeldt, B., Ferré, T., Troch, P., Merz, B., Wziontek, H. & Güntner, A. 2012 Total water storage dynamics in response to climate variability and extremes: inference from long-term terrestrial gravity measurement. *Journal of Geophysical Research: Atmospheres* **117** (D8), D08112.
- Dralle, D. N., Hahm, W. J., Rempe, D. M., Karst, N. J., Thompson, S. E. & Dietrich, W. E. 2018 Quantification of the seasonal hillslope water storage that does not drive streamflow. *Hydrological Processes* **32** (13), 1978–1992.
- Farrick, K. K. & Branfireun, B. A. 2014 Soil water storage, rainfall and runoff relationships in a tropical dry forest catchment. *Water Resources Research* **50** (12), 9236–9250.
- Fenicia, F., Savenije, H. H. G., Matgen, P. & Pfister, L. 2006 *Is the Groundwater Reservoir Linear? Learning From Data in Hydrological Modelling*.
- Fenicia, F., Kavetski, D., Savenije, H. H., Clark, M. P., Schoups, G., Pfister, L. & Freer, J. 2014 Catchment properties, function, and conceptual model representation: is there a correspondence? *Hydrological Processes* **28** (4), 2451–2467.
- Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G. & Ye, M. 2012 Towards a comprehensive assessment of model structural adequacy. *Water Resources Research* **48** (8), W08301.
- Hailegeorgis, T. T. & Alfredsen, K. 2015 Comparative evaluation of performances of different conceptualisations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments. *Hydrology Research* **46** (4), 607–628.
- Hogue, T. S., Bastidas, L. A., Gupta, H. V. & Sorooshian, S. 2006 Evaluating model performance and parameter behavior for varying levels of land surface model complexity. *Water Resources Research* **42** (8), W08430.
- Hrachowitz, M. & Clark, M. P. 2017 HESS opinions: the complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth System Sciences* **21** (8), 3953–3973.
- Ilstedt, U., Tobella, A. B., Bazié, H. R., Bayala, J., Verbeeten, E., Nyberg, G., Sanou, J., Benegas, L., Murdiyarso, D., Laudon, H. & Sheil, D. 2016 Intermediate tree cover can maximize groundwater recharge in the seasonally dry tropics. *Scientific Reports* **6** (1), 1–12.
- Jutebring Sterte, E., Lidman, F., Lindborg, E., Sjöberg, Y. & Laudon, H. 2021 How catchment characteristics influence hydrological pathways and travel times in a boreal landscape. *Hydrology and Earth System Sciences* **25** (4), 2133–2158.
- Kalbus, E., Reinstorf, F. & Schirmer, M. 2006 *Measuring Methods for Groundwater, Surface Water and Their Interactions: A Review*.
- Karlsen, R. H., Grabs, T., Bishop, K., Buffam, I., Laudon, H. & Seibert, J. 2016 Landscape controls on spatiotemporal discharge variability in a boreal catchment. *Water Resources Research* **52** (8), 6541–6556.
- Karlsen, R. H., Bishop, K., Grabs, T., Ottosson-Löfvenius, M., Laudon, H. & Seibert, J. 2019 The role of landscape properties, storage and evapotranspiration on variability in streamflow recessions in a boreal catchment. *Journal of Hydrology* **570**, 315–328.
- Kennedy, J., Ferré, T. P., Güntner, A., Abe, M. & Creutzfeldt, B. 2014 Direct measurement of subsurface mass change using the variable baseline gravity gradient method. *Geophysical Research Letters* **41** (8), 2827–2834.
- Kirchner, J. W. 2009 Catchments as simple dynamical systems: catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research* **45** (2), W02429.
- Kozii, N., Laudon, H., Ottosson-Löfvenius, M. & Hasselquist, N. J. 2017 Increasing water losses from snow captured in the canopy of boreal forests: a case study using a 30 year data set. *Hydrological Processes* **31** (20), 3558–3567.
- Krasnostein, A. L. & Oldham, C. E. 2004 Predicting wetland water storage. *Water Resources Research* **40** (10), W10203.
- Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., Greene, S., Macleod, C. J. & Reaney, S. M. 2019 Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. *Hydrology and Earth System Sciences* **23** (10), 4011–4032.

- Laudon, H., Sjöblom, V., Buffam, I., Seibert, J. & Mörth, M. 2007 The role of catchment scale and landscape characteristics for runoff generation of boreal streams. *Journal of Hydrology* **344** (3-4), 198–209.
- Laudon, H., Taberman, I., Ågren, A., Futter, M., Ottosson-Löfvenius, M. & Bishop, K. 2013 The Krycklan catchment study – a flagship infrastructure for hydrology, biogeochemistry, and climate research in the boreal landscape. *Water Resources Research* **49** (10), 7154–7158.
- Laudon, H. & Ottosson Löfvenius, M. 2016 Adding snow to the picture—providing complementary winter precipitation data to the Krycklan catchment study database. *Hydrological Processes* **30** (13), 2413–2416.
- Laudon, H., Sponseller, R. A. & Bishop, K. 2021 From legacy effects of acid deposition in boreal streams to future environmental threats. *Environmental Research Letters* **16** (1), 015007.
- Ledesma, J. L. & Futter, M. N. 2017 Gridded climate data products are an alternative to instrumental measurements as inputs to rainfall–runoff models. *Hydrological Processes* **31** (18), 3283–3293.
- Lidberg, W., Nilsson, M., Lundmark, T. & Ågren, A. M. 2017 Evaluating preprocessing methods of digital elevation models for hydrological modelling. *Hydrological Processes* **31** (26), 4660–4668.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M. & Bergström, S. 1997 Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology* **201** (1-4), 272–288.
- Maneta, M., Soulsby, C., Kuppel, S. & Tetzlaff, D. 2018 Conceptualizing catchment storage dynamics and nonlinearities. *Hydrological Processes* **32**, 3299–3303.
- McGuire, K. J., McDonnell, J. J., Weiler, M., Kendall, C., McGlynn, B. L., Welker, J. M. & Seibert, J. 2005 The role of topography on catchment-scale water residence time. *Water Resources Research* **41**, 5.
- McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., Aulenbach, B. T. & Hooper, R. 2011 Storage as a metric of catchment comparison. *Hydrological Processes* **25** (21), 3364–3371.
- Mendoza-Sanchez, I., Phanikumar, M. S., Niu, J., Masoner, J. R., Cozzarelli, I. M. & McGuire, J. T. 2013 Quantifying wetland–aquifer interactions in a humid subtropical climate region: an integrated approach. *Journal of Hydrology* **498**, 237–253.
- Meriö, L. J., Ala-aho, P., Linjama, J., Hjort, J., Kløve, B. & Marttila, H. 2019 Snow to precipitation ratio controls catchment storage and summer flows in boreal headwater catchments. *Water Resources Research* **55** (5), 4096–4109.
- Mishra, A., Hata, T., Abdelhadi, A. W., Tada, A. & Tanakamaru, H. 2003 Recession flow analysis of the Blue Nile River. *Hydrological Processes* **17** (14), 2825–2835.
- Moore, R. D. 1997 Storage-outflow modelling of streamflow recessions, with application to a shallow-soil forested catchment. *Journal of Hydrology* **198** (1–4), 260–270.
- O’Callaghan, J. F. & Mark, D. M. 1984 The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* **28** (3), 323–344.
- Parra, V., Arumí, J. L. & Muñoz, E. 2019a Identifying a suitable model for low-flow simulation in watersheds of south-central Chile: a study based on a sensitivity analysis. *Water* **11** (7), 1506.
- Parra, V., Arumí, J. L., Muñoz, E. & Paredes, J. 2019b Characterization of the groundwater storage systems of south-central Chile: an approach based on recession flow analysis. *Water* **11** (11), 2324.
- Peralta-Tapia, A., Sponseller, R. A., Ågren, A., Tetzlaff, D., Soulsby, C. & Laudon, H. 2015 Scale-dependent groundwater contributions influence patterns of winter baseflow stream chemistry in boreal catchments. *Journal of Geophysical Research: Biogeosciences* **120** (5), 847–858.
- Peralta-Tapia, A., Soulsby, C., Tetzlaff, D., Sponseller, R., Bishop, K. & Laudon, H. 2016 Hydroclimatic influences on non-stationary transit time distributions in a boreal headwater catchment. *Journal of Hydrology* **543**, 7–16.
- Rezaei-Sadr, H. 2019 Flood hydrograph simulation in ungaged basins: modeling linear–nonlinear storage-discharge relationships in a semiarid mountain region. *Physical Geography* **40** (1), 52–70.
- Riegger, J. & Tourian, M. J. 2014 Characterization of runoff-storage relationships by satellite gravimetry and remote sensing. *Water Resources Research* **50** (4), 3444–3466.
- Sayama, T., McDonnell, J. J., Dhakal, A. & Sullivan, K. 2011 How much water can a watershed store? *Hydrological Processes* **25** (25), 3899–3908.
- Seibert, J. 1997 Estimation of parameter uncertainty in the HBV model: paper presented at the Nordic Hydrological Conference (Akureyri, Iceland-August 1996). *Hydrology Research* **28** (4-5), 247–262.
- Seibert, J. 2000 Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences* **4** (2), 215–224.
- Seibert, J. & Vis, M. J. 2012 Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences* **16** (9), 3315–3325.
- Seibert, J., Bishop, K., Rodhe, A. & McDonnell, J. J. 2003 Groundwater dynamics along a hillslope: a test of the steady state hypothesis. *Water Resources Research* **39** (1), 1014.
- Seibert, J., Staudinger, M. & van Meerveld, H. I. 2019 Validation and over-parameterization - experiences from hydrological modeling. In: *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives* (C. Beisbart & N. J. Saam, eds). Simulation Foundations, Methods and Applications. Springer International Publishing, Cham, Switzerland, pp. 811–834.
- Sitterson, J., Knightes, C., Parmar, R., Wolfe, K., Avant, B. & Muche, M. 2018 *An Overview of Rainfall-Runoff Model Types*.

- Soulsby, C., Tetzlaff, D. & Hrachowitz, M. 2009 Tracers and transit times: windows for viewing catchment scale storage? *Hydrological Processes* **23** (24), 3503–3507.
- Spearman, C. 1904 The proof and measurement of association between two things. *The American Journal of Psychology* **15** (1), 72–101.
- Sprenger, M., Tetzlaff, D., Buttle, J., Carey, S. K., McNamara, J. P., Laudon, H., Shatilla, N. J. & Soulsby, C. 2018 Storage, mixing, and fluxes of water in the critical zone across northern environments inferred by stable isotopes of soil water. *Hydrological Processes* **32** (12), 1720–1737.
- Staudinger, M., Stoelzle, M., Seeger, S., Seibert, J., Weiler, M. & Stahl, K. 2017 Catchment water storage variation with elevation. *Hydrological Processes* **31** (11), 2000–2015.
- Sterte, E. J., Johansson, E., Sjöberg, Y., Karlsen, R. H. & Laudon, H. 2018 Groundwater-surface water interactions across scales in a boreal landscape investigated using a numerical modelling approach. *Journal of Hydrology* **560**, 184–201.
- Stoelzle, M., Weiler, M., Stahl, K., Morhard, A. & Schuetz, T. 2015 Is there a superior conceptual groundwater model structures for baseflow simulation? *Hydrological Processes* **29** (6), 1301–1313.
- Tetzlaff, D., Buttle, J., Carey, S. K., Van Huijgevoort, M. H., Laudon, H., McNamara, J. P., Mitchell, C. P., Spence, C., Gabor, R. S. & Soulsby, C. 2015 A preliminary assessment of water partitioning and ecohydrological coupling in northern headwaters using stable isotopes and conceptual runoff models. *Hydrological Processes* **29** (25), 5153–5173.
- Teutschbein, C., Grabs, T., Karlsen, R. H., Laudon, H. & Bishop, K. 2015 Hydrological response to changing climate conditions: spatial streamflow variability in the boreal region. *Water Resources Research* **51** (12), 9425–9446.
- Teutschbein, C., Grabs, T., Laudon, H., Karlsen, R. H. & Bishop, K. 2018 Simulating streamflow in ungauged basins under a changing climate: the importance of landscape characteristics. *Journal of Hydrology* **561**, 160–178.
- Tiwari, T., Laudon, H., Beven, K. & Ågren, A. M. 2014 Downstream changes in DOC: Inferring contributions in the face of model uncertainties. *Water Resources Research* **50** (1), 514–525.
- Uhlenbrook, S., Seibert, J. A. N., Leibundgut, C. & Rodhe, A. 1999 Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structures. *Hydrological Sciences Journal* **44** (5), 779–797.
- Van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C., Fénicia, F., Kavetski, D. & Lobligeois, F. 2013 The influence of conceptual model structures on model performance: a comparative study for 237 French catchments. *Hydrology and Earth System Sciences* **17** (10), 4227–4239.
- Wang, D. & Alimohammadi, N. 2012 Responses of annual runoff, evaporation, and storage change to climate variability at the watershed scale. *Water Resources Research* **48** (5), W05546.
- Wittenberg, H. 1999 Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes* **13** (5), 715–726.

First received 7 November 2021; accepted in revised form 15 March 2022. Available online 25 March 2022