Scale-dependency in modeling nivo-glacial hydrological systems: the case of the Arolla basin, Switzerland

Anne-Laure Argentin[1], Pascal Horton[2], Bettina Schaefli[2], Jamal Shokory[3], Felix Pitscheider[1], Leona Repnik[3], Mattia Gianini[3], Simone Bizzi[4], Stuart N. Lane[3], and Francesco Comiti[1,5]

[1]Faculty of Agricultural, Environmental and Food Sciences, Free University of Bozen-Bolzano, Italy
[2]Institute of Geography, and Oeschger Center on Climate Change Research, University of Bern, Switzerland
[3]Institute of Earth Surface Dynamics, University of Lausanne, Switzerland
[4]Department of Geosciences, University of Padova, Italy
[5]Department of Land, Environment, Agriculture and Forestry, University of Padova, Italy

Correspondence: Anne-Laure Argentin (aargentin@unibz.it)

Abstract. Hydrological modeling in alpine catchments poses unique challenges due to the complex interplay of meteorological, topographical, ~~glaciological and streamflow generation factors~~geological and glaciological drivers on streamflow generation. A significant issue arises from the limited availability of streamflow data due to the scarcity of high-elevation gauging stations. Consequently, there is a pressing need to assess whether streamflow models that are calibrated with moderate-elevation ~~datasets~~ streamflow data can be effectively transferred to higher-elevation catchments, notwithstanding differences in the relative importance of different streamflow-generation processes. Here, we investigate the spatial transferability of ~~hydrological~~ calibrated temperature-index melt model parameters within a semi-lumped modeling framework. We focus on evaluating the melt model transferability from the main catchment to nested and neighboring subcatchments in the Arolla valley, southwestern Swiss Alps. We use the Hydrobricks modeling framework to simulate streamflow, implementing three variants of a temperature-index snow- and ice melt model (the classical degree-day, aspect-related, and Hock's temperature index). Through an analysis of streamflow simulations, benchmark metrics consisting of ~~bootstrapped discharge~~ resampled and bootstrapped discharge time series, and model performance metrics, we demonstrate that robust parameter transferability and accurate streamflow simulation are possible across diverse spatial scales. This finding is conditional upon the melt model applied, with melt models using more spatial information leading to convergence of the model parameters until we observe ~~overparametrization~~overparameterization. We conclude that simple semi-lumped models ~~, such as Hydrobricks,~~ can be used to extend hydrological simulations to ungauged catchments in alpine regions and improve high-elevation water resource management and planning efforts, ~~which is crucial in the face~~ especially in the context of climate change.

1 Introduction

Understanding the driving factors of nivo-glacial streamflow regimes is essential for managing high ~~alpine~~ Alpine catchments and their water resources under global change. With ongoing warming, the long-, intermediate- and short-term stor-

age capacities of ~~alpine glaciers are~~ Alpine nivo-glacial systems (e.g. storage capacities of subglacial drainage network, snow cover, glacier ice) will be impacted (Jansson et al., 2003; Huss et al., 2008), and high alpine catchments may transition from nivo-glacial streamflow regimes to dominantly nival regimes (Horton et al., 2006). Currently, ~~alpine~~ Alpine glaciated catchments and downstream areas receive a strong surplus of meltwater from snow during spring and early summer (Penna et al., 2017; Engel et al., 2019; Zuecco et al., 2019), gradually switching to glacier meltwater towards the end of summer. The timing and amount of snow- and ~~glaciermelt water~~ glacier-melt is strongly impacted by global warming and related glacier retreat, ~~making~~ leading to changes in streamflow regimes ~~critical~~ (e.g., Singh and Kumar, 1997; Bradley et al., 2006). These changes in streamflow regimes and runoff generation characteristics have important consequences in terms of sediment transport, hydropower production (Gabbud et al., 2016), flood prediction and ecology (Tague et al., 2020).

Despite this, high ~~alpine~~ Alpine catchments often lack discharge monitoring stations due to their sparse population and difficulty of access. In highly glacierized catchments (i.e. glacial cover $>$ 50%), there are very few gauging stations that provide reliable and long-term streamflow records. This makes attributing historical changes in streamflow regimes to glacial sources challenging and inevitably requires recourse to modeling, not just to predict the future but also to understand the past.

Hydrological models are commonly classified into distributed, semi-distributed, semi-lumped and lumped models (Horton et al., 2022). Distributed models compute the storage and mobilization of water at ~~the pixel~~ a grid-cell scale, with parameters that vary in space (fully distributed) or are partially kept constant (semi-distributed). Semi-lumped models define areas of interest based on relevant physical parameters (e.g., elevation, aspect, stream network topology), and lumped models consider the catchment as a single unit. While (semi-)distributed and semi-lumped models allow some spatial variations to be taken into account and provide a more detailed representation of the processes, lumped models have to represent the functioning of the entire system. The advantage and popularity of (semi-)lumped models should not be ~~reduced~~ attributed solely to their computational efficiency, which ~~enables fast and multiple model runs. They~~ facilitates multiple model simulations. They also represent an optimal level of model complexity with respect to available input and output data from a downward model development perspective (Sivapalan et al., 2003); they furthermore operate at a scale at which averaging of small-scale processes enables a reliable representation of dominant hydrological processes (Clark et al., 2016).

However, one of the main drawbacks of (semi-)lumped models is that streamflow can only be modeled reliably at the selected control points (outlets) for which the model parameters have been calibrated against observed streamflow. Simulated streamflow at other locations within or near the catchment might not reliably represent the actual system dynamics. In other words, the calibrated parameters might not be ~~transferrable~~ transferable to other locations of the stream network (subcatchments) within the system. This transferability issue is particularly important in high ~~alpine~~ Alpine catchments for snow and ice melt ~~and storage~~ parameters, as meltwater plays a major role ~~for~~ in streamflow generation processes. The proportion of streamflow that is melt-derived ~~,~~ (either from snow or ice~~,~~) and the dominant drivers of melt (i.e. components of the energy balance) will change as the basin outlet selected for simulation is shifted

upstream or downstream, which means that melt parameters need to be scale-independent. This difficulty is exacerbated in catchments with strong ~~process~~ topographic gradients and spatial heterogeneity, where the complex spatial averaging complicates the extraction of ~~specific process responses~~ melt-driven streamflow simulations at smaller scales, which is typically the case in glaciated catchments. ~~Adding additional calibration using~~ One possible solution to increase the spatial transferability of calibrated models is the use of additional observed data to better constrain model parameters (Efstratiadis and Koutsoyiannis, 2010) and thereby to increase their reliability, in particular in view of input data uncertainty (van Tiel et al., 2020) and in view of simulating change conditions (climate change, land use change). In the field of high Alpine streamflow simulation, the focus is on the value of glacier mass balance or snow data ~~may help constrain snow and ice melt and storage parameters, yet it remains necessary to study their transferability~~ to ensure that the calibrated streamflow is right for the right reasons (Kirchner, 2006) rather than due to compensating effects (see the review by van Tiel et al. (2020)). Examples include the work of Parajka and Blöschl (2008); Şorman et al. (2009); Koboltschnig et al. (2008); Im , which all focus on different aspects of parameter reliability as a function of model calibration strategy. For example, the work of Griessinger et al. (2016) underlines that the incorporation of snow data is especially important for high-elevation catchments and snow-rich years. However, this wealth of literature does not address the question of how to transfer calibrated parameters to other catchments ~~to predict discharge at ungauged locations~~or to subcatchments, which is the focus of hydrological parameter regionalization methods.

Parameter regionalization techniques (Guo et al., 2021) in hydrological modeling have been developed to facilitate the transfer of model parameters from gauged to ungauged locations (e.g., Mosley, 1981; Abdulla and Lettenmaier, 1997; Bardossy and Singh, 2008). Regionalization methods can be divided into two categories (Samaniego et al., 2010): post-regionalization and simultaneous regionalization. ~~Postregionalization~~ Post-regionalization methods calibrate a model in several basins independently and then statistically link the calibrated model parameters to basin predictors (e.g., mean catchment elevation, stream network density, geology, areal proportion of porous aquifers) using a transfer function (e.g., Abdulla and Lettenmaier, 1997; Seibert, 1999; Parajka et al., 2005; Wagener and Wheater, 2006). Simultaneous regionalizations aim to calibrate model parameters for several basins while taking into account transfer functions that link model parameters to catchment characteristics (e.g., Hundecha and Bárdossy, 2004; Götzinger and Bárdossy, 2007; Fernandez et al., 2000; Troy et al., 2008). The second category of methods was developed to add additional spatial constraints to parameter calibration and to avoid artifacts of the optimization algorithm. In all these methods, the need to define a function that links catchment characteristics and model parameters is subject to additional uncertainties~~. The~~ , and snow parameters are often kept constant in such approaches (Götzinger and Bárdossy, 2007; Kling and Gupta, 2009).

Overall, the number of parameter regionalization studies in Alpine areas remains small (Horton et al., 2022), and the spatial transfer of melt model parameters is still a crucial topic for the prediction of streamflow in catchments without observed streamflow (Guo et al., 2021).

Spatial parameter transfer is particularly challenging in data-sparse high-elevation catchments where glacier melt, interannual snow storage, and highly uncertain precipitation ~~input and evapotranspiration output~~ and evapotranspiration can lead to considerable parameterization difficulties (Schaefli and Huss, 2011). A particular challenge in such catchments

is the estimation of snow and glacier melt contributions, which, for practical, data reasons, is often limited to the use of temperature-index melt models (TI) that link melt rates to air temperature (Eq. 1, Rango and Martinec, 1995):

$$M_{\mathrm{TI}}(t) = \begin{cases} a_j(T_a(t) - T_T) & : T_a(t) > T_T \quad \text{with } j \in \text{snow, ice} \\ 0 & : T_a(t) \leq T_T \end{cases} \tag{1}$$

where $M_{\mathrm{TI}}(t)$ is the melt rate at time step t (mm d^{-1}), a_j the degree day factor for ice or snow (mm d^{-1} ℃$^{-1}$), T_a is the air temperature and T_T is the threshold melt temperature.

Although it is commonly admitted that TI models present a good option for extrapolation to larger scales because of the ~~consistency~~ persistency of temperature over large areas (Frenierre and Mark, 2014), the spatial transferability of the related parameters calibrated at the outlet of a catchment to the outlet of a neighboring catchment exhibiting different characteristics (elevation, aspect, glacial cover) can be questioned (Gabbi et al., 2014; Samaniego et al., 2010), and has been rarely investigated. This challenge was exemplified for nested catchments by Comola et al. (2015), who studied the influence ~~of spatial decorrelation of aspect , i.e. the distance beyond which the spatial variance of aspect does not increase anymore, on discharge simulation. They~~ aspect and found significant variability in the calibrated degree-day factors for small catchments (< 7 km^2) ~~due to the spatial correlation of aspects~~ when using a simple temperature-index model.

One option to regionalize melt model parameters is to compute them directly based on in-situ snow observations, which was already presented by Martinec (1960) (for an application to streamflow modelling see the work of Hingray et al. (2010)). Similarly, they can be computed from remotely sensed snow extents: The work of He et al. (2014) shows that such spatially variable melt model parameters can improve streamflow simulations compared to spatially constant parameters. Such approaches are rare because the value of temperature-index melt models is inherently linked to streamflow simulation and most studies therefore calibrate melt parameters against streamflow.

In this study, we investigate the transferability of the melt and ~~runoff~~ streamflow calibrated parameters between sub-catchments and neighboring catchments considering different melt models and with respect to very high-quality discharge measurements ~~. To do this , we~~ obtained from a hydropower company (see Section 2.2, Figure H1). In view of this exceptional discharge data set, we chose to not use remotely-sensed snow data, because the preprocessing of remotely-sensed snow extents and their use for model calibration includes its own share of uncertainty (Parajka and Blöschl, 2006, 20 , which would obscure our analysis.

For our analysis, we calibrate our model for seven catchments, then take the parameters of the largest ~~catchments~~ catchment and transfer them to its three nested watersheds and three other neighboring catchments. We then analyze the loss of accuracy ~~linked~~ related to the transfer of parameters. To ensure that our conclusions hold ~~with~~ for different commonly used objective functions and assess the sensitivity to these objective ~~function~~functions, we use two different metrics: the Nash-Sutcliffe (NSE; Nash and Sutcliffe, 1970) and Kling-Gupta efficiency (KGE; Gupta et al., 2009). These two very common metrics do not translate into one another: the NSE measures the error variance related to the variance of the discharge, while the KGE measures correlation, variability bias and mean bias. NSE and KGE are differently

Table 1. Catchments used in the simulations, and their properties.

Catchment	Abbre-viation	Area (km2)	Elevation (m)			Mean slope	Dominant aspect	Glacier covered catchm.	Debris covered glacier	Type
			Mean	Min	Max					
Bertol Inférieur	BI	26.0	3063	2183	3722	28.7	NW	38.5%	9.9%	Main
Haut Glacier d'Arolla	HGDA	13.2	3014	2582	3677	29.5	NW	32.0%	16.3%	Nested
Tsijiore Nouve	TN	4.8	3180	2289	3789	28.2	N	57.7%	20.4%	Neigh.
Pièce	PI	2.9	3046	2636	3784	27.8	NE	57.6%	17.3%	Neigh.
Bertol Supérieur	BS	2.6	3127	2913	3583	32.4	SW	9.2%	14.3%	Nested
Vuibé	VU	2.2	3036	2730	3722	24.7	NE	54.4%	1.4%	Nested
Douves Blanches	DB	1.5	3218	3097	3364	35.4	W	10.1%	23.9%	Neigh.

sensitive to discharge errors: for example, for high variation regimes that show small errors and large bias, the NSE would give a very promising value while the KGE would not (Knoben et al., 2019). We carry out this transferability assessment with three temperature-index melt models of increasing complexity, and try answering the question: Could incorporating additional spatial information into more complex TI models increase their spatial transferability?

2 Study area: the upper Arolla river basin and its subcatchments

2.1 Presentation of the study area

We use data (Table 3) from the Arolla river basin located in the south-western Swiss Alps (Fig. 1). A local hydropower company provided 15-minute resolution streamflow recordings of very high quality given strict regulatory requirements for monitoring water use (Lane and Nienow, 2019). The Bertol Inférieur (BI) gauging station is fed by water draining from four subcatchments (Table 1): Bertol Supérieur (BS), Haut Glacier d'Arolla (HGDA), Mont Collon (MC) and Vuibé (VU). The BI catchment, with an area of 26.0 km^2, provides a good opportunity to test the transferability of hydrological parameters to nested catchments as there are three subcatchments that are also gauged upstream: BS (2.6 km^2 area), HGDA (13.2 km^2), and VU (2.2 km^2). Remaining drainage to BI comes from the MC catchment or from points located between the BS, HGDA and VU gauges and the BI gauge (Fig. 1). Immediately to the north of the VU catchment is the Pièce catchment (PI), draining an area of 2.9 km^2; and the Tsijiore Nouve (TN) catchment with a drainage area of 4.8 km^2. On the other side of the valley, immediately to the north of the BS catchment is the Douves Blanches (DB) catchment with a drainage area of 1.5 km^2. These catchments allow us to test the transferability of hydrological parameters to neighboring catchments. The elevation of these basins ranges from 2112 m a.s.l. (the elevation of the BI gauging station) to 3838 m a.s.l., the Grand Bouquetins peak, located in the Haut Glacier d'Arolla. At these elevations, it is extremely unusual to have such high-quality streamflow data for small, highly glacier-covered catchments.

The upper Arolla river basin presents a ~~range~~ variety of aspects (Fig. 1b), and its subcatchments have different general orientations (Fig. 1c). The glacial cover within the Arolla basin ~~diminished~~ decreased over the years (Fig. 1d), from 66.5% in 1850 to 38.5% in 2016 for the BI catchment (GLAMOS, 2020)~~.~~ , while remaining relatively constant since 2009. The geology of the study area consists mainly of metamorphic and igneous rocks, extensively covered with till and colluvial deposits (SwissTopo, 2024, Supplementary Figure B1). The geomorphological characteristics of the subcatchments ~~is~~ are generally similar. DB, BS, TN and the northern part of HGDA all feature some rock glaciers (Lambiel et al., 2016), although their relative area is more significant for DB and BS.

Numerous studies have been carried out in the upper Arolla basin over the years, on topics ranging from glacier dynamics, subglacial hydrology, sediment transport to hydrology (Sharp et al., 1993; Brock et al., 2000; Mair et al., 2002, 2003; Swift et al., 2002, 2005; Arnold, 2005; Pellicciotti et al., 2005; Dadic et al., 2010; Gabbud et al., 2015, 2016; Lane and Nienow, 2019), which makes this study area optimal for a technical study on hydrological parameter transferability.

2.2 Hydro-meteorological datasets

We use MeteoSwiss datasets for daily mean precipitation (MeteoSwiss, 2019a) and daily mean temperature available at 1 km resolution for Switzerland (MeteoSwiss, 2019b).

Discharge data were provided by Grande Dixence SA (2024) at seven water intakes. In Switzerland, regulatory standards require hydroelectric power production companies to report water abstraction details to the authorities. In the upper Arolla river basin, discharge data ~~was~~ were thus provided at a 15-minute resolution since ~~1971.~~ 1969. Each basin features a calibrated water level recorder, initially utilizing a chart recorder and later upgraded to a pressure transducer with digital data logging. Water levels are measured across a broad-crested weir, ensuring highly reliable discharge records ($\pm$0.01 m^3/s for regulatory compliance). ~~Discharge is measured in the intake.~~ Under very high flow conditions, the intake overflows and only part of the water is recorded. However, ~~as~~ since any loss of water is a financial loss, the intake has been designed to capture practically all the discharge. Such overflows are therefore possible, but infrequent. Furthermore, ~~a~~ the current ecological minimum flow was only introduced in one of the intakes, BI, as of 2018, after the period used in this study (Tobias et al., 2023). The intakes defining the extent of each subcatchment are sometimes multiple, as with DB and VU, which both present two intakes, and PI, which presents four (See Fig. 1). Thus, the discharge is the sum of the corresponding intakes.

With the exception of ~~the~~ BI, discharge data were already preprocessed by Lane and Nienow (2019) to eliminate drawdown events linked to sediment removal during intake flushing. Since these drawdown periods typically last between 30 and 60 minutes, they can be visually recognized using the method outlined in the work of Lane et al. (2017). After the removal of data portions corresponding to such drawdowns, any missing data points were linearly interpolated. However, for the VU intake, data were unavailable from August 31 to December 31, 2011, due to intake maintenance work. In our study, we excluded this last period for VU and applied the same preprocessing method to the BI discharge time series, removing drawdown events and discarding associated time periods (in blue, Fig. 2). Furthermore, the water from the HGDA, BS and VU intakes is diverted and does not pass through the BI intake. We thus added the records of

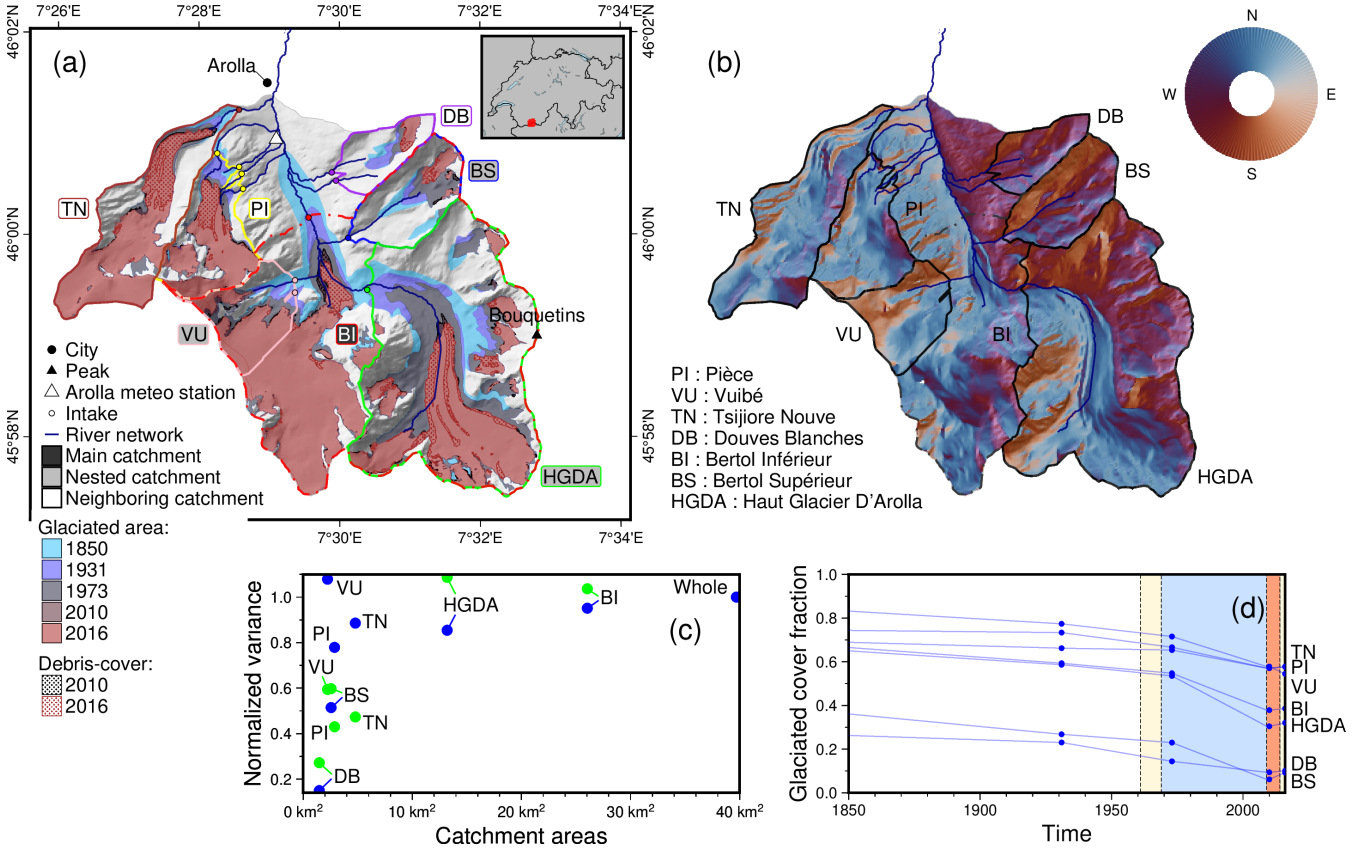

Figure 1. (a) Overview of the modeled catchments and subcatchments in the upper part of the Heremence valley, in the Arolla catchment. Changes in glacial cover through time are indicated in shades of yellow and green. The inset map shows the location of the study catchments in the Swiss Alps. (b) Aspect of the study area. (c) Aspect variogram derived from the aspect variance of each glacier (in blue) and each catchment (in green), normalized by the variance in the total glaciated area/total catchment (Whole), as done by Comola et al. (2015). (d) Glacial cover fraction through time, with the study period highlighted in orange, the available discharge period and meteorological data in blue and yellow, respectively. Topography is obtained from the SwissTopo DHM25 dataset (Swisstopo, accessed 2023) and glacier extents from the GLAMOS inventory (GLAMOS, 2020, Table 3).

its upstream nested intakes to the BI discharge record (the actual measurements in BI is called BIrest, see Fig. 2). We propagated to BI the intake maintenance work of VU and the drawdown removals of BIrest by discarding the affected time periods. The time taken by the water to reach the BI intake from the upstream intakes is approximately 15-30 minutes depending on the days, which is negligible at the daily scale (Supplementary Figure G1). Subsequently, the 15-minutes time step discharge datasets were summed up to daily time step datasets after the preprocessing.

Due to the confidentiality of the original discharge data, these datasets are shown here normalized by the same highest observed discharge values for all catchments (see figures 2, 4, 6, 8, 9, 11 and 14). The normalized dataset is

called either "normalized discharge" when the discharge was expressed in m^3 s^{-1}, and "specific normalized discharge" when the discharge was expressed in mm.

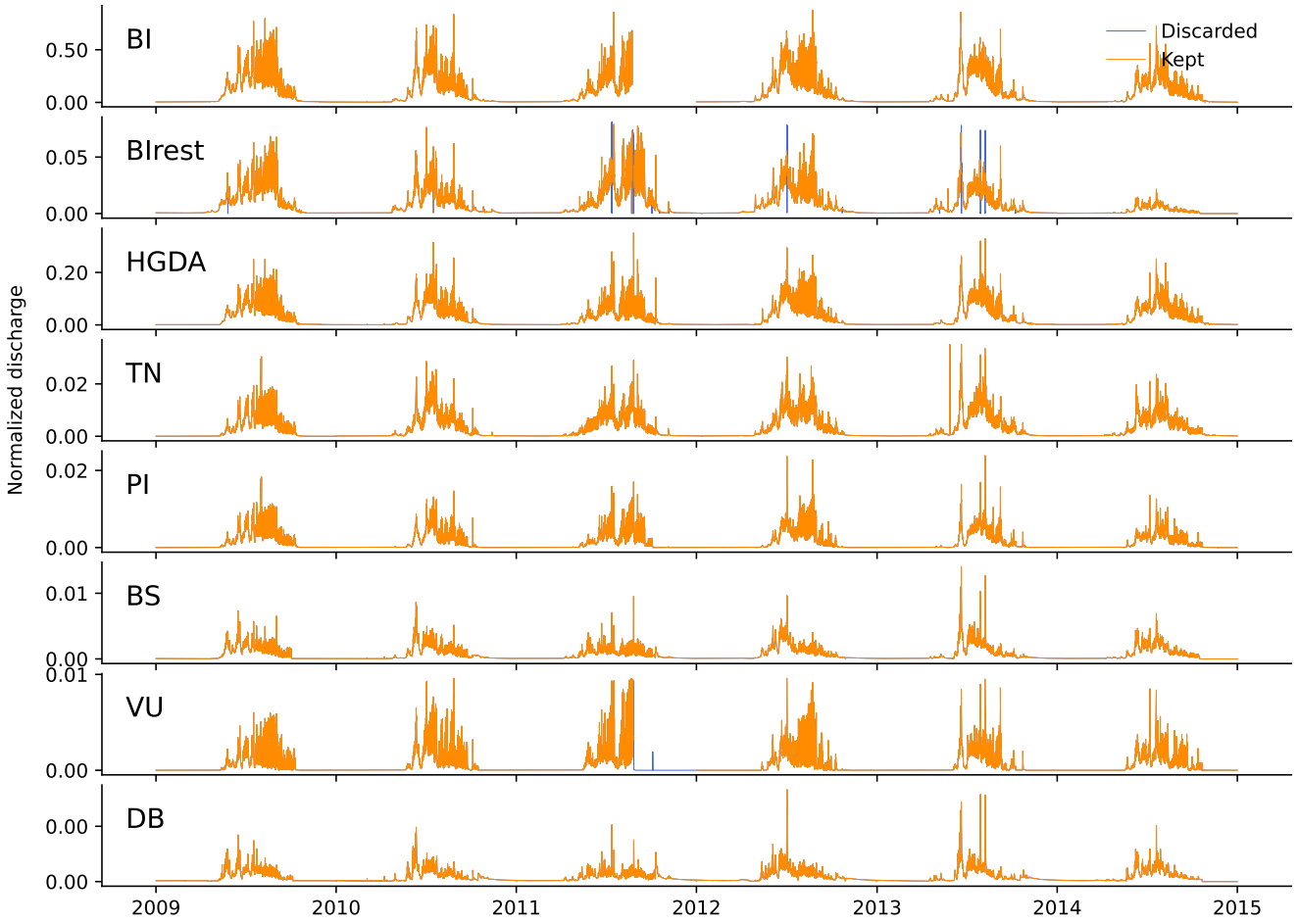

Figure 2. Observed discharge series for all subcatchments: Comparison of the discharge series kept for calibration in Hydrobricks (orange) with the discarded periods (blue). Discharge (unit: m^3 s^{-1}) is normalized ~~to~~ by the same highest ~~value~~observed discharge values for all catchments. The BI discharge corresponds to BIrest + HGDA + BS + VU.

Glacier extents for the years 2010 and 2016 were obtained from the GLAMOS inventory (Fig. 1; GLAMOS, 2020; Linsbauer et al., 2021; Fischer et al., 2014). This inventory specifies the debris cover extent for the year 2016. To obtain older debris cover trends, we used the algorithm developed by Shokory and Lane (2023), now available in an ArcGIS Pro toolbox, and computed the 2010 extents based on Landsat Level 1 imagery (for details, see ~~the~~ Supplementary Material, Section D). We assumed the glacial cover of 2009 to be identical to 2010.

We derived the topography from the SwissTopo DHM25 dataset (Swisstopo, accessed 2023) available at 25 m resolution. From this topography, we automatically extracted the catchment areas, except for VU. VU requires manual correction of its southern extent due to the presence of thick ice cover, which complicated the identification of the drainage divide (Fig. 6 of Bezinge et al., 1989; Hurni, 2021).

3 Methods

3.1 Hydrological modeling with Hydrobricks

Hydrobricks (v0.7.2; Horton and Argentin, 2024) is a hydrological modeling framework that implements the semi-lumped GSM-SOCONT model (Glacier and SnowMelt – SOil CONTribution; Schaefli et al., 2005) to simulate nivo-glacial hydrological regimes. The model consists of two main components: (a) the reservoir-based SOCONT model, which incorporates a linear reservoir method to account for slow storage contribution (emulation of subsurface ground water) and a non-linear reservoir approach to address quick runoff, and (b) the GSM model, which is specifically designed for glacier-covered catchments. The Hydrobricks framework is based on a C++ core integrated into a Python interface, which allows for enhanced computing performances.

The model discretizes the catchment into hydrological response units (HRUs) by elevation, aspect and potential clear-sky direct solar radiation. The HRUs can have fractional land cover types, here 'glacier' for glacier-covered areas, and 'ground' for non-glacier-covered areas. The distinction between debris-covered glacier areas ('glacier_debris') and debris-free glacier areas ('glacier_ice') can also be made (Shokory and Lane, 2023). The processes occurring within the same land cover type but in different HRUs are assigned identical parameters.

Following GSM-SOCONT's original structure, the model behavior differs between the glacier-covered area and the ice-free part. For the ice-free fractional part of a given HRU, surface and subsurface runoff components, along with baseflow from melt and rainfall (Supplementary Figure A1), are computed per HRU and summed across all HRUs to build the non-glacier streamflow component at the outlet. For glacier-covered areas, the liquid water from melt and rainfall produced by each HRU is fed into two lumped parallel linear reservoirs shared by all HRUs. The purpose of these reservoirs, which only apply to glacier surfaces, is to represent the glacier's retention effect on water flow. For a detailed workflow, refer to Supplementary Figure A1 and Schaefli et al. (2005).

The transition from rainfall to snowfall is defined in a fuzzy approach (Schaefli and Huss, 2011) between 0 °C for the lower end ($T_{\text{s-r,min}}$) to 2 °C for the upper end ($T_{\text{s-r,max}}$). Snow and ice start melting at a threshold melt temperature T_T defined at 0 °C, and ice can only melt when ~~not covered anymore~~ it is no longer covered by snow.

We use the SPOTPY library (Houska et al., 2015) provided with Hydrobricks for parameter optimization with the Shuffled Complex Evolution algorithm of the University of Arizona (SCE-UA). The SCE-UA algorithm is designed to prevent remaining stuck in local optima. We use it in combination with the Nash-Sutcliffe (NSE; Nash and Sutcliffe, 1970) and Kling-Gupta efficiency (KGE; Gupta et al., 2009) performance criteria to find the best combination of parameters (Table 2), after 10,000 simulations.

3.2 Hydrobricks developments

In the original version of GSM-SOCONT (Schaefli et al., 2005), precipitation type (snow or rain) is determined by a temperature threshold and melt is calculated through a classic temperature-index melt model (TI). Two new melt models were implemented in Hydrobricks: the aspect temperature-index model (ATI) and the temperature-index model of Hock (HTI).

The aspect temperature-index model (ATI) is based on the discretization of the study area by aspect (North, South, East/West) and the use of a distinct ~~degree day~~ degree-day factor depending on aspect. A more complex model, the temperature-index melt model of Hock (HTI; Hock, 1999), links potential clear-sky direct solar radiation to melt rates (Eq. 2):

$$M_{\mathrm{HTI}}(t) = \begin{cases} (m + r_j I_{\mathrm{pot}})(T_a(t) - T_T) & : T_a(t) > T_T \quad \text{with } j \in \mathrm{snow, ice} \\ 0 & : T_a(t) \leq T_T \end{cases} \tag{2}$$

where M_{HTI} is the melt rate (mm d^{-1}), m is the melt factor common to both ice and snow (mm d^{-1} ℃$^{-1}$), r_j is the radiation factor for ice or snow (mm d^{-1} ℃$^{-1}$ m^2 W^{-1}), I_{pot} is the potential clear-sky direct solar radiation (W m^{-2}), T$_a$ is the air temperature and T$_T$ is the threshold melt temperature. Thus, while the ATI model represents a first attempt at handling spatial differences in melt rates, the HTI model has the benefit of directly taking into account irradiation, which should make it better suited to reproduce melt rates in catchments influenced by aspect and cast shadows (e.g., Gabbi et al., 2014).

The HTI model requires computation of the potential clear-sky direct solar radiation I_{pot}, here implemented using the definition of Hock (1999, Eq. 3):

$$I_{\mathrm{pot}} = I_0 \left(\frac{R_m}{R} \right)^2 \Psi_a^{\left(\frac{P}{P_0 \cos(Z)} \right)} \cos(\theta) \tag{3}$$

where I_0 is the solar constant (1368 W m^{-2}), $(R_m/R)^2$ is the Earth's orbit's eccentricity correction factor, composed of R and R_m the instantaneous and mean Sun-Earth distances, Ψ_a is the mean atmospheric clear-sky transmissivity, P and P_0 the local and the mean sea-level atmospheric pressures, Z the local zenith angle and θ the angle of incidence between the normal to the grid slope and the solar beam. The potential direct solar radiation ~~I~~, computed on a 15-minute interval, is set to 0 when a point is not directly irradiated by sunlight (night time and cast shading brought by surrounding relief) ~~.~~, then summed at the daily scale to obtain the daily potential direct solar radiation I. This 15-min interval allows accounting for changes in sun position, sun rising and setting times during the year.

For the TI model, based on temperature only, the HRUs are evenly spaced elevation bands (Schaefli et al., 2005). For the ATI and HTI models, the HRUs reflect the elevation variations as well as the aspect or the mean annual irradiation variations. To avoid any HRU scaling influence on parameter transferability (Liang et al., 2004; Troy et al., 2008), we use a spacing of 40 m for elevation, 3 categories for aspect (North, South and East-West to group by degree of sun

exposure) and a spacing of 65 W m^{-2} for potential direct solar radiation for all catchments (Supplementary Figure A2). Furthermore, in contrast to earlier studies employing GSM-SOCONT/Hydrobricks, which relied on monitoring station data to derive meteorological lapse rates across the different elevation bands, we needed to derive our meteorological input from gridded datasets. Our study therefore adopts a distinct methodology, extracting meteorological input for each HRU directly from gridded datasets. This involves quantifying how much the different cells in the gridded datasets contribute to each HRU. We do this by downscaling once the grid of the meteorological data (1 km) to the DEM resolution (25 m): we compute the weights representing the contribution of each data cell to each HRU based on their spatial coverage, then use these weights to calculate the mean values for each HRU, for each daily time step. This allows direct use of future climate model outputs often provided as gridded datasets. The evapotranspiration is then computed at the HRU level, from mean values of temperature, following the Hamon equation ~~(Fig 3 Hamon, 1963)~~(Hamon, 1963).

In the case where we differentiate between debris-covered and debris-free glacier coverage, we also have to adapt the melt models by introducing new melt parameters governing the ice melt. The TI model switches from a single parameter (a_{ice}) to two parameters ($a_{\text{deb-free}}$ and $a_{\text{deb-cov}}$). The ATI model goes from three parameters ($a_{\text{ice,j}}$ with $j \in$ N, S, EW) to six parameters ($a_{\text{deb-free,j}}$ and $a_{\text{deb-cov,j}}$ with $j \in$ N, S, EW). The HTI model goes from two parameters (m and r_{ice}) to three parameters (m, $r_{\text{deb-free}}$ and $r_{\text{deb-cov}}$). The new melt parameters respect the same calibration ranges, but an inequality constraint is added to force lower melting rates of debris-covered ice: $a_{\text{deb-cov}} < a_{\text{deb-free}}$, $a_{\text{deb-cov,j}} < a_{\text{deb-free,j}}$ with $j \in N, S, EW$, or $r_{\text{deb-cov}} < r_{\text{deb-free}}$, depending on the chosen melt model.

Under current climate conditions, virtually all snow in our study area melts every summer, making it unnecessary to model the firn separately.

3.3 Modeling approach

Our study of the transferability of nivo-glacial parameters from the TI, ATI and HTI melt models to nested subcatchments and neighboring catchments (cf. section 2) can be divided into 4 steps:

1. Calibration runs: We calibrate the model on all subcatchments and neighboring catchments independently and compare them.

2. Transfer runs in nested catchments: We transfer the parameters calibrated on the main catchment, the Bertol Inférieur (BI), to its nested subcatchments (BS, HGDA and VI) to simulate their streamflow and compare the results to observed streamflow.

3. Transfer runs in neighboring catchments: As previous step but we transfer the parameters calibrated on the main catchment, the Bertol Inférieur (BI), to the neighboring catchments (TN, PI and DB).

4. Increased model complexity run: We repeat the three above points but calibrate and run the model with differentiation of debris-covered glacier and debris-free glacier areas.

For the first step of our study, we calibrate the model for all catchments individually using daily observed streamflow over the years 2009-2014. We chose this simulation period because glacier cover remains relatively stable (Fig 1d). For performance metric assessment, the first simulation year is discarded since it is assumed to initialize the system. These runs are called "calibration runs" as the whole period is used for calibration, and no validation is carried out.

For the second and third steps of our study, we transfer the calibrated parameters from the calibration run of the main catchment to nested and neighboring subcatchments. As for the calibration runs, the first year is discarded. These runs do not include any calibration procedure and are called "transfer runs".

To analyze the effect of differentiating between bare ice melt and debris-covered glacier melt, we complete all of the above steps twice: once assuming bare ice for the entire glacier area and once accounting for debris-cover.

3.4 Benchmark metrics

A key challenge in model calibration is to assess how good a calibrated model actually is since ~~the commonly used~~ commonly-used metrics do not have an absolute meaning (Schaefli and Gupta, 2007). Here, we assess how good the transferred runs are by assessing if they outperform 1) exhaustively resampled and 2) bootstrapped time series. Bootstrapping is a statistical resampling method that generates independent samples by repeated random draw with filling (Efron, 1979), on the basis of which statistics can then be calculated. ~~Since normal bootstrapping~~ In hydrology, bootstrap methods have been used mostly to assess uncertainty (Vogel and Shallcross, 1996; Ebtehaj et al., 2010; Clark et al., 2021), sometimes in the performance metrics used (Ritter and Muñoz-Carpena, 2013; Clark and Slater, 2006) and to simulate multi-site discharge time-series (Srinivas and Srinivasan, 2005; Clark et al., 2021). Since the original bootstrapping procedure (Efron, 1979) assumes independent and identically distributed data, which would destroy the temporal correlation of our discharge data, ~~we~~ hydrologists (e.g., Srinivas and Srinivasan, 2005; Ebtehaj et al., 2010; Clark et al., 2021) usually opt for block ~~boostrapping, which~~ bootstrapping (Carlstein, 1986; Künsch, 1989), which preserves the internal data structure and is typically used for time-series. ~~Thus, we perform block bootstrapping on the discharge series during the evaluation~~ In the case of hydrological data strongly impacted by snow and glacier melt, the block size is annual, starting in January during low flow, to coincide with the annual periodic structure of the discharge (Ebtehaj et al., 2010). As the evaluation of our hydrological simulation takes place over 4 years only, we calculate a derived metric where we exhaustively resample the years of the evaluation period (2010-2014~~) with yearly block sizes. This process is repeated 100 times, ensuring that the streamflow data from a particular year is not used for that same year's prediction.~~ ; 3125 combinations), based on the same principles as block bootstrapping. This resampling has the advantage of not being affected by long-term processes such as global warming, and being easily reproducible because of its non-stochasticity. For comparison, we perform block bootstrapping on the entire discharge series (1969-2014), stochastically drawing 3125 combinations. We then compute the NSE and the KGE on each exhaustively resampled and bootstrapped series and average them to obtain benchmark metrics. In the following article, we call "benchmark" NSE and KGE the metrics computed over 5 years, and "long-term benchmark" the metrics computed over 46 years. The benchmark NSE and KGE correspond to the prediction potential of the discharge dataset itself.

Table 2. Parameters used in the simulations and their a priori range of values.

Parameter (set)	Unit	Description	Set value	Melt model
$T_{\text{s-r, min}}$	°C	lower temperature threshold of the snow-rain fuzzy transition	0	all
$T_{\text{s-r, max}}$	°C	upper temperature threshold of the snow-rain fuzzy transition	2	all
T_T	°C	threshold melt temperature	0	all

Parameter (calibrated)	Unit	Description	Condition	Range	Melt model
MELT MODEL-DEPENDENT PARAMETERS					
a_{ice} or $a_{\text{deb-free}}$, $a_{\text{deb-cov}}$	mm d^{-1} °C^{-1}	ice degree-day factor, independent (ice) or dependent on ice cover (debris-covered or debris-free)	$a_{\text{deb-cov}} < a_{\text{deb-free}}$	5 - 20	TI
a_{snow}	mm d^{-1} °C^{-1}	snow degree-day factor	$a_{\text{snow}} < a_{\text{ice}}$ or $a_{\text{deb-cov}}$	2 - 12	TI
$a_{\text{ice,j}}$ or $a_{\text{deb-free,j}}$, $a_{\text{deb-cov,j}}$ with $j \in$ N, S, EW	mm d^{-1} °C^{-1}	ice degree-day factor, independent (ice) or dependent on ice cover (debris-covered or debris-free) and dependent on aspect (North, South, East/West)	$a_{\text{deb-cov,j}} < a_{\text{deb-free,j}}$ with $j \in$ N, S, EW	5 - 20, 0 - 20 (North)	ATI
$a_{\text{snow,N}}$, $a_{\text{snow,S}}$, $a_{\text{snow,EW}}$	mm d^{-1} °C^{-1}	snow degree-day factor, dependent on aspect (North, South, East/West)	$a_{\text{snow,j}} < a_{\text{ice,j}}$ or $a_{\text{deb-cov,j}}$ with $j \in$ N, S, EW	2 - 12, 0 - 12 (North)	ATI
m	mm d^{-1} °C^{-1}	melt factor		0 - 12	HTI
r_{ice} or $r_{\text{deb-free}}$, $r_{\text{deb-cov}}$	mm d^{-1} °C^{-1} m^2 W^{-1}	ice radiation factor, independent (ice) or dependent on ice cover (debris-covered or debris-free)	$r_{\text{deb-cov}} < r_{\text{deb-free}}$	0 - 1	HTI
r_{snow}	mm d^{-1} °C^{-1} m^2 W^{-1}	snow radiation factor	$r_{\text{snow}} < r_{\text{ice}}$	0 - 1	HTI
RUNOFF TRANSFORMATION PARAMETERS					
k_{ice}	d^{-1}	ice outflow coefficient			
k_{snow}	d^{-1}	snowpack outflow coefficient	$k_{\text{snow}} < k_{\text{ice}}$		
k_{quick}	d^{-1}	surface runoff outflow coefficient			
A	mm	slow storage capacity		0 - 3000	
k_{slow_1}	d^{-1}	slow storage outflow coefficient	$k_{\text{slow}_1} < k_{\text{quick}}$		
k_{slow_2}	d^{-1}	baseflow storage outflow coefficient	$k_{\text{slow}_2} < k_{\text{slow}_1}$		
ρ_{perc}	mm d^{-1}	slow storage percolation rate to the baseflow storage		0 - 10	

4 Results

4.1 Calibration runs

For the calibration runs without accounting for debris-cover, the NSE and KGE values are better than ~~those obtained from bootstrapping~~ the benchmark values for all catchments (Fig. 3; section 3.4), implying a consistent enhancement in streamflow modeling with Hydrobricks compared to a simple temporal transfer of observed data. This improvement is

Table 3. Data used in the simulations, with corresponding source.

Dataset	Description	Acquisition period	Provider
Mean temperature	daily interval, 1 km resolution gridded dataset	since 1961	MeteoSwiss
Mean precipitation	daily interval, 1 km resolution gridded dataset	since 1961	MeteoSwiss
Discharge	15-minute sampling time interval, measured at water intakes	since 1969	Grande Dixence SA
Topographic data	25 m resolution DEM (DHM25 dataset)	-	SwissTopo
Clean glacier extents	shapefiles of glacier extents	1850, 1931, 1973, 2010, 2016	GLAMOS inventory
Debris-covered glacier extents	shapefiles of debris-covered glacier extents	2016	GLAMOS inventory
Landsat imagery	Level 1 Landsat 7 imagery at 30 m resolution for debris-free ice mapping	06/09/2009	Landsat

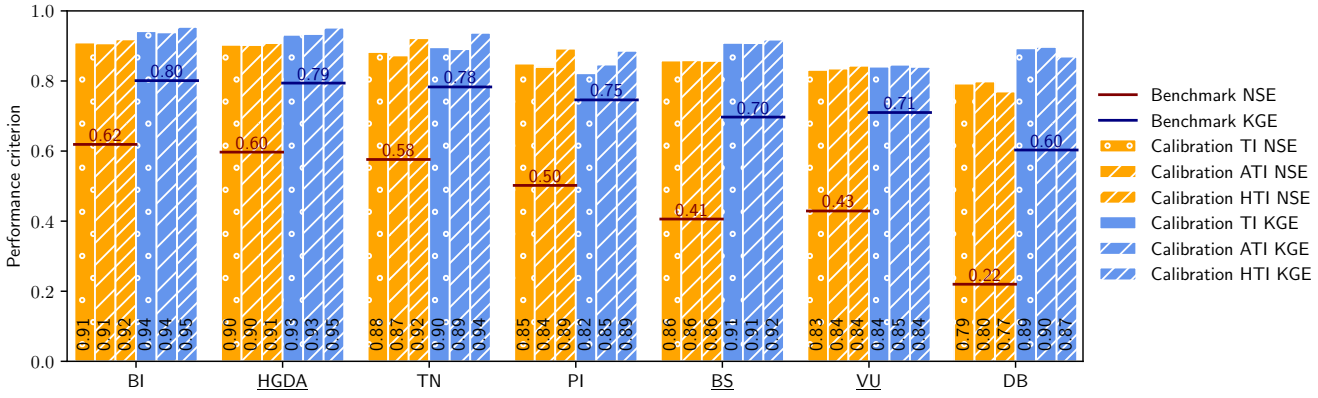

Figure 3. Comparison of the performance of the three melt models on the seven catchments for the period 2010-2014, quantified either by the Nash–Sutcliffe efficiency (NSE, orange bars) or by the Kling-Gupta efficiency (KGE, blue bars). For comparison, the benchmark NSE and KGE are computed and plotted as red and dark blue thresholds. The model is calibrated by running 10,000 times over the years 2009 - 2014, where 2009 is discarded for model initialization. Catchments are ordered by area, from BI (largest) to DB (smallest). BI's nested catchments are underlined.

more pronounced in smaller catchments, as the values of the benchmark metrics decrease with decreasing catchment area, while the Hydrobrick simulation scores remain consistently high. This suggests that in smaller catchments, discharge time-series present more variable and marked yearly signals than in bigger catchments, which can be effectively replicated using Hydrobricks, but not with ~~bootstrapping~~resampling. Given that, in general, KGE values tend to be higher than NSE values (Knoben et al., 2019), this trend is particularly apparent with the benchmark NSE and slightly less with

the benchmark KGE. Thus, even though the achieved NSE values are often lower than those of KGE, the improvement they represent compared to the benchmark metric values is much bigger.

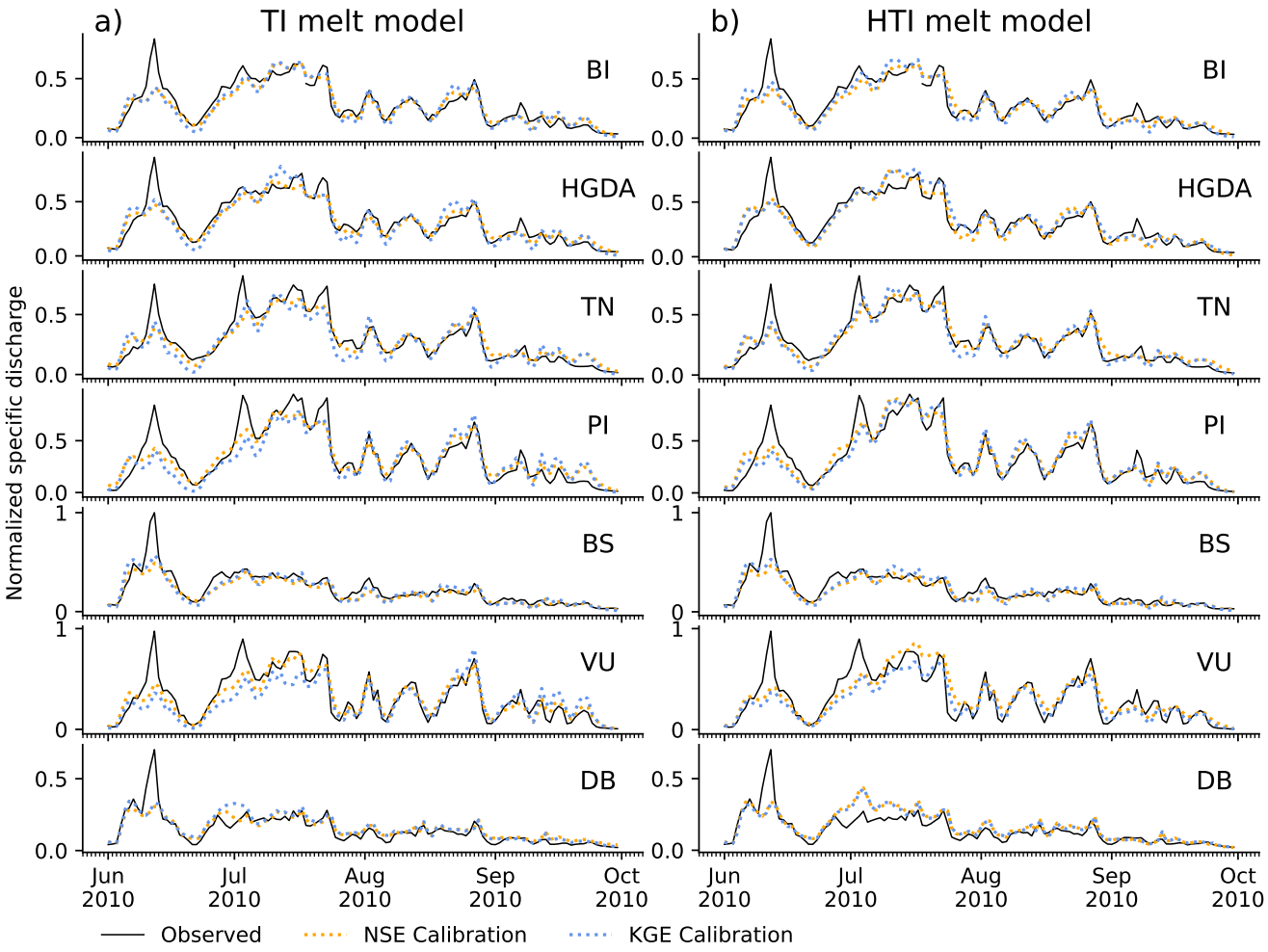

Figure 4. Observed and simulated hydrographs for all catchments for 2010 with the a) TI and b) HTI melt models. Observed discharge (black solid line) is compared to the calibration run using NSE (dotted orange) and KGE (dotted blue). Specific discharge (unit: mm) is normalized to the highest value.

To illustrate how well the calibrated discharge simulations fit the observed discharge, Fig. 4 shows the corresponding hydrographs. The calibration runs based on NSE and KGE (Fig. 4) result in globally similar hydrographs for the TI and HTI melt models, and for the ATI model (Fig. C1). While overall discharge dynamics are well simulated, some discharge peaks are smoothed and thus not adequately reproduced. This is particularly evident in the case of the "spring event" occurring early- to mid-June. This first prominent peak during the melting season results from the melt of the supraglacial

and hillslope snowpacks, that occurred due to an unusually strong foehn that blew on the 9th and 10th of June (Zbinden et al., 2010). This foehn event is ~~partially recorded~~ only partially captured in the temperature records of the period~~(Fig. 14),~~ , and none of the melt models account for wind, which is why ~~it~~ this peak could not be reproduced entirely.

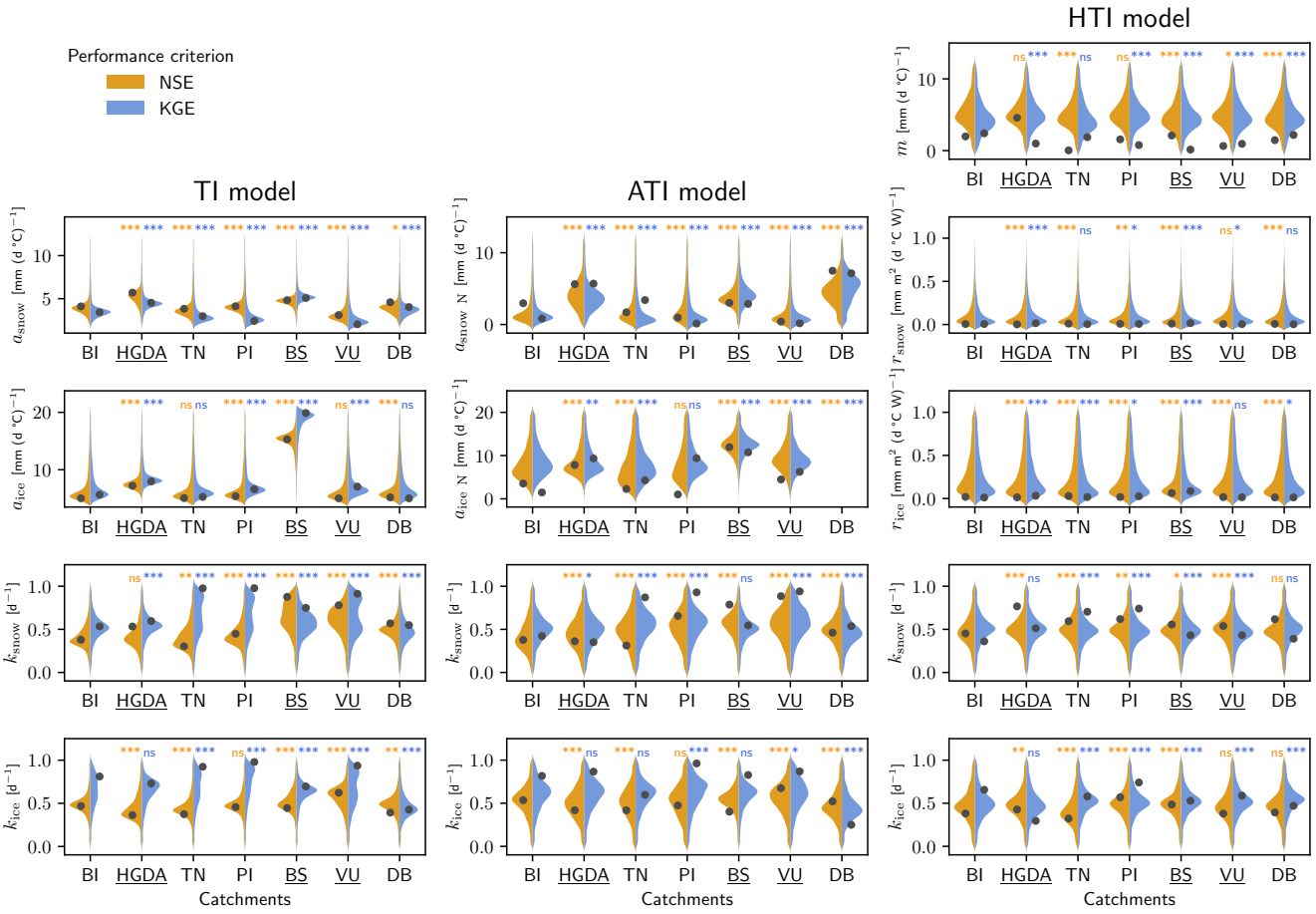

Figure 5. Calibrated ice and snow melt parameters for all simulated NSE and KGE runs for all catchments, with the three melt models and the two performance criteria. The parameter sets achieving the best NSE and KGE scores are plotted on top with a dot. Catchments are ordered by area, from BI (largest) to DB (smallest). The significance of the parameter distribution difference between BI and its neighboring and nested catchments is denoted as follows: *** for p < 0.001, ** for $p < 0.01$, * for $p < 0.05$, and ns for non-significant (Kruskal-Wallis test). BI's nested catchments are underlined.

~~Depending on the model used, the calibrated parameters are more or less similar~~ The variability of calibrated parameters between performance criteria and ~~across catchments~~ between catchments depends on the melt model used (Kruskal-Wallis test; Fig. 5). Some models produce more pronounced variations, while others generate more limited variations. With the TI model, the calibrated parameter sets show very different values, depending on the catchment and the perfor-

mance criterion used. With the ATI model, the degree-day factors and outflow coefficients obtained with KGE and NSE tend to converge to similar values, but these values still show a certain spread between the catchments. With the HTI model, the all parameter values are more consistent both between the two performance criteria and across the seven catchments. For example, the HGDA catchment shows high similarity with the BI catchment for both the snow radiation factor r_{snow} and the ice radiation factor r_{ice} in NSE calibration (non significant distribution change - ns; Fig. 5).

Thus, refining the representation of the melt process leads to increased spatial coherence of the melt parameters. The parameters showing no significant distribution changes between catchments could be assumed to be transferable between these catchments without calibration. The HTI model can thus be assumed to be suitable to model the melt processes occurring in neighboring or nested catchments.

4.2 Transfer runs in nested catchments: spatial parameter transferability

To assess the spatial parameter transferability to nested subcatchments, we apply, for all melt models, the calibrated parameter set obtained for the BI catchment to model the discharge of its nested subcatchments BS, HGDA and VI. The results for the TI and HTI ~~model~~ models are shown in Fig. 6, and the ATI model in Fig. C2; the transfer runs closely match the observed discharge ~~closely~~ for the TI and HTI models. For both models, the simulated discharges of the HGDA and BS catchments show slightly underestimated low flow periods and peaks. While this can be observed for HGDA throughout the summer, it is especially true for BS in the late summer. VU, on the contrary, shows slightly overestimated discharge in the low flows and the peaks, starting July, and overall, its discharge is best reproduced by the TI model. Close inspection reveals some differences between the transfer runs based on NSE versus KGE calibration, but no systematic differences.

As expected, the transfer runs reproduce the observed discharge less closely than the calibration runs for each of the catchments. The performance metric values of the transfer runs (Fig. 7) are, nevertheless, high compared to the benchmark values. Globally, the performance drop is ~~bigger~~ larger for KGE than for NSE but given the different sensitivities of the two metrics, they cannot be directly compared (see Section 4.1). Although all subcatchments and models experience a drop in KGE values, VU is the only catchment whose results drop just below the KGE benchmark value with ~~the ATI model~~all models. With the exception of the BS catchment, the best results are obtained with the TI and the HTI models.

We tested the ~~conservativity~~ consistency of the model across scales by checking whether the simulated discharges of the subcatchments (VU, HGDA and BS) were coherent across subcatchments and with the discharge of the main catchment (BI) (Fig. 8). This test is partial, as the discharge generated by the Mont Collon (MC) area is not monitored, and thus the added discharges of VU, HGDA and BS do not account for all of BI's discharge. We thus expect the sum of the three subcatchments' discharges ~~to always be~~ (dotted black line, Fig. 8) to be always lower than the discharge of the main catchment BI (dotted purple), and to have a consistent overestimation/underestimation of the flow across the different subcatchments. This is true for the high flow event trigered by the 9th-10th June 2010 foehn event (Zbinden et al., 2010), which is consistently underestimated in all catchments. Similarly, we find that for the low flow periods end of June and early September, which are slightly overestimated in BI, the stacked discharges (dotted black) are consistent and stay below the BI discharge (dotted purple).

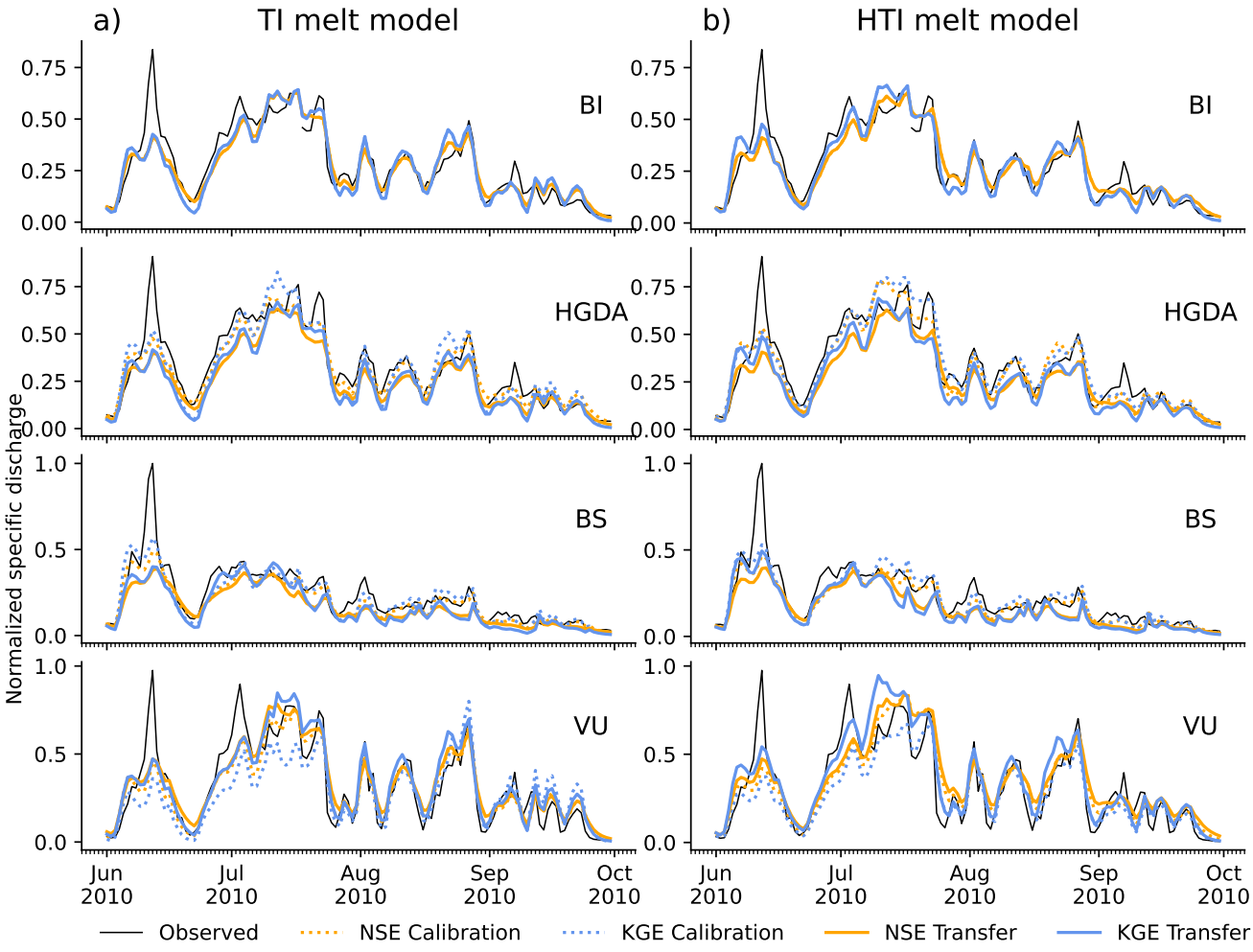

Figure 6. Observed and simulated hydrographs of the BI catchment and its nested subcatchments for 2010 with a) TI and b) HTI melt models. Observed discharge (solid black line) is compared to the calibration runs and to the transfer runs with the calibrated parameters of BI: Shown are the results for NSE (orange) and KGE (blue); dotted lines show the calibration runs, solid lines show the transfer runs. For BI, the calibration and transfer runs are identical. Specific discharge (unit: mm) is normalized to the highest value.

4.3 Transfer runs in neighboring catchments: spatial parameter transferability

380 The results of transferring the calibrated parameters of BI to neighboring catchments (Fig. 9~~t~~, C3) demonstrate for 2010 high accuracy in discharge simulation, both with the TI and HTI models, with minimal performance loss compared to calibration runs. The simulated discharge changes resulting from this forcing are more pronounced for the TI model than for the HTI model. Again, the discharge dynamics are relatively ~~well reproduced~~ well-reproduced but with a significant underestimation of the initial June discharge peak for all catchments and the July ones for the TN and PI catchments.

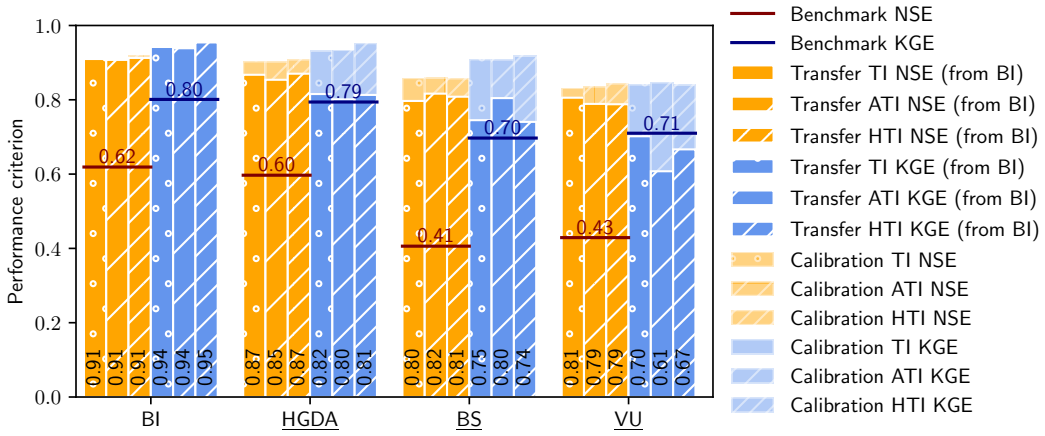

Figure 7. As Figure 3 but comparing calibration and transfer runs for the subcatchments of BI for the period 2010-2014. Shown are NSE values (orange) and KGE values (blue), along with the benchmark NSE value (red line) and KGE value (dark blue line) values. The performance values for the corresponding calibration run are shown in more transparent color. BI's nested catchments are underlined.

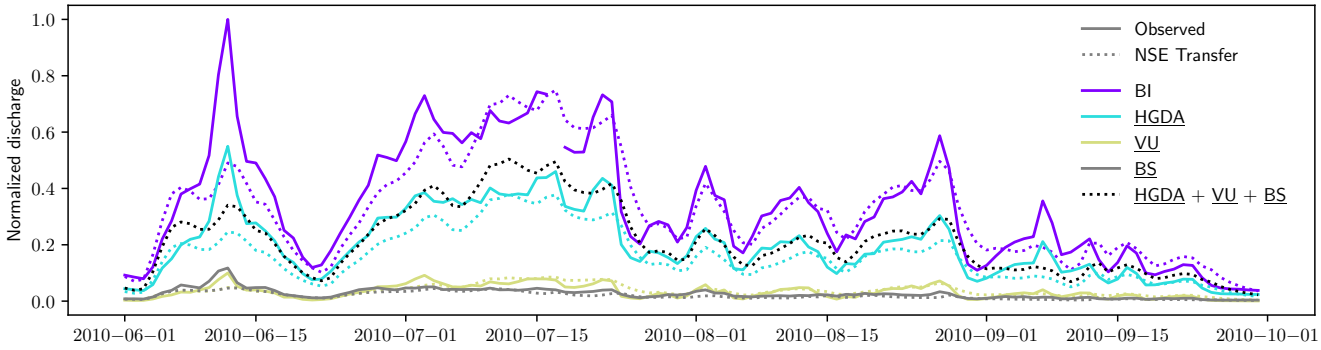

Figure 8. Observed and simulated hydrographs with the HTI model for the Bertol Inférieur (BI) and its subcatchments (VU, HGDA and BS) for the summer of 2010. For the subcatchments, the simulations (dotted lines) are the NSE-transfer runs, for the main catchment, BI, the dotted line corresponds to the calibration run. The solid lines are the observed hydrographs. The dotted black line shows the sum of the transfer runs of the three subcatchments. Discharge (unit: $m^3\ s^{-1}$) is normalized to the highest value. BI's nested catchments are underlined.

385 Interestingly, the July peaks are absent in the observed discharge of the DB catchment (as they are absent in the observed discharge of the BS catchment, see Figure 6). As seen in the nested forcing results (Figure 6), the NSE calibration run fits the discharge peak sometimes better than the KGE calibration run, such as in catchment DB in early July with the TI model.

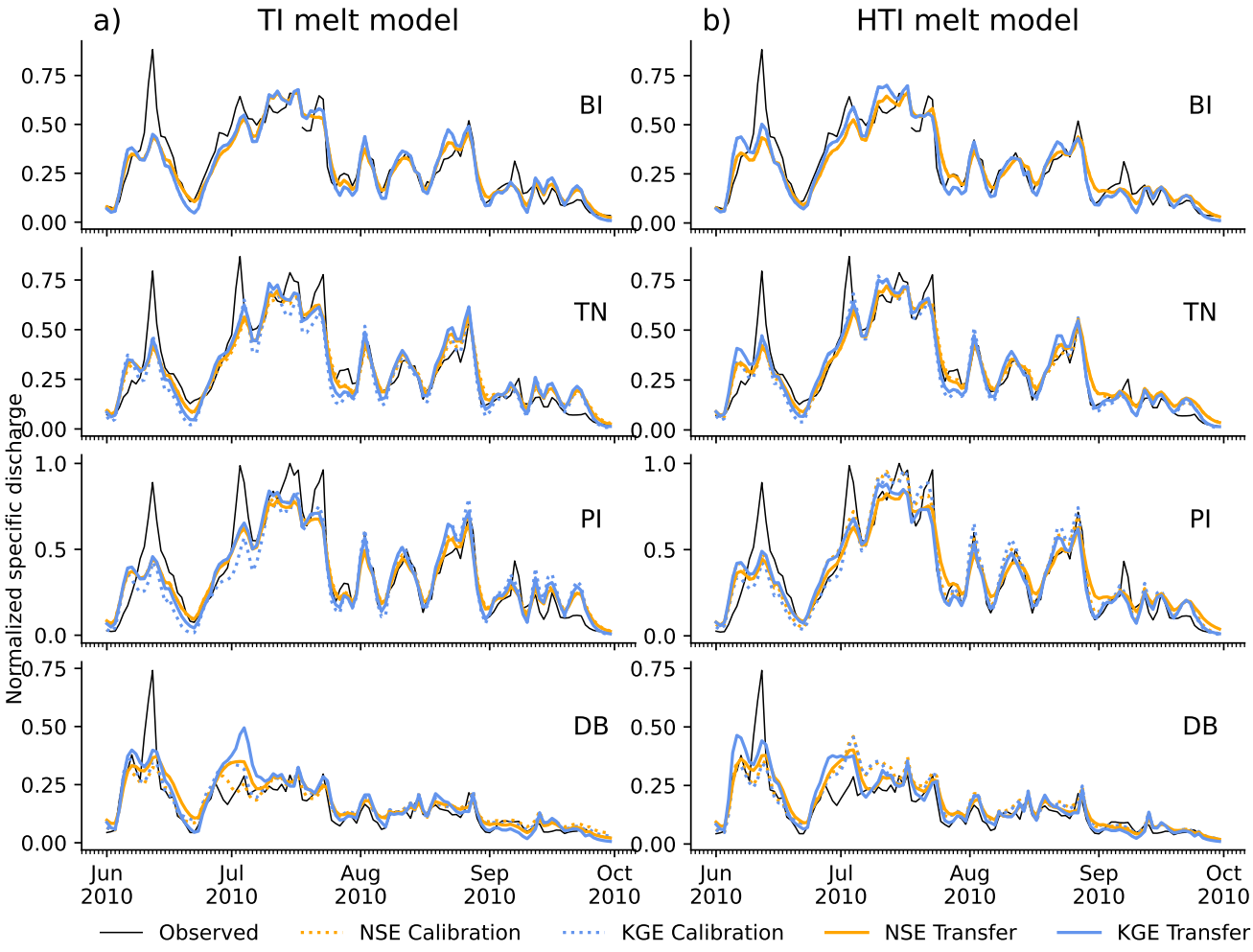

Figure 9. Observed and simulated hydrographs of the BI catchment and its neighboring catchments for 2010 with a) TI and b) HTI melt models. Observed discharge (solid black line) is compared to the calibration run and to the transfer runs with the calibrated parameters of BI: Shown are the results for NSE (orange) and KGE (blue); dotted lines show the calibration runs, solid lines show the transfer runs. For BI, the calibration and transfer runs are identical. Specific discharge (unit: mm) is normalized to the highest value.

We observe similar drops in NSE and KGE values when applying the BI parameters to the neighboring catchments as for the transfer runs in the BI subcatchments (compare Figs. 7 and 10). Nevertheless, with the exception of the DB catchment using the ATI melt model, the performance decreases are less pronounced compared to the nested catchments. In all neighboring catchments, the simulations exhibit NSE and KGE values that ~~surpass those obtained through bootstrapping~~exceed those of the benchmark. Thus, the catchments whose discharges are the least well simulated with BI's calibrated parameters are the BI's nested subcatchments: VU, BS and HGDA.

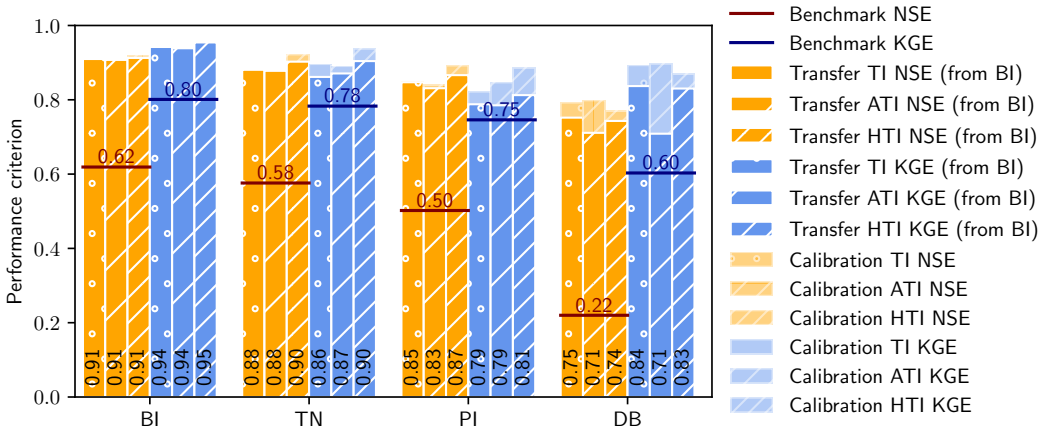

Figure 10. As Figure 7 but comparing calibration and transfer runs for the neighboring catchments of BI for the period 2010-2014. Shown are NSE values (orange) and KGE values (blue), along with the benchmark NSE value (red line) and KGE value (dark blue line) values. The performance values for the corresponding calibration runs are shown in more transparent color.

Analyzing monthly discharge hydrographs (Fig. 11) can yield additional insights into KGE performance, since this metric is by construction more sensitive to model biases than NSE, and such biases can become more apparent in monthly values compared to daily values. The monthly hydrographs (Fig. 11) clearly show the monthly discharge patterns that contribute to decreases in KGE, especially notable in 2012: In the VU catchment, discharge is overestimated, whereas in the HGDA and BS catchments, underestimations are observed.

4.4 Increased model complexity run: accounting for debris-cover

In an attempt to investigate ~~if~~ whether the melt model could be missing an important driving factor, we ~~thus~~ tried to attribute the performance decrease between calibration and transfer simulations to catchment characteristics. We chose to focus on the nested catchments, which show the ~~biggest~~ largest drop in performance (Fig. 12). The catchment area and mean catchment elevation do not show any obvious relations to performance decreases. However, the percentage of glacier debris cover and the mean catchment and glacier slopes show more consistent relations. When the slope is steeper than in BI, discharge is underestimated, whereas when the slope is flatter, discharge is overestimated. In a similar way, when the debris coverage of the glacier is smaller than in BI, discharge is underestimated. ~~Accordingly, in a~~ Consequently, in the next step, we tested the transferability of model versions that differentiate between debris-covered and debris-free glacier areas.

With model versions that apply different melt and radiation factors to simulate melt from debris-covered and debris-free glacier areas, we obtain better model performances in the calibration phase (see Supplementary Material, Figure E). However, for the transfer runs, the performances are lower than for model versions that do not account for debris cover

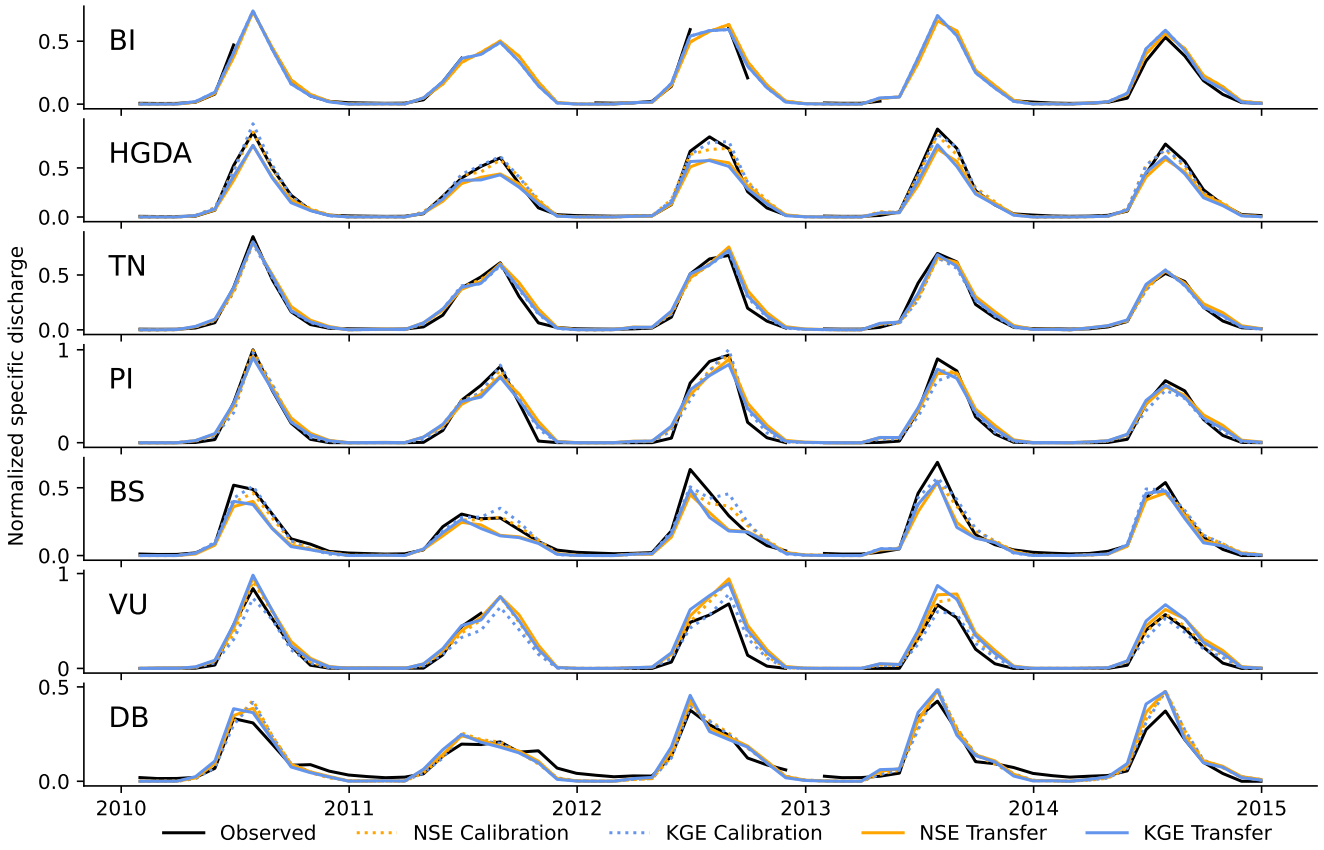

Figure 11. Observed and simulated monthly hydrographs for the HTI melt model on the seven catchments, calibrated or transferred with the calibrated parameter set of BI. Observed discharge (solid black line) is compared to the calibration run and to the transfer runs with the calibrated parameters of BI: shown are the results for NSE (orange) and KGE (blue); dotted lines show the calibration runs, solid lines show the transfer runs. For BI, the calibration and transfer runs are identical. Observed monthly discharges with missing values are not shown. Specific discharge (unit: mm) is normalized to the highest value.

(Fig. 13b), i.e. the transferability of the model parameters decreases. This is especially noticeable for the VU and TN catchments.

4.5 Regionalization of the melt model

We showed that with the TI and HTI models it is possible to simulate the discharge of nested and neighboring catchments with parameters calibrated at the main local outlet (BI), albeit with a small decrease in performance. The ensuing question is to know whether or not these parameters can be used to infer conclusions about the physical processes and dynamics occurring in the neighboring and subcatchments. In our simulations, the studied catchments have very similar meteorological drivers, in terms of precipitation (Fig. 14a) and temperature (Fig. 14b). The meteorological data is

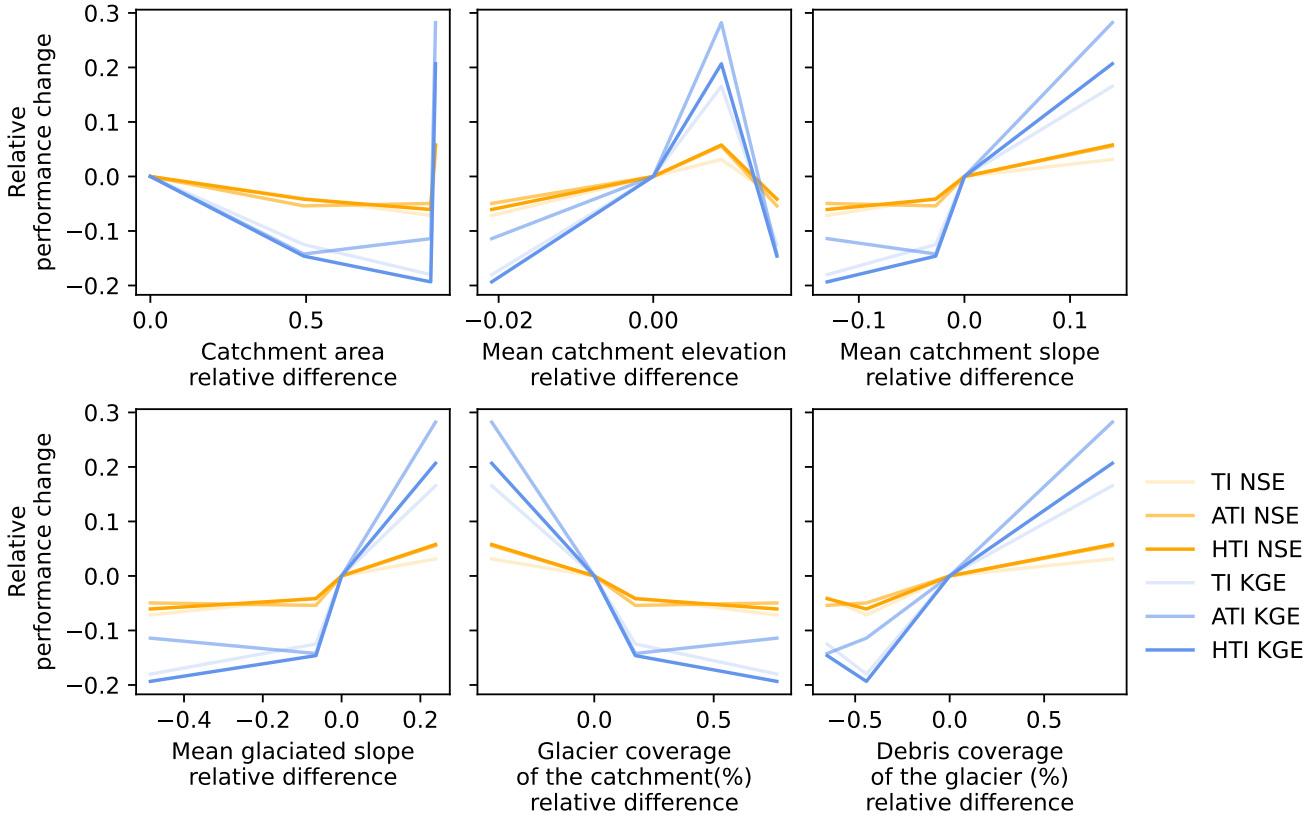

Figure 12. Comparison of selected physiographic characteristics of the nested catchments with the relative performance change between the calibration and transfer discharge simulations. The relative performance change is the relative drop in KGE and NSE performance criterion, calculated as follow: $(\text{calibration run} - \text{transfer run})/\text{calibration run} \times c$ with c a visually assessed coefficient reflecting the overestimation (1; VU) or the underestimation of simulated discharge (-1; HGDA and BS). The relative differences are calculated as follow, with the catchment are as example: $(\text{BI catchment area} - \text{catchment area})/\text{BI catchment area}$. The point showing performance changes and relative differences of 0 is BI.

interpolated based on ~~the~~ ground-based observations from a few rather low elevation measuring stations, with the only station in our study area being that of Arolla at 2005 m a.s.l. (Fig. 1a) and the highest station in the ~~bigger~~ surrounding area being located 29 km SW at Col du Grand St-Bernard, at 2472 m a.s.l.. Thus, the actual weather patterns may be more different between the studied subcatchments than ~~what~~ is suggested by the interpolated weather data. Indeed, the studied catchments' discharge patterns show clear differences between DB and BS and the other catchments. DB and BS show for all years on record a single melt-induced discharge peak in early summer, followed by low discharge. All other catchments show the same discharge peak in June, followed by even higher discharges in the subsequent summer months (Fig. 14c).

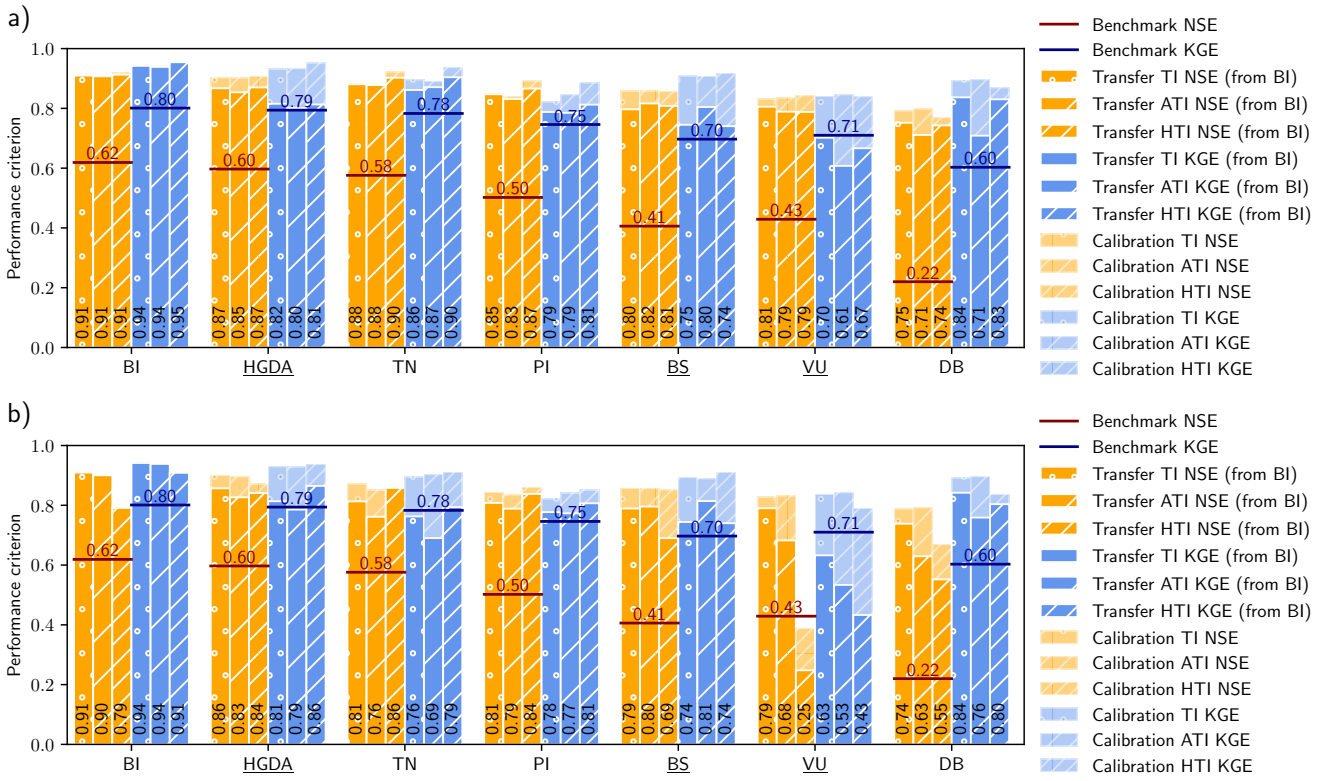

Figure 13. As Figure 3 but comparing calibration and transfer runs for all seven catchments, a) without and b) with debris cover separation, for the period 2010-2014. Shown are NSE values (orange) and KGE values (blue), along with the benchmark NSE value (red line) and KGE value (dark blue line) values. The performance values for the corresponding calibration runs are shown in more transparent color. BI's nested catchments are underlined.

To assess the quality and uncertainty of the precipitation inputs fed into the model, we compared the mean precipitation for each catchment with the measured precipitation from four weather stations: the Arolla station (2005 m a.s.l., since end of 2011 only), the Orzival (2640 m a.s.l.) and Tracuit (2590 m a.s.l.) stations from the Anniviers valley just to the East, and the Col du Grand St-Bernard station (2472 m a.s.l.) farther to the west. The precipitation inputs of our catchments globally fall within the same range as the measured precipitation of the Arolla station (Supplementary Figure F1c) in terms of annual precipitation amounts. The modeled interannual trend also matches the interannual trend observed in the neighboring Anniviers valley. At the daily scale, the peaks are globally well reproduced in terms of timing, but are sometimes more uncertain in terms of amount (July 2012 peak; Supplementary Figure F1b). This discrepancy is explained by the high variability of precipitation in high alpine areas, which is well illustrated by the precipitation differences at large spatial ~~scale~~ scales between the Col du Grand St-Bernard and the Anniviers stations (Supplementary Figure F1a, c), and at smaller spatial ~~scale~~ scales between the two Anniviers stations (April 2012 peak; Supplementary Figure F1b).

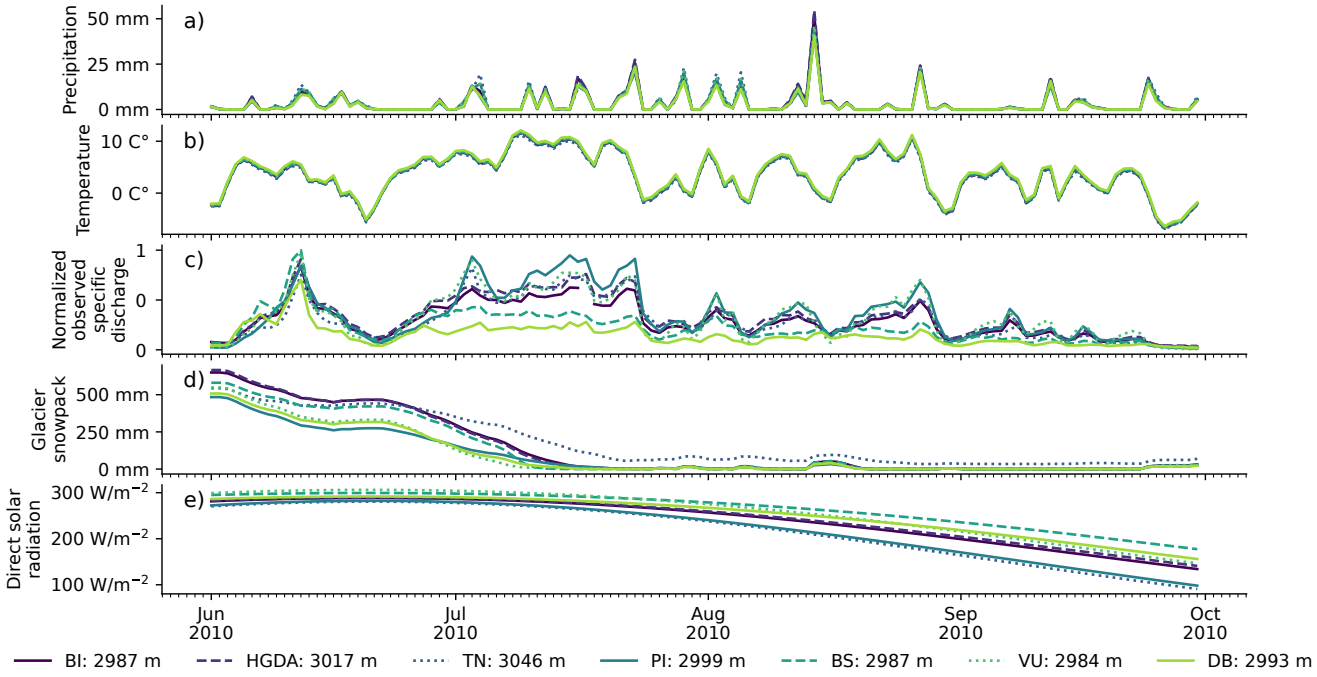

Figure 14. Patterns of a) precipitation, b) temperature, c) observed discharges, d) simulated glacier snowpack thickness and e) potential clear-sky direct solar radiation for the different catchments during the summer 2010. The values are mean values computed over the entire catchments from the RhiresD and TabsD MeteoSwiss datasets. The mean elevation of the catchment is indicated in the legend. Specific discharge (unit: mm) is normalized to the highest value.

This small spatial scale variability, however, has no impact on the global interannual trends (Supplementary Figure F1c). There is little variation in the amount of precipitation between our catchments (Supplementary Figure F1b), which leads us to believe that the spatial variability of our precipitation input is probably underestimated. This minimized variability of the input precipitation might explain the model's difficulty in reproducing specific discharge peaks, and the distinctive hydrological regimes exhibited by DB and BS. ~~However, we argue that as the discharge is calibrated on the whole 2010-2014 period, this underestimation of variability has a minimal impact on global spatial transferability of our model to sub- and neighboring catchments~~Thus, some effects from rainfall variability that are not captured in the input data are possible.

~~A further~~ Another reason may relate to the glacial coverage of the DB and BS catchments. Based on the simulated snow water content, we find that the lower discharges exhibited by BS and DB in July-August cannot be ~~explained entirely~~ entirely explained by snow exhaustion, as BS and DB still show more than 10 cm of mean simulated snow depth at the beginning of July. However, ~~it can~~ they may be explained by the intra-annual pattern of snow and ice melt - when snow melt slows due to depletion, glaciers become snow-free and ice melt begins. DB and BS show the lowest glacial coverage ($< 10.1\%$; Table 1), so when the snow has disappeared, very little ice melt ~~ensues~~occurs. In other catchments, ice cover

is much greater ($> 32.0\%$), and in late July and early August, discharge is at its highest due to high rates of ice melt. We also note that TN is the only catchment for which the simulated snow cover did not melt completely during summer 2010 (Fig. 14d).

5 Discussion

5.1 ~~A new~~ An exhaustive block resampling benchmark

In this study, we have introduced a ~~new~~ benchmark based on the bootstrapping method, where we ~~use~~ exhaustively resample discharge data from the 5 years 2010-2014 ~~to simulate 100~~ in yearly blocks to simulate 3125 discharge time-series, calculate the NSE and KGE objective functions for each of these ~~100~~ time-series and average them. This ~~bootstrapping~~ resampling method was retained because it is an easy metric to compute and it gives a good idea of the model fit in comparison to a sample of discharge with the same hydrologic regime. ~~If the sampled years exhibit significantly different dynamics, this bootstrapping method is less meaningful. However, our~~ Our catchments are all dominated by recurrent summer snowmelt and icemelt dynamics (Supplement Figure ~~??~~I5), albeit with some temporal variability. ~~By taking a large number of boostrapped combinations, we compensate for outlier years . We decided against averaging the discharges for the day of the year, as done by Schaefli and Gupta (2007), because we compute the benchmark over a relatively short 5-year period. Averaging discharges is particularly interesting for long time periods where it gives a robust representation of the average seasonal signal. In our case, the modeling period is short and~~ To ensure that our benchmark is not affected by outlier years exhibiting significantly different dynamics, we computed a long-term benchmark over the ~~resulting average of the day of the year could be strongly influenced by a single year, which can be avoided with the retained bootstrapping method.~~ whole available dataset (46 years; Table I1), which show lower values of NSE and KGE than the 5-year benchmark, demonstrating the relative consistency of the dynamics in recent years. ~~We heuristically choose 100 combinations, as motivated by the origin of the bootstrapping method, that has a strong random component to it.~~

5.2 Discharge predictability: catchment size matters

We first discuss the influence of catchment size on discharge predictability, notably on our benchmark metrics. The benchmark metrics show that the predictability of discharge from past discharge signals alone is less high for small catchments than for larger ones (Fig. 3). However, Hydrobricks shows similarly high model performances for small and larger catchments, highlighting the added value of a hydrological model in small catchments. This outcome can be directly explained by spatial relations. Large catchments exert a stronger averaging effect on spatio-temporal processes than small catchments. Indeed, the discharge in small catchments is driven by a localized and likely uniform meteorological patterns, while larger catchments draw from multiple local meteorological patterns, leading to a certain averaging. This is well illustrated by the precipitation events in the Anniviers valley (Supplement Figure F1b), which are not always recorded by both stations. This complexity obscures the correlation between meteorology and discharge in larger catch-

ments, but not in smaller ones. Furthermore, the difference in catchment areas is here closely linked to differences in stream order, which results in a different balance between water travel times in unchanneled states (hillslopes, surface runoff) and in channeled states (in-stream) (Rinaldo et al., 2006; Michelon et al., 2021). Longer in-stream flow paths lead hereby to a stronger dampening effect of hillslope- and glacier-scale runoff variability. This geomorphological dispersion of the discharge waves traveling downstream (Rinaldo et al., 1991) can also be observed when comparing the discharge patterns recorded in the BI, VU and HGDA intakes (Supplementary G1): the smaller catchments (VU and HGDA) are relatively much more affected by the precipitation event than BI, even though they are hydrologically similar. Given the inherent year-to-year variability in meteorological patterns, and the close link between meteorology and discharge, it ensues that in small catchments, the discharge patterns from previous years are poor predictors of the current discharge. In contrast, even simple meteorology-based hydrological models such as Hydrobricks deliver much better results. ~~An ideal benchmark should not depend on scale; however, we do not see at this stage how to construct such a benchmark.~~

Additionally, we note that the NSE benchmark metric values tends to decrease much more strongly than the KGE benchmark metric values with decreasing catchment size (from ~~0.52 to 0.05~~ 0.62 to 0.22 for NSE and from ~~0.75 to 0.50~~ 0.80 to 0.60 for KGE; Fig. 3). This difference in decrease is explained by the ~~bootstrapping~~ resampling approach used to produce the discharge data.

The NSE assesses the fit of one series to another solely based on the squared difference between the two time series. The KGE, on the other hand, uses a linear combination of correlation between the two series, variability error (ratio of the standard deviations), and bias error (ratio of the means). Given the definition of KGE and NSE, the correlation term is linearly related to NSE, ~~while~~ whilst the variability term and the bias term have a quadratic relation to NSE (Clark et al., 2021). As a result, the NSE is much more sensitive to changes in ~~bias, changes in~~ variability or shifted yearly patterns than the KGE (see Supplementary Material, I1; Knoben et al., 2019). Furthermore, since the KGE evaluates bias, variability, and correlation independently, two good components (e.g., bias and variability) can offset a suboptimal third component (e.g., correlation). This is not the case in the NSE. Thus, the benchmark KGE ~~is~~ values past discharges much more highly than the NSE, which makes it a much harder ~~criteria~~ criterion to meet for simulated discharges than the benchmark NSE. We thus expect the simulated hydrographs to outperform the benchmark NSE and match the benchmark KGE.

5.3 Satisfactory hydrograph predictions across nested and neighboring catchments

As discussed previously, we expect the simulated hydrographs to outperform the benchmark NSE and match the benchmark KGE. When transferring the parameters calibrated for the largest subcatchment (BI) to model the discharge in all other nested (Fig. 7) and neighboring catchments (Fig. 10), we observe that despite exhibiting slightly inferior performance compared to the direct calibration, most catchments still show satisfactory results, even for the smallest ones. For all catchments, the transfer simulations with transferred parameters match both the benchmark KGE and NSE, at the exception of VU~~with the ATI model~~. The decreases in KGE for the HGDA, BS and VU observed in all melt models is not accompanied by a similar NSE decrease, which hints toward an amplitude change in the discharge signal (Supp. Mat.

I1). This amplitude change is produced by the underestimation (HGDA, BS) and overestimation (VU) in discharges that we observed in Figure 6.

The lowest NSE score obtained through a HTI transfer run is 0.76, and the biggest decrease with respect to the calibration run reaches 0.05. As a comparison, the best fits achieved by Parajka et al. (2005) for regionalization over Austrian catchments yielded a NSE decrease from 0.72 to 0.67, and globally, the mean and maximum NSE in European catchments reach 0.72 and 0.91, respectively (Guo et al., 2021). Furthermore, the discharges in nested catchments are consistent with each other (Fig. 8), as the sum of the nested discharges does not exceed the main catchment's discharge.

The performance decrease between calibration and transfer simulations in nested catchments could be attributed to the slope of the terrain and the debris coverage of the glacier (Fig. 12). When the catchment and glaciated slopes are steeper, the models tend to underestimate discharge, maybe because steeper slopes lead to faster runoff and higher discharge rates that are not fully captured by the model. Conversely, for flatter slopes, the models overestimate discharge, possibly due to slower runoff and more significant water retention. Additionally, lower debris coverage on glaciers leads to underestimated discharge, potentially because debris cover, in our case, increases rather than decreases melt rates. By taking into account these relations between discharge and physiography explicitly in the model, we could potentially improve its transferability.

5.4 Possible explanations for slightly over- and underestimated discharges

Several factors could contribute to the overestimation of the VU discharge and to the underestimation of the HGDA and BS discharges observed in the monthly hydrographs (Fig. 11). Firstly, the meteorological forcing could be incorrect, as it is highly variable in such alpine environments, difficult to observe, and available at a coarse resolution (1 km). Calibrated parameters are known to compensate for such input error effects (Bárdossy and Das, 2008) and transferred parameters might thus induce biases. Secondly, the delineation of some of the catchments is uncertain, considering the uncertainty of water flow paths beneath glaciers. This might in particular be the case for VU, where Bezinge et al. (1989) suggest a potentially smaller catchment area to the north. Thirdly, the melt model could be missing an important driving factor. Temperature-index based methods are known to yield good results in environments where melt is mainly driven by incoming longwave radiation and sensible heat flux (Ohmura, 2001), which is typically the case in ~~alpine~~ Alpine catchments (Thibert et al., 2018).

We also observe a worsened result when accounting for debris cover on glaciers. The following reasons might explain this result: i) an overfitting of local specificities of the model, ii) a spatially inconsistent effect of debris cover on ice-melt or iii) difficulties making the high number of parameters converge given the amount of reference information contained in observed discharge (which does not provide enough constraints on the parameters). These reasons are linked. The inconsistent effect of debris (ii) when the "local specificity" is, for example, a strong melting or protective effect of the debris cover, which is not shared by other catchments is a specific example of model overfitting (i). Convergence difficulties (iii) lead to the difficulty of finding a single set of parameters to explain the results, which in turn can lead to an overfitting of local specificities if a solution is slightly better (i). All these reasons fall under the overparametrisation problem.

We observe that in both the cases of the ATI and the debris cover in TI or HTI, we introduce differential melt rates that cannot be independently calibrated as we only use the discharge to calibrate against. Compared to the simple TI model (without accounting for debris), the introduction of debris cover within the TI model results in a greater drop in goodness of fit than using the simple ATI model. Thus, we can conclude that this decrease in goodness of fit is probably due to the differential behavior of debris ~~coverage~~cover.

5.5 Enhanced parameter transferability through improved melt model accounting for potential solar radiation

Melt rates per positive degree-day are sensitive to a number of characteristics that influence the surface energy balance, and which include elevation, direct solar radiation input, albedo, wind speed and seasonality (Hock, 2003; Ismail et al., 2023). These ~~explain~~ imply that for a given glacier, the degree-day factors of ice and snow are different, with ice, being less reflective than snow, melting more per positive degree-day. This variability of the link between positive air temperature and melt can also be found at a local-scale, within snow and ice patches (Gabbud et al., 2015). This sensitivity of melt rates tends to limit the transferability of melt parameters from one catchment to a neighboring or nested catchment for the basic TI model ~~.~~

~~By discretizing the study catchment by aspect and solar radiation and implementing the ATI and HTI models~~and motivated the computation of spatially variable degree-day factors from snow data in previous work (He et al., 2014; Hingray et al., 20 . Here, we tested the influence of aspect and radiation on melt model parameter transferability by discretizing the study catchment according to aspect and solar radiation and implementing the ATI and HTI models. According to the work of Comola et al. (2015), local-scale degree-day factors become stable (and therefore transferable) at scales at which the variability of local hillslopes' orientation does not further increase (less than 7 km^2, in their study). In this case study, we have five catchments that are small enough to be affected by their dominant aspect, but only two of them show low aspect variance also on their glaciated surfaces (DB and BS, Fig. 1c). However, all TI and ATI calibrated melt and radiation factors are highly inconsistent across catchments, whereas we find strong parameter ~~overlap~~ overlaps between catchments for the HTI model (Fig. 5). Similar to the work of Comola et al. (2015), we find that taking into account solar radiation patterns using the HTI model does not fully explain the hydrological response variability at smaller catchment scales. Indeed, BS and HGDA tend to have higher melt parameters when calibrated alone (Fig. 5), and produce slightly underestimated discharge when transferred with BI's parameters (Fig. 6), responsible for their lower KGE values (Fig. 7a). The transfer results obtained from the HTI model do not demonstrate improvements in terms of simulated hydrographs compared to the TI model (Fig. 6 and 9), suggesting that the radiation as computed by Hock (1999) may not be enough to explain the KGE differences. However, both the TI and the HTI models show good transferability in terms of metrics to even the smaller catchments (Fig. 7 and 10). ~~We elaborate on Comola et al. (2015)~~ Although Comola et al. (2015) showed that aspect significantly influences the calibration of degree-day factors for small catchments (< 7 km^2) with the TI model, we elaborate to conclude that the ~~obtained hydrographs~~ hydrographs obtained through regionalization are still very good fits for smaller, nested catchments~~, and~~ . We find that parameter transferability to catchments below 7 km^2 in the TI and HTI

model is a reasonable approximation, but that the HTI model should be preferred due to the more consistent parameter calibration.

These differences could also be due to variations in ice albedo. Indeed, the glaciers in the studied catchments are not identical in terms of debris cover (Fig. 1). We thus tried to take the debris into account, but failed to obtain better results in transfer runs (Fig. 13b), which hints towards a non-consistent behavior of the debris cover. This is very possible, as debris cover is known to either shield or amplify melting (Gabbud et al., 2015). We do not have information about the debris thicknesses in our study area, and the contribution of two processes so close as debris-free ice melt and debris-covered ice melt would be hard to constrain from discharge data only (Pokhrel et al., 2008). Thus, debris cover related parameters are less transferable than the global ice parameters.

5.6 Implications

Our analysis underlines the value of including potential radiation in the temperature-index model for spatial transferability of melt model parameters, a topic that has been neglected in previous studies and is worth being pursued further. It remains to be ~~seen~~ shown whether similar transferability of the HTI melt model can be expected in larger catchments (i.e. >500 km2; e.g., Hingray et al., 2010; Fatichi et al., 2015), although we expect this to be true for all catchments whose hydrological regimes are strongly influenced by snow and glaciers, provided, of course, that other geomorphological characteristics are relatively constant. ~~These results are conditioned by the melt models tested here and it would be good practice to subject any new models to such transferability tests to see whether they perform better or worse than the HTI model before being transferred.~~ Most likely, the added value of potential radiation decreases for increasing catchment sizes since the variability of aspects with increasing scale tends to a limit value that averages out the effect on melt (Comola et al., 2015). In addition, future work could focus on the benefit from including snow data to jointly calibrate the melt model parameters and other hydrological process parameters. This multi-signal calibration, on both flow and snow data, could reduce parameter equifinality (though not eliminate it (Finger et al., 2011)) and reduce parameter interdependency, as shown in the work by Ruelland (2024). This could further enhance the transferability of glacio-hydrological models (Carenzo et al., 2009). This only holds if the melt model gives an adequate representation of the melt processes for a streamflow simulation model: A large body of literature on temperature-index modelling focuses on how to further improve such models, e.g. by analyzing how degree-day factors or melt factors vary along the melt season (Ismail et al., 2023, and references therein), by improving the parameterisation of sub-daily melt dynamics (Tobin et al., 2013), or by linking melt to other variables rather than average daily air temperature (Follum et al., 2019; Nasab and Chu . For all these interesting developments, the question of how melt model improvements increase (or decrease!) the spatial transferability of the model should receive much more attention than in the past.

6 Conclusions

In this study, we tested the spatial transferability of melt models incorporating progressively more spatial information: a classical temperature-index melt model (TI), a temperature-index melt model based on aspect (ATI) and the temperature-index melt model of Hock (HTI). To do so, we calibrated each melt model over seven different catchments, then transferred the calibrated parameters of the main catchment to the three nested catchments, and the three neighboring catchments.

The results show that for high ~~alpine~~ Alpine catchments, it is possible to spatially transfer relatively simple semi-lumped glacio-hydrological models. We have demonstrated that our semi-lumped model (Hydrobricks) can successfully simulate discharge at several upstream points of the catchment after calibration to a single downstream observed discharge time series. This makes multi-point discharge simulation possible.

Although the best results in terms of transferability are achieved with the TI and HTI models, the highest consistency between parameters is achieved with the HTI model. This better convergence of parameters is witnessed both between the two performance metrics, as is also the case for the ATI model, but also between the seven catchments. The inclusion of debris cover on glaciers does not produce better results, and leads to model overparameterization. The NSE metric gives better calibration results than KGE when trying to fit the discharge peaks, but the benchmark KGE shows to be a harder, thus more significant, criterion to meet, and reproduces better the observed peaks. Thus, we find that the best framework to transfer parameters calibrated in the biggest local catchment to subcatchments and neighboring ones is by using the HTI model without debris cover.

Our simulations highlighted the possible influence of catchment and glaciated slopes, as well as debris cover percentage on the overestimation and underestimation of discharge in transfer runs. Since the inclusion of debris cover led to overparametrization, future research should focus on the integration of these characteristics in more spatially-informed ways.

Code availability. The software used to carry out this study is available at Horton and Argentin (2024).

Appendix A: Model structure of Hydrobricks and study area discretization

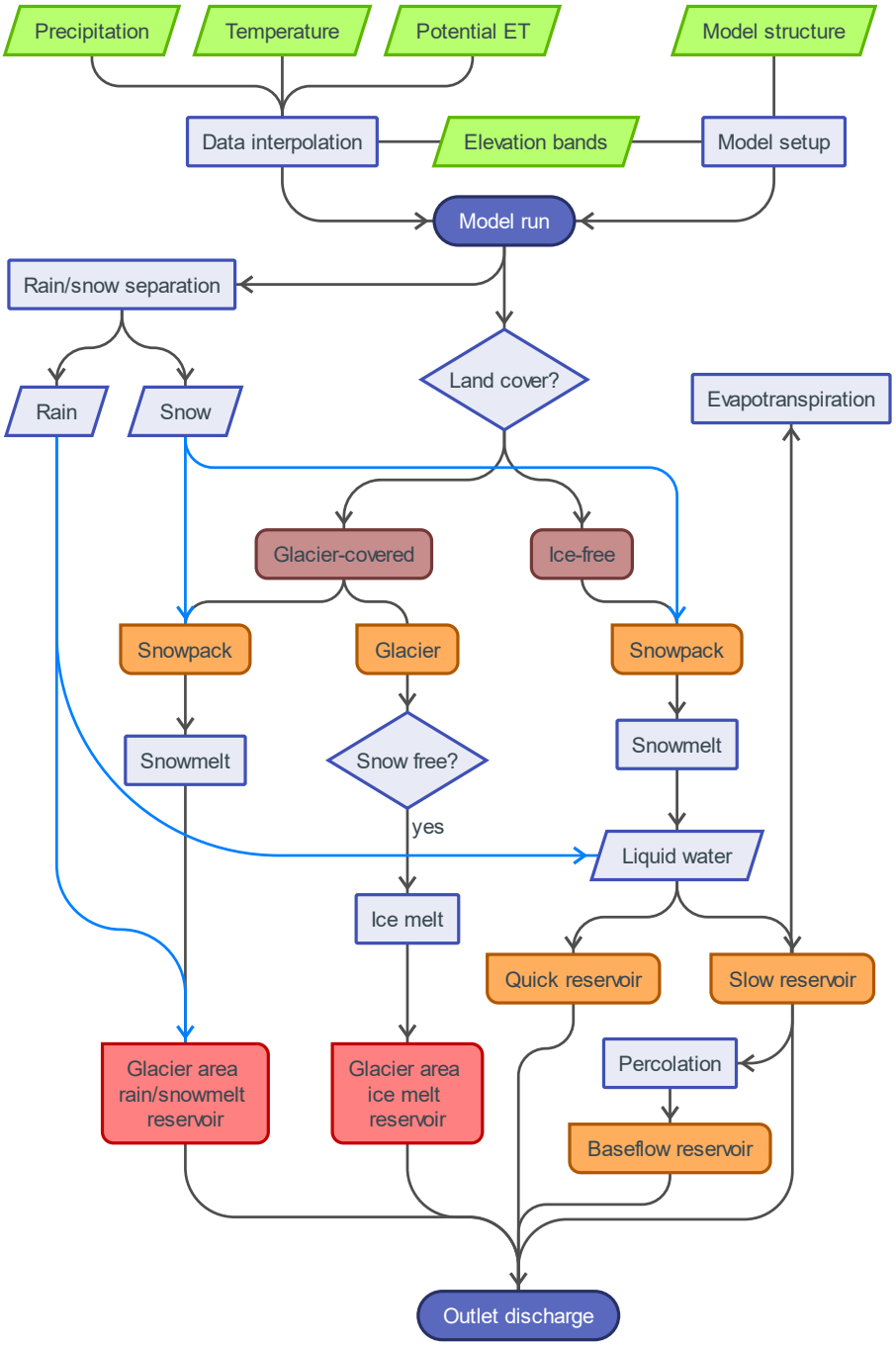

Figure A1. Illustration of the Hydrobricks model workflow used in this study. The glacier-covered part illustrates the behavior of both the bare ice and debris-covered glaciers. Orange reservoirs are distributed over all elevation bands and red reservoirs are lumped over the catchment. Figure taken from Shokory et al. (2023).

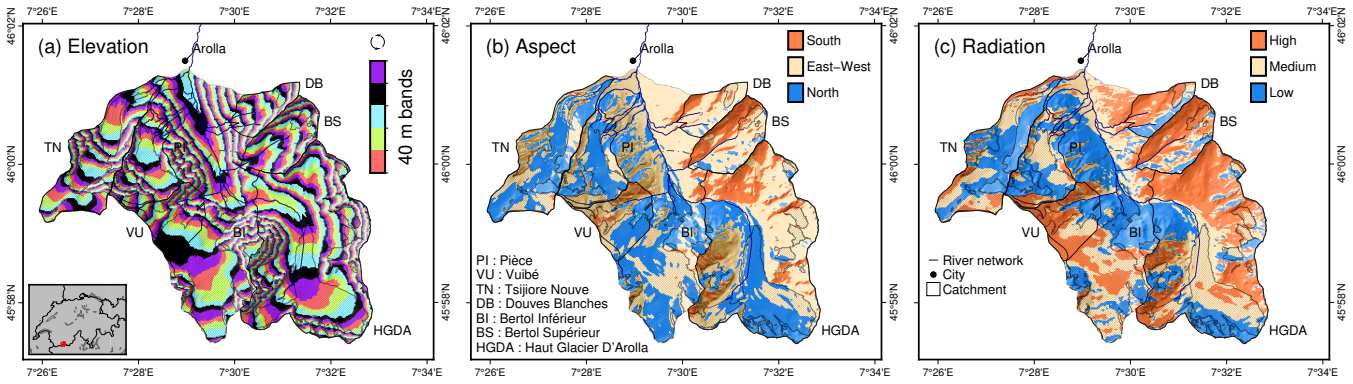

Figure A2. Hydrobricks' hydrological units for the whole catchment, discretized (a) according to elevation to use in the classic temperature-index (TI) model, (b) according to aspect to use in combination with elevation discretization in the aspect temperature-index (ATI) model (c) according to mean annual potential clear-sky direct solar radiation with cast shadows to use in combination with elevation in the Hock temperature-index (HTI) model.

Appendix B: Geology of the study area and characteristics of the catchments

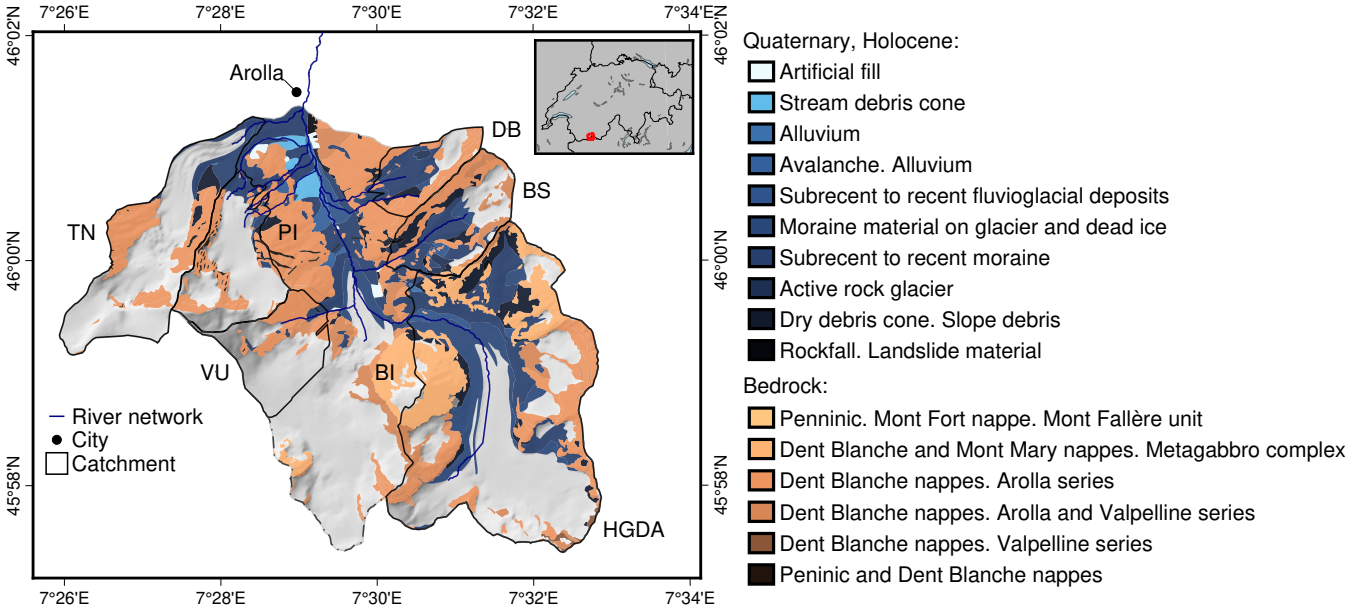

Figure B1. Geological cover of the study area, and of the different subcatchments, extracted from the GeoCover V2 product (SwissTopo, 2024).

Catchment	Minimum elevation (m)	Maximum elevation (m)	Mean elevation (m)	Standard deviation elevation (m)	Mean catch. slope
Whole	2183	3789	3085	289	29.3
BI	2183	3722	3063	229	28.7
HGDA	2582	3677	3014	191	29.5
TN	2289	3789	3180	443	28.2
PI	2636	3784	3046	266	27.8
BS	2913	3583	3127	117	32.4
VU	2730	3722	3036	152	24.7
DB	3097	3364	3218	58	35.4

Table B1. Basic statistics on the topography of the glaciated areas (2016) of each catchment.

Catchment	Mean (glaciated)	SD (glaciated)	Mean (catchment)	SD (catchment)
Whole	356.3	60.6	341.1	93.7
BI	347.9	59.1	323.4	95.4
HGDA	343.1	56.0	304.3	97.7
TN	12.6	57.1	15.4	64.5
PI	24.7	53.5	28.7	61.5
BS	238.9	43.5	237.7	72.4
VU	38.4	63.0	64.9	72.2
DB	249.7	23.4	259.2	48.9

Table B2. Circular means and standard deviations of the aspect over the glaciated areas (2016) and total areas of each catchment, computed with the zonal statistics of ArcGIS. SD: Standard deviation.

Catchment	Debris cover area (km^2)	Glacier area (km^2)	Debris coverage percentage
BI	1.00	10.04	9.9%
HGDA	0.69	4.22	16.3%
TN	0.56	2.76	20.4%
PI	0.29	1.66	17.3%
BS	0.03	0.24	14.3%
VU	0.02	1.21	1.4%
DB	0.04	0.15	23.9%

Table B3. Debris cover areas, glaciated areas and debris cover percentage for each catchment for the year 2016.

Appendix C: Additional results with the ATI melt model

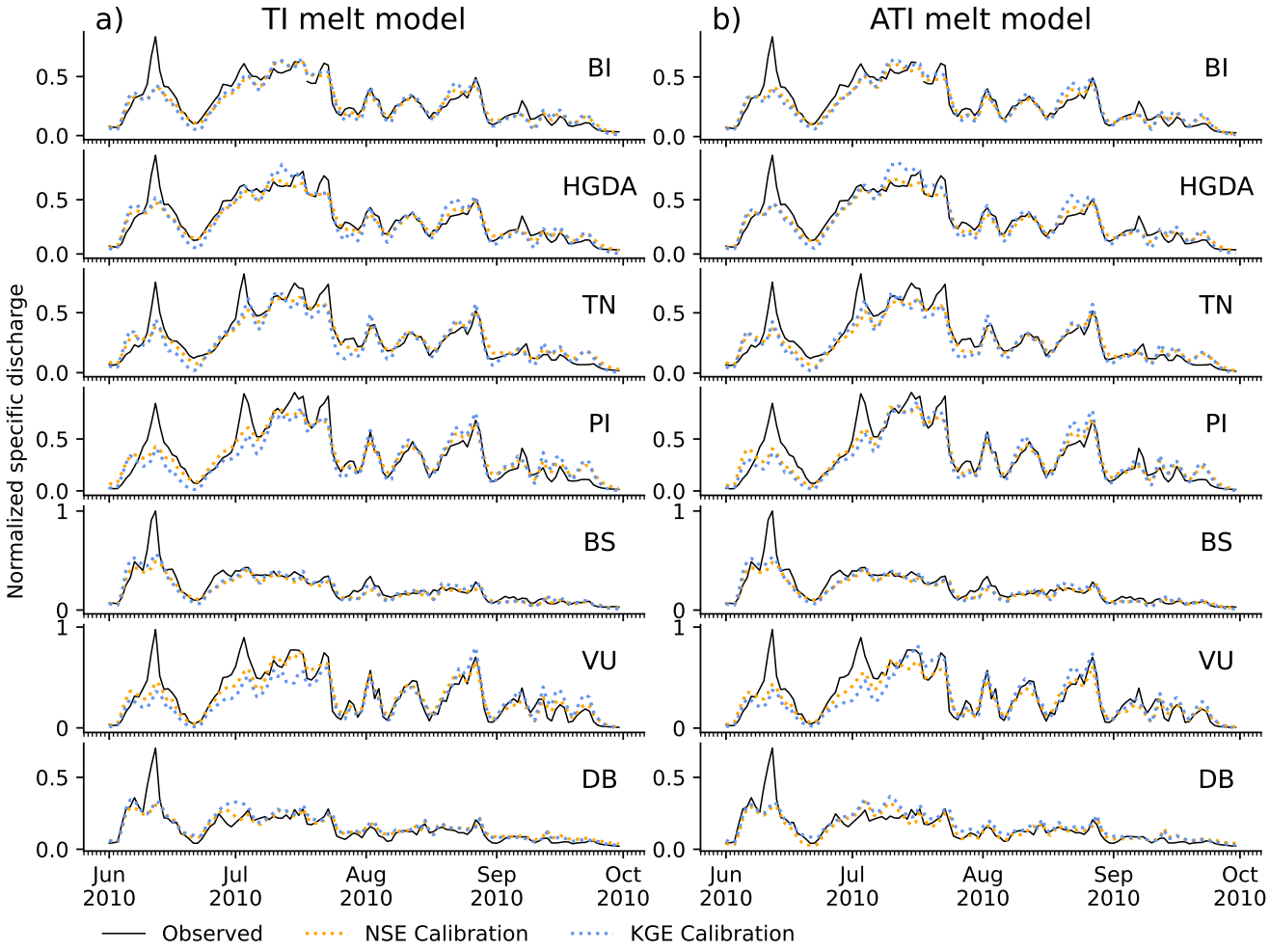

Figure C1. Same as figure 4, but with the a) TI and b) ATI melt models.

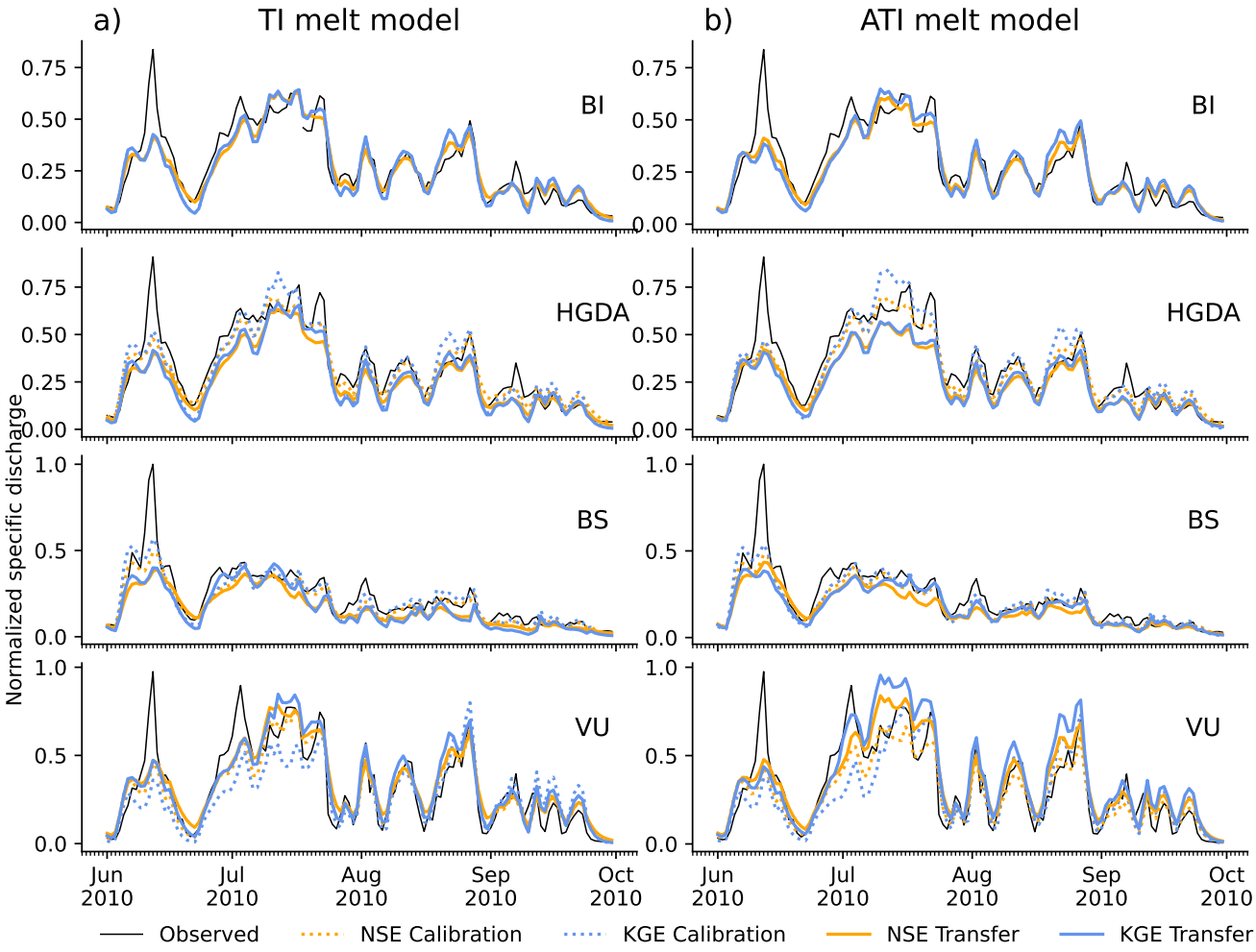

Figure C2. Same as figure 6, but with a) TI and b) ATI melt models.

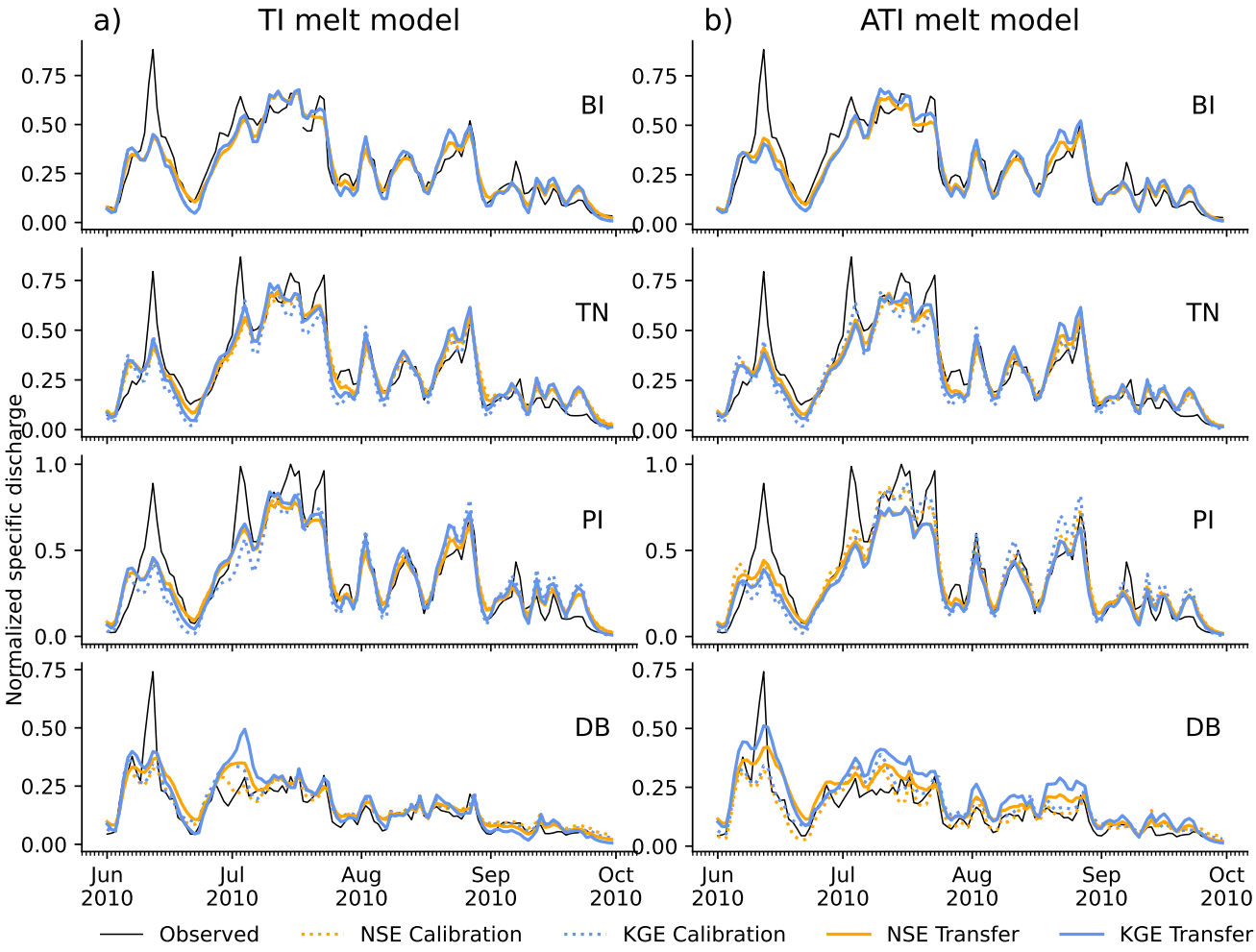

Figure C3. Same as figure 9, but with a) TI and b) ATI melt models.

Appendix D: Debris cover mapping

The GLAMOS dataset offers both debris-free ice extent and glacier extent for 2016, but only glacier extent for 2010 and previous years (Linsbauer et al., 2021; Fischer et al., 2014). To obtain the debris-free ice extent trend since 2010, we relied on the debris-free ice detection algorithm from (Shokory and Lane, 2023), now available under ArcGIS Pro. We applied it to compute the corresponding debris-free ice extents for the 2010 GLAMOS dataset, thus allowing us to infer the debris cover evolution from 2010 to 2016. No estimates of debris cover thicknesses were available.

Given the suboptimal conditions of Landsat 7 images in 2010 for mapping, we opted for an image from 2009. Two images, dated 06/09/2009 and 22/09/2009, displayed minimal cloud cover and limited snow patches. Between the two, the 06/09/2009 image displayed the smallest swath gaps. We corrected the Landsat 7 Level 1 near infrared (NIR)-B4 and shortwave Infrared (SWIR)-B5 bands, both available at 30m resolution, for top of atmosphere reflectance with solar angle correction. To do so, we applied the radiometric rescaling coefficients given in the associated metadata files provided with the Landsat Level-1 NIR and SWIR bands. We then applied the methodology of Shokory and Lane (2023) that uses the condition $\frac{NIR}{SWIR} \geq t$, with NIR representing the Near Infrared band, SWIR the Shortwave Infrared band, and t denoting the threshold condition for debris-free ice delineation. We tested incremental thresholds with steps of 0.05 between 1.00 and 3.00 and determined that a threshold value t of 2.00 provided the best results in the transition areas between debris-free ice and debris-covered ice (in brown, Fig. D1). We nonetheless had to manually correct for the influence of the swath gaps (in red, Fig. D1).

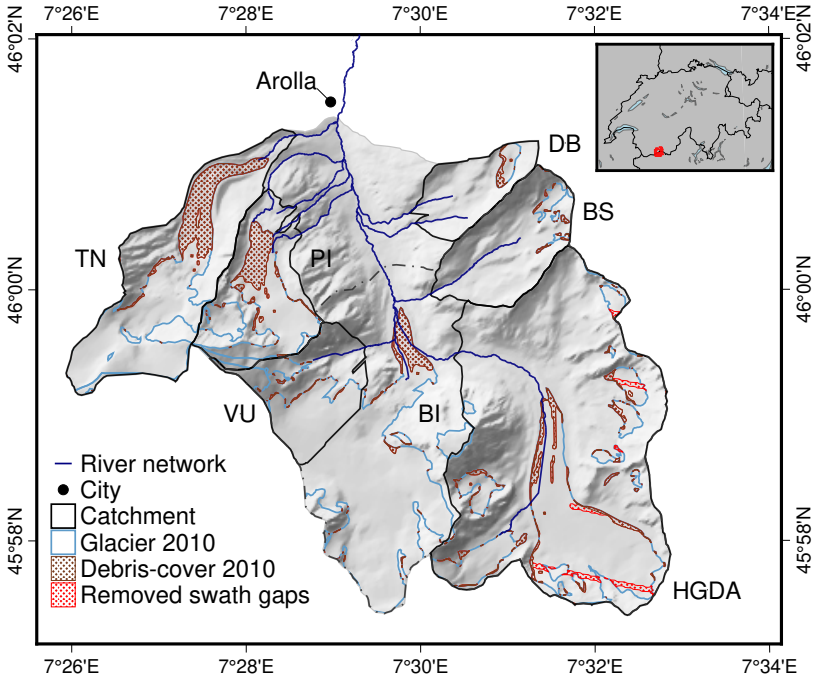

Figure D1. The mapped 2010 debris cover extent is indicated in brown, and the GLAMOS 2010 glacier extent in blue (Fischer et al., 2014). The manually removed debris linked to the swath gaps are indicated in red.

Appendix E: Additional results with debris cover

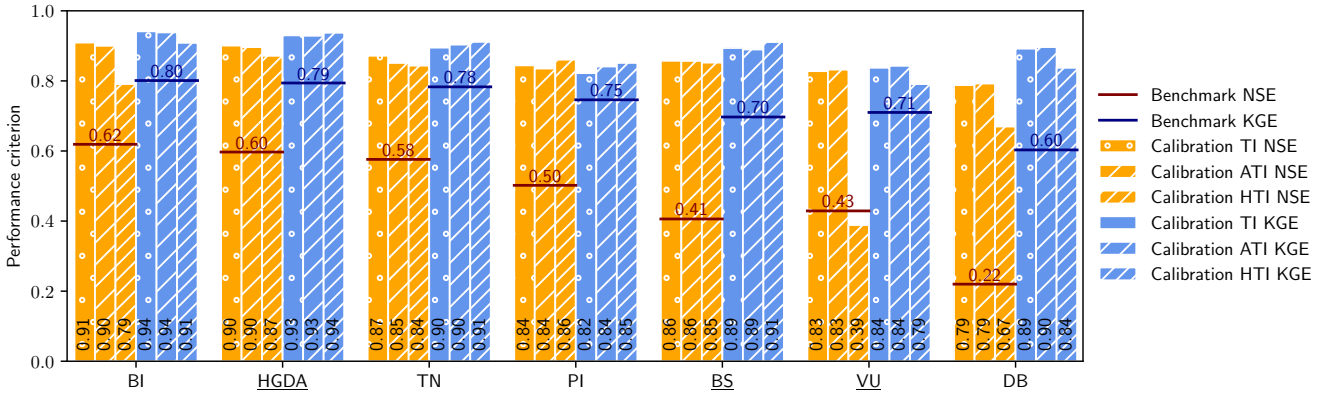

Figure E1. Comparison of the performance of the three melt models on the seven catchments, quantified either by the Nash–Sutcliffe efficiency (NSE, orange bars) or by the Kling-Gupta efficiency (KGE, blue bars) performance criteria of observed and simulated discharges for the period ~~2009-2014~~2010-2014. For comparison, the benchmark NSE and KGE are computed and plotted as red and dark blue thresholds, respectively. The simulations are run 10000 times over the years 2009 - 2014, with 2009 the calibration year. Catchments are ordered by area, from BI (largest) to DB (smallest). All performance criteria are computed on the 2010-2014 time period.

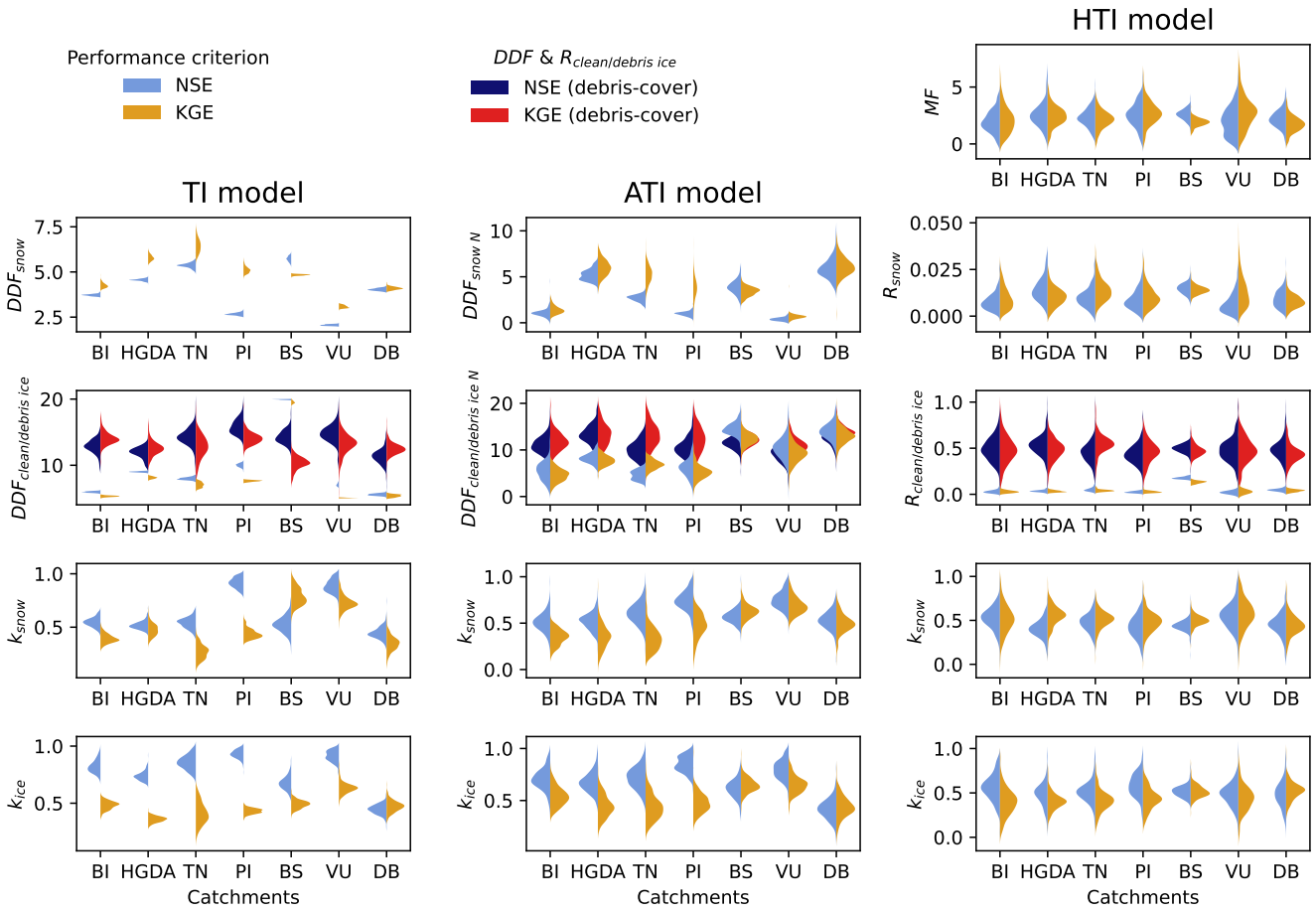

Figure E2. Obtained ice and snow parameters for the best 5% NSE and KGE scores for all catchments, with the three melt models and the two performance criteria.

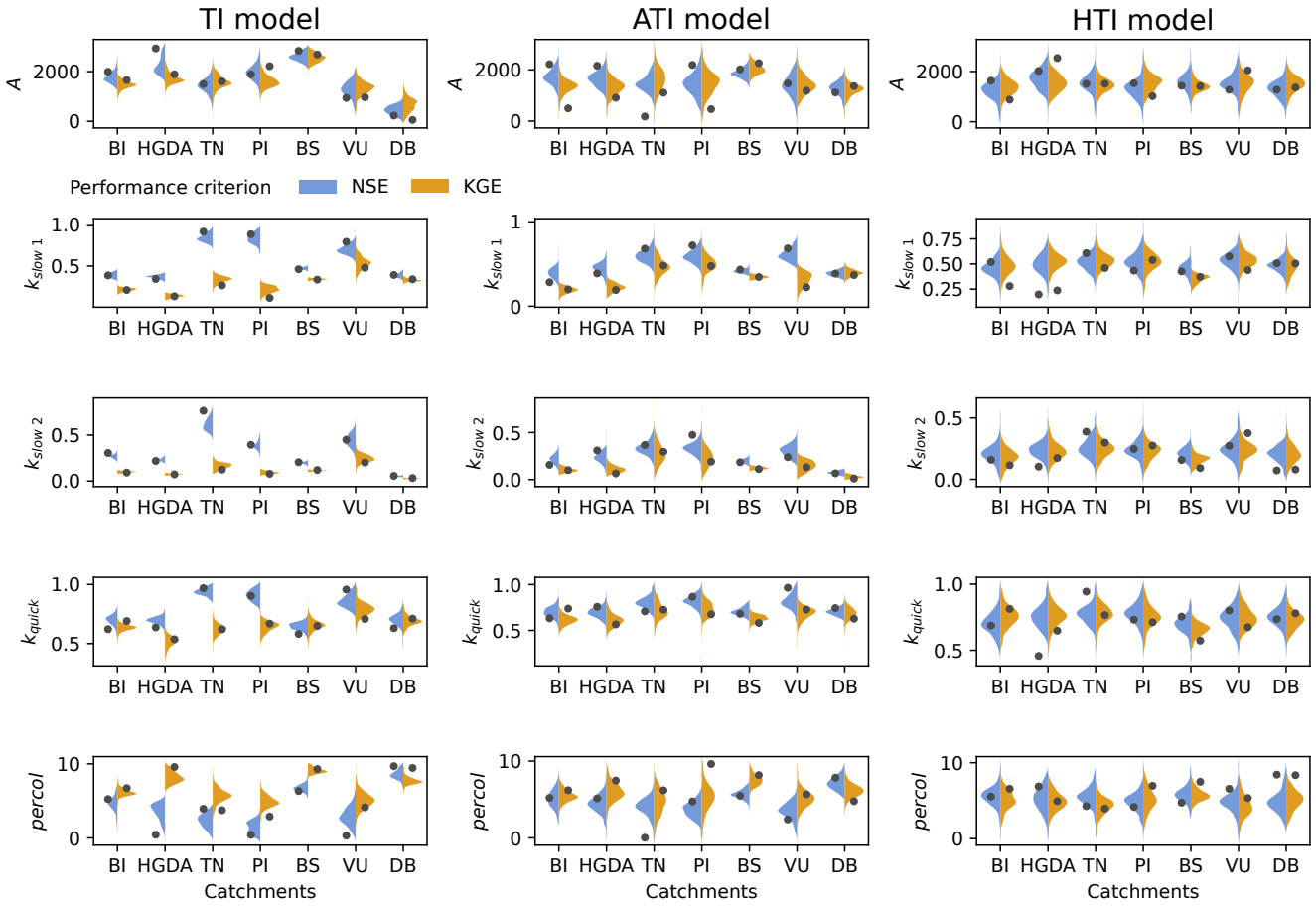

Figure E3. Obtained ground parameters for the best 5% NSE and KGE scores for all catchments, with the three melt models and the two performance criteria. The parameter set values achieving the best NSE and KGE scores are plotted on top with a dot. Catchments are ordered by area, from BI (largest) to DB (smallest).

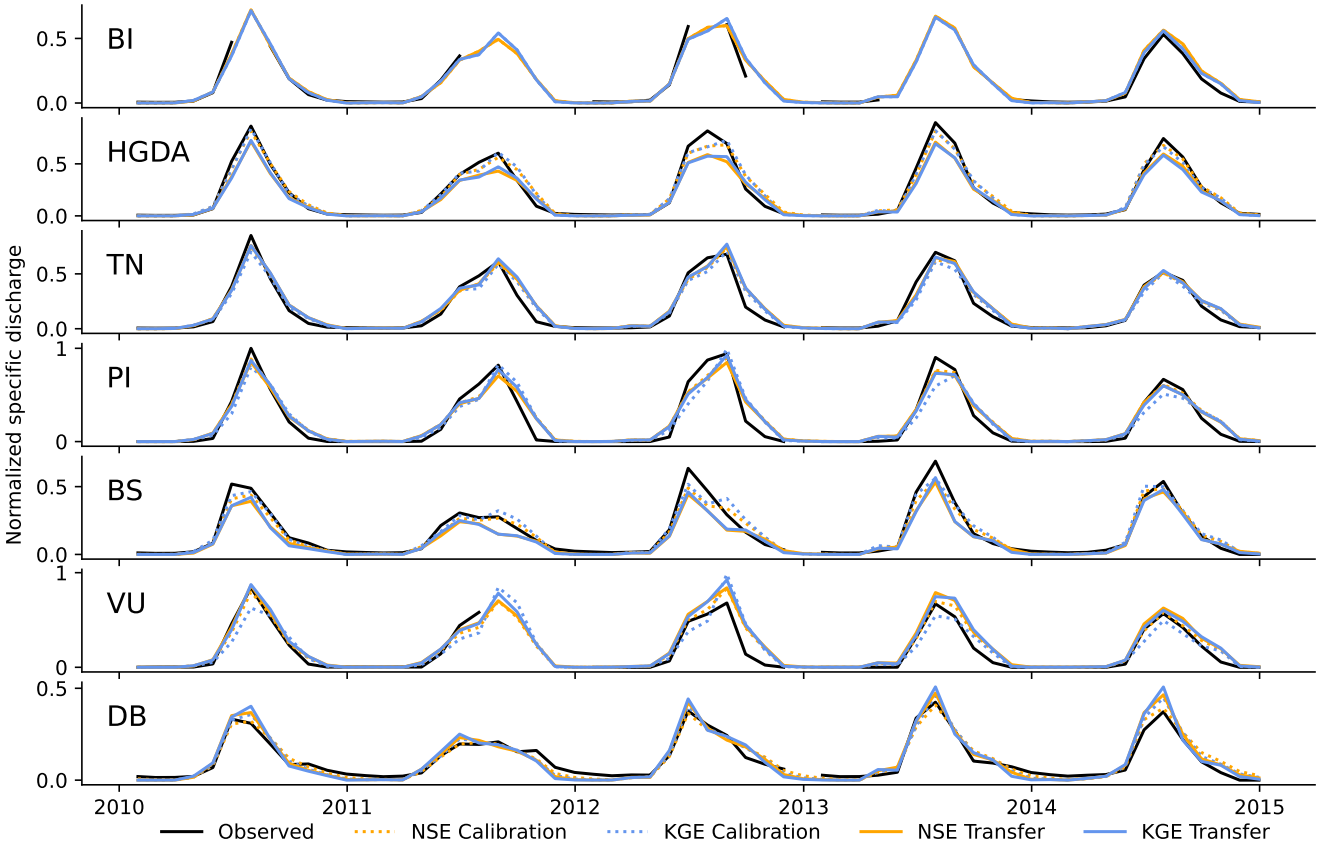

Figure E4. Monthly water hydrographs for the TI melt model on the seven catchments, calibrated or transferred through the application of the parameter set of BI. The original observed dataset (black) is compared to the calibration run using the NSE (dotted orange) and the KGE (dotted blue), and with the transfer run with the calibrated parameters found in the BI catchment with the NSE (orange) and the KGE (blue). Observed monthly yields with missing discharge values are not computed. Specific discharge (unit: mm) is normalized to the highest value.

Appendix F: Precipitation patterns

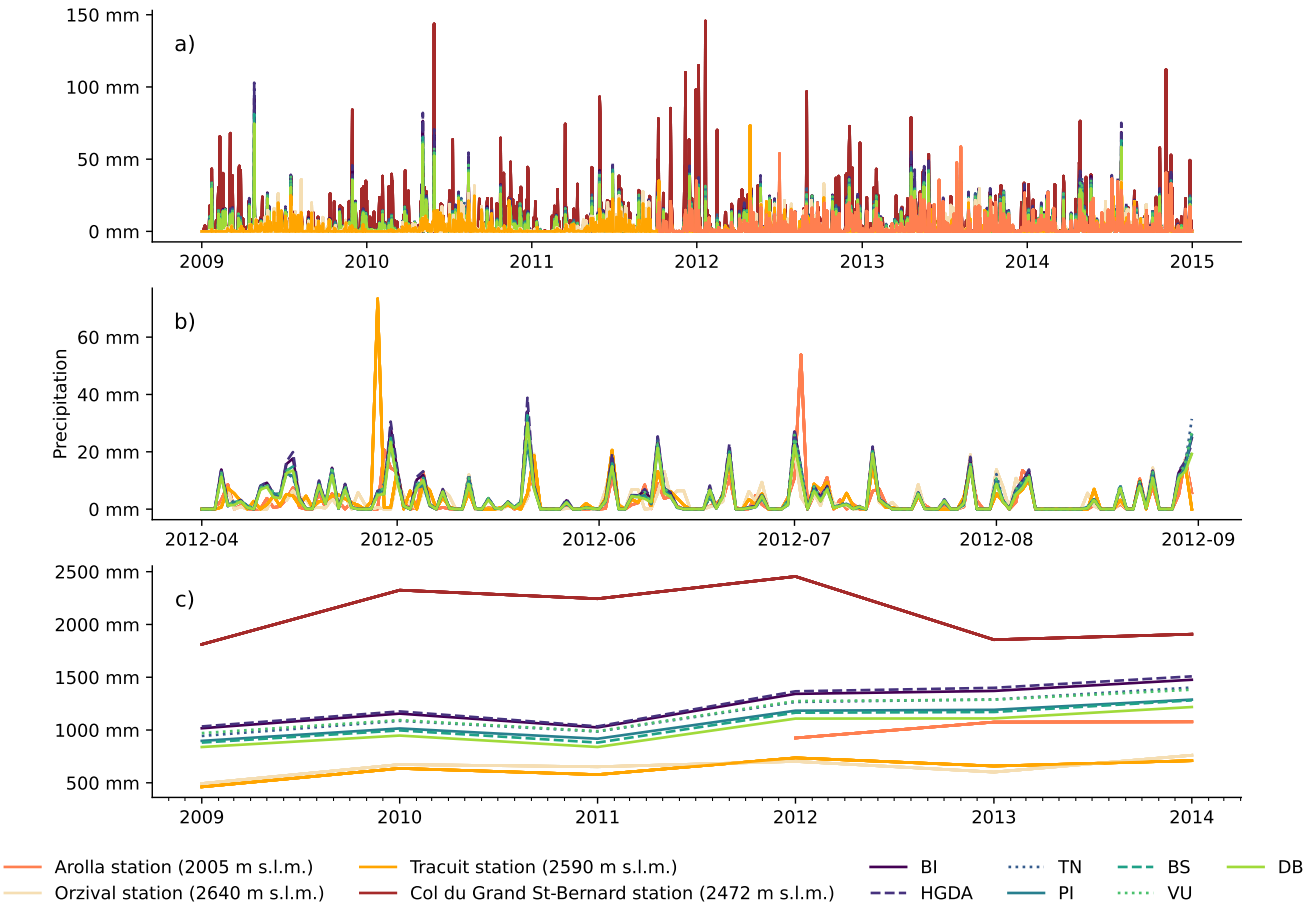

Figure F1. Comparison of the precipitation patterns at the Arolla station with the daily mean precipitation patterns for each catchment. Comparison at the daily scale (a) over the model time period, with the Arolla station data only becoming available end of 2011, (b) over the year 2012, and (c) at the annual scale, over the model time period. Orzival station: 20 km NNE of Arolla station. Tracuit station: 18 km NE of Arolla station. Col du Grand St-Bernard station: 29 km SW of Arolla station.

Appendix G: Discharge patterns

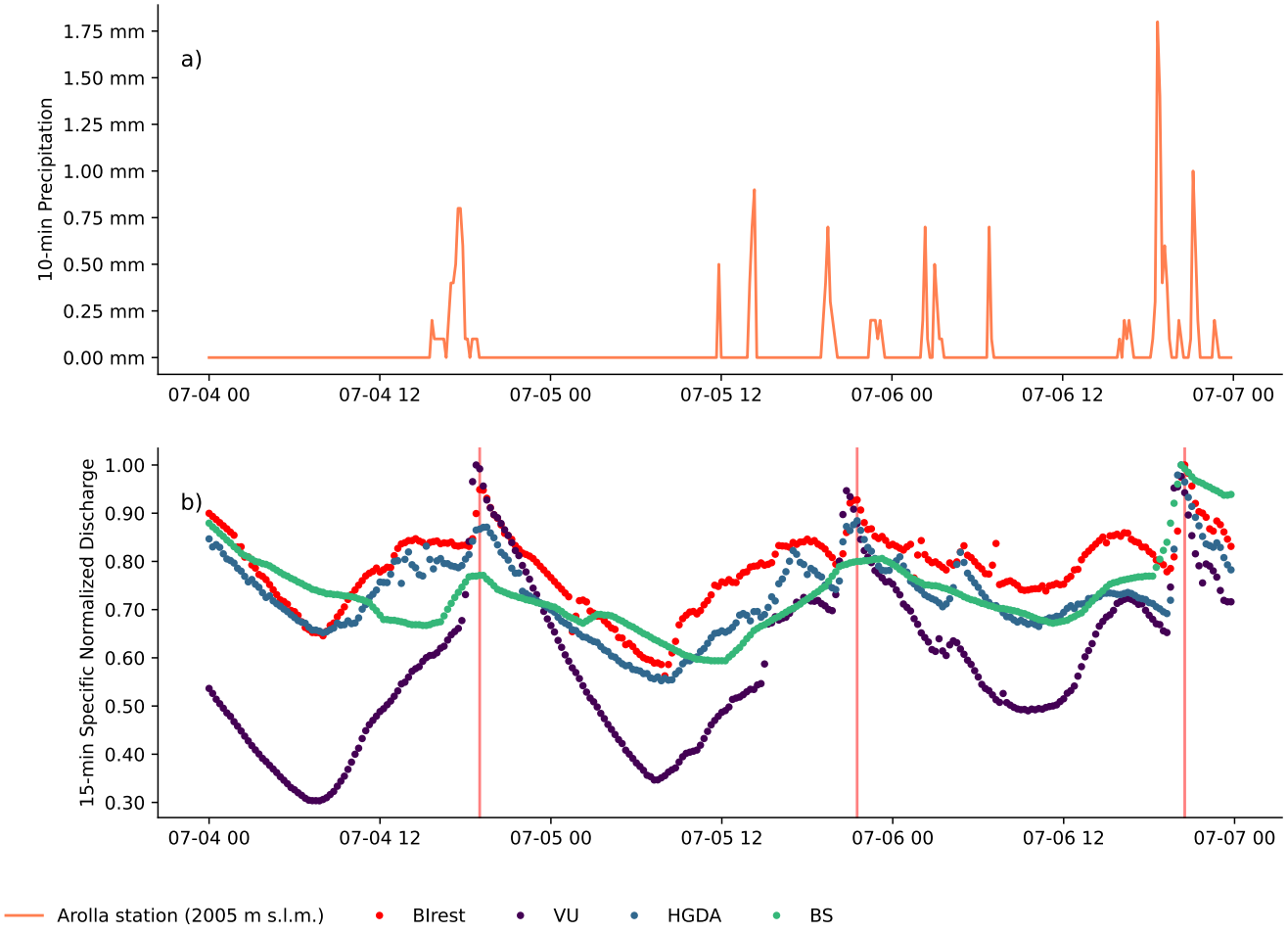

Figure G1. Comparison of (a) the 10-minute precipitation patterns at the Arolla station with (b) the 15-minute normalized specific discharge patterns for the BI intake (called BIrest) and its upstream intakes: VU, HGDA and BS. The vertical lines indicate the time of maximum discharge for the BI intake for each day, and highlight that the water takes around 15-30 minutes to reach the BI intake from the upstream intakes.

~~Comparison of bootstrapped discharge time-series with observed discharge time-series for all catchments. Bootstrapped time-series are obtained by randomly replacing each year's discharge with the discharge observed in another year of the 2010-2014 period in the same catchment.~~

Appendix H: Patterns of water fluxes and retention

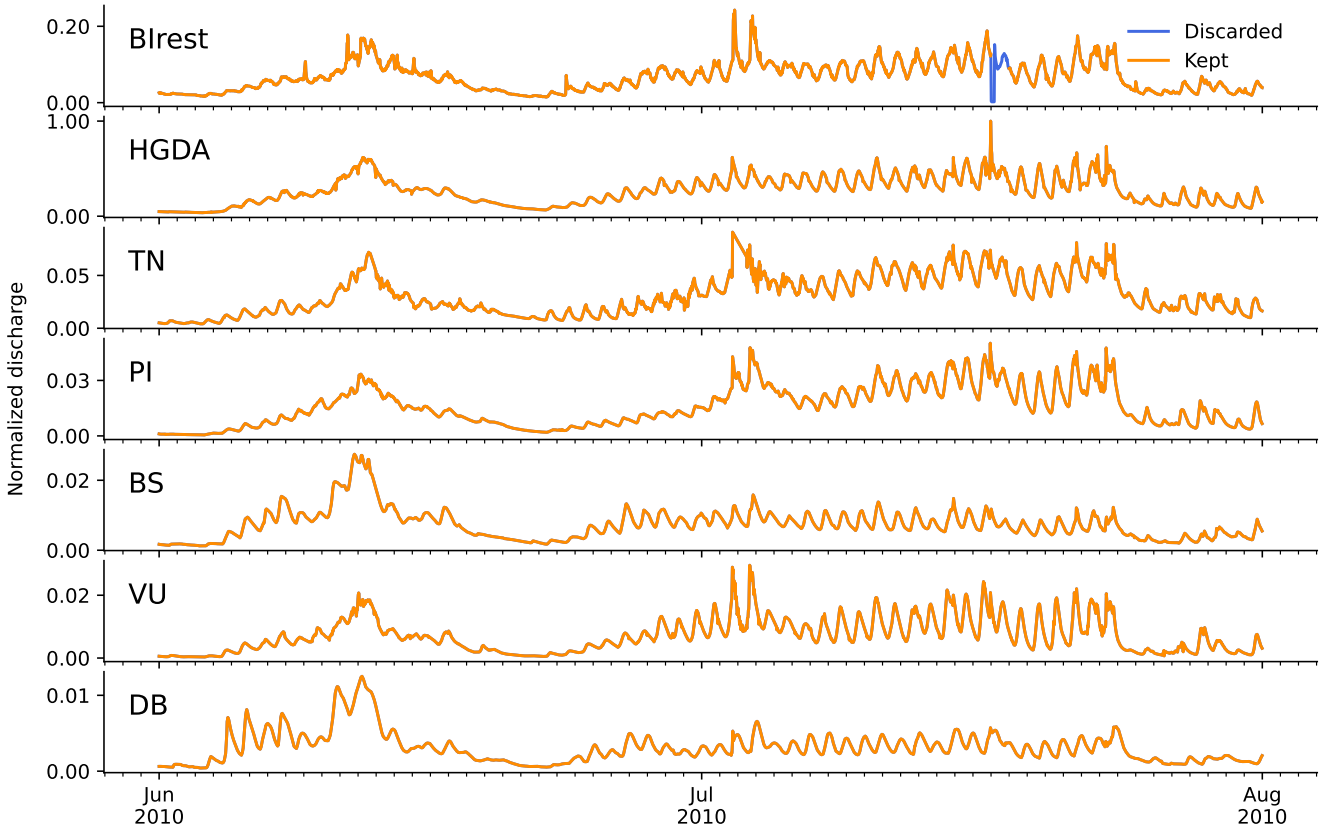

Figure H1. Comparison of the discharge series kept for calibration in Hydrobricks (orange) with the discarded periods (blue), over the summer 2010. Analysis according to Swift et al. (2005) leads to the interpretation that glacial snowpack was removed from mid-June on, allowing diurnal discharge patterns to take on a peaked shape. Discharge (unit: $m^3\ s^{-1}$) is normalized to the highest value.

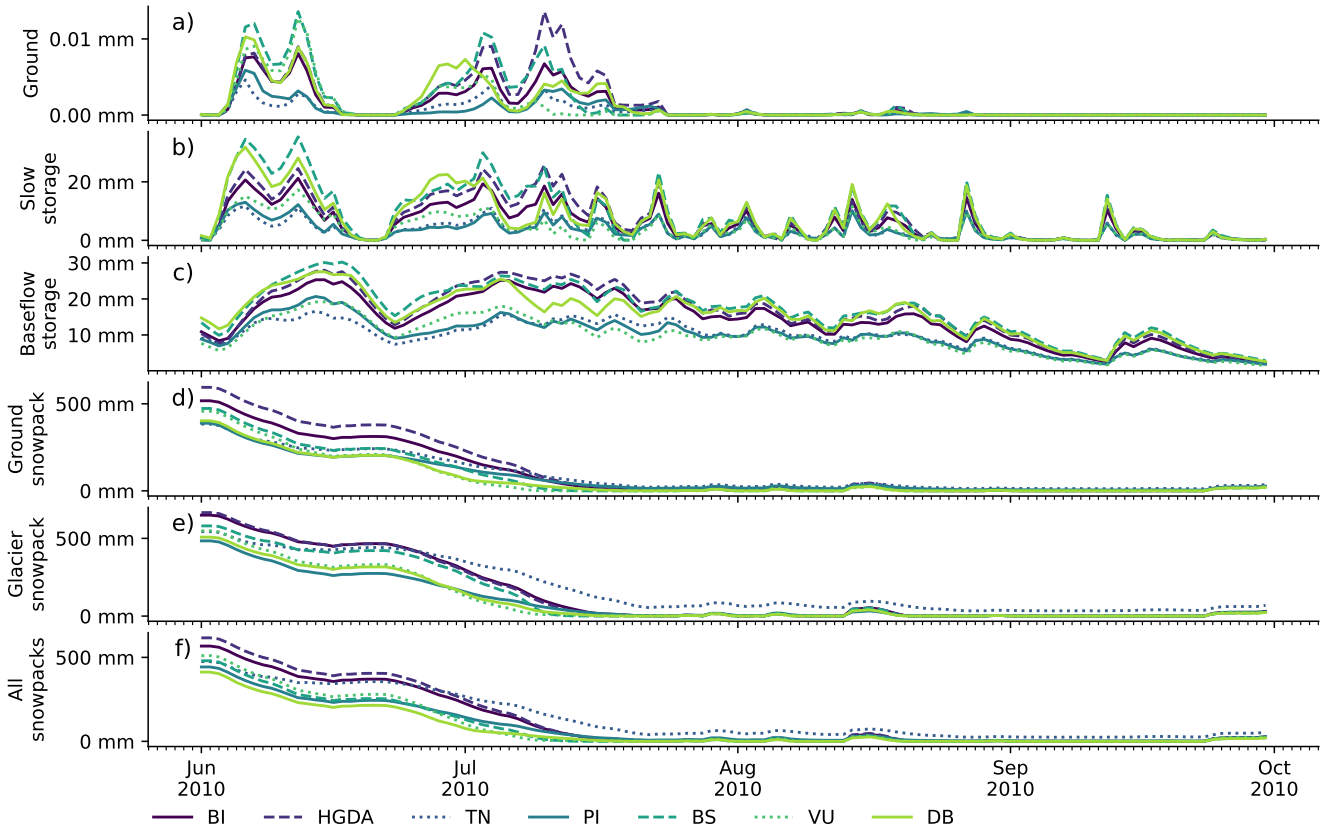

Figure H2. ~~Water~~ Modeled water content heights in the a) ground, b) slow storage and c) baseflow reservoir, and modeled snow water equivalent on the d) ground, e) glacier and f) ground and glacier during the summer 2010. Water heights are computed on their respective areas.

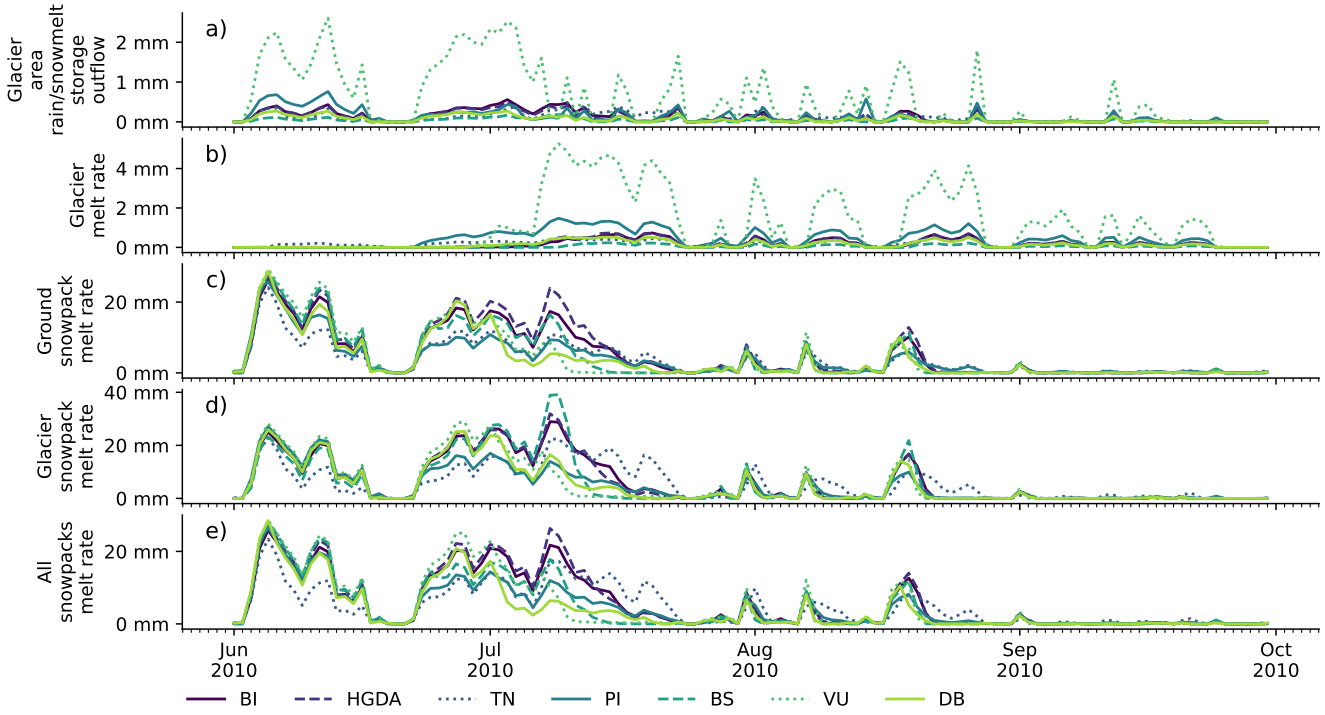

Figure H3. ~~Water~~ Modeled water fluxes rates due to a) outflow of the glacier area rain/snowmelt storage, b) glacier melt rate, c) ground snowpack melt rate, d) glacier snowpack melt rate and e) global snow melt rate during the summer 2010. Melt rates are computed on their respective areas.

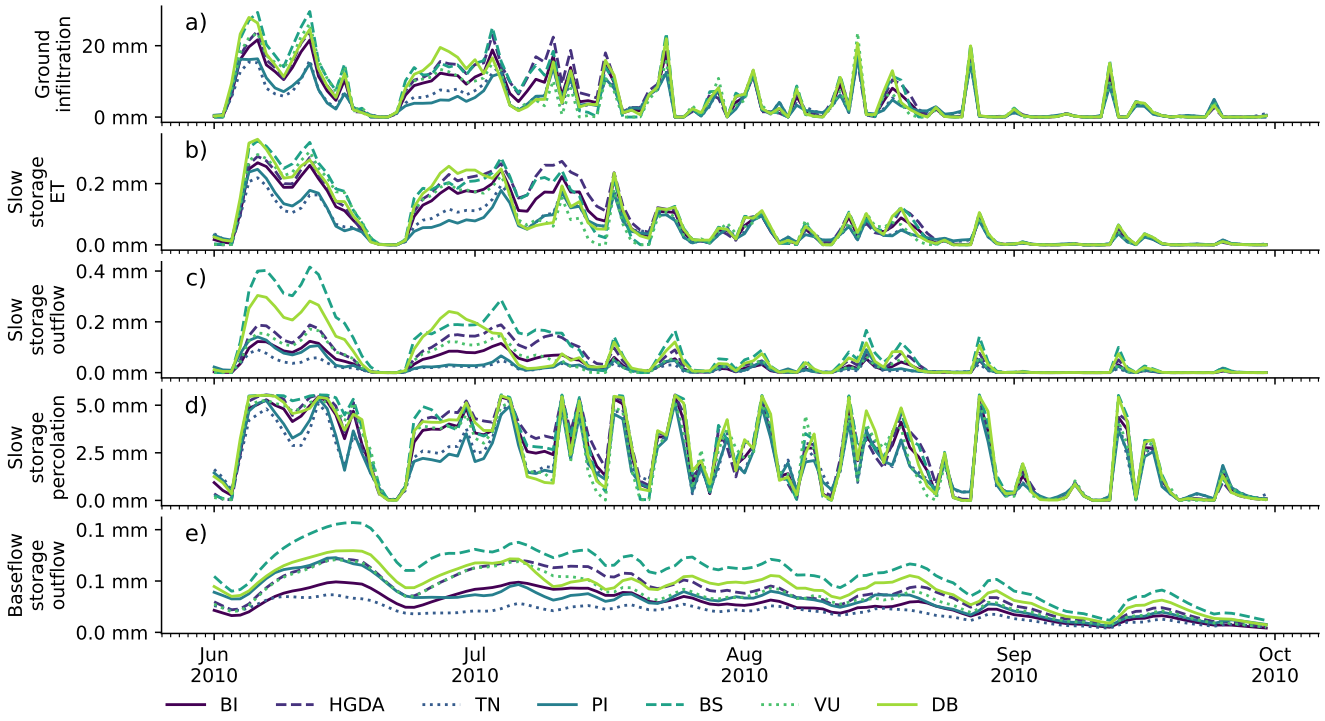

Figure H4. ~~Water~~ Modeled water fluxes rates due to a) ground infiltration into the slow storage, b) evapotranspiration and c) runoff out of the slow storage, d) percolation from the slow storage, into the baseflow storage and e) runoff out of the baseflow storage during the summer 2010. All these rates are computed on the ground areas only.

Appendix I: ~~Aid to understand~~ NSE and KGE ~~behavior~~metrics

I1 Aid to understand NSE and KGE behavior

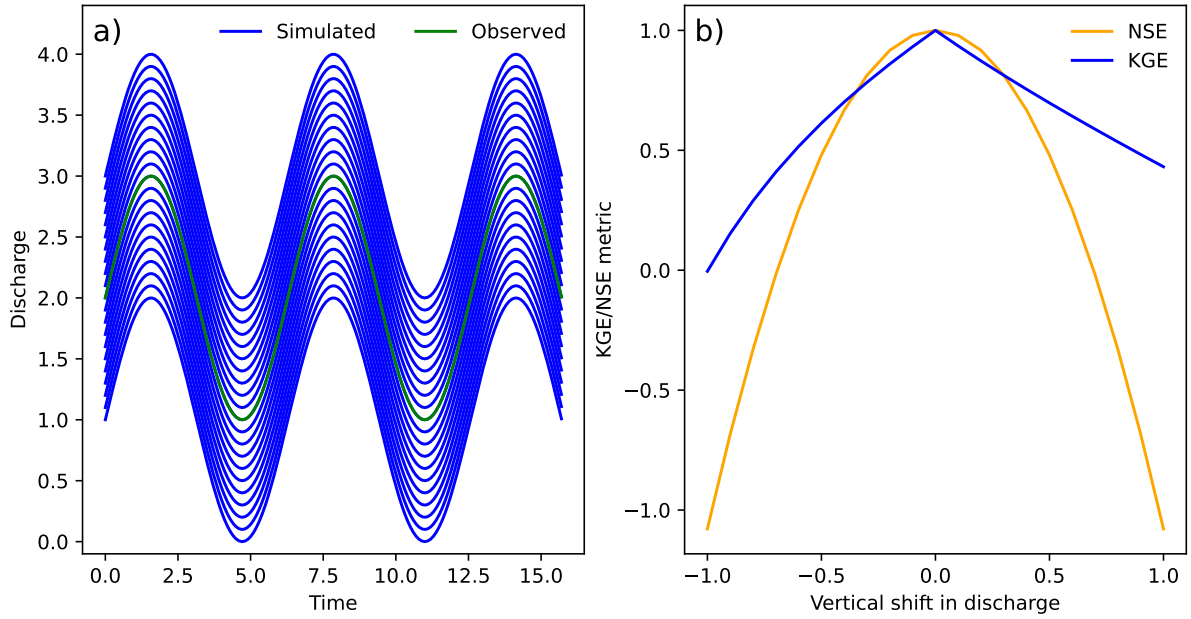

Figure I1. a) Vertical shift between the observed and simulated discharges, and b) the associated changes in NSE and KGE.

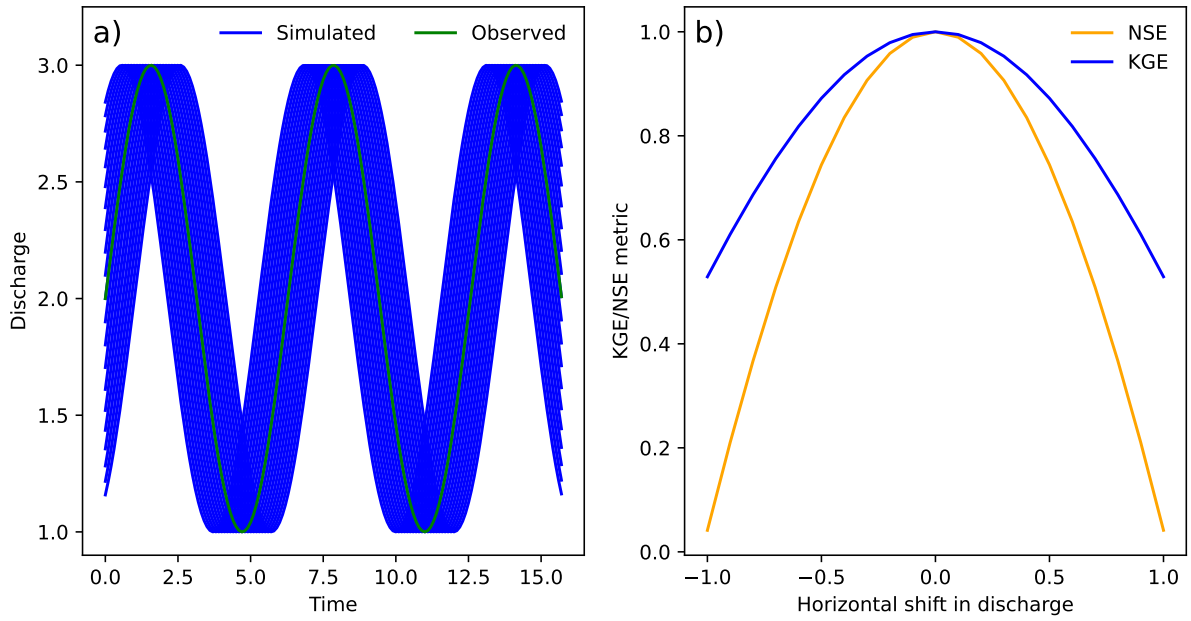

Figure I2. a) Horizontal shift between the observed and simulated discharges, and b) the associated changes in NSE and KGE.

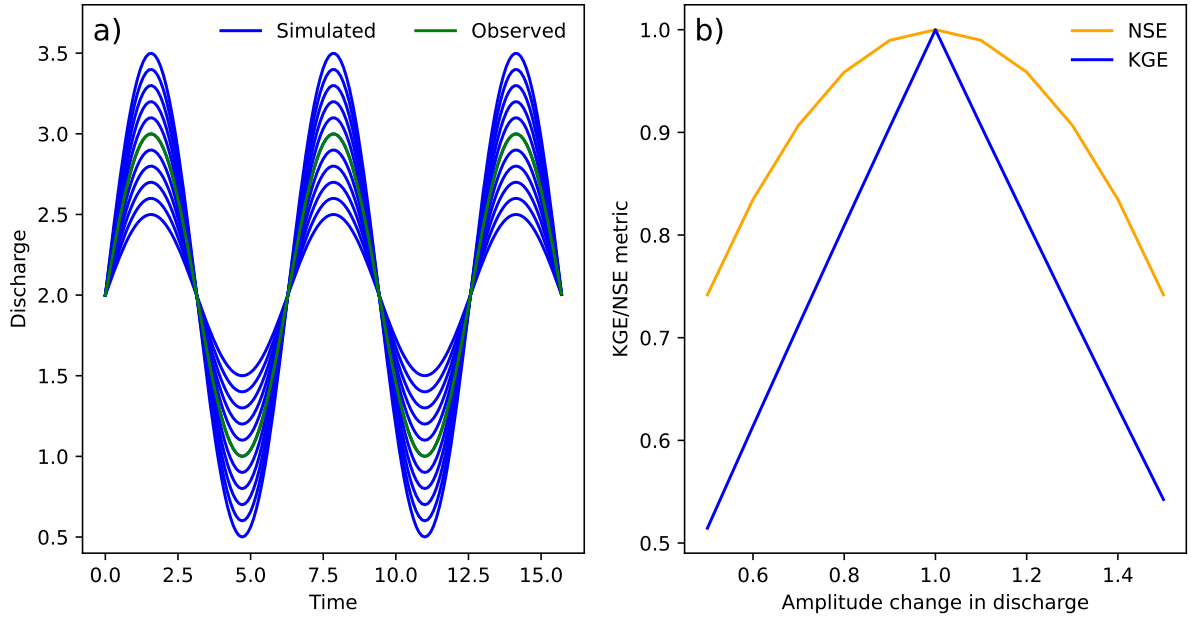

Figure I3. a) Amplitude change between the observed and simulated discharges, and b) the associated changes in NSE and KGE.

I2 KGE vs NSE scoring

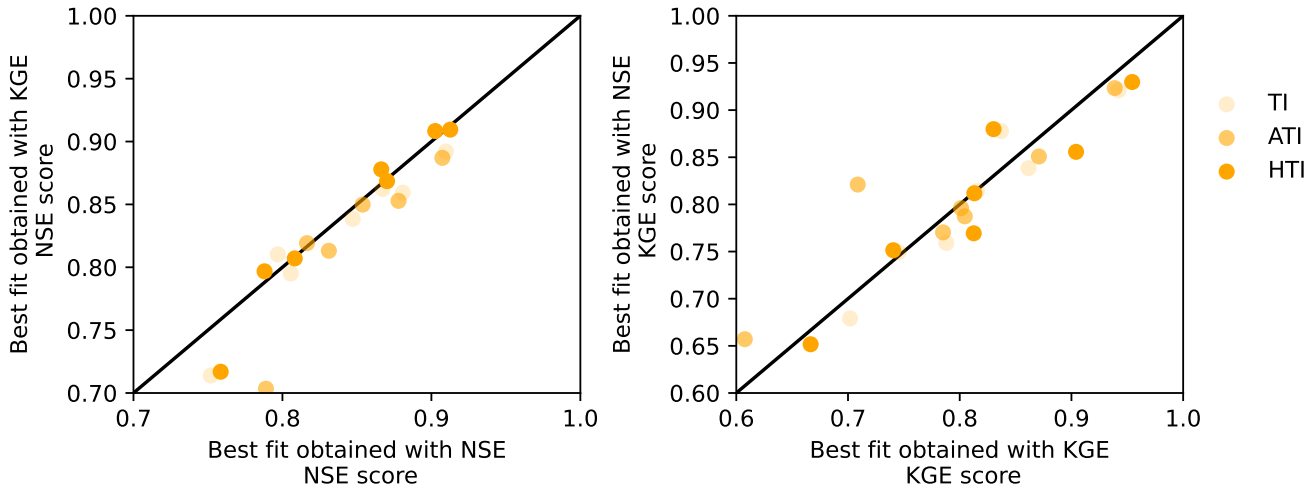

Figure I4. Comparison of the performance of the two NSE and KGE performance criteria in finding the calibrated parameters that are then transferred onto the different catchments. In x the NSE/KGE score when transferred with parameters obtained through NSE/KGE, rep. calibration, and in y the opposite.

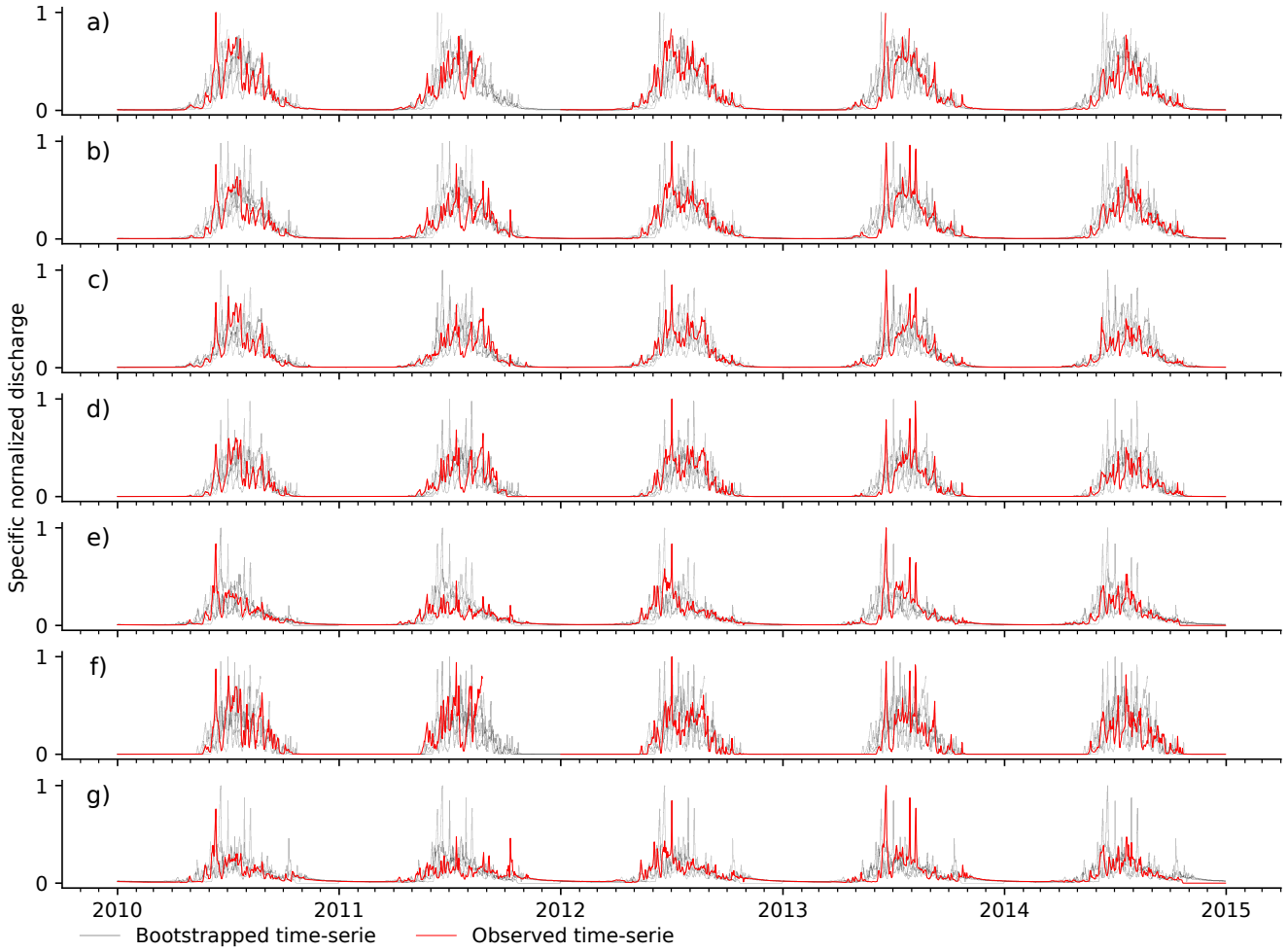

Figure I5. Comparison of resampled discharge time-series with observed discharge time-series for all catchments. Resampled time-series are obtained by exhaustively replacing each year's discharge with the discharge observed during one year of the 2010-2014 period in the same catchment.

I4 Comparison of different benchmark metrics

	NSE		KGE	
Catchment	5-year resample	46-year bootstrap	5-year resample	46-year bootstrap
BI	0.619	0.516	0.801	0.759
HGDA	0.597	0.578	0.794	0.731
TN	0.576	0.531	0.783	0.701
PI	0.502	0.474	0.746	0.689
BS	0.406	0.405	0.697	0.670
VU	0.429	0.372	0.710	0.662
DB	0.220	0.193	0.603	0.422

Table I1. Results of the different benchmarking methods for the NSE and the KGE. The 5-year exhaustive resampling is done by using all possible combinations of the simulated years 2010-2014 (5 years: 3125 combinations). The 46-year bootstrapping is done by randomly selecting 3125 combinations from the 1969-2014 discharge data.

Author contributions. Conceptualization: A.-L.A., S.L., F.C.; Methodology: S.L., P.H., B.S., A.-L.A., J.S.; Software: P.H., B.S., A.-L.A., J.S.; Validation: A.-L.A., P.H.; Formal analysis: A.-L.A., B.S.; Investigation: A.-L.A.; Resources: S.L., P.H., B.S.; Data curation: A.-L.A., P.H.; Writing—original draft preparation: A.-L.A.; Writing—review and editing: A.-L.A., P.H., B.S., J.S., F.P., L.R., M.G., S.B., S.L., F.C.; Visualization: A.-L.A.; Supervision: S.L., F.C., P.H.; Project administration: S.L., F.C., A.-L.A.; Funding acquisition: S.L., F.C.. All authors have read and agreed to the published version of the manuscript.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements. We thank Grande Dixence SA and HydroExploitation for providing the discharge data and the national weather service MeteoSwiss for providing the meteorological time series. This study was carried out in the context of the ALTROCLIMA project, a Joint Project Alto Adige-Switzerland 2022 funded by the Provincia Autonoma di Bolzano – Alto Adige (Autonomous Province of Bolzano-Bozen) and the Swiss National Science Foundation that analyses climate change impacts on bedload transport in the Alps. We acknowledge the use of AI to enhance the clarity of our English phrasing.

References

Abdulla, F. A. and Lettenmaier, D. P.: Development of regional parameter estimation equations for a macroscale hydrologic model, Journal of Hydrology, 197, 230–257, 1997.

Arnold, N.: Investigating the sensitivity of glacier mass-balance/elevation profiles to changing meteorological conditions: Model experiments for Haut Glacier D'Arolla, Valais, Switzerland, Arctic, Antarctic, and Alpine Research, 37, 139–145, https://doi.org/10.1657/1523-0430(2005)037[0139:ITSOGE]2.0.CO;2, 2005.

Bardossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrology and Earth System Sciences, 12, 1273–1283, 2008.

Bezinge, A., Clark, M. J., Gurnell, A. M., and Warburton, J.: The management of sediment transported by glacial melt-water streams and its significance for the estimation of sediment yield, Annals of Glaciology, 13, 1989.

Bradley, R. S., Vuille, M., Diaz, H. F., and Vergara, W.: Threats to Water Supplies in the Tropical Andes, https://doi.org/10.1126/science.1114856, 2006.

Brock, B. W., Willis, I. C., and Sharp, M. J.: Measurement and parameterization of albedo variations at haut glacier d'arolla, Switzerland, Journal of Glaciology, 46, 675–688, https://doi.org/10.3189/172756500781832675, 2000.

Bárdossy, A. and Das, T.: Influence of rainfall observation network on model calibration and application, Hydrol. Earth Syst. Sci, 12, 77–89, www.hydrol-earth-syst-sci.net/12/77/2008/, 2008.

Carenzo, M., Pellicciotti, F., Rimkus, S., and Burlando, P.: Assessing the transferability and robustness of an enhanced temperature-index glacier-melt model, 2009.

Carlstein, E.: The Use of Subseries Values for Estimating the Variance of a General Statistic from a, Source: The Annals of Statistics, 14, 1171–1179, 1986.

Clark, M. P. and Slater, A. G.: Probabilistic Quantitative Precipitation Estimation in Complex Terrain, 2006.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Dan Moore, R., Istanbulluoglu, E., and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, Water Resources Research, 52, 2350–2365, https://doi.org/10.1002/2015wr017910, 2016.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., and Papalexiou, S. M.: The Abuse of Popular Performance Metrics in Hydrologic Modeling, Water Resources Research, 57, https://doi.org/10.1029/2020WR029001, 2021.

Comola, F., Schaefli, B., Ronco, P. D., Botter, G., Bavay, M., Rinaldo, A., and Lehning, M.: Scale-dependent effects of solar radiation patterns on the snow-dominated hydrologic response, Geophysical Research Letters, 42, 3895–3902, https://doi.org/10.1002/2015GL064075, 2015.

Dadic, R., Mott, R., Lehning, M., and Burlando, P.: Wind influence on snow depth distribution and accumulation over glaciers, Journal of Geophysical Research: Earth Surface, 115, https://doi.org/10.1029/2009JF001261, 2010.

Ebtehaj, M., Moradkhani, H., and Gupta, H. V.: Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling, Water Resources Research, 46, https://doi.org/10.1029/2009WR007981, 2010.

Efron, B.: Bootstrap methods: another look at the jackknife, The Annals of Statistics, 7, 1–26, 1979.

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

Engel, M., Penna, D., Bertoldi, G., Vignoli, G., Tirler, W., and Comiti, F.: Controls on spatial and temporal variability in streamflow and hydrochemistry in a glacierized catchment, Hydrology and Earth System Sciences, 23, 2041–2063, https://doi.org/10.5194/hess-23-2041-2019, 2019.

Fatichi, S., Rimkus, S., Burlando, P., Bordoy, R., and Molnar, P.: High-resolution distributed analysis of climate and anthropogenic changes on the hydrology of an Alpine catchment, Journal of Hydrology, 525, 362–382, https://doi.org/10.1016/j.jhydrol.2015.03.036, 2015.

Fernandez, W., Vogel, R. M., and Sankarasubramanian, A.: Calage régional d'un modèle de bassin hydrologique, Hydrological Sciences Journal, 45, 689–707, https://doi.org/10.1080/02626660009492371, 2000.

Finger, D., Pellicciotti, F., Konz, M., Rimkus, S., and Burlando, P.: The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model, Water Resources Research, 47, https://doi.org/10.1029/2010WR009824, 2011.

Fischer, M., Huss, M., Barboux, C., and Hoelzle, M.: The new Swiss Glacier Inventory SGI2010: Relevance of using high-resolution source data in areas dominated by very small glaciers, Arctic, Antarctic, and Alpine Research, 46, 933–945, https://doi.org/10.1657/1938-4246-46.4.933, 2014.

Follum, M. L., Niemann, J. D., and Fassnacht, S. R.: A comparison of snowmelt-derived streamflow from temperature-index and modified-temperature-index snow models, Hydrological Processes, 33, 3030–3045, https://doi.org/10.1002/hyp.13545, 2019.

Frenierre, J. L. and Mark, B. G.: A review of methods for estimating the contribution of glacial meltwater to total watershed discharge, Progress in Physical Geography, 38, 173–200, https://doi.org/10.1177/0309133313516161, 2014.

Gabbi, J., Carenzo, M., Pellicciotti, F., Bauder, A., and Funk, M.: A comparison of empirical and physically based glacier surface melt models for long-term simulations of glacier response, Journal of Glaciology, 60, 1199–1207, https://doi.org/10.3189/2014JoG14J011, 2014.

Gabbud, C., Micheletti, N., and Lane, S. N.: Instruments and methods: Lidar measurement of surface melt for a temperate Alpine glacier at the seasonal and hourly scales, Journal of Glaciology, 61, 963–974, https://doi.org/10.3189/2015JoG14J226, 2015.

Gabbud, C., Micheletti, N., and Lane, S. N.: Response of a temperate alpine valley glacier to climate change at the decadal scale, Geografiska Annaler, Series A: Physical Geography, 98, 81–95, https://doi.org/10.1111/geoa.12124, 2016.

GLAMOS: Swiss Glacier Inventory 2016, release 2020, Glacier Monitoring Switzerland, doi:10.18750/inventory.2016.r2020, 2020.

Grande Dixence SA: Discharge measurements, Personal communications, location: Arolla valley, 2024.

Griessinger, N., Seibert, J., Magnusson, J., and Jonas, T.: Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments, Hydrology and Earth System Sciences, 20, 3895–3905, https://doi.org/10.5194/hess-20-3895-2016, 2016.

Guo, Y., Zhang, Y., Zhang, L., and Wang, Z.: Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review, Wiley Interdisciplinary Reviews: Water, 8, https://doi.org/10.1002/wat2.1487, 2021.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Gyawali, D. R. and Bárdossy, A.: Development and parameter estimation of snowmelt models using spatial snow-cover observations from MODIS, Hydrology and Earth System Sciences, 26, 3055–3077, https://doi.org/10.5194/hess-26-3055-2022, 2022.

Götzinger, J. and Bárdossy, A.: Comparison of four regionalisation methods for a distributed hydrological model, Journal of Hydrology, 333, 374–384, https://doi.org/10.1016/j.jhydrol.2006.09.008, 2007.

Hamon, W. R.: Estimating potential evapotranspiration, Transactions of the American Society of Civil Engineers, 128, 324–338, 1963.

He, Z. H., Parajka, J., Tian, F. Q., and Blöschl, G.: Estimating degree-day factors from MODIS for snowmelt runoff modeling, Hydrology and Earth System Sciences, 18, 4773–4789, https://doi.org/10.5194/hess-18-4773-2014, 2014.

Hingray, B., Schaefli, B., Mezghani, A., and Hamdi, Y.: Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments, Hydrological Sciences Journal, 55, 1002–1016, https://doi.org/10.1080/02626667.2010.505572, 2010.

Hock, R.: A distributed temperature-index ice- and snowmelt model including potential direct solar radiation, Journal of Glaciology, 45, 101–111, https://doi.org/10.3189/s0022143000003087, 1999.

Hock, R.: Temperature index melt modelling in mountain areas, Journal of Hydrology, 282, 104–115, https://doi.org/10.1016/S0022-1694(03)00257-9, 2003.

Horton, P. and Argentin, A.-L.: Hydrobricks: v0.7.2, https://doi.org/10.5281/zenodo.11082505, 2024.

Horton, P., Schaefli, B., Mezghani, A., Hingray, B., and Musy, A.: Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty, Hydrological Processes, 20, 2091–2109, https://doi.org/10.1002/hyp.6197, 2006.

Horton, P., Schaefli, B., and Kauzlaric, M.: Why do we have so many different hydrological models? A review based on the case of Switzerland, Wiley Interdisciplinary Reviews: Water, 9, https://doi.org/10.1002/wat2.1574, 2022.

Houska, T., Kraft, P., Chamorro-Chavez, A., and Breuer, L.: SPOTting model parameters using a ready-made python package, PLoS ONE, 10, https://doi.org/10.1371/journal.pone.0145180, 2015.

Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, Journal of Hydrology, 292, 281–295, https://doi.org/10.1016/j.jhydrol.2004.01.002, 2004.

Hurni, M.: Assimilation of snow depth data from Sentinel-1 to improve the hydrological model of the Borgne catchment (VS), Ph.D. thesis, EPFL, CREALP, 2021.

Huss, M., Farinotti, D., Bauder, A., and Funk, M.: Modelling runoff from highly glacierized alpine drainage basins in a changing climate, Hydrological Processes, 22, 3888–3902, https://doi.org/10.1002/hyp.7055, 2008.

Immerzeel, W. W., Droogers, P., de Jong, S. M., and Bierkens, M. F.: Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing, Remote Sensing of Environment, 113, 40–49, https://doi.org/10.1016/j.rse.2008.08.010, 2009.

Ismail, M. F., Bogacki, W., Disse, M., Schäfer, M., and Kirschbauer, L.: Estimating degree-day factors of snow based on energy flux components, Cryosphere, 17, 211–231, https://doi.org/10.5194/tc-17-211-2023, 2023.

Jansson, P., Hock, R., and Schneider, T.: The concept of glacier storage: A review, Journal of Hydrology, 282, 116–129, https://doi.org/10.1016/S0022-1694(03)00258-0, 2003.

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

Kling, H. and Gupta, H.: On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, Journal of Hydrology, 373, 337–351, https://doi.org/10.1016/j.jhydrol.2009.04.031, 2009.

Knoben, W. J., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019, 2019.

Koboltschnig, G. R., Schöner, W., Zappa, M., Kroisleitner, C., and Holzmann, H.: Runoff modelling of the glacierized Alpine Upper Salzach basin (Austria): Multi-criteria result validation, Hydrological Processes, 22, 3950–3964, https://doi.org/10.1002/hyp.7112, 2008.

Konz, M. and Seibert, J.: On the value of glacier mass balances for hydrological model calibration, Journal of Hydrology, 385, 238–246, https://doi.org/10.1016/j.jhydrol.2010.02.025, 2010.

Künsch, H. R.: The Jackknife and the Bootstrap for General Stationary Observations, Source: The Annals of Statistics, 17, 1217–1241, 1989.

Lambiel, C., Maillard, B., Kummert, M., and Reynard, E.: Geomorphology of the Hérens valley (Swiss Alps), Journal of Maps, 12, 160–172, https://doi.org/10.1080/17445647.2014.999135, 2016.

Lane, S. N. and Nienow, P. W.: Decadal-Scale Climate Forcing of Alpine Glacial Hydrological Systems, Water Resources Research, 55, 2478–2492, https://doi.org/10.1029/2018WR024206, 2019.

Lane, S. N., Bakker, M., Gabbud, C., Micheletti, N., and Saugy, J. N.: Sediment export, transient landscape response and catchment-scale connectivity following rapid climate warming and Alpine glacier recession, Geomorphology, 277, 210–227, https://doi.org/10.1016/j.geomorph.2016.02.015, 2017.

Liang, X., Guo, J., and Leung, L. R.: Assessment of the effects of spatial resolutions on daily water flux simulations, Journal of Hydrology, 298, 287–310, https://doi.org/10.1016/j.jhydrol.2003.07.007, 2004.

Linsbauer, A., Huss, M., Hodel, E., Bauder, A., Fischer, M., Weidmann, Y., Bärtschi, H., and Schmassmann, E.: The New Swiss Glacier Inventory SGI2016: From a Topographical to a Glaciological Dataset, Frontiers in Earth Science, 9, https://doi.org/10.3389/feart.2021.704189, 2021.

Mair, D., Nienow, P., Sharp, M., Wohlleben, T., and Willis, I.: Influence of subglacial drainage system evolution on glacier surface motion: Haut Glacier d'Arolla, Switzerland, Journal of Geophysical Research: Solid Earth, 107, https://doi.org/10.1029/2001jb000514, 2002.

Mair, D., Willis, I., Fischer, U. H., Hubbard, B., Nienow, P., and Hubbard, A.: Hydrological controls on patterns of surface, internal and basal motion during three "spring events": Haut Glacier d' Arolla, Switzerland, Journal of Glaciology, 49, 2003.

Martinec, J.: The degree-day factor for snowmelt-runoff forecasting, in: IAHS Publication, in: IUGG General Assembly of Helsinki, Helsinki, Finland, vol. 51, pp. 468–477, 1960.

MeteoSwiss: Documentation of MeteoSwiss Grid-Data Products: Daily Precipitation (final analysis): RhiresD, Tech. rep., MeteoSwiss, 2019a.

MeteoSwiss: Documentation of MeteoSwiss Grid-Data Products: Daily Precipitation (final analysis): TabsD, Tech. rep., MeteoSwiss, 2019b.

Michelon, A., Benoit, L., Beria, H., Ceperley, N., and Schaefli, B.: Benefits from high-density rain gauge observations for hydrological response analysis in a small alpine catchment, Hydrology and Earth System Sciences, 25, 2301–2325, https://doi.org/10.5194/hess-25-2301-2021, 2021.

Mosley, M.: Delimitation of New Zealand hydrologic regions, Journal of Hydrology, 49, 173–192, 1981.

Nasab, M. T. and Chu, X.: Do sub-daily temperature fluctuations around the freezing temperature alter macro-scale snowmelt simulations?, Journal of Hydrology, 596, https://doi.org/10.1016/j.jhydrol.2020.125683, 2021.

Nash, J. E. and Sutcliffe, J. V.: River Flow Forecasting Through Conceptual Models Part I - A Discussion of Principles, Journal of Hydrology, 10, 282–290, 1970.

Ohmura, A.: Physical basis for the temperature-based melt-index method, Journal of Applied Meteorology, 40, 753–761, https://doi.org/10.1175/1520-0450(2001)040<0753:PBFTTB>2.0.CO;2, 2001.

Parajka, J. and Blöschl, G.: Validation of MODIS snow cover images over Austria, Hydrol. Earth Syst. Sci, 10, 679–689, www.hydrol-earth-syst-sci.net/10/679/2006/, 2006.

Parajka, J. and Blöschl, G.: The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, Journal of Hydrology, 358, 240–258, https://doi.org/10.1016/j.jhydrol.2008.06.006, 2008.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrology and Earth System Sciences, 9, 157–171, https://doi.org/10.5194/hess-9-157-2005, 2005.

Pellicciotti, F., Brock, B., Strasser, U., Burlando, P., Funk, M., and Corripio, J.: An enhanced temperature-index glacier melt model including the shortwave radiation balance: Development and testing for Haut Glacier d'Arolla, Switzerland, Journal of Glaciology, 51, 573–587, https://doi.org/10.3189/172756505781829124, 2005.

Penna, D., Engel, M., Bertoldi, G., and Comiti, F.: Towards a tracer-based conceptualization of meltwater dynamics and streamflow response in a glacierized catchment, Hydrology and Earth System Sciences, 21, 23–41, https://doi.org/10.5194/hess-21-23-2017, 2017.

Pokhrel, P., Gupta, H. V., and Wagener, T.: A spatial regularization approach to parameter estimation for a distributed watershed model, Water Resources Research, 44, https://doi.org/10.1029/2007WR006615, 2008.

Rango, A. and Martinec, J.: Revisiting the degree-day method for snowmelt computations, Water Resources Bulletin, 31, 1995.

Rinaldo, A., Marani, A., and Rigon, R.: Geomorphological dispersion, Water Resources Research, 27, 513–525, https://doi.org/10.1029/90WR02501, 1991.

Rinaldo, A., Botter, G., Bertuzzo, E., Uccelli, A., Settin, T., and Marani, M.: Hydrology and Earth System Sciences Transport at basin scales: 1. Theoretical framework, Hydrology and Earth System Sciences, 10, 19–29, www.copernicus.org/EGU/hess/hess/10/19/, 2006.

Ritter, A. and Muñoz-Carpena, R.: Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments, Journal of Hydrology, 480, 33–45, https://doi.org/10.1016/j.jhydrol.2012.12.004, 2013.

Ruelland, D.: Potential of snow data to improve the consistency and robustness of a semi-distributed hydrological model using the SAFRAN input dataset, Journal of Hydrology, 631, https://doi.org/10.1016/j.jhydrol.2024.130820, 2024.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, https://doi.org/10.1029/2008WR007327, 2010.

Schaefli, B. and Gupta, H. V.: Do Nash values have value?, Hydrological Processes, 21, 2075–2080, https://doi.org/10.1002/hyp.6825, 2007.

Schaefli, B. and Huss, M.: Integrating point glacier mass balance observations into hydrologic model identification, Hydrology and Earth System Sciences, 15, 1227–1241, https://doi.org/10.5194/hess-15-1227-2011, 2011.

Schaefli, B., Hingray, B., Niggli, M., and Musy, A.: A conceptual glacio-hydrological model for high mountainous catchments, Hydrology and Earth System Sciences, 9, 95–109, www.copernicus.org/EGU/hess/hess/9/95/SRef-ID:1607-7938/hess/2005-9-95EuropeanGeosciencesUnion, 2005.

Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, Agricultural and Forest Meteorology, 98-99, 279–293, 1999.

Sharp, M., Richards, K., Willis, I., Arnold, N., Nienow, P., Lawson, W., and Tison, J.: Geometry, bed topography and drainage system structure of the haut glacier d'Arolla, Switzerland, Earth Surface Processes and Landforms, 18, 557–571, https://doi.org/10.1002/esp.3290180608, 1993.

Shokory, J. A. N. and Lane, S. N.: Patterns and drivers of glacier debris-cover development in the Afghanistan Hindu Kush Himalaya, Journal of Glaciology, 69, 1260–1274, https://doi.org/10.1017/jog.2023.14, 2023.

Shokory, J. A. N., Horton, P., Schaefli, B., and Lane, S. N.: Glacier-influenced hydrological regimes of Afghanistan Hindu Kush Himalaya (AHKH) under current and future climate, Ph.D. thesis, UNIL, 2023.

Singh, P. and Kumar, N.: Impact assessment of climate change on the hydrological response of a snow and glacier melt runoff dominated Himalayan river, 1997.

Sivapalan, M., Blöschl, G., Zhang, L., and Vertessy, R.: Downward approach to hydrological prediction, Hydrological Processes, 17, 2101–2111, https://doi.org/10.1002/hyp.1425, 2003.

Srinivas, V. V. and Srinivasan, K.: Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows, Journal of Hydrology, 302, 307–330, https://doi.org/10.1016/j.jhydrol.2004.07.011, 2005.

Swift, D. A., Nienow, P. W., Spedding, N., and Hoey, T. B.: Geomorphic implications of subglacial drainage configuration: rates of basal sediment evacuation controlled by seasonal drainage system evolution, Sedimentary Geology, 149, 5–19, www.elsevier.com/locate/sedgeo, 2002.

Swift, D. A., Nienow, P. W., Hoey, T. B., and Mair, D. W.: Seasonal evolution of runoff from Haut Glacier d'Arolla, Switzerland and implications for glacial geomorphic processes, Journal of Hydrology, 309, 133–148, https://doi.org/10.1016/j.jhydrol.2004.11.016, 2005.

SwissTopo: SwissTopo - Geological vector datasets (GeoCover), SwissTopo - Swiss Federal Office of Topography, https://www.swisstopo.admin.ch/en/geological-model-2d-geocover, version 2., 2024.

Swisstopo: The digital height model of Switzerland, accessed 2023.

Tague, C. L., Papuga, S. A., Gerlein-Safdi, C., Dymond, S., Morrison, R. R., Boyer, E. W., Riveros-Iregui, D., Agee, E., Arora, B., Dialynas, Y. G., Hansen, A., Krause, S., Kuppel, S., Loheide, S. P., Schymanski, S. J., and Zipper, S. C.: Adding our leaves: A community-wide perspective on research directions in ecohydrology, Hydrological Processes, 34, 1665–1673, https://doi.org/10.1002/hyp.13693, 2020.

Thibert, E., Sielenou, P. D., Vionnet, V., Eckert, N., and Vincent, C.: Causes of Glacier Melt Extremes in the Alps Since 1949, Geophysical Research Letters, 45, 817–825, https://doi.org/10.1002/2017GL076333, 2018.

Tiwari, D., Trudel, M., and Leconte, R.: On optimization of calibrations of a distributed hydrological model with spatially distributed information on snow, Hydrology and Earth System Sciences, 28, 1127–1146, https://doi.org/10.5194/hess-28-1127-2024, 2024.

Tobias, W., Manfred, S., Klaus, J., Zappa, M., and Bettina, S.: The future of Alpine Run-of-River hydropower production: Climate change, environmental flow requirements, and technical production potential, Science of the Total Environment, 890, https://doi.org/10.1016/j.scitotenv.2023.163934, 2023.

Tobin, C., Schaefli, B., Nicótina, L., Simoni, S., Barrenetxea, G., Smith, R., Parlange, M., and Rinaldo, A.: Improving the degree-day method for sub-daily melt simulations with physically-based diurnal variations, Advances in Water Resources, 55, 149–164, https://doi.org/10.1016/j.advwatres.2012.08.008, 2013.

Troy, T. J., Wood, E. F., and Sheffield, J.: An efficient calibration method for continental-scale land surface modeling, Water Resources Research, 44, https://doi.org/10.1029/2007WR006513, 2008.

van Tiel, M., Stahl, K., Freudiger, D., and Seibert, J.: Glacio-hydrological model calibration and evaluation, Wiley Interdisciplinary Reviews: Water, 7, https://doi.org/10.1002/wat2.1483, 2020.

Vogel, R. M. and Shallcross, A. L.: The moving blocks bootstrap versus parametric time series models, Water Resources Research, 32, 1875–1882, https://doi.org/10.1029/96WR00928, 1996.

Wagener, T. and Wheater, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, Journal of Hydrology, 320, 132–154, https://doi.org/10.1016/j.jhydrol.2005.07.015, 2006.

Zbinden, P., Gehrig, R., Herren, T., Bader, S., Gaillard, M. G., Umbricht, A., and Wetterdienst: 147. Jahrgang 2010 Annalen Annales Annali, https://www.meteosuisse.admin.ch/assets/weather-archive/annalen-2010.pdf, Accessed: 04.06.2024, 2010.

Zuecco, G., Carturan, L., Blasi, F. D., Seppi, R., Zanoner, T., Penna, D., Borga, M., Carton, A., and Fontana, G. D.: Understanding hydrological processes in glacierized catchments: Evidence and implications of highly variable isotopic and electrical conductivity data, Hydrological Processes, 33, 816–832, https://doi.org/10.1002/hyp.13366, 2019.

Şorman, A. A., Şensoy, A., Tekeli, A. E., Şorman, A. , and Akyürek, Z.: Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey, Hydrological Processes, 23, 1031–1040, https://doi.org/10.1002/hyp.7204, 2009.