He, Z., Vorogushyn, S., Unger-Shayesteh, K., Gafurov, A., Kalashnikova, O., Omorova, E., Merz, B. (2018): The Value of Hydrograph Partitioning Curves for Calibrating Hydrological Models in Glacierized Basins. - *Water Resources Research*, *54*, 3, pp. 2336—2361.

DOI: http://doi.org/10.1002/2017WR021966

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2017WR021966

The Value of Hydrograph Partitioning Curves for Calibrating Hydrological Models in Glacierized Basins

Zhihua He[1], Sergiy Vorogushyn[1], Katy Unger-Shayesteh[1], Abror Gafurov[1], Olga Kalashnikova[3], Elvira Omorova[4], and Bruno Merz[1,2]

[1]GFZ German Research Centre for Geosciences, Section 5.4: Hydrology, Potsdam, Germany, [2]Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany, [3]Department Climate, Water and Natural Resources, CAIAG Central Asian Institute of Applied Geosciences, Bishkek, Kyrgyzstan, [4]Department of Hydrological Forecasting, Kyrgyz Hydromet Service, Bishkek, Kyrgyzstan

Key Points:
- The HPC-based method (1) delivers model-internal consistency comparably to the tri-data set calibration method
- The HPC-based method improves the stability of calibrated parameter values across various calibration periods
- The HPC-based method estimates the contributions of runoff components similarly to the tri-data set calibration method

Correspondence to:
Z. He,
zhihuahe@gfz-potsdam.de

Citation:
He, Z., Vorogushyn, S., Unger-Shayesteh, K., Gafurov, A., Kalashnikova, O., Omorova, E., et al. (2018). The value of hydrograph partitioning curves for calibrating hydrological models in glacierized basins. *Water Resources Research*, *54*, 2336–2361. https://doi.org/10.1002/2017WR021966

Received 29 SEP 2017
Accepted 17 FEB 2018
Accepted article online 22 FEB 2018
Published online 30 MAR 2018

Abstract This study refines the method for calibrating a glacio-hydrological model based on Hydrograph Partitioning Curves (HPCs), and evaluates its value in comparison to multidata set optimization approaches which use glacier mass balance, satellite snow cover images, and discharge. The HPCs are extracted from the observed flow hydrograph using catchment precipitation and temperature gradients. They indicate the periods when the various runoff processes, such as glacier melt or snow melt, dominate the basin hydrograph. The annual cumulative curve of the difference between average daily temperature and melt threshold temperature over the basin, as well as the annual cumulative curve of average daily snowfall on the glacierized areas are used to identify the starting and end dates of snow and glacier ablation periods. Model parameters characterizing different runoff processes are calibrated on different HPCs in a stepwise and iterative way. Results show that the HPC-based method (1) delivers model-internal consistency comparably to the tri-data set calibration method; (2) improves the stability of calibrated parameter values across various calibration periods; and (3) estimates the contributions of runoff components similarly to the tri-data set calibration method. Our findings indicate the potential of the HPC-based approach as an alternative for hydrological model calibration in glacierized basins where other calibration data sets than discharge are often not available or very costly to obtain.

1. Introduction

1.1. Calibration of Hydrological Models in Glacierized Basins

It is generally accepted that single-data set calibration solely based on discharge observations is insufficient to guarantee model-internal consistency (Konz & Seibert, 2010; Le Lay et al., 2008; Madsen, 2000; Seibert & McDonnell, 2002). Strong intercompensation between runoff generation processes can result in a good simulation of discharge at the basin outlet (Tarasova et al., 2016). Multidata set calibration using discharge and additional measurements of internal catchment variables are therefore required to improve internal process identification and model parameter identifiability (Efstratiadis & Koutsoyiannis, 2010; Parajka et al., 2009; Seibert, 2000).

In glacierized and snow-dominated basins, snow and glacier melt provide important sources for river runoff. Hence, constraining the glacier and snow melt dynamics in hydrological modeling is essential for preventing internal process compensation. For example, multiple snow cover area (SCA) images from MODIS and AVHRR sensors were used by Parajka and Blöschl (2008a) and Duethmann et al. (2014), respectively, in addition to streamflow to improve the model performance for both simulations of SCA and discharge. Konz and Seibert (2010) highlighted the value of annual glacier mass balance (GMB) data for identifying model parameters. The particular importance of seasonal GMB for constraining melt parameters and achieving high model-internal consistency was pointed out by Schaefli and Huss (2011) and Mayr et al. (2013).

In spite of the fact that hydrological model performance is typically improved when including additional observations, the practical application of multidata set calibration is often restricted by the lack of observational data, especially in glacierized basins. GMB measurements are extremely laborious and are carried out usually on a few glaciers only within vast mountain regions. Although various methods for regionalization

of point measurements of glacier mass balances have been developed (e.g., Huss, 2012), the spatial variability of mass balances and related uncertainty remain high (Farinotti et al., 2015). Modeled GMB data based on differencing of multi-temporal digital elevation models (e.g., Aizen et al., 2007; Shangguan et al., 2015) or laser altimetry (Kääb et al., 2012) can provide spatially resolved assessments; however, they are severely limited in temporal resolution delivering the integral mass change over long time spans, typically over years and decades.

Although processed satellite SCA products become increasingly available in mountain regions for past periods (Peters et al., 2015; Zhou et al., 2013), their applicability is heavily impacted by cloud cover. Some cloud removal algorithms have been developed for the processing of operational satellite data (e.g., Gafurov et al., 2016; Parajka & Blöschl, 2008b), but it is typically time consuming to run the cloud removal algorithms to obtain daily time series of SCA data for model calibration. Moreover, hydrological models typically reproduce the snow water equivalent (SWE) in the basin. Observed SWE data are rather rare in high mountain basins (Hamlet et al., 2005), and thus transforming approaches from simulated SWE to simulated SCA are often required for the calibration against satellite SCA (Duethmann et al., 2014; Parajka & Blöschl, 2008a). These transforming approaches however result in an additional uncertainty in the model evaluation.

In contrast, sophisticated calibration approaches making use of additional information extracted from only discharge data bear a potential to improve model-internal consistency and reduce parameter uncertainty. Discharge observations represent integrated information of all runoff processes in a catchment (Guse et al., 2016; Nearing & Gupta, 2015; Reusser et al., 2009), and are the most commonly available observational constraint for parameter calibration in hydrological modeling. Previously proposed hydrograph signatures provide a perspective to extract hydrological information from a discharge time series (Richter et al., 1996; Sivapalan et al., 2003). Evaluating model performance against hydrograph signatures offers diagnostic insights into the catchment behavior (Gupta et al., 2008; Hrachowitz et al., 2013), and is supposed to increase model consistency (Fovet et al., 2015). For example, Shafii and Tolson (2015) proposed a framework based on a scoring method transforming 13 hydrological signatures to a calibration objective function to improve model consistency. Coxon et al. (2014) improved model performance using diagnostic signatures, including interannual and intraannual runoff variabilities. Other signatures, including regime curve, base flow index, and seasonal water balance, were used to indicate how specific system functions have been represented in the model (Hingray et al., 2010; Jothityangkoon et al., 2001; Liucci et al., 2014; Luo et al., 2012; Ye et al., 2012). However, it is still not clear how hydrograph signatures perform in constraining the simulations of multiple internal hydrological variables other than discharge, such as snow and glacier coverage patterns, in glacierized basins. The extent to which the calibration methods based on extracted information from discharge can be comparable to multidata set calibration methods has not been examined to the knowledge of the authors.

1.2. Hydrograph Partitioning Curves

Given the temporal variations of runoff generation processes, different parts of the catchment hydrograph can be dominated by different hydrological processes (Pfannerstill et al., 2014). Correspondingly, different hydrograph parts possess varying power for the identification of specific model parameters. This is particularly relevant in glacierized basins exhibiting a distinct hydrological process seasonality (Razavi & Tolson, 2013). For example, summer discharge tends to be less important to identify parameters describing snow accumulation (Singh & Bardossy, 2012), while winter discharge tends to be irrelevant for parameters characterizing glacier melt rate. Relating different hydrograph features to corresponding sensitive model parameters can help to increase the robustness in parameter values (Nearing & Gupta, 2015).

Several previous studies exploited the temporal variation in parameter sensitivity for model calibration, including Hingray et al. (2010), who progressively calibrated model parameters on the rising and falling limbs of the hydrograph, as well as Schaefli et al. (2005), who estimated different parameters from different hydrograph signatures in a stepwise way. Other studies applied a stepwise calibration for different subsets of parameters to different subperiods of flow time series, by splitting the flow series into dry/wet periods or high/low flows (Boyle et al., 2000; Hay et al., 2006; Willems, 2014). However, few of these calibrations exploited the relation between hydrograph signatures and model parameters using the dominance of runoff processes. Along this line, He et al. (2014, 2015) used temperature and precipitation time series to partition the flow hydrograph into several segments corresponding to the dominant runoff processes in

different seasons. They proposed the hydrograph partitioning curves (HPC) for model calibration in glacierized basins. In the HPC-based calibration, model parameters were related to individual hydrograph partitions according to their controls on the simulations of runoff processes, and calibrated in a stepwise way (He et al., 2015). However, their HPC extraction procedure needs to be refined to distinguish summer snowmelt in glacier-covered areas from glacier melt. The value of HPCs for improving model-internal consistency and parameter stability in time has not been investigated in relation to the multidata set calibration approach.

1.3. Objectives of the Paper

Starting from these motivations, we further refine the approach in He et al. (2015) by extracting HPCs using the annual cumulative curve of the difference between average daily temperature and melt threshold temperature (ADT) over the basin, as well the annual cumulative curve of average daily snowfall (ASF) on the glacierized areas to identify the starting and end dates of snow and glacier ablation periods. The objective of this paper is to investigate the value of the extracted hydrological information represented by HPCs from discharge observation for constraining the simulations of snow and glacier coverage patterns in glacierized basins. We contrast the performance of the HPC-based method with multidata set calibration, which uses different combinations of discharge series, annual GMB data, and SCA images as constraining observations. Specific research questions are: (1) Can the HPC-based method achieve model-internal consistency comparably to multidata set calibration methods, without using additional calibration data sets besides discharge? (2) Does the HPC-based method improve the stability of model parameters in time compared to the multidata set calibration approaches? (3) Does the HPC-based method compute contributions of runoff components to discharge comparably to the multidata set methods? We expect our results to shed light on model calibration and identification of runoff components in glacierized basins where discharge is the only available observation for model calibration.

The paper is organized as follows: a brief description of the study area and available data is provided in section 2. In section 3, we describe the used hydrological model and the extraction of HPCs. Section 4 presents the simulation performance of various calibration methods and corresponding parameter uncertainty. The limitations of the method are discussed in section 5 followed by conclusions.

2. Study Area and Data

The Ala-Archa basin is located in Kyrgyzstan, Central Asia, near the capital city Bishkek (Figure 1). The basin area is 233 km^2, approximately 17% of which is glacierized. Downstream irrigated agriculture and potable water supply for the capital are heavily dependent on the freshwater availability in the Ala-Archa river. Basin elevation ranges from 1,560 to 4,864 m above sea level (a.s.l.), with mean elevation of about 3,300 m a.s.l. The long-term annual mean precipitation and temperature during 1970–2000 measured at the Alplager station are 560 mm and 2.6°C, respectively. Precipitation reaches its peak in spring-summer months and mainly occurs as snowfall in winter (Aizen et al., 2000, 2007). Snow and glacier melt strongly affect river discharge during March–September.

Daily precipitation and temperature as well as humidity and radiation time series are available since the 1970s at Baitik (1,580 m, a.s.l.) and Alplager (2,100 m, a.s.l.) climate stations run by the Kyrgyz Hydromet Service (Figure 1). Mean daily streamflow observations starting in the 1960s are available from the Ala-Archa hydrologic station (close to Baitik climate station) at the basin outlet. The Advanced Very High Resolution Radiometer (AVHRR) SCA daily image products with a spatial resolution of 1.1 km $\times$ 1.1 km processed by Dietz et al. (2014) were used in this study. Strict quality control procedures have been implemented for the SCA images, discarding abnormal images, similarly to Duethmann et al. (2014). AVHRR SCA images from the snowmelt season (March–August) during 1987–2000 were selected for multidata set model calibration. We focused on the snow melt season as SCA is most sensitive to the snow degree-day parameterization during this period. The glacier-covered area in the 1970s delineated from Landsat satellite images was used as the initial glacier coverage for the glacier evolution modeling. The glacier mass balance of the Golubin glacier has been measured during 1973–1993 on the annual basis using glaciological method (Figure 1). The GMB measurements were carried out in the end of the ablation period (usually in the late summer) based on glaciological stake readings, and were reported by Kyrgyz researchers to the World Glacier Monitoring Service (WGMS) under the standard protocol. The Golubin glacier is of summer accumulation type, where the

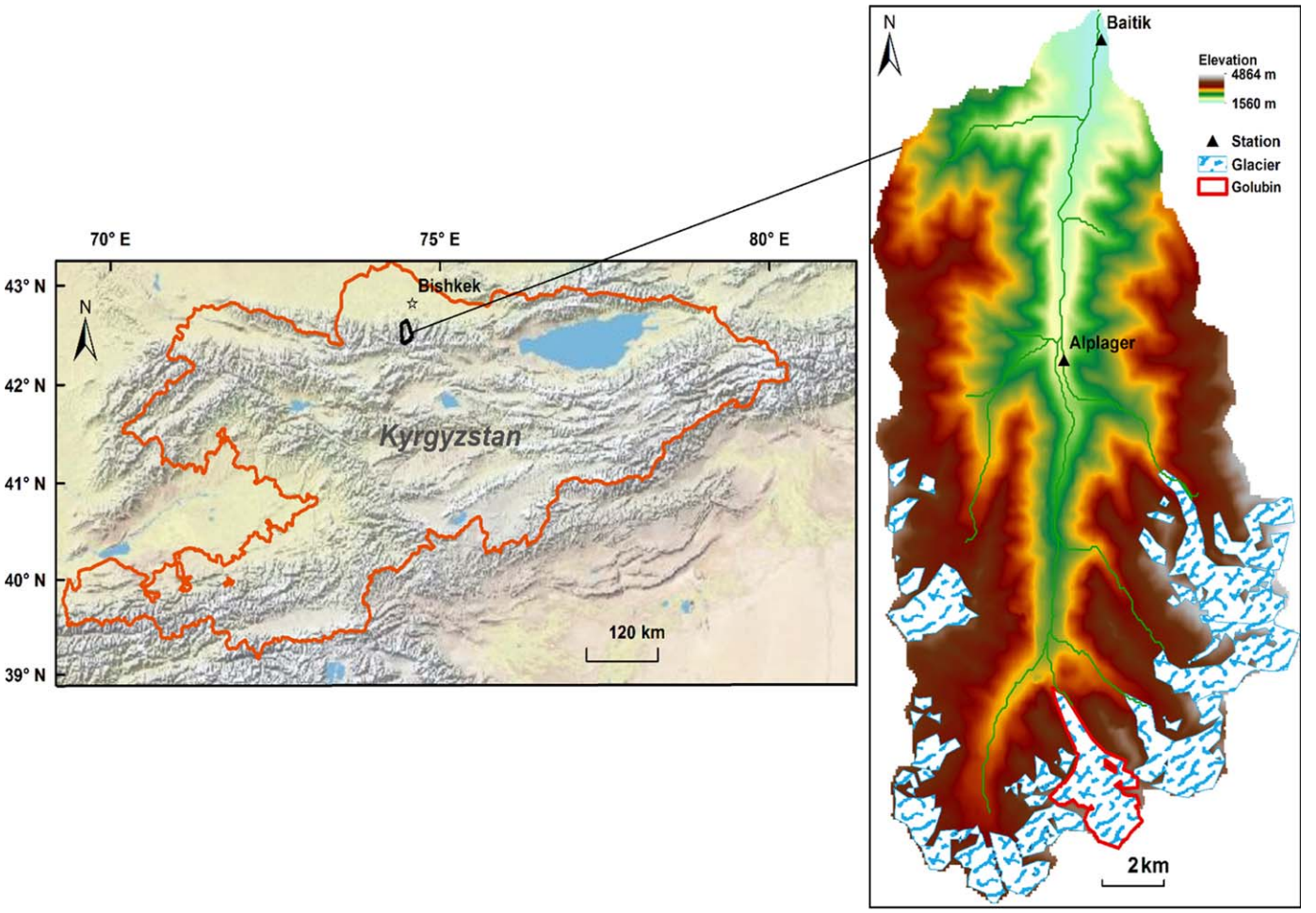

Figure 1. Location of the Ala-Archa basin in Kyrgyzstan (derived from the ESRI World Topographic Map). Black triangles indicate the two climate stations: Baitik and Alplager. The Ala-Archa streamflow gauging station is located in the proximity of the Baitik climate station. The annual glacier mass balance measurements were carried out on the Golubin glacier (red polygon).

highest portion of precipitation comes in spring and early summer. The accumulation and ablation take place nearly in parallel, which would improve the representation of the annual GMB for the glacier dynamics in the study basin.

3. Methodology

3.1. Hydrological Model

The semi-distributed model WASA (Model of Water Availability in Semi-Arid Environments) developed by Güntner and Bronstert (2004) was adopted in this study. Previously, it has been extended for snow and glacier melt processes and successfully applied in Central Asian mountain basins for the evaluation of various precipitation data sets (Duethmann et al., 2013), and for the climate change impact assessment on water resources (Duethmann et al., 2016). The WASA model is based on the discretization of the landscape into model units according to soil properties, land cover, and elevation band height. The WASA model includes the following runoff generation components:

Overland Surface Runoff. The rain and melt water infiltration into the soil is simulated using the Green-Ampt approach (Green & Ampt, 1911). Soil hydraulic conductivities (parameters *k_sat_f* and *kf_corr_f*, see Table 1) control the percolation processes and infiltration excess surface runoff. The generation of saturation excess surface runoff is characterized by the fraction of saturated area (parameter *frac_riparian*), and the current soil moisture state in each model unit (as a function of parameter *sat_area_var*).

Table 1
Parameters Used in the WASA Model (Duethmann et al., 2013), *Which are Grouped Into Four Categories According to the Runoff Processes They Control*

Group	Parameter	Calibrated HPC	Calibration Step in HPC-based method	Unit	Value range	Description
Groundwater	*frac2gw*	Q_{GW}	1	-	0–1	Recharge fraction factor from interflow to groundwater
	interflow delay factor	Q_{GW}	1	days	10–100	Outflow delay factor for interflow
	groundwater delay factor	Q_{GW}	1	days	30–400	Outflow delay factor for groundwater
Snowmelt	snowmelt factor	$Q_{GW} + Q_{SM}$	2	mm/°C/d	0–7	Degree-day factor for snowmelt
Glaicer melt	glacier melt factor	$Q_{GW} + [Q_{SM}] + Q_{GM}$	3	mm/°C/d	1–10	Degree-day factor for glacier melt
Rainfall	*frac_riparian*	$Q_{GW} + [Q_{SM}] + [Q_{GM}] + Q_{RF}$	4	-	0-0.05	Fraction of saturated area
surface	*sat_area_var*	$Q_{GW} + [Q_{SM}] + [Q_{GM}] + Q_{RF}$	4	-	0-0.3	Spatial variability of saturated areas within a model unit
runoff	*kf_corr_f*	$Q_{GW} + [Q_{SM}] + [Q_{GM}] + Q_{RF}$	4	-	0.01–100	Soil hydraulic conductivities correction factor
	k_sat_f	$Q_{GW} + [Q_{SM}] + [Q_{GM}] + Q_{RF}$	4	-	0.01–100	Soil hydraulic conductivities
-	*precip_f*	all	5	-	0.0–5.0	Precipitation correction factor

Subsurface Flow. The WASA model simulates interflow and groundwater components. The water release from these two storages are controlled by two outflow delay factors (*interflow delay factor* and *groundwater delay factor* in Table 1). The percolation from the interflow subsurface reservoir into the groundwater storage is controlled by the fraction factor *frac2gw*.

Snow and Glacier melt. Snowmelt and glacier melt are calculated using a temperature-index approach (Hock, 2003), with two different degree-day factors for snow and glacier (Table 1). To differentiate between rainfall and snowfall, a threshold temperature (T_m) was used. The threshold temperature (T_o) in the degree-day module was set to the same value as T_m. The SCA within each model spatial unit is characterized by a snow depletion curve with the parameter *sdmax* (Liston, 2004), calculating the SCA as a function of the ratio between the current SWE and the maximum SWE during the accumulation season (Duethmann et al., 2014). The SCA of the entire basin is estimated by the ratio of the sum of snow cover areas from all model units to the basin area. The initial glacier ice thickness distribution within each glacierized model unit was estimated using the Glabtop2 model introduced by Linsbauer et al. (2012) and Frey et al. (2014). GMB is calculated annually by subtracting snowmelt and glacier melt from snowfall on the glacierized areas. The Δh-approach was implemented in the WASA model to account for mass balance redistribution with glacier elevation range and account for glacier depth and area changes (Huss et al., 2010).

Considering the insufficient representation of spatial precipitation derived from sparse ground gauge in high mountain areas, a precipitation correction factor (*precip_f*) was introduced in the model to correct the precipitation input. Precipitation forcing was derived by multiplying the observed precipitation with the correction factor (Duethmann et al., 2013). In total, 10 model parameters controlling the four main runoff generation processes were calibrated in the various calibration methods (Table 1). To facilitate the comparison of calibration methods, the value of T_m was optimized based on the simulation of the entire discharge series prior to the comparison, and the parameter *sdmax* was only optimized in the calibration methods which use the SCA. The WASA model was set up for the Ala-Archa basin using the climate data series from two available climate stations (Figure 1). Monthly lapse rates of precipitation and temperature were derived from observed data series and used to estimate daily precipitation and temperature in each model unit based on elevation.

3.2. Extracting Hydrograph Partitioning Curves

To identify the snow and glacier ablation periods, we introduced two cumulative curves: the annual cumulative curve of average daily snowfall on the glacierized area (ASF) and the annual cumulative curve of the difference between average daily temperature and T_m of the entire basin (ADT). ASF is used to identify the starting and end dates of snow accumulation across the glacier-covered model units, and ADT presents the period when the average temperature over the entire basin exceeds T_m. Figure 2 shows an example for the extraction of HPCs in the year 1987. The minimum (P1) of the ADT curve indicates the assumed starting date of the snowmelt period, and the maximum (P2) indicates the assumed end date of the glacier melt.

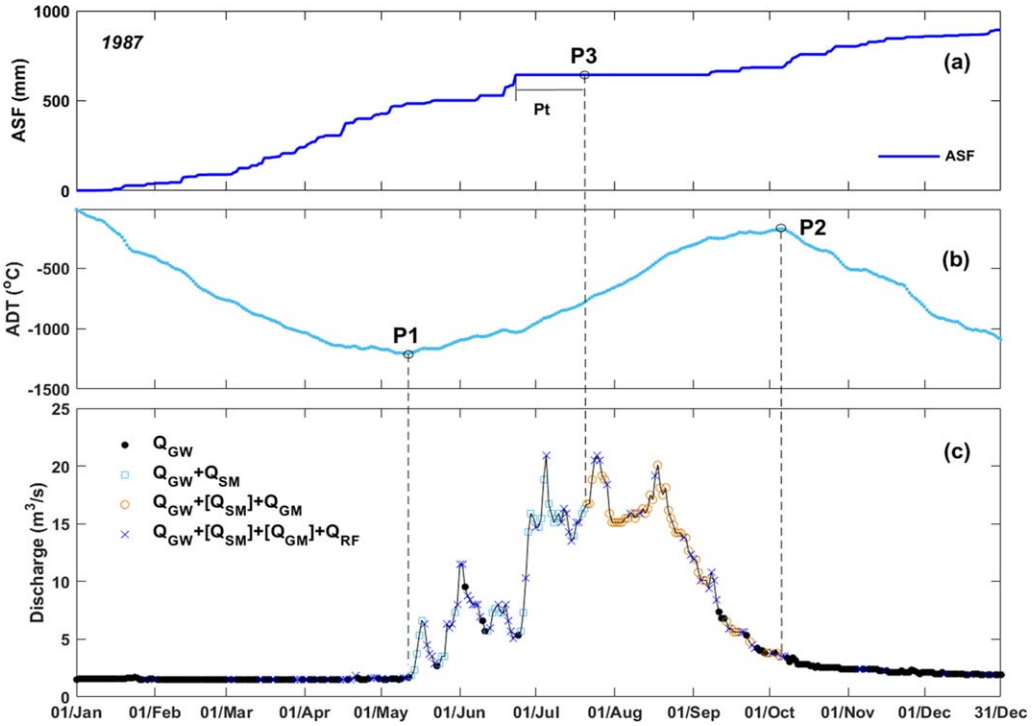

Figure 2. (a) Cumulative curve of average daily snowfall on glacier (ASF) and (b) cumulative curve of the difference between average daily temperature and melt threshold temperature over the basin (ADT) in 1987 used to extract (c) hydrograph partitioning curves (HPCs). Q_{GW} represents the groundwater dominated hydrograph part, $Q_{GW} + Q_{SM}$ refers to the groundwater and snowmelt dominated hydrograph part, $Q_{GW} + [Q_{SM}] + Q_{GM}$ stands for the groundwater and meltwater (glacier melt and possible snowmelt) dominated hydrograph part, and $Q_{GW} + [Q_{SM}] + [Q_{GM}] + Q_{RF}$ refers to the hydrograph part controlled by groundwater and rainfall, as well as possible meltwater.

The turning point on the ASF curve followed by the longest nonsnowy period indicates the end of snow accumulation on the glacierized areas. A period after this point indicated by the Pt value is adopted to represent the persistence of snow accumulation on glaciers after the last snowfall event. We predefined the Pt value based on our knowledge of the catchment conditions. Five Pt values were adopted to indicate the assumed starting date of glacier melt, ranging from 1 to 5 weeks. Figure 2 shows the extraction of HPCs using a Pt value of 4 weeks in the year 1987. The refined method identifies the snowmelt and glacier melt contribution periods more appropriately compared to that in He et al. (2015), which used a specific elevation threshold to identify the ablation period. Their method did not consider the snow cover on glacierized areas and was not able to distinguish summer snowmelt from glacier melt.

Subsequently, three runoff component indices S_i, G_i, and R_i, are defined as step functions between 0 and 1 indicating the dominant sources of river runoff for each day using the ASF and ADT curves (equations (1–3)):

$$S_i = \begin{cases} 1, & \text{for } \underset{j=1\to nn}{mean}(T_j^i) > T_m \,\&\, i \in [P1, P3] \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

$$G_i = \begin{cases} 1, & \text{for } \underset{k=1\to n}{mean}(T_k^i) > T_m \;\&\; i \in [P3, P2] \\ 0, & \text{otherwise} \end{cases}, \tag{2}$$

$$R_i = \begin{cases} 1, & \text{for } \underset{j=1\to nn}{\max}((T_j^i - T_m)\cdot P_j^i) > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

where S_i, G_i, and R_i are the indices indicating the dominance of snowmelt, glacier melt, and rainfall surface runoff on day i, respectively. The index value of 1 indicates the presence of the respective

contribution to streamflow on that day, while 0 indicates its absence. T_j^i is the daily temperature on model unit j; nn indicates the total number of model units in the basin; T_k^i is the daily temperature on the glacierized model unit k; n is the total number of glacierized units; and P_j^i is the daily precipitation on model unit j.

According to the daily values of the runoff component indices, the observed hydrograph at the basin outlet can be stratified into four HPCs by equation (4). Each HPC indicates the dominant period of a specific combination of runoff components:

$$Q=\begin{cases} Q_{GW}, & for\ S_i+R_i+G_i=0 \\ Q_{GW}+Q_{SM}, & for\ S_i-R_i-G_i=1 \\ Q_{GW}+[Q_{SM}]+Q_{GM}, & for\ G_i-R_i=1 \\ Q_{GW}+[Q_{SM}]+[Q_{GM}]+Q_{RF}, & for\ R_i=1 \end{cases} \tag{4}$$

where Q_{GW} stands for the groundwater dominant period, i.e., days when both melt water and rainfall runoff do not occur ($S_i = R_i = G_i = 0$). This partition is used for the calibration of groundwater parameters. $Q_{GW} + Q_{SM}$ represents the groundwater and snowmelt dominant period, i.e., days when both rainfall and glacier melt do not occur ($S_i = 1$, $R_i = G_i = 0$). This partition is used for the calibration of the snowmelt parameter. $Q_{GW} + [Q_{SM}] + Q_{GM}$ refers to the groundwater and melt water (glacier melt and possible remaining snowmelt on nonglacierized areas) dominant period on nonrainy days ($G_i = 1$, $R_i = 0$). This partition is used for the calibration of the glacier melt parameter. $Q_{GW} + [Q_{SM}] + [Q_{GM}] + Q_{RF}$ indicates rainy days ($R_i = 1$) with groundwater, rainfall, and possible melt water. This partition is used for the calibration of rainfall surface runoff parameters. The "[]" notation indicates the possible but not necessary presence of melt runoff. We assume that inflow is transformed into discharge at the gauge site entirely within the same day. The subsurface storage effects on the HPC process are considered by separating outflow into groundwater runoff and surface runoff, and that any delays due to surface and network storage could be a relatively minor effect. The application of the HPC-based method in large catchment will be discussed in section 4.7.

3.3. HPC-Based and Multidata Set Calibration

The WASA model was calibrated using five different approaches: (1) stepwise and iterative approach based on HPCs, (2) single-data set calibration based on the discharge time series, (3) bi-data set calibration based on discharge records and glacier mass balance series (GMB), (4) bi-data set calibration based on discharge records and remotely sensed snow cover information (SCA), and (5) tri-data set calibration based on discharge, GMB and SCA. We used the average value of Nash-Sutcliffe efficiency (*NSE*, equation (5)) and logarithmic *NSE* (*logNSE*, equation (6)) as the objective function to measure the performance for the simulation of discharge (hereafter abbreviated as *avNSE*), the *NSE* for the simulation of each HPC, the relative SCA error (*SE*, equation (7)) for the simulation of SCA, and the volumetric deviation efficiency (*VE*, equation (8)) to assess the simulation performance for GMB.

$$NSE=1-\frac{\sum_{i=1}^{n}(Q_{obs}(i)-Q_{sim}(i))^2}{\sum_{i=1}^{n}(Q_{obs}(i)-\overline{Q_{obs}})^2},\ for\ discharge\ and\ HPC \tag{5}$$

$$\log NSE=1-\frac{\sum_{i=1}^{n}(\log Q_{obs}(i)-\log Q_{sim}(i))^2}{\sum_{i=1}^{n}(\log Q_{obs}(i)-\overline{\log Q_{obs}})^2},\ for\ discharge \tag{6}$$

$$SE=1-\sqrt{\frac{\sum_{k=1}^{m}(S_{obs}(k)-S_{sim}(k))^2}{m}},\quad for\ SCA \tag{7}$$

$$VE = 1 - \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(M_{obs}(t) - M_{sim}(t)\right)^2}}{|\overline{M_{obs}}|}, \quad for\ GMB \tag{8}$$

where $Q_{obs}(i)$ and $Q_{sim}(i)$ are the observed and simulated discharge on day i, respectively; $\overline{Q_{obs}}$ is the average observed daily discharge during the calibration period with n days; $S_{obs}(k)$ and $S_{sim}(k)$ are the remotely sensed and simulated SCA over the basin on day k, respectively; $M_{obs}(t)$ and $M_{sim}(t)$ are the observed and simulated GMB on the Golubin glacier in year t, respectively; $\overline{M_{obs}}$ is the average observed GMB on the Golubin glacier during the N evaluated years.

All five calibration approaches were automatically carried out using the ε-NSGAII algorithm developed by Deb et al. (2002) and Kollat and Reed (2006). In the multidata set calibration methods, we carried out a multiobjective ε-NSGAII algorithm. The objective functions for the simulation of discharge, SCA and GMB were weighted equally by using the same Epsilon values. In the HPC-based method, the ten parameters were calibrated in five steps. In each step, a single-objective ε-NSGAII algorithm was adopted to constrain the corresponding parameters by optimizing the *NSE* value for an individual HPC. All the calibration approaches were carried out in two trials to investigate the impacts of the initialization of the ε-NSGAII algorithm. In the first trial, we ran the ε-NSGAII algorithm using initial population size of 8 and maximum generation per run of 250. In the second trail, the initial population size was 16.

In the HPC-based calibration method, parameters grouped in Table 1 were constrained on different HPCs in a stepwise and iterative way. The calibration procedure in the first trial is summarized as follows:

1. Each step consists of a ε-NSGAII process, using an initial population size of 8, maximum population size of 4,000, and maximum generations per run of 250. For each step, a single parameter vector is chosen, corresponding to the highest performance, and then fixed for the subsequent step.
2. In step 1, parameters *frac2gw*, *interflow delay factor* and *groundwater delay factor* controlling groundwater outflow were first calibrated on the Q_{GW} partitioning curve. In step 2, the snowmelt factor was calibrated on the $Q_{GW} + Q_{SM}$ partitioning curve. The glacier melt factor was calibrated on the $Q_{GW} + [Q_{SM}] + Q_{GM}$ partitioning curve in the third step. In step 4, parameters determining surface runoff (*k_sat_f*, *kf_corr_f*, *frac_riparian* and *sat_area_var*) were calibrated on the $Q_{GW} + [Q_{SM}] + [Q_{GM}] + Q_{RF}$ partitioning curve. Finally, the precipitation correction factor (*precip_f*) was calibrated on all four HPCs. The *NSE* value was used as the objective function in the first four steps, while the Euclidean distance of the *NSE* values for four HPCs was used as the objective function in the last step.
3. The entire five-step process was repeated 8 times in total to consider the parameter interaction. This iterative procedure, using parameters calibrated in previous steps, considers the effects of the parameter initialization on the simulation of each HPC. All parameters were updated in each iteration when five calibration steps were finished. The impact of the parameter initialization could be reduced after several iterative procedures.
4. The final calibrated parameter set is based on the composite parameter vector optimized during the last iteration. The parameters that produced the highest objective-function value in each step in the last iteration were selected for the combination. To simulate the entire discharge, the annual GMB on the Golubin glacier and SCA over the entire basin, we re-ran the model using the composite parameter vector.

For comparison with the multidata set calibration, the model calibration and validation procedure was implemented in three scenarios: a baseline scenario including five methods, i.e., single-data set, bi-data set using discharge and GMB (hereafter abbreviated as QM-based bi-data set), bi-data set using discharge and SCA (hereafter abbreviated as QS-based bi-data set), tri-data set, and HPC-based; and two scenarios to investigate parameter stability. In the latter two scenarios, three calibration methods (single-data set, QM-based bi-data set or QS-based bi-data set, and HPC-based) were applied due to the limited data availability.

Baseline Scenario (S1). Five calibration methods were implemented in the baseline calibration period 1987–1993. The calibrated parameters were validated in three periods (1973–1979, 1980–1986, and 1994–2000). Since SCA data are available after 1986, and GMB data are not available during 1994–2000, we investigated the performance for the simulations of discharge and GMB during 1973–1979 and 1980–1986, and the performance for the simulations of discharge and SCA during 1994–2000. The parameter *sdmax* was optimized by the tri-data set method in this scenario.

Scenario 2 (S2). The model was calibrated for 1973–1979 and validated in the periods 1980–1986, 1987–1993, and 1994–2000. The QM-based bi-data set calibration method was carried out in this scenario.

Scenario 3 (S3). The model was calibrated for 1994–2000 and validated in the periods 1973–1979, 1980–1986, and 1987–1993. The QS-based bi-data set calibration method was carried out in this scenario. The parameter *sdmax* was optimized by the bi-data set method.

3.4. Calibration Using Other Hydrograph Signatures

We also adopted other hydrograph signatures for the model calibration to contrast the ability of HPCs in constraining multiple runoff processes. The chosen signatures include the intraannual runoff variability and the flow duration curve (FDC). These signatures were selected for comparison due to their foci on runoff seasonality, which was similarly emphasized by the HPC. The objective function to quantify the reproduction of the intraannual runoff variability is defined in equation (9), similar to Duethmann et al. (2015). Following Shafii and Tolson (2015), we divided the FDC into three segments to examine the simulation of the overall FDC. The corresponding objective functions are given in equations (10–12). A three-objective calibration procedure based on these signatures was conducted in the ε-NSGAII algorithm. The first objective-function f_1 was the *avNSE* for the overall discharge series, the latter two objective functions were f_2 and f_3 derived from equations (9) and (13).

$$f_2=\frac{1}{4}\left(\frac{NSE_{DJF}+NSE_{DJF}^{\log}}{2}+\frac{NSE_{MAM}+NSE_{MAM}^{\log}}{2}+\frac{NSE_{JJA}+NSE_{JJA}^{\log}}{2}+\frac{NSE_{SON}+NSE_{SON}^{\log}}{2}\right) \tag{9}$$

where NSE_{DJF} (NSE_{MAM}, NSE_{JJA}, NSE_{SON}) and NSE_{DJF}^{log} are the *NSE* and logarithmic *NSE* calculated from daily observed and simulated discharge in winter (spring, summer, and autumn). The perfect value of objective-function f_2 is 1.

$$g_1=\sum_{l=1}^{n1}[Q_{obs}(l)-Q_{sim}(l)]/\sum_{l=1}^{n1}Q_{obs}(l),\ \textit{for FDC high-segment volume} \tag{10}$$

where *l* is the flow index within the high-flow segment, which contains high flows with exceedance probability lower than 0.2. *n1* is the total number of daily discharge falling into the high segment.

$$g_2=\frac{[\log Q_{obs}(m1)-\log Q_{obs}(m2)]-[\log Q_{sim}(m1)-\log Q_{sim}(m2)]}{\log Q_{obs}(m1)-\log Q_{obs}(m2)},\ \textit{for FDC midsegment slope} \tag{11}$$

where *m*1 and *m*2 are the lowest and highest flow exceedance probabilities within the midsegment of FDC. Here we used 0.2 and 0.7 as thresholds, respectively:

$$g_3=\frac{\sum_{z=1}^{nk}[\log Q_{obs}(z)-\log Q_{obs}(M)]-\sum_{z=1}^{nk}[\log Q_{sim}(z)-\log Q_{sim}(M)]}{\sum_{z=1}^{nk}[\log Q_{obs}(z)-\log Q_{obs}(M)]},\ \textit{for FDC low-segment volume} \tag{12}$$

where *z* is the flow index in the low-flow segment (with exceedance probability larger than 0.7). *M* is the index of the minimum flow, *nk* is the total number of daily discharge falling into the low segment. The perfect value of objective-function f_3 is also 1.

$$f_3=1-\frac{abs(g_1)+abs(g_2)+abs(g_3)}{3} \tag{13}$$

4. Results

4.1. Hydrograph Partitioning Curves

Figure 2 illustrates the extracted HPCs assuming Pt of 4 weeks for the year 1987. The starting date of the ablation period (P1) was estimated on 10 May, the end date (P2) on 4 October, and the starting date of glacier melt (P3) was estimated on 20 July. HPCs for other years were constructed analogously. The effect of the Pt value on the HPC-based calibration method will be discussed in section 4.8.

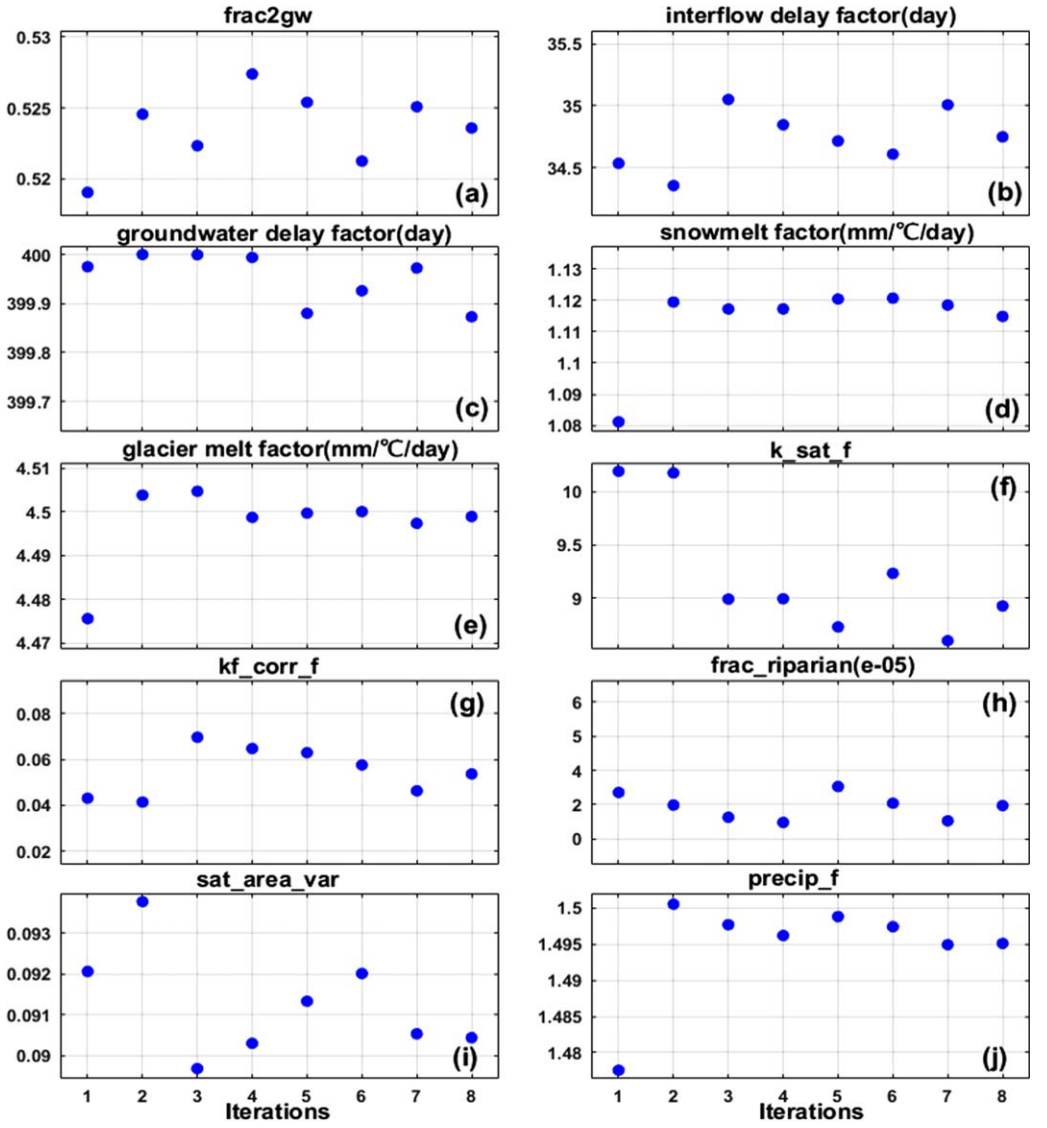

Figure 3. Variability of parameter values during the stepwise-iterative calibration procedure in the HPC-based method.

Observed daily streamflow values were assigned to the HPCs representing the dominant runoff components identified by equation (4). For example, the Q_{GW} partitioning curve in Figure 2c shows the dominance of groundwater in winter months when both meltwater and rainfall rarely occur. The Q_{GW} +Q_{SM} partitioning curve indicates that snowmelt plays a crucial role in the total streamflow during May, June, and July. The dominant periods of the different runoff components identified by the HPCs are consistent with the prior knowledge about the seasonal dominance of different water sources in the Ala-Archa basin (Aizen et al., 1996). In the entire calibration period 1987–1993, the number of days contained in the Q_{GW} and the $Q_{GW} + Q_{SM}$ partitioning curves are 1417 and 225, respectively. Two hundred and fifty six days are included in the $Q_{GW} + [Q_{SM}] + Q_{GM}$ partitioning curve and the remaining 659 days constitute the $Q_{GW} + [Q_{SM}] + [Q_{GM}] + Q_{RF}$ partitioning curve.

4.2. Comparison with Multidata Set Calibration Approaches

Figure 3 illustrates the variability of the 10 parameters during the iterations using the HPC-based stepwise-iterative calibration procedure during 1987–1993. Eight iterations were finished to test the variability of parameter values. Parameter values in the first two iterations tend to be significantly different from those in the later iterations. The variance of parameter values in the last four iterations tends to be small.

Table 2 summarizes the performance for the simulations of discharge, GMB and SCA in calibration scenario S1, driven by individual behavioral parameter sets in two trials. For the single-data set calibration method,

Table 2
Summary of the Performance for Discharge(avNSE), GMB(VE), and SCA(SE) Simulations by Various Calibration Methods in Two Trails During the Calibration Period 1987–1993

	Trial 1			Trial 2		
	avNSE	*SE*	*VE*	*avNSE*	*SE*	*VE*
Sing.	0.8781	0.8020	0.6012	0.8790	0.8002	0.5862
Bi-QM.	0.8735	0.8314	0.6425	0.8718	0.8327	0.6433
Bi-QS.	0.8653	0.8597	0.6118	0.8654	0.8597	0.6147
Trip.	0.8623	0.8604	0.6407	0.8626	0.8604	0.6406
HPC.	0.8620	0.8608	0.6388	0.8619	0.8607	0.6391
Ex-HPC.	0.8628	0.8605	0.6392	0.8629	0.8606	0.6393

Note. "Sing." stands for performance produced by the single-data set calibration method, "Bi-QM." and "Bi-QS." refer to the bi-data set calibration methods using discharge/GMB and discharge/SCA, respectively. "Trip." indicates the tri-data set method, "HPC." represents the HPC-based method, "Ex-HPC." presents the extended HPC-based method in large catchments.

the behavioral parameter set was defined as the parameter set that produced the highest *avNSE* value for discharge. For the bi-data set and tri-data set calibration methods, the behavioral parameter sets were selected from the Pareto-optimal front by taking the smallest Euclidean distance from the perfect point. For the HPC-based method, the behavioral parameter set was derived from the combination of the calibrated parameters in five individual steps. The parameters that produced the highest objective-function value in each step during the last iteration were selected for the combination.

Figures 4–6 compare the performance of five calibration methods in the calibration period 1987–1993, driven by the individual behavioral parameter sets of trial 1. The single-data set calibration produced the highest *NSE* and *logNSE* values for the simulation of discharge (Figure 4a). The QM-based bi-data set performed similarly to the single-data set method, while the QS-based bi-data set, as well as the HPC-based and tri-data set methods obtained lower *NSE* and *logNSE* values. Figure 4b shows the agreement between the simulated and observed FDCs. All calibration approaches overestimated moderate flows, while significantly underestimating high flows. Figures 4c and 4d present simulated bias values for various discharge ranges. We ranked the daily streamflow using seven quantiles (5%, 10%, 25%, 50%, 75%, 90%, and 95% in Figure 4c). Compared to the single-data set method, the HPC-based method achieved lower bias values for low flow (lower than 1.66 m^3/s), but larger values for flow ranging from 1.66 to 7.1 m^3/s. Figure 4d shows the simulated bias values for five flow quantiles from the FDC. The HPC-based method outperformed other methods on the 25% and 75% quantiles, but produced larger overestimation on the median flow.

As expected, the single-data set calibration yielded the lowest *VE* value for GMB, followed by the QS-based bi-data set calibration (Table 2 and Figure 5a). Bias values for the simulated GMB over the entire GMB series indicate significantly lower performance of the single-data set and QS-based bi-data set methods (Figure 5b). Observed GMB has not been used as constraining observation in these two calibration methods. Thus, a stronger internal compensation between glacier melt and other runoff components occurred compared to the other methods. The HPC-based method performed similarly well as the tri-data set and QM-based bi-data set methods in terms of *VE* values, and achieved the lowest bias value for the entire GMB series (Figure 5b). Figure 6 shows that the single-data set method performed worst with respect to the simulation of SCA in snowmelt seasons, closely followed by the QM-based bi-data set method, due to the absence of SCA data as constraining observation. The HPC-based method performed similarly to the tri-data set and QS-based bi-data set methods (see *SE* values in Table 2). Both the HPC-based and tri-data set methods obviously outperformed the single-data set and QM-based bi-data set methods on the simulation of low snow fraction events (lower bias values shown in Figure 6b).

In general, the HPC-based method achieved comparable performance to the tri-data set method, despite solely using hydrograph partitions as constraining observation. The HPC-based method even produced lower bias values for GMB and SCA than all the other four methods. The results of the second trial (summarized in Table 2) also support these findings. For the sake of brevity, all further results discussed hereafter are from the first trial and they are similar to those produced in the second trial.

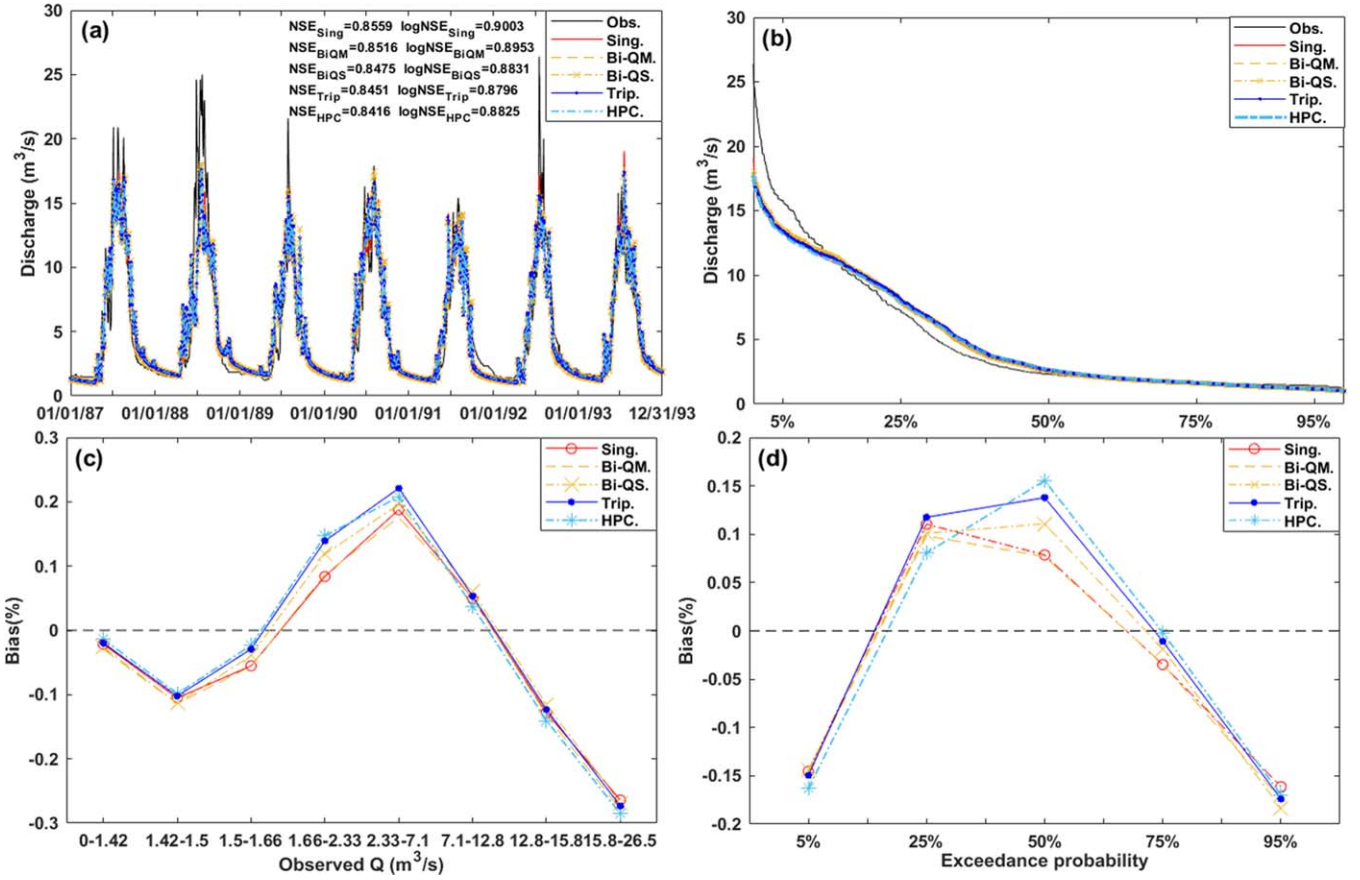

Figure 4. (a) Simulations of daily discharge in the calibration period 1987–1993 achieved with various calibration methods. "Sing." stands for the single-data set calibration, "Bi-QM." ("Bi-QS.") refers to the bi-data set calibration using discharge and GMB (discharge and SCA), "Trip." indicates the tri-data set calibration, and "HPC." represents the HPC-based calibration. (b) Simulation of the Flow Duration Curve (FDC). (c) Simulated bias values for various observed discharge ranges. (d) Simulated bias values for flow quantiles from the FDC.

Figure 7 shows the pairwise trade-off between two simulation data sets (discharge versus SCA and discharge versus GMB) for the calibration and validation periods in three calibration scenarios (S1–3, section 3.3). The performance criteria values were produced by the behavioral parameter sets of various calibration methods. The Euclidean distance to the perfect point (1, 1) was used to evaluate the performance in each subplot. In calibration scenario S2, the HPC-based method performed similarly to the QM-based bi-data set method in the calibration period 1973–1979 (i.e., similar Euclidean distance to the perfect point, Figure 7a), and outperformed the bi-data set and single-data set methods in all three validation periods (Figures 7b–7e). In scenario S1, the HPC-based method performed similarly well to the tri-data set method, and outperformed the single-data set method in the calibration period 1987–1993 (Figures 7h and 7i). As expected, the QM-based (QS-based) bi-data set method shows lower performance than the tri-data set method in Figure 7h (7i). In the three validation periods (Figures 7f, 7g, and 7j), both the HPC-based and tri-data set methods tended to outperform the other methods. In scenario S3, the HPC-based method performed similarly to the QS-based method in the calibration period 1994–2000 (Figure 7o), and outperformed the bi-data set and single-data set methods in the validation periods (Figures 7k–7m). In all the three calibration scenarios, the single-data set method performed poorly in capturing the dynamics in snow and glacier observations. The HPC-based method performed comparably to the multidata set calibration methods for the reproducing of SCA and GMB, in spite of that SCA and GMB observations were not included in this calibration procedure.

4.3. Parameter Stability and Identifiability

Figure 8 compares the behavioral parameter values obtained by different calibration methods across the three calibration scenarios. For all 10 parameters, the HPC-based method tended to achieve more stable parameter values across different calibration periods, compared to the single-data set and bi-data set

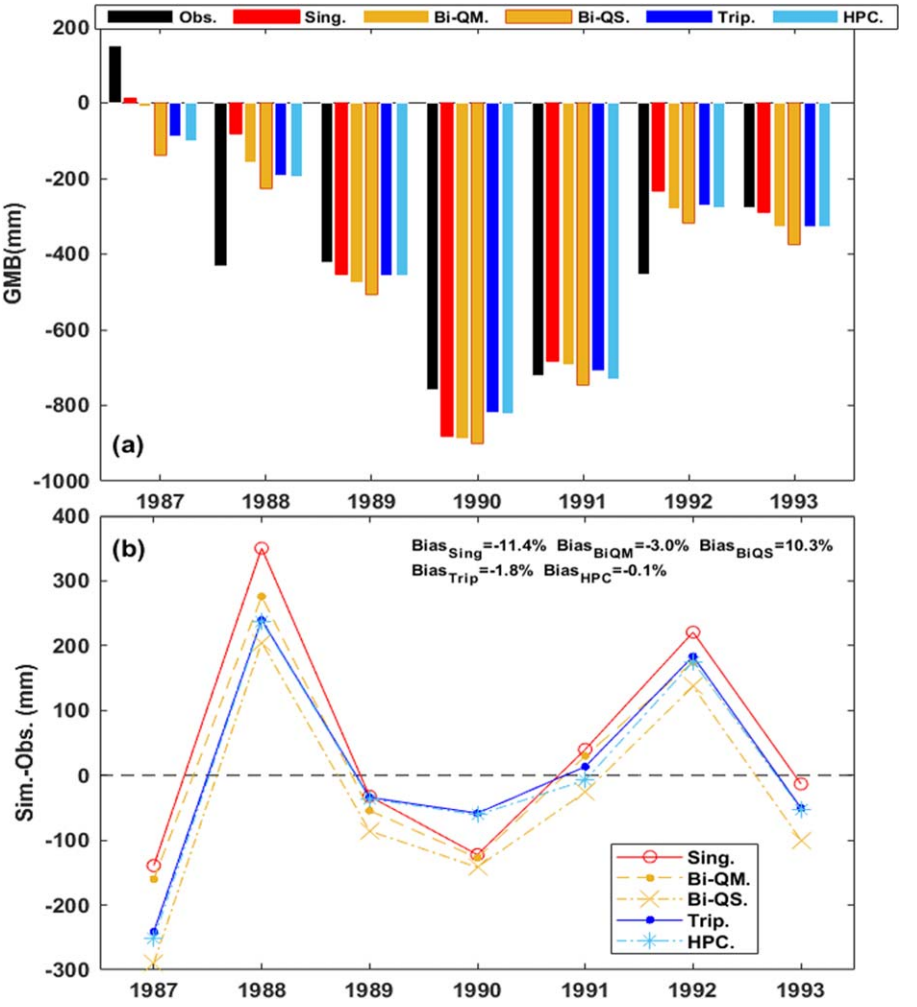

Figure 5. (a) Simulations of annual glacier mass balance (GMB) in the calibration period 1987–1993 by various calibration methods. (b) Error (simulation minus observation) between the simulated and observed GMB in each year, and the bias values for the entire annual GMB series.

methods. The single-data set method yielded the largest variation in many parameter values. For example, the value of glacier melt factor ranges from 3.41 to 9.99 when calibrated by the single-data set method (Figure 8e), while this value ranges from 4.45 to 4.54 in the HPC-based method, from 4.59 to 5.45 in the QM-based bi-data set method, and from 5.03 to 5.17 in the QS-based bi-data set method. Figures 7 and 8 show that the behavioral parameter sets produced by the HPC-based method achieved stable values and high performance for the model cross validation, suggesting a better parameter identifiability with reduced parameter uncertainty.

To investigate whether the improvement in parameter stability stems from the information provided by the HPCs or from the stepwise-iterative procedure adherent to the HPC-based calibration, we ran two additional calibration scenarios: a stepwise-iterative calibration procedure using discharge and SCA/GMB data sets, and a four-objective calibration based on four HPCs. In the stepwise-iterative calibration scenario, we conducted four calibration steps. In the first step, we calibrated the parameters belonging to the groundwater and surface runoff groups on the whole discharge series (see Table 1 for parameter groups). In the second step, we calibrated the snowmelt factor on SCA, and in the third step, the glacier melt factor was calibrated on GMB. Finally, we calibrated the precipitation correction factor to minimize the Euclidean distance of the av*NSE* values for discharge and *SE* (*VE*) for SCA (GMB). In periods when only discharge and SCA were available, the glacier melt factor was calibrated in the first step, and when only discharge and GMB were

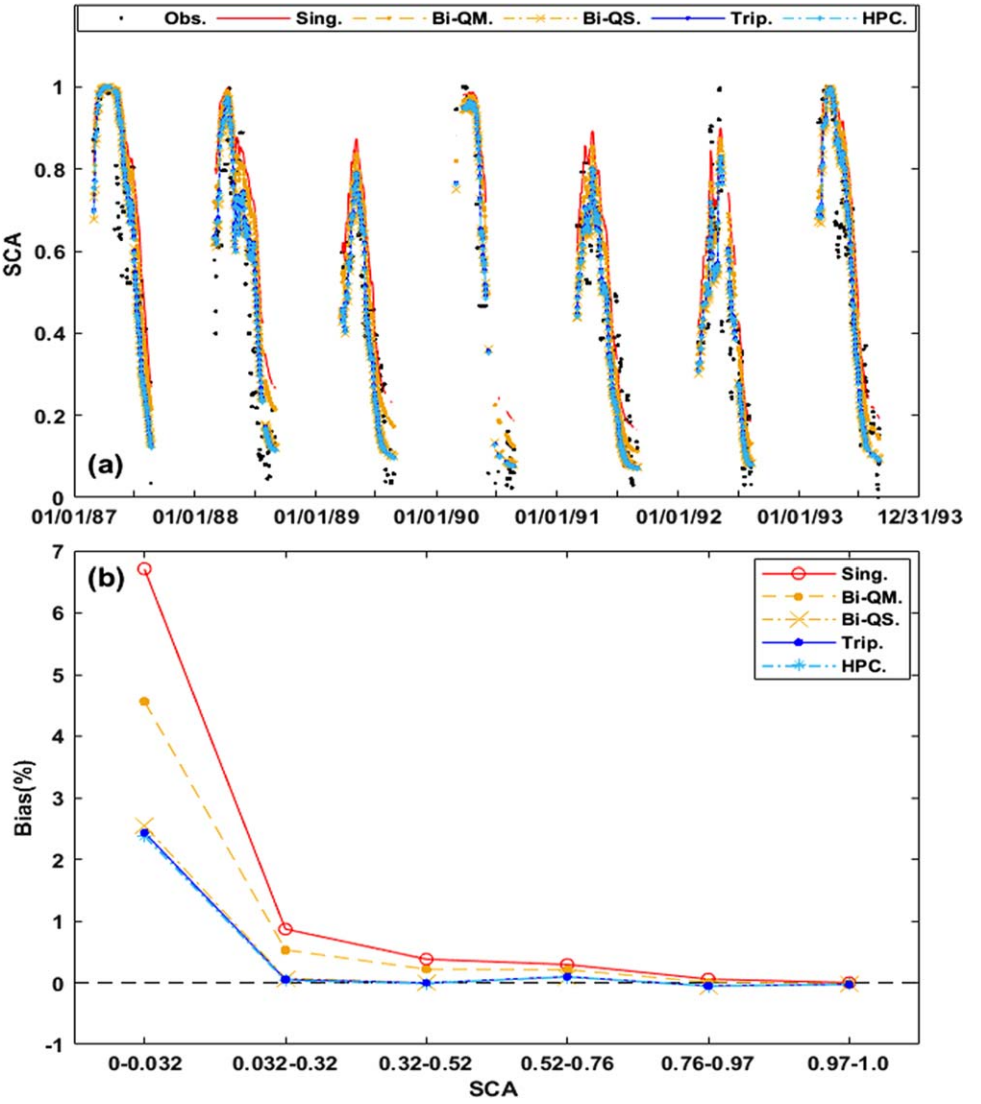

Figure 6. (a) Simulations of snow cover area (SCA) during the snowmelt season (March–August) in the calibration period 1987–1993 by various calibration methods. (b) Bias values for the simulated SCA ranges.

available, both snow and glacier melt factors were calibrated on the GMB data set. The behavioral parameter set for this scenario was derived from the combination of the calibrated parameters in individual steps. In the four-objective scenario, the Pareto-optimal parameter set that achieved the smallest Euclidean distance for the *NSE* values of the four HPCs was defined as the behavioral parameter.

Figure 9 compares the behavioral parameter values obtained by the two calibration scenarios in three calibration periods. The stepwise-iterative procedure using discharge and GMB produced stable values for the glacier melt factor and *frac_riparian* across the periods 1973–1979 and 1987–1993 (Si-QM in Figures 9e and 9h). The stepwise-iterative procedure using discharge and SCA delivered considerably stable values for groundwater parameters, the glacier melt factor, *frac_riparian* and *sat_area_var* across the periods 1987–1993 and 1994–2000 (Si-QS in Figures 9a–9c, 9e, 9h, and 9i). However, the stepwise-iterative procedures using separated discharge and SCA/GMB data sets tend to show only minor (or negative) improvement in parameter stability, compared to the standard bi-data set calibration in Figure 8. The four-HPC method improved the stability in groundwater parameters and the snowmelt factor (Figures 9a–9d), compared to the standard QM-based bi-data set calibration, and yielded comparable parameter stability to the standard QS-based bi-data set calibration. The four-HPC method shows ability to deliver stable values for most parameters (F-HPC in Figures 9a, 9c–9f, and 9h), while the standard HPC-based method further improved the parameter

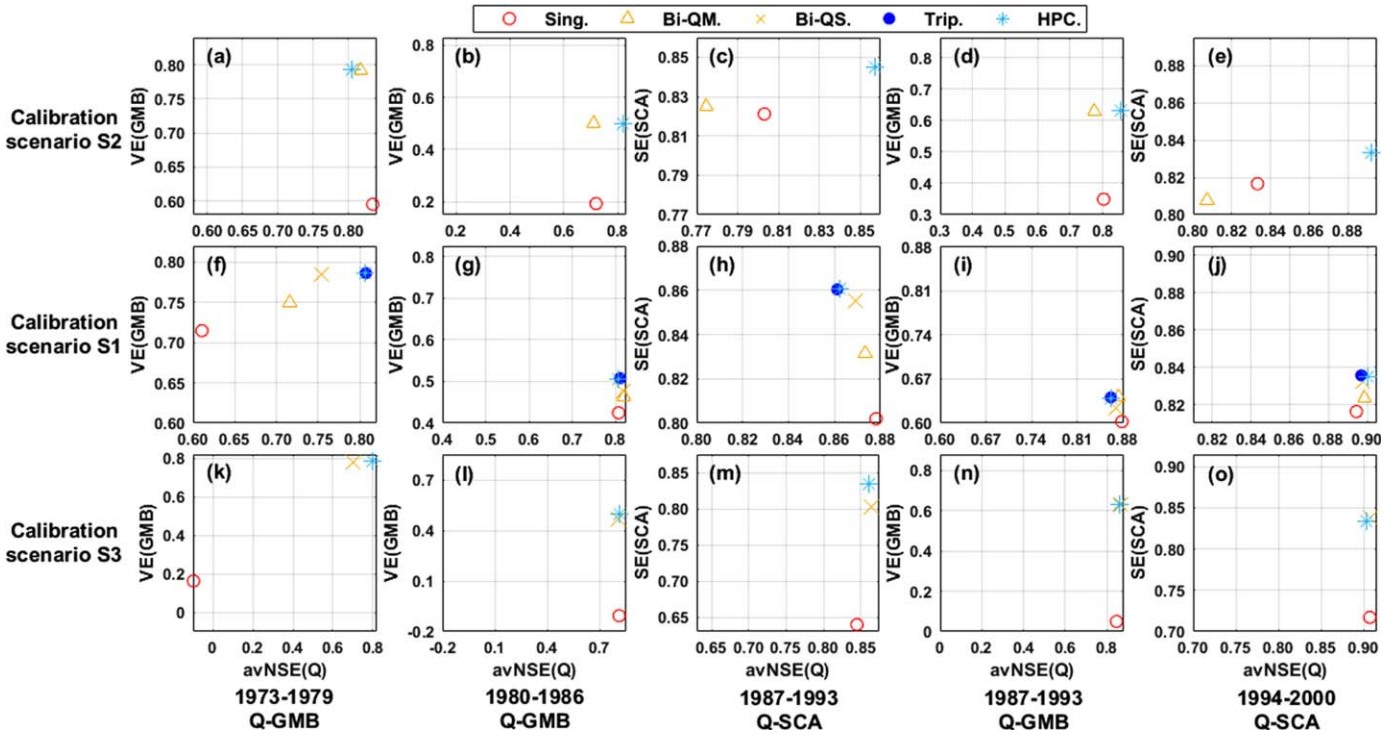

Figure 7. Pairwise trade-off between the performances for multiple data sets (Q versus SCA or Q versus GMB) in the calibration and validation periods in three calibration scenarios (S1–3) yielded by the behavioral parameter sets of five calibration methods. The performance is evaluated by the Euclidean distance to the perfect point (1, 1). The abbreviations for the five calibration methods are the same to those in Figure 4.

stability using the stepwise-iterative procedure (HPC. in Figure 9). For example, the HPC-based method reduced the variations of the parameters *kf_corr_f* and *sat_area_var* across the calibration periods, compared to the four-HPC method (Figures 9g and 9i). The values for the parameter *k_sat_f* produced by the HPC-based method are significantly different from those produced by the four-HPC method (Figure 9f). These could be attributed to the different parameter optimization procedures in the four-HPC and HPC-based methods. In the four-HPC method, model parameters were optimized by the simultaneous simulation of the four HPCs, while the HPC-based method optimized the parameter values by the separated simulation of each HPC in each step. Therefore, there is a stronger chance for compensation between model parameters in the four-HPC method, enhancing the equifinality in parameter values. The stepwise-iterative optimization procedure in the HPC-based method separately calibrated parameters on the corresponding sensitive HPCs to reduce the internal compensation of runoff processes, and thus helped to reduce the parameter equifinality.

4.4. Contributions of Runoff Components

We ran the model for the period 1973–2000 using behavioral parameter sets obtained in scenario S1 to estimate the contributions of runoff components to the river discharge. Table 3 summarizes the annual contributions of runoff components simulated by the different calibration methods in three representative years. From the annual precipitation series during 1973–2000, the wettest year 1979 (annual precipitation 602 mm), the median wet year 1983 (annual precipitation 549 mm), and the driest year 1997 (annual precipitation 392 mm) were selected. There is a strong interannual variability in runoff components in the study basin. As expected, the contributions of groundwater, snowmelt, and rainfall decrease with decreasing annual precipitation. In contrast, the contribution of glacier melt increases with decreasing annual precipitation due to lower snow accumulation on glacierized areas and potentially longer ablation period. The five calibration methods mainly differed in the estimation of the glacier melt contribution. For example, in the median wet year 1983, the HPC-based and tri-data set methods computed the glacier melt contribution as 17–18%, the QS-based and QM-based bi-data set methods estimated the glacier melt contributions to

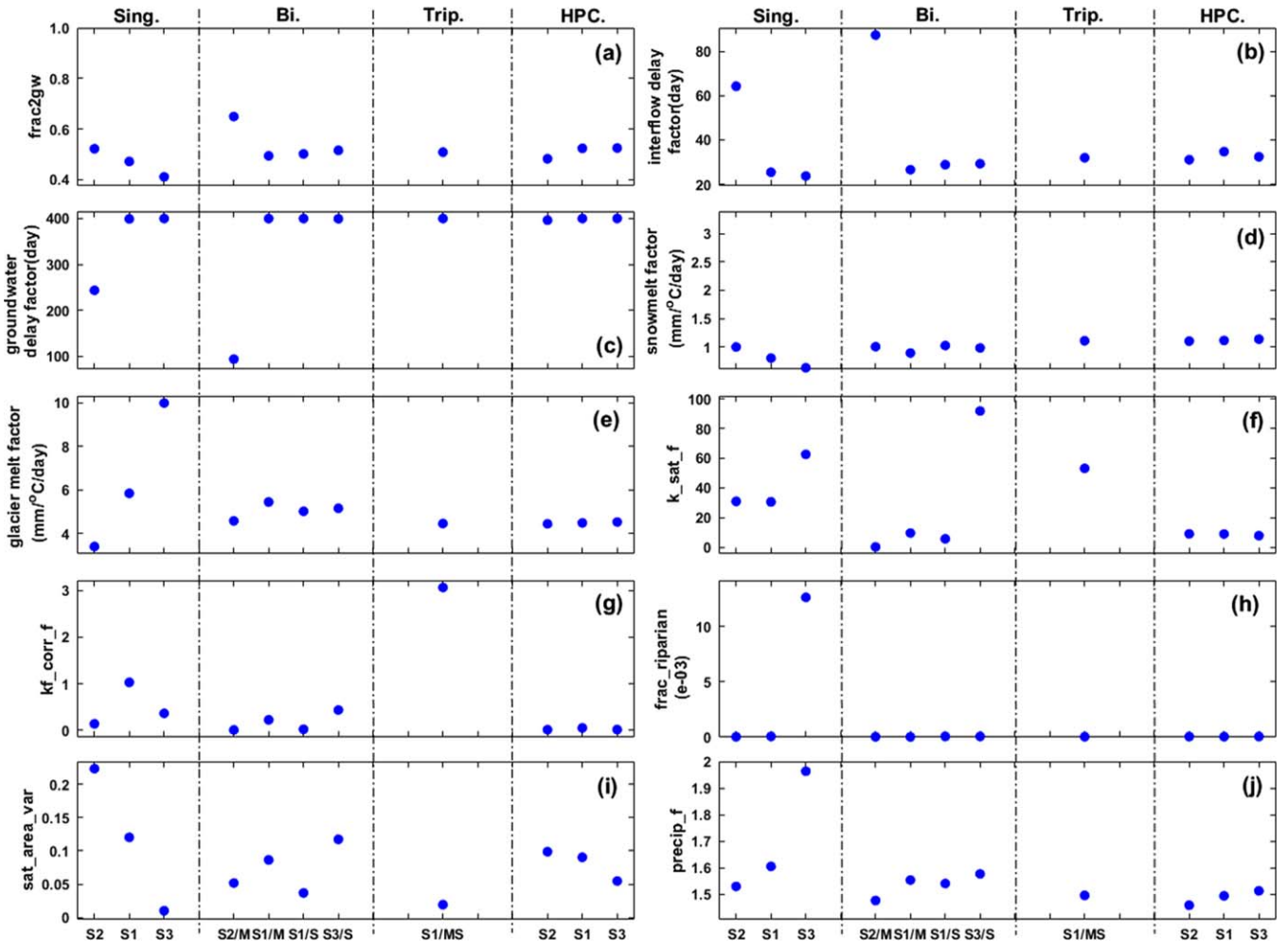

Figure 8. Behavioral parameter values calibrated by various methods in three calibration scenarios. "/M" indicates the bi-data set calibration using discharge and GMB; "/S" indicates bi-data set calibration using discharge and SCA; S1: parameter values calibrated during 1987–1993; S2: calibrated during 1973–1979; S3: calibrated during 1994–2000.

15% and 13%, and the single-data set method obtained a glacier melt contribution of 10%. Both the HPC-based and tri-data set methods tended to compute the lowest contributions of groundwater, rainfall, and snowmelt, but the highest contribution of glacier melt. The opposite holds true for the single-data set method. The HPCs based approach delivered similar contributions to the well-constrained tri-data set approaches despite being informed only by discharge observation (additionally temperature and precipitation). As the tri-data set method was constrained by all three calibration data sets, we assumed the contributions of runoff components to discharge estimated by this method to be the most reliable (similar to Finger et al., 2015 and Tarasova et al., 2016) and used this approach as a benchmark.

Figure 10 shows the contributions of runoff components at the seasonal scale. The contributions simulated by the five calibration methods show visible differences in the melt water in summer and autumn. For example, in the median wet year 1983, the single-data set method obtained 10% lower contribution of glacier melt in summer and 8% lower in autumn (Figures 10g and 10k) compared to the HPC-based and tri-data set methods. The QM-based bi-data set method computed 8% lower contribution of glacier melt in summer in the driest year, and the QS-based bi-data set method computed 4% lower contribution of glacier melt in autumn in the median wet year, compared to the HPC-based and tri-data set methods. The lower contributions of glacier melt estimated by the single-data set and bi-data set methods are compensated by the higher contributions of snowmelt and rainfall.

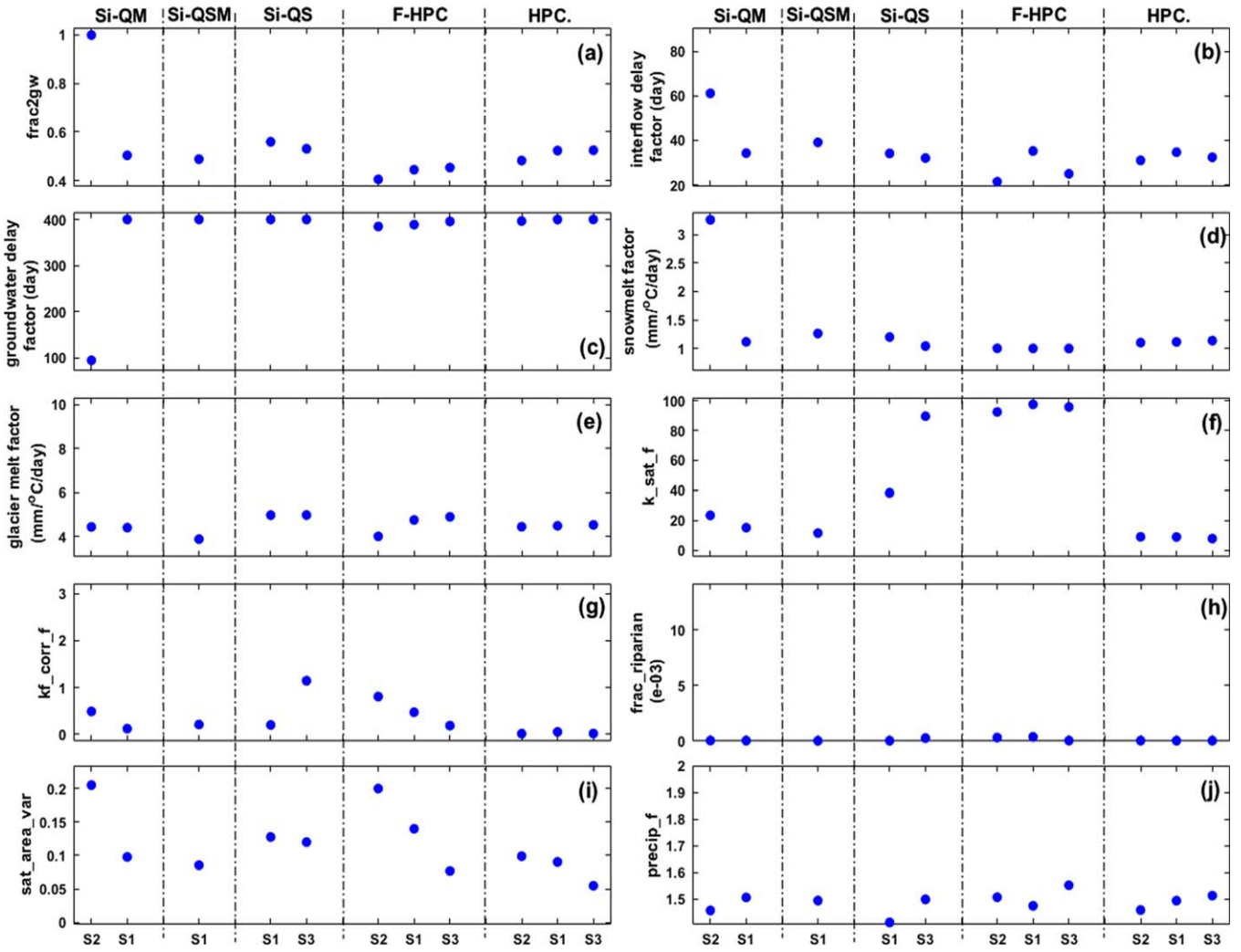

Figure 9. Behavioral parameter values calibrated by the multiple data sets based method using a stepwise-iterative approach and the four-objective method using four HPCs. Si-QM and Si-QS refer to the stepwise-iterative approach using discharge + GMB and discharge + SCA; Si-QMS: refers to the stepwise-iterative approach using discharge + GMB + SCA; F-HPC: refers to the four-objective method using four HPCs; HPC.: standard stepwise-iterative method using four HPCs.

4.5. Uncertainty Analysis

As the aim of our study is to compare the performance of various calibration methods for estimating model parameters and achieving reasonable performance criteria, the parameter uncertainty analysis is of main concern. Previous studies on hydrological modeling in glacierized basins have emphasized the importance of the calibration approach (Finger et al., 2015; Tarasova et al., 2016), and demonstrated that increasing model structure complexity does not necessarily help to increase the model performance. The uncertainty in precipitation input has been implicitly considered in all calibration procedures by adjusting the precipitation correction factor (*precip_f*) (Duethmann et al., 2013; Immerzeel et al., 2015). Considering that all the calibration methods were tested using the same model structure and driven by the same input climate data sets, the analysis of structural uncertainty and input data uncertainty has not been performed here. This would go beyond the scope of the presented work.

To quantify the parameter uncertainty, we added three trials for each calibration method with the initial population sizes for the ε-NSGAII algorithm of 24, 32, and 40, respectively. To estimate the posterior distributions of parameter values calibrated by the various methods in each trial, we selected the 100 best parameter sets from the Pareto-optimal fronts produced by the ε-NSGAII algorithm. For the single-data set calibration method, the parameter sets that produced the 100 highest *avNSE* values for discharge were

Table 3
Annual Contributions (%) of Runoff Components to River Discharge Simulated by Five Calibration Methods in Three Representative Years

Year	Runoff components	Sing.	Bi-QM.	Bi-QS.	Trip.	HPC.
1979 (wettest year)	GW	43	42	41	41	41
	SM	34	34	33	33	33
	GM	3	4	7	7	7
	RF	20	20	19	19	19
1983 (median wet year)	GW	37	36	35	34	34
	SM	33	32	31	30	30
	GM	10	13	15	17	18
	RF	20	19	19	19	18
1997 (driest year)	GW	31	30	28	28	29
	SM	30	27	25	24	24
	GM	21	25	29	30	29
	RF	18	18	18	18	18

Note. GW: groundwater; SM: snowmelt; GM: glacier melt; RF: rainfall.

selected. For the bi-data set and tri-data set calibration methods, the parameter sets that produced the 100 smallest Euclidean distances from the perfect point were selected. For the HPC-based method, the 100 best parameter sets were derived from the combination of the calibrated parameters in five individual steps. The ε-NSGAII algorithm optimized the parameter values from more than 20,000 samples in each calibration method. For example, in the trial with initial population size of 8, the QM-based bi-data set method optimized parameter values from 3,108 * 8 = 24,864 samples. In the absence of Monte Carlo analysis method,

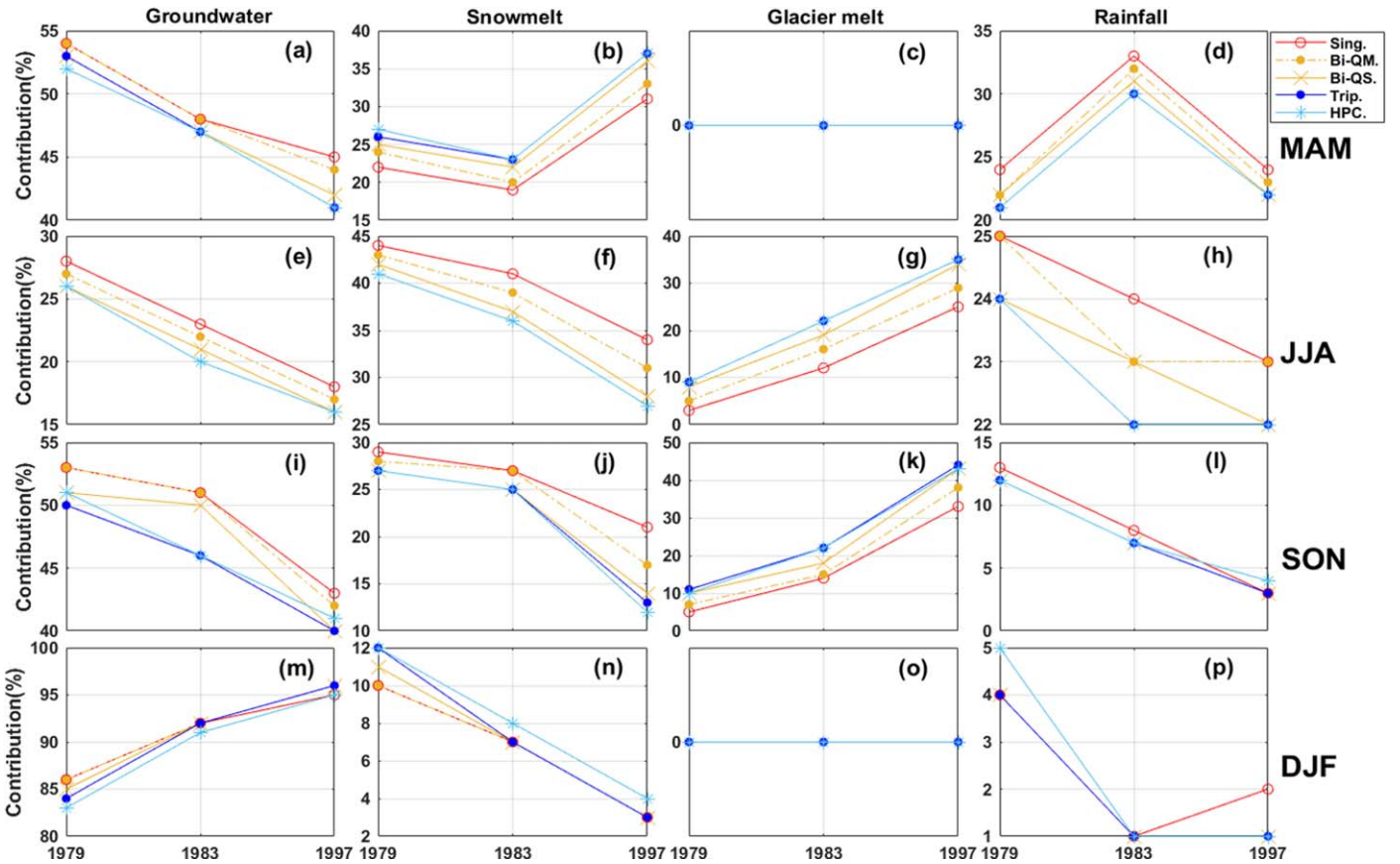

Figure 10. Seasonal contributions of runoff components computed by various calibration methods in three representative years.

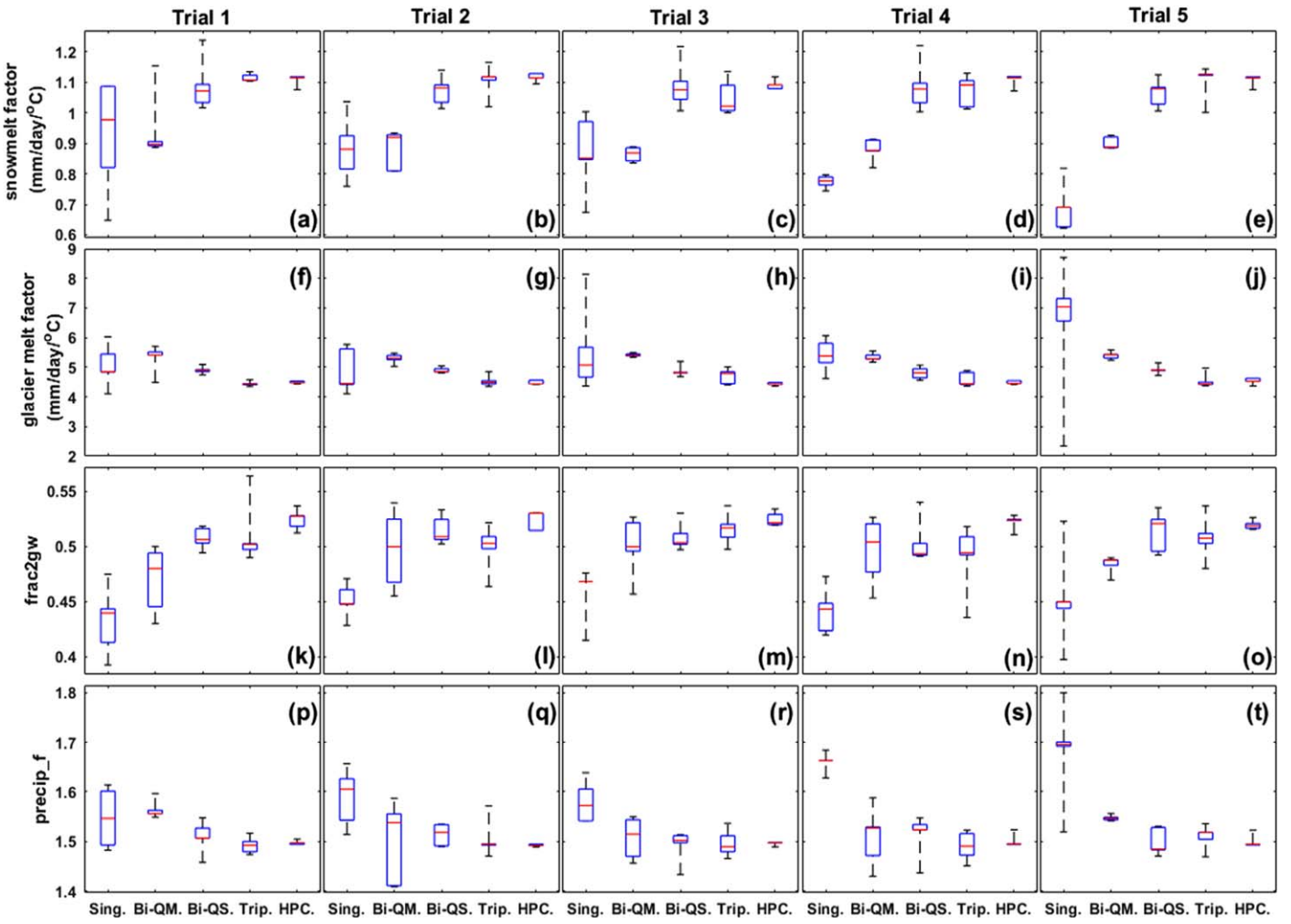

Figure 11. Posterior distributions of the 100 best parameter sets calibrated by various calibration methods in scenario S1 in five trials.

we assumed the selected 100 best parameter sets from the ε-NSGAII algorithm test samples could adequately represent parameter uncertainty, similar to Soulsby et al. (2015).

Figure 11 shows the posterior distributions of the 100 best parameter sets calibrated by the various methods in scenario S1 for five trials. To facilitate the presentation, only the posterior distributions of the four most sensitive parameters, namely the snowmelt and glacier melt factors, the groundwater parameter *frac2gw* and the precipitation correction factor *precip_f*, were included in this figure. The parameters calibrated by the single-data set method tend to show the largest uncertainty ranges. The bi-data set and tri-data set methods helped to narrow the ranges for most parameters. The HPC-based method tended to produce the narrowest uncertainty ranges for all four analyzed parameters. The analysis also indicates that the HPC-based method produced the most stable median parameter values across all trials. This further corroborates the findings with regards to the superior parameter identifiability in the HPC-based calibration.

The performance of the 100 best parameter sets in Figure 11 for both calibration and validation periods was compared in Figure 12. Again, we use the Euclidean distance to the perfect point (1, 1) to evaluate the performance in each subplot. In the calibration period 1987–1993 (Figures 12a–12j), the HPC-based method performed similarly to the QS-based (QM-based) bi-data set and the tri-data set methods, and outperformed the QM-based (QS-based) bi-data set and single-data set methods for the simultaneous simulation of discharge and SCA (GMB). In the validation period 1973–1979 (Figures 12k–12o), the HPC-based method performed similarly to the tri-data set method, and outperformed the other three methods. In the validation period 1994–2000 (Figures 12p–12t), the HPC-based method performed similarly to the QS-based bi-data

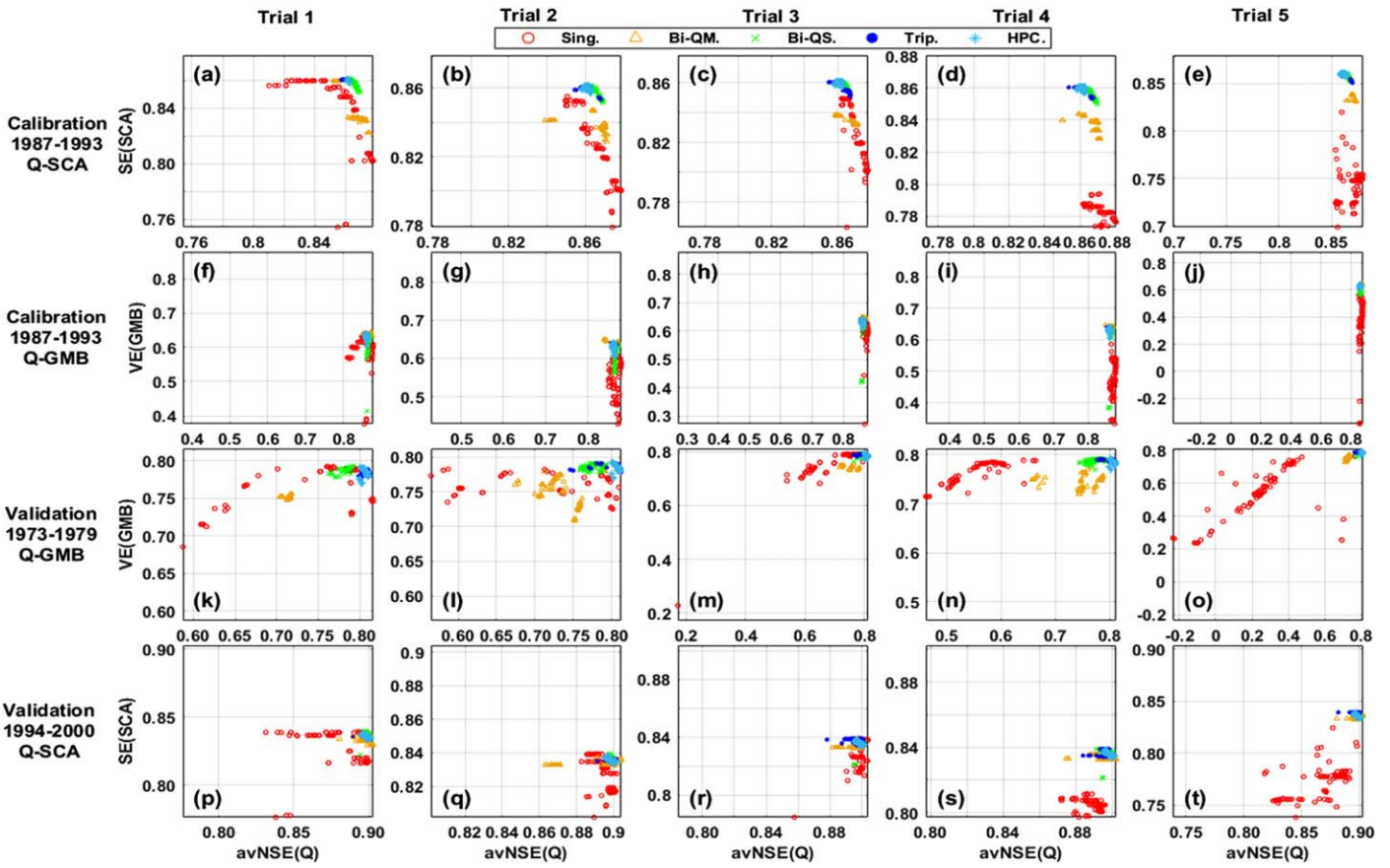

Figure 12. Performance for the simulations of discharge, SCA and GMB in both calibration and validation periods produced by the 100 best parameter sets calibrated in scenario S1 in five trials.

set and the tri-data set methods, and outperformed the other two methods. We can conclude that even considering the (fairly small) parameter uncertainty in the HPC-based calibration method, it performed better or similarly to the bi-data set and tri-data set methods.

4.6. Comparison With Other Hydrograph Signatures

In order to contrast the calibration power of the HPC-based method with other typical hydrograph signatures, we tested the flow duration curve (FDC) segments and intraannual variability for the model calibration. Combined with the *avNSE* for discharge, we carried out a three-objective calibration (abbreviated as Sgn-based method, see section 3.4). The parameter sets that produced the 100 smallest Euclidean distances from the perfect point (1, 1, 1) were selected to estimate the posterior parameter distributions in this method. Figure 13 compares the performance produced by the 100 best parameter sets in 5 trials during the calibration period 1987–1993. In all the five trials, the Sgn-based method failed to capture the dynamics in SCA and GMB, in spite of its good performance for discharge. The low performance for SCA and GMB in Figure 13 is partly caused by the overestimated contribution of precipitation, which results in significant underestimation in snow and glacier melt runoff. The analysis also indicates a high sensitivity of this method to the initial population size of the ε-NSGAII algorithm, which is in contrast to the HPC-based approach. Considering the low performance of the Sgn-based method, this method was not included for comparison in Figures 11, 12.

4.7. Applicability in Large Catchments

In the current study, the HPC-based method was evaluated in a moderate basin with an area of 233 km^2. To extract the HPCs from the observed hydrograph, we adopted the average basin temperature, as well as the average snowfall in the glacierized areas. This seems reasonable for a moderate basin like Ala-Archa, but implies larger uncertainty in large basins where precipitation and temperature exhibit stronger spatial

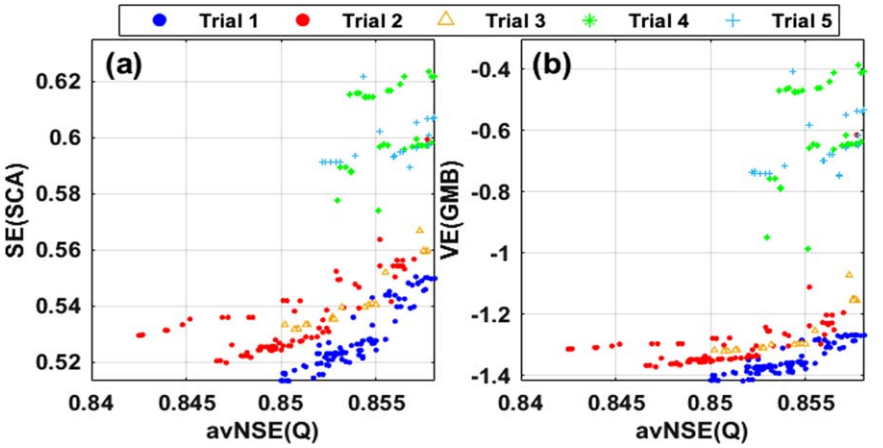

Figure 13. Performance of the Sgn-based calibration method in calibration period 1987–1993 in five trials.

variability. To test the extension of the HPC approach to large basins, we divided the study basin into four subbasins in a first step. The daily values of the runoff component indices (i.e., S_i, G_i, and D_i) were defined for each subbasin using individual ASF and ADT curves. Considering the cases that one runoff process is dominant for the river runoff in one subbasin, while others are dominant in other subbasins, we used the inverse distance weighted subbasin runoff component indices to identify the dominant runoff process for the hydrograph at the entire basin outlet. The distance here refers to the channel length between the subbasin outlet and the entire basin outlet. The identified dominant periods of runoff processes were then used for the extraction of HPCs at the entire basin outlet.

The performance of the extended HPC-based method in the first two trials is summarized in Table 2 (Ex-HPC.). As expected, the extended HPC-based method shows higher performance than the lumped HPC-based method, due to the consideration of spatial variability in dominant runoff components. The results in Table 2 indicate the potential suitability of the extended HPC-based method in large catchments. To facilitate the description of the method, we treated the study basin as one lumped catchment in earlier sections. We acknowledge that more case studies in other basins are needed to prove the robustness of the HPC-basin method, however, such comparisons require ground GMB measurements for large basins which are rather limited.

4.8. Sensitivity of the HPC-Based Calibration Method to the Pt Value

In the HPC-based method, the starting date of glacier melt is determined by the choice of the Pt value indicating the persistence of snow after the end of accumulation. The Pt value defines the $Q_{GW} + Q_{SM}$ and $Q_{GW} + [Q_{SM}] + Q_{GM}$ partitioning curves, and potentially influences the calibration results. We analyzed the sensitivity of the HPC-based method to the choice of the Pt value during the period 1987–1993, and compared it to the results of the tri-data set method (Figure 14). Five Pt values from 1 to 5 weeks were analyzed. Figures 14a–14d show the calibrated values of the four most sensitive parameters (the same parameters in Figure 11) using different Pt values. When using larger Pt values, more (fewer) summer peak flows are included in the $Q_{GW} + Q_{SM}$ ($Q_{GW} + [Q_{SM}] + Q_{GM}$) partitioning curve, resulting in a higher (lower) snow (glacier) melt factor (Figures 14a and 14b). The parameters *frac2gw* and *precip_f* tend to decrease with increasing Pt values (Figures 14c and 14d). It is also noted that the parameter values calibrated using Pt as 5 weeks are obviously different from those calibrated using Pt as 1–4 weeks, partly due to the incorrect division between snowmelt and glacier melt.

Figures 14e–14h present the performance produced by the corresponding parameter sets in Figures 14a–14d. For Pt = 5 weeks, the HPC-based method yielded the lowest performance for all three criteria, indicating the incorrect division between snowmelt and glacier melt, partly due to the fact that glacier melt starts earlier than 5 weeks after the snow accumulation. Pt = 1 week produced the highest performance for discharge (Figure 14e), but relatively low performance for SCA and GMB. Pt using 4 weeks generally yielded the highest model performance and matched better with the performance of the tri-data set calibration. Pt values of 2 and 3 weeks produced similar performance to Pt of 4 weeks (in term of the Euclidean distance

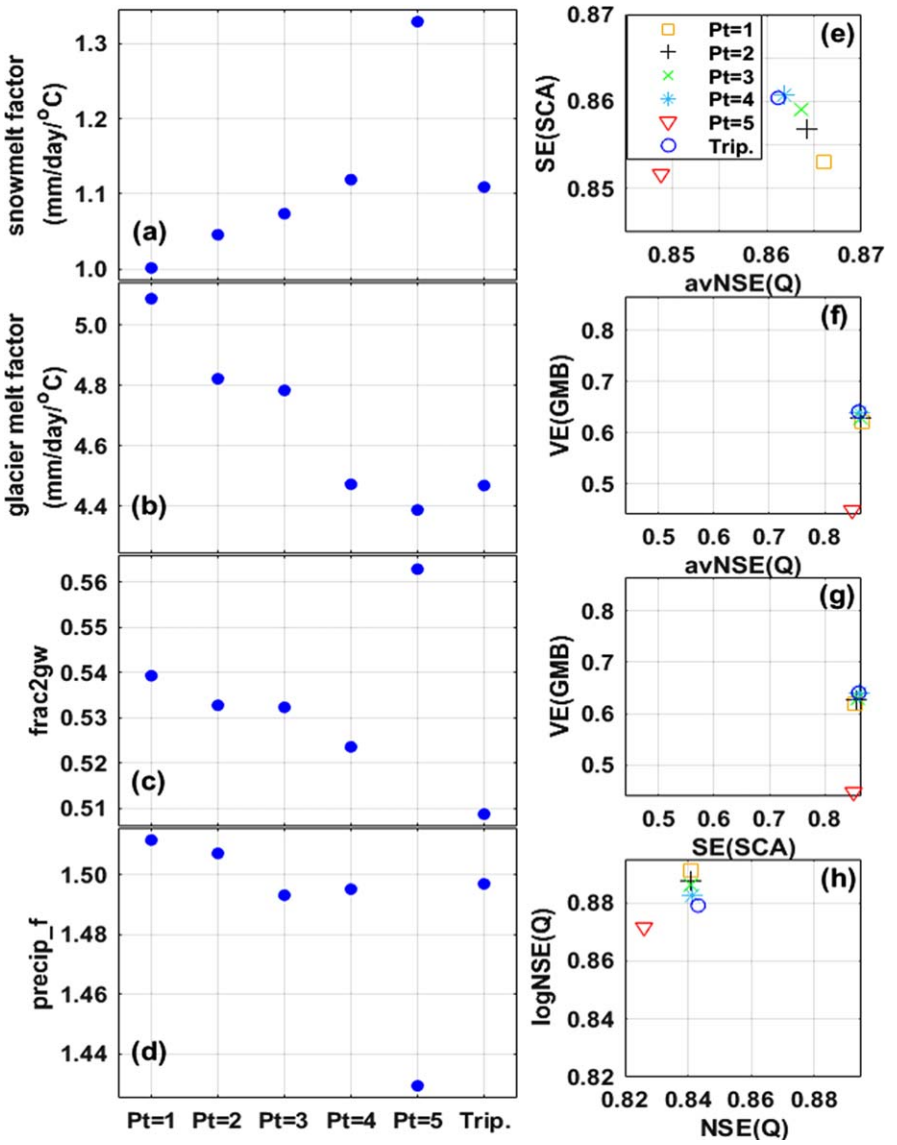

Figure 14. Effects of the Pt value on the calibrated parameter values and the model performance for the HPC-based method. Pt = 1 refers to the model results assuming that glacier melt starts to feed river streamflow one week after the last summer snowfall event in the glacierized areas, analogously for Pt = 2, 3, 4, and 5. "Trip." refers to the tri-data set calibration.

to point [1, 1]). In our case study, using Pt values ranging from 2 to 4 weeks should be acceptable for the implementation of the HPC-based method. In this range, the model performance is relatively insensitive to the selection of the Pt value. We then suggest to estimate it a priori based on the knowledge of the catchment conditions.

5. Discussion

5.1. The Stepwise-Iterative Procedure for Model Calibration

The stepwise calibration approach has been successfully applied for hydrological modeling in catchments characterized by various terrain and climate settings (Boyle et al., 2000; Hay et al., 2006; Hingray et al., 2010; Schaefli et al., 2005; Willems, 2014). They demonstrated that calibrating parameters on corresponding sensitive hydrograph features in a stepwise way can help to increase the robustness in parameter values

(Nearing & Gupta, 2015). In this study, the stepwise-iterative procedure using information represented in the HPCs significantly improved the parameter stability and identifiability.

In spite of the interaction of runoff processes, each HPC represents the dominant period of specific runoff process or of their combinations. The HPCs provide separated observations for the constraining of groundwater, snow, and glacier melt and rainfall surface runoff parameters. In the stepwise-iterative calibration procedure, model parameter groups were separately constrained to reduce the internal compensation of runoff processes. In each calibration step, a small group of model parameters was optimized by the simulation of the sensitive HPC, resulting in the reduction of degrees of freedom in the parameter space. The stepwise calibration procedure reduced the internal compensation of model parameters and narrowed the parameter variation ranges, helping to resolve the equifinality issue in model calibration. As shown in Figure 9, the HPC-based method narrowed the parameter variations in different time periods and improved the parameter identifiability using the stepwise-iterative procedure, compared to the traditional four-HPC calibration method. Hence, the stepwise-iterative procedure allows the information in the HPC to be more effectively utilized. Given the persistent need in the effective calibration approaches to reduce parameter uncertainty, the stepwise-iterative procedure shows the potential to be applied for hydrological modeling in other catchments beyond glacierized regions.

5.2. Limitations

The Pt value should be predefined for the implementation of the HPC-based method, which indicates the starting date of glacier melt in glacierized areas. We assumed that glacier melt starts several weeks after the end of snow accumulation on the glacierized area. Five different Pt values have been tested for our study basin. In the range of Pt values between 2 and 4 weeks the model performance was found relatively insensitive and this range was considered suitable to define the starting date of glacier melt. Theoretically, Pt value can be included into the calibration procedure, but it would heavily increase the already high computation burden. For each Pt value a separated HPC-curve pattern would need to be derived. Hence, we suggest that using a predefined Pt value based on knowledge of the catchment conditions, i.e., how long it approximately takes from ceased snow accumulation till the start of glacier ablation, is sufficient to identify this value.

Another limitation of the HPC-based method is the computation cost. We ran the ε-NSGAII algorithm on a Linux Cluster with 16 nodes. The automatic algorithm stopped before reaching the maximum runs, once the optimization was finished. In the HPC-based method, the algorithm finished the optimization with fewer computation runs in each step, since fewer parameters were optimized in each step. In steps 1–3 and 5, the automatic algorithm finished the optimization with 320 runs (each run contains eight initialization samples in the first trial) per step, on average. In the fourth step, calibrating four surface runoff parameters, the automatic algorithm required on average 600 runs. In total, 1,880 runs ($320 * 4 + 600$) were needed per iteration. The single-data set and tri-data set methods required 7,632 and 7,820 runs, respectively. Runs (3,108 and 5,896) were finalized for the QM-based and QS-based bi-data set methods. To investigate the variability in parameter values, we ran the HPC-based method for eight iterations, costing 15,040 runs. However, we suggest to run fewer iterations for the application of the HPC-based method. Figure 3 illustrates that the calibrated parameter values show small variance during the last four iterations, producing minor difference in the objective-function values for each calibration step. When running the HPC-based method for four iterations for practical application, its computation cost is then comparable to the multidata set calibration.

The objective of this study is to investigate the performance of the HPC-based method for constraining snow and glacier dynamics simulations without using additional observations besides discharge for calibration. Our results demonstrated that the HPC-based method is superior to the single-data set and bi-data set calibration methods, and is comparable to the tri-data set calibration method, in terms of the simultaneous simulation of discharge, annual GMB, and SCA. One superiority of the HPC-based method is that it does not require additional and costly snow cover and glacier mass balance observations. Moreover, the HPC-based method helps to improve the stability of parameter values calibrated across different periods, and to narrow the variation ranges in parameter values, delivering higher parameter identifiability. Therefore, the computation cost is worth to bear, given the increasing computation resources and the need for effective methods to reduce parameter uncertainty in hydrological modeling.

6. Conclusions

We further advanced a calibration method which disaggregates discharge time series into Hydrograph Partitioning Curves (HPCs) to represent the dominant periods of the different runoff processes. Model parameters were successively estimated by relating specific parameter groups to individual HPCs. The performance of the HPC-based calibration method for constraining the simulations of snow and glacier dynamics was investigated in comparison to multidata set methods, which used different combinations of observed discharge, annual glacier mass balance (GMB), and satellite snow cover images (SCA). The main results are summarized as follows:

1. The HPC-based method performs comparably to the tri-data set methods for the simultaneous simulation of discharge, annual GMB, and SCA. Cross validation in different periods indicates consistently good model performance in reproducing the dynamics of discharge, GMB, and SCA. The HPC-based method delivers good model-internal consistency without using additional constraining observations like SCA and GMB.
2. Parameter values calibrated by the HPC-based method show higher stability across different calibration periods, compared to the single-data set and bi-data set methods. Model parameters were separately calibrated on the corresponding HPC which represents the dominant period of a specific runoff process or of their combinations. The stepwise calibration procedure helps to reduce internal compensation of runoff processes, and thus to improve the parameter identifiability.
3. The estimation of contributions of runoff components to discharge heavily depends on the calibration method, especially for the glacier melt in summer. Given that the tri-data set calibration uses information from discharge and from both SCA and GMB, it is reasonable to assume that this method yields the most appropriate estimation of runoff components. The HPC-based method produced rather similar contributions of runoff components to the tri-data set method, without using additional SCA and GMB data for parameter calibration.

In glacierized basins, reasonable reproduction of snow and glacier dynamics is the highest importance for the hydrological model to properly quantify the contributions of runoff components. However, the glacier mass balance data for model calibration are usually not available in many glacierized basins. Discharge is commonly the most convenient observation for model calibration. Our results demonstrated that extracting hydrological information from the only available discharge observation in the form of HPCs, and running single-objective optimization procedure in a stepwise way, could achieve comparable model performance to a tri-data set calibration approach, which additionally requires SCA and GMB data sets. The HPC-based method could achieve a fairly low uncertainty in calibrated parameter values. This strongly favors the application of the HPC-based method in catchments without in situ monitoring of glacier mass balance and saves the complex processing of satellite snow cover information.

Acknowledgments
This work has been carried out in the frame of the GlaSCA (grant 01DK15002A) and GlaSCA-V (grant 88 501) projects funded by the German Federal Ministry for Science and Education (BMBF) and Volkswagen Foundation, respectively. The climate data to drive the model, as well as the calibrated model parameters are available at: https://doi.org/10.6084/m9.figshare.5457292.v1. We thank the Kyrgyz Hydromet Service for providing the meteorological and discharge data in our study area.

References

Aizen, V., Aizen, E., Glazirin, G., & Loaiciga, H. A. (2000). Simulation of daily runoff in Central Asian alpine watersheds. *Journal of Hydrology*, *238*, 15–34. https://doi.org/10.1016/S0022-1694(00)00319-X

Aizen, V. B., Aizen, E. M., & Melack, J. M. (1996). Precipitation, melt and runoff in the northern Tien Shan. *Journal of Hydrology*, *186*, 229–251. https://doi.org/10.1016/S0022-1694(96)03022-3

Aizen, V. B., Kuzmichenok, V. A., Surazakov, A. B., & Aizen, E. M. (2007). Glacier changes in the Tien Shan as determined from topographic and remotely sensed data. *Global and Planetary Change*, *56*, 328–340. https://doi.org/10.1016/j.gloplacha.2006.07.016

Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, *36*, 3663–3674. https://doi.org/10.1029/2000WR000207

Coxon, G., Freer, J., Wagener, T., Odoni, N. A., & Clark, M. (2014). Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments. *Hydrological Processes*, *28*, 6135–6150. https://doi.org/10.1002/hyp.10096

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*, 182–197. https://doi.org/10.1109/4235.996017

Dietz, A. J., Conrad, C., Kuenzer, C., Gesell, G., & Dech, S. (2014). Identifying changing snow cover characteristics in Central Asia between 1986 and 2014 from Remote Sensing Data. *Remote Sensing*, *6*, 12752–12775. https://doi.org/10.3390/rs61212752

Duethmann, D., Bolch, T., Farinotti, D., Kriegel, D., Vorogushyn, S., Merz, B., et al. (2015). Attribution of streamflow trends in snow and glacier melt-dominated catchments of the Tarim River, Central Asia. *Water Resources Research*, *51*, 4727–4750. https://doi.org/10.1002/2014WR016716

Duethmann, D., Menz, C., Jiang, T., & Vorogushyn, S. (2016). Projections for headwater catchments of the Tarim River reveal glacier retreat and decreasing surface water availability but uncertainties are large. *Environmental Research Letters*, *11*, 054024. https://doi.org/10.1088/1748-9326/11/5/054024

Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., & Güntner, A. (2014). The value of satellite- derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. *Water Resources Research*, *50*, 2002–2021. https://doi.org/10.1002/2013WR014382

Duethmann, D., Zimmer, J., Gafurov, A., Güntner, A., Kriegel, D., Merz, B., et al. (2013). Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling. *Hydrology and Earth System Sciences*, *17*, 2415–2434. https://doi.org/10.5194/hess-17-2415-2013

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

Farinotti, D., Longuevergne, L., Moholdt, G., Duethmann, D., Mölg, T., Bolch, T., et al. (2015). Substantial glacier mass loss in the Tien Shan over the past 50 years. *Nature Geoscience*, *8*, 716–722. https://doi.org/10.1038/ngeo2513

Finger, D., Vis, M., Huss, M., & Seibert, J. (2015). The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments. *Water Resources Research*, *51*, 1939–1958. https://doi.org/10.1002/2014WR015712

Fovet, O., Ruiz, L., Hrachowitz, M., Faucheux, M., & Gascuel-Odoux, C. (2015). Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models. *Hydrology and Earth System Sciences*, *19*, 105–123. https://doi.org/10.5194/hess-19-105-2015

Frey, H., Machguth, H., Huss, M., Huggel, C., Bajracharya, S., Bolch, T., et al. (2014). Estimating the volume of glaciers in the Himalayan-Karakoram region using different methods. *The Cryosphere*, *8*, 2313–2333. https://doi.org/10.5194/tc-8-2313-2014

Gafurov, A., Lüdtke, S., Unger-Shayesteh, K., Vorogushyn, S., Schöne, T., Schmidt, S., et al. (2016). MODSNOW-Tool: An operational tool for daily snow cover monitoring using MODIS data. *Environmental Earth Sciences*, *75*, 1078. https://doi.org/10.1007/s12665-016-5869-x

Green, W. H., & Ampt, G. (1911). Studies of soil physics, part I—The flow of air and water through soils. *Journal of Agricultural Science*, *4*, 1–24.

Güntner, A., & Bronstert, A. (2004). Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas. *Journal of Hydrology*, *297*, 136–161. https://doi.org/10.1016/j.jhydrol.2004.04.008

Gupta, H. V., Wagener, T., & Liu, Y. Q. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*, 3802–3813. https://doi.org/10.1002/hyp.6989

Guse, B., Pfannerstill, M., Gafurov, A., Fohrer, N., & Gupta, H. (2016). Demasking the integrated information of discharge—Advancing sensitivity analysis to consider different hydrological components and their rates of change. *Water Resources Research*, *52*, 8724–8743. https://doi.org/10.1002/2016WR018894

Hamlet, A. F., Mote, P. W., Clark, M. P., & Lettenmaier, D. P. (2005). Effects of temperature and precipitation variability on snowpack trends in the western United States. *Journal of Climate*, *18*, 4545–4561. https://doi.org/10.1175/Jcli3538.1

Hay, L. E., Leavesley, G. H., Clark, M. P., Markstrom, S. L., Viger, R. J., & Umemoto, M. (2006). Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin. *Journal of the American Water Resources Association*, *42*, 877–890. https://doi.org/10.1111/j.1752-1688.2006.tb04501.x

He, Z. H., Parajka, J., Tian, F. Q., & Blöschl, G. (2014). 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

He, Z. H., Tian, F. Q., Gupta, H. V., Hu, H. C., & Hu, H. P. (2015). Diagnostic calibration of a hydrological model in a mountain area by hydrograph partitioning. *Hydrology and Earth System Sciences*, *19*, 1807–1826. https://doi.org/10.5194/hess-19-1807-2015

Hingray, B., Schaefli, B., Mezghani, A., & Hamdi, Y. (2010). 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

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

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

Huss, M. (2012). Extrapolating glacier mass balance to the mountain-range scale: The European Alps 1900–2100. *The Cryosphere*, *6*, 713–727. https://doi.org/10.5194/tc-6-713-2012

Huss, M., Jouvet, G., Farinotti, D., & Bauder, A. (2010). Future high-mountain hydrology: A new parameterization of glacier retreat. *Hydrology and Earth System Sciences*, *14*, 815–829. https://doi.org/10.5194/hess-14-815-2010

Immerzeel, W. W., Wanders, N., Lutz, A. F., Shea, J. M., & Bierkens, M. F. P. (2015). Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff. *Hydrology and Earth System Sciences*, *19*, 4673–4687. https://doi.org/10.5194/hess-19-4673-2015

Jothityangkoon, C., Sivapalan, M., & Farmer, D. L. (2001). Process controls of water balance variability in a large semi-arid catchment: Downward approach to hydrological model development. *Journal of Hydrology*, *254*, 174–198. https://doi.org/10.1016/S0022-1694(01)00496-6

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., & Arnaud, Y. (2012). Contrasting patterns of early twenty first-century glacier mass change in the Himalayas. *Nature*, *488*, 495–498. https://doi.org/10.1038/nature11324

Kollat, J. B., & Reed, P. M. (2006). Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Advances in Water Resources*, *29*, 792–807. https://doi.org/10.1016/j.advwatres.2005.07.010

Konz, M., & Seibert, J. (2010). 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

Le Lay, M., Saulnier, G. M., Galle, S., Seguis, L., Metadier, M., & Peugeot, C. (2008). Model representation of the Sudanian hydrological processes: Application on the Donga catchment (Benin). *Journal of Hydrology*, *363*, 32–41. https://doi.org/10.1016/j.jhydrol.2008.09.006

Linsbauer, A., Paul, F., & Haeberli, W. (2012). Modeling glacier thickness distribution and bed topography over entire mountain ranges with GlabTop: Application of a fast and robust approach. *Journal of Geophysical Research-Earth Surface*, *117*, F03007. https://doi.org/10.1029/2011jf002313

Liston, G. E. (2004). Representing subgrid snow cover heterogeneities in regional and global models. *Journal of Climate*, *17*, 1381–1397. https://doi.org/10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2

Liucci, L., Valigi, D., & Casadei, S. (2014). A new application of flow duration curve (FDC) in designing run-of-river power plants. *Water Resources Management*, *28*, 881–895. https://doi.org/10.1007/s11269-014-0523-4

Luo, Y., Arnold, J., Allen, P., & Chen, X. (2012). Baseflow simulation using SWAT model in an inland river basin in Tianshan Mountains, Northwest China. *Hydrology and Earth System Sciences*, *16*, 1259–1267. https://doi.org/10.5194/hess-16-1259-2012

Madsen, H. (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, *235*, 276–288. https://doi.org/10.1016/S0022-1694(00)00279-1

Mayr, E., Hagg, W., Mayer, C., & Braun, L. (2013). Calibrating a spatially distributed conceptual hydrological model using runoff, annual mass balance and winter mass balance. *Journal of Hydrology*, *478*, 40–49. https://doi.org/10.1016/j.jhydrol.2012.11.035

Nearing, G. S., & Gupta, H. V. (2015). The quantity and quality of information in hydrologic models. *Water Resources Research*, *51*, 524–538. https://doi.org/10.1002/2014WR015895

Parajka, J., & Blöschl, G. (2008a). 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

Parajka, J., & Blöschl, G. (2008b). Spatio-temporal combination of MODIS images—Potential for snow cover mapping. *Water Resources Research*, *44*, W03406. https://doi.org/10.1029/2007WR006204

Parajka, J., Naeimi, V., Bloschl, G., & Komma, J. (2009). Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria. *Hydrology and Earth System Sciences*, *13*, 259–271. https://doi.org/10.5194/hess-13-259-2009

Peters, J., Bolch, T., Gafurov, A., & Prechtel, N. (2015). Snow cover distribution in the Aksu Catchment (Central Tien Shan) 1986–2013 based on AVHRR and MODIS data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8–11*, 5361–5375. https://doi.org/10.1109/JSTARS.2015.2477108

Pfannerstill, M., Guse, B., & Fohrer, N. (2014). Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *Journal of Hydrology*, *510*, 447–458. https://doi.org/10.1016/j.jhydrol.2013.12.044

Razavi, S., & Tolson, B. A. (2013). An efficient framework for hydrologic model calibration on long data periods. *Water Resources Research*, *49*, 8418–8431. https://doi.org/10.1002/2012WR013442

Reusser, D., Blume, T., Schaefli, B., & Zehe, E. (2009). Analysing the temporal dynamics of model performance for hydrological models. *Hydrology and Earth System Sciences*, *13*, 999–1018. https://doi.org/10.5194/hess-13-999-2009

Richter, B. D., Baumgartner, J. V., Powell, J., & Braun, D. P. (1996). A method for assessing hydrologic alteration within ecosystems. *Conservation Biology*, *10*, 1163–1174. https://doi.org/10.1046/j.1523-1739.1996.10041163.x

Schaefli, B., Hingray, B., Niggli, M., & Musy, A. (2005). A conceptual glacio-hydrological model for high mountainous catchments. *Hydrology and Earth System Sciences*, *9*, 95–109. https://doi.org/10.5194/hess-9-95-2005

Schaefli, B., & Huss, M. (2011). 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

Seibert, J. (2000). Multi-dataset calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, *4*, 215–224. https://doi.org/10.5194/hess-4-215-2000

Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, *38*(11), 1241. https://doi.org/10.1029/2001WR000978

Shafii, M., & Tolson, B. A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*, *51*, 3796–3814. https://doi.org/10.1002/2014WR016520

Shangguan, D. H., Bolch, T., Ding, Y. J., Krohnert, M., Pieczonka, T., Wetzel, H. U., et al. (2015). Mass changes of Southern and Northern Inylchek Glacier, Central Tian Shan, Kyrgyzstan, during similar to 1975 and 2007 derived from remote sensing data. *Cryosphere*, *9*, 703–717. https://doi.org/10.5194/tc-9-703-2015

Singh, S. K., & Bardossy, A. (2012). Calibration of hydrological models on hydrologically unusual events. *Advances in Water Resources*, *38*, 81–91. https://doi.org/10.1016/j.advwatres.2011.12.006

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

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

Tarasova, L., Knoche, M., Dietrich, J., & Merz, R. (2016). Effects of input discretization, model complexity, and calibration strategy on model performance in a data-scarce glacierized catchment in Central Asia. *Water Resources Research*, *52*, 4674–4699. https://doi.org/10.1002/2015WR018551

Willems, P. (2014). Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes—Part 1: Step-wise model-structure identification and calibration approach. *Journal of Hydrology*, *510*, 578–590. https://doi.org/10.1016/j.jhydrol.2014.01.017

Ye, S., Yaeger, M., Coopersmith, E., Cheng, L., & Sivapalan, M. (2012). Exploring the physical controls of regional patterns of flow duration curves—Part 2: Role of seasonality, the regime curve, and associated process controls. *Hydrology and Earth System Sciences*, *16*, 4447–4465. https://doi.org/10.5194/hess-16-4447-2012

Zhou, H., Aizen, E., & Aizen, V. (2013). Deriving long term snow cover extent dataset from AVHRR and MODIS data: Central Asia case study. *Remote Sensing of Environment*, *136*, 146–162. https://doi.org/10.1016/j.rse.2013.04.015