

This is the accepted manuscript version of the contribution published as:

He, Z., Unger-Shayesteh, K., Vorogushyn, S., **Weise, S.M.**, Kalashnikova, O., Gafurov, A., Duethmann, D., Barandun, M., Merz, B. (2019):
Constraining hydrological model parameters using water isotopic compositions in a glacierized basin, Central Asia
J. Hydrol. **571**, 332 - 348

The publisher's version is available at:

<http://dx.doi.org/10.1016/j.jhydrol.2019.01.048>

Constraining hydrological model parameters using water isotopic compositions in a glacierized basin, Central Asia

Zhihua He^{1,a}, Katy Unger-Shayesteh^{1,b}, Sergiy Vorogushyn¹, Stephan M. Weise³, Olga Kalashnikova⁴, Abror Gafurov¹, Doris Duethmann⁵, Martina Barandun⁶, Bruno Merz^{1,2}

¹GFZ German Research Centre for Geosciences, Section Hydrology, Telegrafenberg, Potsdam, Germany.

²University of Potsdam, Institute of Earth and Environmental Science, Potsdam, Germany

³UFZ Helmholtz Centre for Environmental Research UFZ, Department Catchment Hydrology, Halle Germany

⁴CAIAG Central Asian Institute of Applied Geosciences, Department Climate, Water and Natural Resources, Bishkek, Kyrgyzstan

⁵Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology (TU Wien), Vienna, Austria

⁶Department of Geosciences, University of Fribourg, Fribourg, Switzerland

^aNow at Centre for Hydrology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

^bNow at German Aerospace Center (DLR), International Relations, Linder Höhe, Cologne, Germany

Manuscript submitted to *Journal of Hydrology*

Abstract

Water stable isotope signatures can provide valuable insights into the catchment internal runoff processes. However, the ability of the water isotope data to constrain the internal apportionments of runoff components in hydrological models for glacierized basins is not well understood. This study developed an approach to simultaneously model the water stable isotopic compositions and runoff processes in a glacierized basin in Central Asia. The fractionation and mixing processes of water stable isotopes in and from the various water sources were integrated into a glacio-hydrological model. The model parameters were calibrated on discharge, snow cover and glacier mass balance data, and additionally isotopic composition of streamflow. We investigated the value of water isotopic compositions for the calibration of model parameters, in comparison to calibration methods without using such measurements. Results indicate that: 1) The proposed isotope-hydrological integrated modeling approach was able to reproduce the isotopic composition of streamflow, and improved the model performance in the evaluation period; 2) Involving water isotopic composition for model calibration reduced the model parameter uncertainty, and helped to reduce the uncertainty in the quantification of runoff components; 3) The isotope-hydrological integrated modeling approach quantified the contributions of runoff components comparably to a three-component tracer-based end-member mixing analysis method for summer peak flows, and required less water tracer data. Our findings demonstrate the value of water isotopic compositions to improve the quantification of runoff components using hydrological models in glacierized basins.

1. Introduction

Glacierized basins substantially provide freshwater for the downstream agriculture and potable water supply, especially in warm and dry years (Prasch et al., 2013). However, changes in climate are altering the short- and long-term characteristics of runoff processes in these basins (Stahl and Moore, 2006; Stahl et al., 2008; Duethmann et al., 2015). Understanding changes in runoff generation processes is therefore critical to the downstream water resource utilization, considering the particular vulnerability of snow and glacier-dominated environments to changing climatic conditions (Barnett et al., 2005; Penna et al., 2014). Sound quantification of glacier melt, snowmelt, rainfall and groundwater contributions to the streamflow helps to understand the changes in freshwater availability provided by glacierized headwater basins (Jost et al., 2012).

Hydrological modeling is an effective tool to quantify changes in individual runoff components, providing insights into the catchment dynamics. A number of studies have modeled the contributions of runoff components to streamflow in glacierized basins (Jost et al., 2012; Lutz et al., 2014; He et al., 2015). For example, Verbunt et al. (2003) investigated the contributions of snowmelt and glacier melt to total runoff using a spatial-gridded hydrological model in high alpine catchments in Switzerland. Engelhardt et al. (2014) analyzed the spatial variations and temporal evolution of the water sources, including snowmelt, glacier melt and rainfall, by means of a distributed hydrological model in a Norway glacierized basin. Zhao et al. (2013) estimated the contribution of glacier melt to runoff using hydrological modeling for a glacierized catchment in Central Asia. However, the large calibration uncertainty in hydrological model parameters could lead to incorrect estimation of the runoff component shares (Nolin et al., 2010; Finger et al., 2015). In glacierized basins, the hydrological model typically integrates complex runoff generation processes. In addition, uncertainties in the forcing data for the hydrological models are high due to the commonly sparse climatic gauge

network (Immerzeel et al., 2014; Tarasova et al., 2016). Both effects can lead to considerable compensation between different runoff components in the hydrological model. A hydrological model could produce rather similar performance on the simulation of the total runoff by overestimating (underestimating) one runoff component to compensate the underestimation (overestimation) of the other runoff component. For example, a hydrological model could overestimate (underestimate) the glacier melt runoff and underestimate (overestimate) the precipitation-triggered runoff (i.e., sum of rainfall and snowmelt) to produce high performance for the simulation of total runoff, due to the insufficient accuracy in model inputs (Duethmann et al., 2013; Duethmann et al., 2014; He et al., 2014; Immerzeel et al., 2015). Therefore, multi-criteria calibration methods using coupled hydrological observations, including glacier mass balance, satellite remotely-sensed snow cover area and observed discharge, were adopted to reduce the model uncertainty (Koboltschnig et al., 2008; Parajka and Blöschl, 2008; Konz and Seibert, 2010; Schaefli and Huss, 2011; Duethmann et al., 2014; Duethmann et al., 2015; Finger et al., 2015).

However, there is still significant uncertainty in the quantification of runoff components even when calibrating the hydrological model using both discharge and glacier/snow cover observations (Duethmann et al., 2015). Water stable isotopic compositions provide insights into the dominance of various runoff processes on total runoff, bearing the potential to further constrain the internal apportionments of runoff components (Soulsby and Tetzlaff, 2008; van Huijgevoort et al., 2016).

Recently, water stable isotope data have been increasingly integrated into hydrological models to improve the understanding of dynamics in runoff processes. Ala-aho et al. (2017a) simultaneously simulated the flux, storage and mixing of water and water isotopes in three snow-influenced catchments. Some other studies in non-glacierized basins used water isotopes to reduce the model calibration uncertainty, attempting to achieve the 'right answers for the

right reasons' (Seibert et al., 2003; Weiler et al., 2003; Stadnyk et al., 2005; Dunn and Bacon, 2008; Stadnyk et al., 2013). For instance, Birkel et al. (2010) highlighted that isotope-tracer data helped to constrain the acceptable behavioral parameter sets and reduce model's degrees of freedom. Capell et al. (2012) found that using isotope-tracer data reduced parameter uncertainties and helped improving the model structure for reproducing spatial variabilities of runoff processes. Tetzlaff et al. (2015) demonstrated that integrating water isotopes into hydrological models could help to test the conceptualizations of physical processes in the model. However, the value of the water stable isotope measurements for model calibration in glacierized basins has not been investigated in previous studies. As far as we know, water isotopic compositions have not yet been integrated into glacio-hydrological models, partly due to the logistical challenges related to long-term fieldwork and water sampling in cold and high-altitude environments. Therefore, the extent to which the water isotope data can constrain the internal apportionments of runoff components in hydrological models for glacierized basins is not clear.

The tracer-based end-member mixing analysis approach is another widely used method for the quantification of runoff components in glacierized basins (Dahlke et al., 2014; Engel et al., 2016). For instance, Pu et al. (2017) estimated the contribution of glacier and snow meltwater to streamflow in a glacierized basin in the southeast margin of the Tibetan Plateau using the signatures of $\delta^{18}\text{O}$ and $\delta^2\text{H}$. Maurya et al. (2011) employed a three-component end-member mixing method to estimate the fraction of glacier melt runoff in a Himalayan river using $\delta^{18}\text{O}$ and electrical conductivity. However, uncertainties in the contributions of runoff components estimated by the tracer-based end-member mixing method are typically large, partly due to the strong spatial and temporal variability in the water tracers (Joerin et al., 2002). Difficulties in field sampling and seasonal inaccessibility of water samples in glacierized basins enhance the uncertainty in the quantification of runoff components (Rahman et al., 2015).

Against this background, we integrated the water isotopic compositions of runoff components into the WASA hydrological model (model of Water Availability in Semi-Arid environments, Güntner and Bronstert 2004) and developed an isotope-hydrological integrated modeling approach in a glacierized basin in Central Asia. The objective of this study is to investigate the value of the water stable isotope data for model calibration, as well as the ability of the proposed modeling approach for constraining the internal apportionments of runoff components for a glacierized basin. To achieve this goal, we compared an isotope-aided calibration approach with a common multi-criteria calibration approach which used discharge, remotely-sensed snow cover area and glacier mass balance data in the study basin. Specific questions addressed are two-fold: 1) What benefits can be gained for model calibration by incorporating the simulation of water isotopic composition into hydrological model in glacierized basins? 2) Is the isotope-hydrological integrated modeling approach superior to a three-component tracer-based end-member mixing analysis method (abbreviated as tracer-based mixing method hereinafter) for the quantification of runoff components in summer peak flow periods?

The paper is organized as follows: A brief description of the study area and data collection is provided in section 2. In section 3, we describe the proposed isotope-hydrological integrated modeling approach, as well as the model calibration methods. Section 4 presents the performance of the isotope-hydrological model, parameter uncertainties, and the contributions of runoff components. We discuss our results in relation to the literature and the limitations of this study in section 5, followed by conclusions in section 6.

2. Study area and data collection

2.1 Study area

The Ala-Archa basin is located in Kyrgyzstan, Central Asia, at the northern edge of the Tien Shan mountain system (74°24' E–74°34' E; 42°25' N–42°39' N, Fig. 1). It has an area of 233 km², approximately 17% of which is glacierized. The basin spreads throughout an elevation range from 1560 to 4864 m above sea level (a.s.l.). Approximately 83% of the Ala-Archa glacierized area consists of large valley glaciers, with about 76% of the total glacierized area located between 3700 to 4100 m a.s.l. (Aizen et al., 2007). The annual mean vegetation coverage is around 28% in this basin. Snowmelt runoff feeds the river streamflow during the melt period from March to September. Glacier melt runoff mainly feeds the river streamflow from July to September. The largest daily discharge generally occurs in July-August with a magnitude of around 25-40 m³/s, while the lowest daily discharge in January-February is only around 1.9 m³/s. The long-term annual mean precipitation and temperature are 560 mm and 2.6 °C, respectively, based on the data series recorded at the Alplager station at 2100 m a.s.l. for the period 1970-2000. Precipitation mainly occurs in the form of snowfall in winter and has peak storms in the spring-summer months (Aizen et al., 1995, 2000 and 2007). Water availability in the Ala-Archa River is critical for the downstream irrigated agriculture and the potable water supply.

2.2 Hydrometeorological and cryosphere data

Since the 1970s, daily precipitation, air temperature, humidity and global radiation have been recorded at the Baitik (1580 m a.s.l.) and Alplager (2100 m a.s.l.) meteorological stations run by Kyrgyz Hydromet Service (Fig. 1). The recorded relative humidity in air at the Baitik and Alplager station are generally larger than 50% in both summer and cold seasons. Mean daily streamflow has been measured at the Ala-Archa hydrologic station (close to the Baitik meteorological station) at the basin outlet since the 1960s.

The annual glacier mass balance (GMB) of the Golubin glacier (red polygon in Fig. 1, ranging from 3320 to 4350 m a.s.l.) has been measured during 1973-1993 and 2011-present

(Hoelzle et al., 2017). The glaciological ablation and accumulation measurements were distributed over the entire glacier surface and repeated annually. The point measurements were extrapolated to the glacier surface using a model-based inverse distance algorithm (e.g. Huss et al., 2009; Barandun et al., 2015) and the area weighted mean provides the annual glacier-wide mass balance. Additionally, the mass balance was calculated for every 100 m elevation band and used to calculate the mass balance gradient with elevation (Hoelzle et al. 2017). Here, we used specific mass balance calculated for each elevation band for hydrological model calibration. The data was collected, analyzed and reported to the World Glacier Monitoring Service (WGMS) under the standard protocol (Hoelzle et al., 2017).

The Moderate Resolution Imaging Spectroradiometer (MODIS) daily snow cover products (MOD10A1 and MYD10A1) with a spatial resolution of 500 m processed by the cloud removal procedure MODSNOW-Tool (Gafurov et al., 2016), were also explored for model calibration.

2.3 Isotope data

Weekly streamflow samples have been directly collected from the river channel near the Baitik (Ala-Archa hydrologic station) and Alplager meteorological stations since July 2013. The streamflow grab samples were collected by station operators from the river around noon (but maybe not exactly at the same time) every Wednesday. Precipitation samples have been collected at the Baitik and Alplager meteorological stations since January 2013. The precipitation events were collected from plastic rain collectors at the meteorological stations and accumulated over one month in a rain container, from where a mixed sample was collected for monthly isotopic analysis. To act against the effect of evaporation, the rain container for monthly precipitation accumulation was filled by a thin mineral oil layer before the collection of precipitation and stored in a shade room. Precipitation volumes were collected as immediately as possible after the rainfall/snowfall event.

Glacier melt grab samples were collected annually during the summer field campaigns. Flowing meltwater on the Golubin glacier surface at various elevation bands in the ablation zone were collected to consider the spatial variability (i.e., points G1-4 in Fig. 1, with elevation ranging from 3280 m to 3520 m).

Snowmelt and groundwater grab samples were occasionally collected during the warm season (March to October), due to the heavy snow coverage and limited catchment accessibility in the cold season. At each sampling site (points S1-3 in Fig. 1), snow samples were collected by using a pure polyethylene plastic tube. We drilled the pure plastic tube into the snowpack to collect the whole snow layer. All the snow collected by the tube was poured into a plastic bag and stored in a cooling box. Snowmelt samples were then collected from the meltwater inside the bag when all the snow had melted out. The depth of the sampled snow layers differed at the sampling sites, ranging from 10 cm to 150 cm. Groundwater samples were collected from one spring draining to the river at the foot of a hillslope at point W1 (Fig. 1).

All samples were collected in pure polyethylene plastic bottles and stored at 4 °C before the analysis in the Helmholtz Center for Environmental Research (UFZ) laboratory (Department of catchment hydrology, Halle in Germany). Isotopic compositions were analyzed with Laser-based infrared spectrometry (LGR TIWA 45, Picarro L1102-i) and calibrated against the common VSMOW scale with a precision of $\delta^{18}\text{O}$: ± 0.25 ‰ and $\delta^2\text{H}$: ± 0.4 ‰, respectively. Electrical Conductivity (EC) in the water samples was measured in situ or in laboratory using portable EC/PH/TDS meters. Strict quality control procedures have been implemented for the isotope and EC data, discarding abnormal values caused by evaporation effects and sampling errors (for example, samples that experienced significant evaporation before the lab isotope analysis, and abnormal EC values caused by sediments from the sampling sites and significant laboratory measurement error were discarded).

The measured tracer characteristics of various water sources are summarized in Table 1. The $\delta^{18}\text{O}$ of precipitation generally shows the largest variability with a coefficient of variation (CV) of 0.53 at the Baitik station, followed by the snowmelt $\delta^{18}\text{O}$ which exhibits a CV of 0.26. As expected, the glacier melt and snowmelt show the most depleted $\delta^{18}\text{O}$ among the water sources. The groundwater shows only minor variability in the $\delta^{18}\text{O}$ as can be expected. Snow and glacier melt present the lowest EC value, while groundwater shows the highest EC. The seasonality of EC characteristics of various water sources is presented in Fig. 2. The EC of streamflow reached its highest value in winter, and showed low value in summer.

3. Methodology

3.1 Isotope-hydrological integrated modeling approach

The semi-distributed hydrological model WASA (Model of Water Availability in Semi-Arid Environments) originally developed by Güntner and Bronstert (2004) was adopted in this study. It has been extended for snow and glacier melt processes and successfully applied in Central Asian mountain basins (Duethmann et al., 2013; Duethmann et al. 2015). The WASA model is based on the discretization of the landscape into model units according to soil properties, land cover and elevations. To set up the WASA model, the whole Ala-Archa basin was divided into 404 units, with an average size of around 0.58 km² (about twice of the size of a MODIS pixel). Snowmelt and glacier melt on model units were calculated using a temperature-index approach (Hock, 2003), with two different degree-day factors for snow and glacier (Table 2). To differentiate between rainfall and snowfall, a threshold temperature (T_m , set as 2.23 °C from He et al. 2018) was used. The threshold temperature for melting (T_o) in the degree-day module was set to the same value as T_m . Annual GMB was calculated for the hydrological years from September to August by subtracting snowmelt and glacier melt from snowfall on the glacierized areas, in order to compare it with the GMB measured in the field

typically in late summer. The Δh -approach (parameterization for the changes in glacier thickness) was implemented in the WASA model to account for mass balance redistribution with glacier elevation range and account for glacier thickness and area changes (Huss et al., 2010). More details on the model structure can be found in He et al. (2018). Monthly lapse rates of precipitation and temperature were derived from the observed time series at the two meteorological stations (Fig. 1), and were used to estimate daily precipitation and temperature in each model unit based on elevation.

We then extended the WASA model for the simulation of the water isotopic composition of streamflow. Fig. 3 shows a schematic representation of the extended model structure. We integrated the simulation of the δ -notation (‰) of ^{18}O into the WASA hydrological model structure (hereafter abbreviated as IsoWASA model). We simulated the $\delta^{18}\text{O}$ signature instead of $\delta^2\text{H}$ in the model because the higher precision of the $\delta^{18}\text{O}$ measurements (± 0.25 ‰) compared to the $\delta^2\text{H}$ (± 0.4 ‰) in our laboratory and both signatures are strongly correlated. The $\delta^{18}\text{O}$ of streamflow was simulated by two modules: the initialization module of $\delta^{18}\text{O}$ inputs for water sources and the fractionation-mixing module of the $\delta^{18}\text{O}$ of various runoff components along with the runoff generation processes.

For the initialization of $\delta^{18}\text{O}$ inputs, we assumed that $\delta^{18}\text{O}$ of precipitation, glacier melt and initial groundwater are linearly related to basin elevation (Allen et al., 2018). The elevation effect on isotopic compositions of runoff components has been reported in previous studies. For example, the elevation-gradient for precipitation isotopic composition was measured in the Himalaya areas by Dalai et al. (2002). Field data in areas in the Central Andean measured by Ohlanders et al. (2013) demonstrated the elevation effect on the isotopic composition of snowmelt. The $\delta^{18}\text{O}$ value of glacier ice-melt water generated from higher elevations tended to be more depleted in the Himalaya foothills, as measured by Maurya et al. (2011). Mark and McKenzie (2007) used an elevation-gradient to estimate the isotopic compositions of glacier

melt dominated streamflow in an Andean sub-basin. Measurements from spring samples in a Greece mountain basin showed that $\delta^{18}\text{O}$ of groundwater decreased with the elevation increase (Payne, 1978).

Based on the sampling data in our study basin, we estimated the elevation-gradients for the isotopic compositions of precipitation, glacier meltwater and groundwater in Table 1. The elevation-gradient for the isotopic composition of precipitation (LPI) was estimated by the monthly isotope measurements at the Baitik (1580 m a.s.l.) and Alplager (2100 m a.s.l.) meteorological stations. The elevation-gradient for the isotopic composition of glacier melt (LGI) was estimated by the isotopic compositions of summer samples from flowing water on the Golubin glacier surface. The LGI was only applied in the ablation zone to calculate the isotopic composition of glacier melt. The elevation range of the ablation zone was defined as around 3200 - 4200 m a.s.l., according to our field investigations in summer. The isotopic composition of glacier ice in the accumulation zone was not considered in this study. Regarding the isotopic composition of groundwater, we assumed that streamflow in January was fed by groundwater alone, which is the release of subsurface water storage recharged by runoff components in the warm season. Baseflow draining to the streamflow was defined as groundwater in this study basin. The elevation-gradient for the isotopic composition of groundwater (LGWI) was estimated by the January streamflow isotopic compositions measured at the Baitik and Alplager stations. To be noted, the LGWI was only used for the initialization of the groundwater isotopic composition in January when the model starting to run. The isotopic composition of groundwater was updated daily in the model along with the generation and infiltration of surface runoff components. The elevation range to apply the LGWI was set as 1580-3200 m a.s.l., since 3200 m is the lowest elevation of the glacierized area in the basin. The terrain at altitudes higher than 3200 m a.s.l. is extremely steep in the study basin. Soil layers in these areas are very shallow. Contribution of groundwater from these areas is therefore

assumed minor. The isotopic composition of subsurface runoff at altitudes higher than 3200 m are dominated by the isotopic compositions of rainfall and melt water.

The assumptions for fractionation-mixing of $\delta^{18}\text{O}$ of different runoff components are as follows. $\delta^{18}\text{O}$ of snowfall and rainfall were assumed to be the same as that of precipitation. The changes in water isotopic composition caused by the rain out and snowfall out processes were not considered in the model. The initial snowpack $\delta^{18}\text{O}$ was estimated as the $\delta^{18}\text{O}$ of the first snowfall in the modeling period. $\delta^{18}\text{O}$ of snowpack accumulation ($\delta^{18}\text{O}_{\text{SW}}$) was then updated by the mixing of $\delta^{18}\text{O}$ of new snowfall and $\delta^{18}\text{O}$ of the existing snowpack. $\delta^{18}\text{O}$ of flowing water on snowpack surface was a mixture of the $\delta^{18}\text{O}$ of snowmelt and potential rain on snow. $\delta^{18}\text{O}$ of snowmelt ($\delta^{18}\text{O}_{\text{SM}}$) was simulated by a Rayleigh fractionation method given in Eq. (1) (Hindshaw et al., 2011). $\delta^{18}\text{O}$ of the remaining snowpack ($\delta^{18}\text{O}_{\text{SR}}$) was then updated by the melt-out $\delta^{18}\text{O}$ of the snowmelt (Eq. 2). The Rayleigh fractionation factor (α) was assumed to be a function of temperature (Eq. 3, see also Majoube, 1971; Wolfe et al., 2007). A correction factor (CF_s , Eqs. 1-2) for the fractionation factor was adopted to improve the simulation of the snowmelt $\delta^{18}\text{O}$.

$$\delta^{18}\text{O}_{\text{SM}} = (\delta^{18}\text{O}_{\text{SW}} + 1000) \cdot \frac{1 - f^{CF_s \cdot (1/\alpha - 1) + 1}}{1 - f} - 1000 \quad (1)$$

$$\delta^{18}\text{O}_{\text{SR}} = (\delta^{18}\text{O}_{\text{SW}} + 1000) \cdot f^{CF_s \cdot (1/\alpha - 1)} - 1000 \quad (2)$$

$$\ln \alpha = -0.00207 + \frac{-0.4156}{T} + \frac{1137}{T^2} \quad (3)$$

$$f = 1 - \frac{SM}{SW} \quad (4)$$

where SM is the out water amount generated by snowmelt, SW is the original snow water equivalent (SWE) in the snowpack, and SR is the remaining SWE after the snowmelt. f is the ratio of the remaining SWE to the original SWE in the snowpack. T (K) is the air temperature in the corresponding model spatial unit.

The $\delta^{18}\text{O}$ of direct runoff was derived from the mixing of $\delta^{18}\text{O}$ of rainfall, snowmelt, glacier melt and shallow subsurface flow (Fig. 3). $\delta^{18}\text{O}$ of glacier meltwater at specific elevation band was calculated by the LGI and meltwater $\delta^{18}\text{O}$ measured at the glacier tongue. The $\delta^{18}\text{O}$ of shallow subsurface flow was updated by the mixing of the $\delta^{18}\text{O}$ of the infiltrated water from rainfall and meltwater and the $\delta^{18}\text{O}$ of existing soil water. $\delta^{18}\text{O}$ of soil water was initialized as the $\delta^{18}\text{O}$ of the first infiltrated melt or rain water in the model period, and then updated by the $\delta^{18}\text{O}$ of the infiltrated water and modulated by the evaporation process. Rayleigh fractionation processes in $\delta^{18}\text{O}$ caused by evaporation in surface runoff and soil water storage were similarly considered by Eqs. (1)-(4). We adopted another correction factor (CF_e) for the evaporation fractionation. The $\delta^{18}\text{O}$ of groundwater was derived by the mixing of initial $\delta^{18}\text{O}$ of groundwater and the $\delta^{18}\text{O}$ of the percolated soil water storage. Finally, the $\delta^{18}\text{O}$ of river streamflow was derived from the mixing of $\delta^{18}\text{O}$ of direct runoff and groundwater outflow.

3.2 Model calibration

The calibration procedure for the IsoWASA model was implemented in four methods: (1) a benchmark single-dataset calibration using only discharge (observed at the Baitik station, same for the following), (2) a bi-dataset calibration using discharge and $\delta^{18}\text{O}$ of streamflow, (3) a tri-dataset calibration using discharge, MODIS snow cover area (SCA) and GMB, and (4) a four-dataset calibration based on the tri-dataset calibration used additionally $\delta^{18}\text{O}$ of streamflow. The bi-dataset calibration was carried out to investigate the power of the $\delta^{18}\text{O}$ of streamflow to constrain the simulations of snow and glacier dynamics in comparison to the single-dataset calibration. The four-dataset calibration was used to test the benefits of the additional use of streamflow $\delta^{18}\text{O}$ to reduce parameter uncertainty, in comparison to the tri-dataset calibration. The model was evaluated at the daily time step for discharge, and at time steps according to the sampling interval for the other variables.

The model parameters calibrated in all the four calibration methods are summarized in Table 2. Initial values of the calibrated and non-calibrated parameters were adopted from He et al. (2018). For the fairness of comparison of the calibration methods, the correction factors CF_s and CF_e for water isotope fractionation were estimated prior to the assessment of the four calibration methods, because these two factors have no effects on the performance of the single- and tri-dataset calibration methods. We first calibrated all the model parameters, including the correction factors CF_s and CF_e , only on the simulation of streamflow $\delta^{18}\text{O}$. The calibrated values of CF_s and CF_e were then fixed in the four calibration methods at 1.46 and 2.13, respectively. The multiple-objective ϵ -NSGAI automatic algorithm (Deb et al., 2002; Kollat and Reed, 2006) was run for the optimization of parameter values in all the four calibration methods, with initial population size of eight and maximum number of function evaluations of 40,000. The average value of Nash–Sutcliffe efficiency (NSE , Eq. 5) and logarithmic NSE ($\log NSE$, Eq. 6) was used as the objective function to evaluate the simulation of discharge (hereafter abbreviated as $avNSE$), the root mean squared error (SE , Eq. 7) for the simulations of SCA and $\delta^{18}\text{O}$ of streamflow, and the volumetric deviation efficiency at elevation bands (VE , Eq. 8) to assess the simulation for the GMB. The different observation datasets were weighted equally by using the same Epsilon values in the ϵ -NSGAI algorithm for the different objective functions. Since the streamflow isotope data is only available from July 2013, we selected the years 2015-2016 as the calibration period, and 2013-2014 as the evaluation period. The model was warmed up using two-year data before the calibration. In each calibration method, the model was run for four years. The first two-year warm-up run was used to update the initial variable estimates. The values of model parameters were optimized based on the fitness between the simulated and observed datasets in the last two years. Forcing data in the first two years were simply repeated from the last two years in each calibration method. To identify the behavioral parameter sets from the Pareto-optimal fronts generated by the ϵ -NSGAI algorithm

in each calibration method, we defined acceptable threshold values for the performance metrics for the four observation datasets. For the simulation of discharge, we defined the acceptable threshold value for the *avNSE* metric as 0.8. The threshold values for *SE* and *VE* values for the simulations of $\delta^{18}\text{O}$, SCA and GMB were defined as 0.8, 0.8 and 0.95, respectively. In calibration methods that involved multiple observation datasets for the parameter optimization, only the parameter sets which produced performance metrics higher than the defined threshold values for all the observation datasets were identified as behavioral. The choices of the threshold values do not affect the conclusions, as the same threshold values were used for all the four calibration methods.

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

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

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

$$VE = 1 - \frac{\left| \sum_{t=1}^N \sum_{l=1}^{NB} (M_{obs}^l(t) - M_{sim}^l(t)) \right|}{\left| \sum_{t=1}^N \sum_{l=1}^{NB} M_{obs}^l(t) \right|}, \text{ for GMB} \quad (8)$$

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 (or analyzed and simulated $\delta^{18}\text{O}$ of streamflow) over the basin on day k , respectively; m is the total number of measured SCA (or $\delta^{18}\text{O}$) data used for the calibration during the calibration period; $M_{obs}^l(t)$ and $M_{sim}^l(t)$ are the

observed and simulated GMB (mm) on the Golubin glacier at elevation band l in year t , respectively; NB is the total number of elevation bands, and N is the total number of the evaluated years.

3.3 Quantifying contributions of runoff components to streamflow using the isotope-hydrological integrated model and a tracer-based mixing method

The contributions of runoff components to the streamflow were quantified by two approaches in this study. We first estimated the contributions of runoff components with a tracer-based mixing method using measurements of water stable isotopic compositions and EC. The adopted three-component tracer conservative mixing method with rainfall, meltwater and groundwater as end members is described in Eq. (9). Since the water tracer characteristics of snowmelt and glacier melt were very similar during the investigated periods, these two runoff components were treated as one component. We only used the $\delta^{18}\text{O}$ measurements for the mixing method, due to the high correlation between $\delta^{18}\text{O}$ and $\delta^2\text{H}$. To consider the spatial variability in the rainfall $\delta^{18}\text{O}$, we used the volume-weighted average rainfall $\delta^{18}\text{O}$ across the basin instead of the measured precipitation $\delta^{18}\text{O}$ at the meteorological stations. The spatial distribution of rainfall was estimated by the lapse rates of precipitation and temperature. The spatial distribution of precipitation $\delta^{18}\text{O}$ was estimated by the elevation-gradient LPI. Streamflow $\delta^{18}\text{O}$ measured at the two meteorological stations were used for the mixing method. Groundwater samples collected from the sampled spring were used in the tracer-based mixing method. The stream water samples in January were only used for the initialization of the groundwater isotopic composition in the hydrological model. The uncertainty in the mixing method was estimated by an error propagation method adopted in Genereux (1998) and Phillips and Gregg (2001).

$$\begin{cases} Q = RF + GW + MW, & \text{for water balance} \\ \delta^{18}O_q \cdot Q = \delta^{18}O_{RF} \cdot RF + \delta^{18}O_{GW} \cdot GW + \delta^{18}O_{MW} \cdot MW, & \text{for } \delta^{18}O \\ EC_q \cdot Q = EC_{RF} \cdot RF + EC_{GW} \cdot GW + EC_{MW} \cdot MW, & \text{for EC} \end{cases} \quad (9)$$

where, RF refers to the runoff generated by rainfall, GW stands for the runoff generated by groundwater and MW is the total meltwater generated by snowmelt and glacier melt.

Second, we estimated the contributions of runoff components based on the IsoWASA model outputs produced by the four calibration methods. The total simulated streamflow was first separated into two parts, i.e. groundwater and direct runoff. Direct runoff here is composed of the shallow subsurface flow and surface runoff triggered by rainfall and meltwater (Fig. 3). The groundwater storage was recharged by the percolation of shallow subsurface flow using the fraction factor $frac2gw$ (Table 2), and the groundwater outflow was simulated by the groundwater delay factor. The contribution of groundwater was directly estimated as the ratio between the simulated groundwater and the total simulated streamflow. Direct runoff was estimated by subtracting the groundwater contribution from the total simulated streamflow, and was further assigned into three parts based on the proportions of the simulated runoff components (i.e. snowmelt, glacier melt and rainfall). Glacier melt was defined as melt from the glacier ice in this study.

The comparison of the quantification of runoff components produced by the IsoWASA model and the tracer-based mixing method was carried out for two summer peak flow periods, because the snow and glacier melt runoff contribute the largest portion to streamflow in summer. The selected peak flow 1 extends from July 13th to August 19th 2015, and peak flow 2 from July 10th to August 10th 2016. As snow and glacier melt samples were not available during the period of peak flow 1, the $\delta^{18}O$ and EC values for snow and glacier melt during this peak flow were adopted from the samples during the period of peak flow 2. Snowmelt and glacier melt during the two peak flow periods should occur in similar elevation range in the two contiguous years.

We assumed that the glacier melt water occurring at the same elevation band have similar water tracer characteristics. This assumption is supported by field data reported in other glacierized basins. For example, the sampling data in an Italian glacierized basin collected by Penna et al. (2017) demonstrated that the isotopic compositions of glacier melt in two contiguous years are rather similar, and the EC in glacier meltwater showed low variability in space and time. Sampling data in the Himalaya foothills used by Maurya et al. (2011) showed that glacier melt occurred at the same latitude and altitude tended to have similar isotopic compositions. For snowmelt that occurred in the two peak flow periods, we assumed they presented the similar water tracer characteristics also because the water tracer characteristic of summer snowmelt was related to that of snowpack which was formed by snowfall in the last winter during the study period. Some carryover effect can be expected. Our field data show that the winter precipitation in the years of 2014 and 2015 have rather similar isotopic compositions and EC values (see also Figs. 2 and 8b and 8f).

Good model performance for the simulation of the peak flows is a prerequisite for the quantification of contributions of runoff components to the peak streamflow. However, hydrological models often failed to capture summer peak flows in high-elevation mountain basins (Aizen et al., 2000; Jasper et al., 2002; He et al., 2017), partly due to the incorrect climate inputs derived from the sparse climatic gauge stations (Li and Williams, 2008). Correcting the precipitation input for hydrological modeling in high elevation glacierized basins has been commonly reported in previous studies (e.g., Duethmann et al., 2013; Immerzeel et al., 2015). We thus added a precipitation (C_P) and a temperature (C_T) correction factor to adjust the model inputs in the two peak flow periods. We corrected both the precipitation and temperature inputs to 1) compare the performance of various calibration methods to correct the model inputs based on the constraints of various observation datasets; 2) evaluate the ability of various calibration methods to reduce the compensation between temperature-triggered glacier melt runoff and

precipitation-triggered runoff during the summer peak flow periods. The model parameters and the two additional correction factors were recalibrated in the whole calibration period to obtain high performance for the simulation of the two peak flows.

4. Results

4.1 Model calibration and evaluation

The values of the objective functions for the four observation datasets produced by the behavioral parameter sets in both calibration and evaluation periods are compared in Fig. 4 and Table 3. In the calibration period, the single-dataset method produced the highest performance for the simulation of discharge (median *avNSE* value is higher than 0.86, Fig. 4a and Table 3), while producing lower performance for the simulations of $\delta^{18}\text{O}$, SCA and GMB (see *SE* and *VE* values in Figs. 4b-d and Table 3). The bi-dataset method yielded the highest performance for the simulation of $\delta^{18}\text{O}$ (Fig. 4b), while producing lower performance for the simulation of GMB compared to the tri- and four-dataset methods (Fig. 4d). The tri-dataset method produced the highest performance for the simulations of SCA and GMB, while tending to produce *SE* values lower than 0.8 for the simulation of $\delta^{18}\text{O}$. The four-dataset method produced high performance for the simulations of $\delta^{18}\text{O}$ and GMB, despite of its lower performance for the simulation of discharge than the single-dataset method and the lower performance for SCA than the tri-dataset method.

In the evaluation period (Figs. 4e-h and Table 3), the model performance for the four observation datasets tended to increase with increasing observation data involved in the calibration methods. For example, the model performance for discharge produced by the multi-dataset calibration methods were higher than that produced by the calibration to discharge only (Fig. 4e). The four-dataset method produced the highest performance for all the four

observation datasets in terms of the median values of the objective functions, followed by the tri-dataset method.

Fig. 5 shows the uncertainty ranges for the simulations produced by the behavioral parameter sets of the four calibration methods in the calibration period (similar situation in the evaluation period, not shown for the sake of brevity). The single-dataset calibration produced the largest uncertainty ranges, followed by the bi-dataset calibration. The four-dataset calibration generally produced the narrowest simulation ranges. The bi-dataset calibration reduced the uncertainty for the simulations of SCA and GMB compared to the single-dataset calibration (Figs. 5e-f and 5m-n), even though the information of SCA and GMB was not used in this calibration approach. The additionally used $\delta^{18}\text{O}$ in the bi-dataset calibration helped to constrain the simulations of SCA and GMB. The tri-dataset calibration method narrowed the uncertainty for the simulation of streamflow $\delta^{18}\text{O}$ compared to the single-dataset calibration method (Figs. 5i and 5k), despite the $\delta^{18}\text{O}$ data was not used. This can be explained by the fact that the dynamics of isotopic composition of streamflow are typically dominated by the snow and glacier melt runoff in glacierized basins.

4.2 Parameter uncertainty and identifiability

Fig. 6 shows the uncertainty ranges related to the values of the behavioral parameters calibrated by the four methods. The single-dataset calibration produced the largest uncertainty ranges for all the ten parameters. The bi- and the tri-dataset calibrations narrowed the parameter ranges compared to the single-dataset calibration. It is worth noting that the bi-dataset calibration narrowed the uncertainty ranges for five parameters even stronger than the tri-dataset calibration (Figs. 6c-d and 6h-j), even though only two datasets were involved in this calibration method. The four-dataset calibration generally produced the smallest uncertainty ranges for all the parameters. This underscores the additional power of water isotopic data to constrain model parameters.

Fig. 7 examines the sensitivity of the model performance for the simulation of the streamflow $\delta^{18}\text{O}$ to the individual parameter values. In each subplot, the model performance was tested on the value range of one specific parameter. The values for the other remaining parameters were fixed using the optimized values from the four-dataset calibration. In terms of the curve slopes (see also the third column of Table 4), the simulation of the streamflow $\delta^{18}\text{O}$ shows the strongest sensitivity to the values of the snowmelt melt factor (Fig. 7a), followed by the values of parameters groundwater delay factor and $p_correct_f$ (Figs. 7g and 7j). The SE value for the streamflow $\delta^{18}\text{O}$ also shows significant sensitivity to the parameters controlling the generation of direct runoff (e.g., parameters kf_corr_f and sat_area_var in Figs. 7d and 7i) and glacier melt (Fig. 7b). These can be expected, as the dynamics in $\delta^{18}\text{O}$ signature are controlled by the dominance of snowmelt and glacier melt runoff, as well as the partitioning between direct runoff and groundwater. The parameter values that produced SE value higher than 0.8 for the simulation of $\delta^{18}\text{O}$ are assumed behavioral and labeled in red in Fig. 7 (the normalized behavioral ranges are summarized in Table 4). Data show that the $\delta^{18}\text{O}$ data has varied power for constraining the behavioral ranges of model parameters, with the highest identifiability on the behavioral values of parameters snowmelt factor, kf_corr_f and $p_correct_f$ (Figs. 7a, 7d and 7j).

4.3 Simulations of $\delta^{18}\text{O}$ of runoff components

Fig. 8 compares the simulated and measured $\delta^{18}\text{O}$ of runoff components in both calibration and evaluation periods. The simulated $\delta^{18}\text{O}$ series were produced by the four-dataset calibration method using the behavioral parameter sets. Figs. 8a and 8e present the simulated $\delta^{18}\text{O}$ of groundwater in the model unit where the sampled spring is located. The simulated groundwater $\delta^{18}\text{O}$ are generally stable throughout the year, apart from the slight increase in the summer months caused by the increased temperature. The simulated groundwater $\delta^{18}\text{O}$ tend to match the measured values well in the years 2015 and 2013 (Figs. 8a and 8e), while showing

certain difference from the measured values in July-August of 2016 and in August-September of 2014. Measured groundwater $\delta^{18}\text{O}$ in the years 2016 and 2014 show large variability, partly because some of these samples were collected on rainy days. The sampled water from the spring on rainy days could be a mixture of groundwater and rainwater, leading to the abnormal $\delta^{18}\text{O}$ measurements.

Figs. 8b and 8f show the simulated and measured rainfall $\delta^{18}\text{O}$ in the calibration and evaluation periods, respectively. Given the rainfall $\delta^{18}\text{O}$ in the model units was estimated by the LPI based on site measurements of precipitation $\delta^{18}\text{O}$, we presented the rainfall $\delta^{18}\text{O}$ in all the model spatial units across the basin. Both the simulated and measured rainfall $\delta^{18}\text{O}$ reach the highest values in the warm season and the lowest values in the cold season. The simulated rainfall $\delta^{18}\text{O}$ ranges generally capture the measured values in both calibration and evaluation periods. Because the two meteorological stations, where precipitation samples were collected, are both located in the downstream area of the basin, the simulated rainfall $\delta^{18}\text{O}$ in most model units appears to be lower than the measured rainfall $\delta^{18}\text{O}$ due to the elevation effect.

Figs. 8c-d and 8g-h present the simulated and measured snowmelt and glacier melt $\delta^{18}\text{O}$ on the days when samples were collected. Only the simulated $\delta^{18}\text{O}$ of meltwater in the model units, where the sampled sites are located, were presented. In the spring months of 2016 and 2014, the simulated snowmelt $\delta^{18}\text{O}$ generally matched well with the measurements. However, in the spring months of 2015 and August of 2016, the simulated snowmelt $\delta^{18}\text{O}$ are lower than the measurements partly caused by the errors in the simulations of snow accumulation and melt processes. The simulated glacier melt $\delta^{18}\text{O}$ tend to match the measurements well in both the calibration and evaluation periods (Figs. 8d and 8h). To be noted, the IsoWASA model was not calibrated on the simulations of $\delta^{18}\text{O}$ of runoff components. Only the $\delta^{18}\text{O}$ of streamflow was used in the model calibration procedure.

4.4 Contributions of runoff components quantified by different calibration methods

Fig. 9 compares the contributions of runoff components at annual and seasonal scales quantified by the four calibration methods in the calibration period. The uncertainty ranges of the contributions were caused by the uncertainty in the behavioral parameters in Fig. 6.

At the annual scale (uncertainty ranges based on the four-dataset calibration), the runoff contributions to streamflow during the calibration period were 13.1 - 18.1% from rainfall (Fig. 9a), 17.6 - 24.7% from snowmelt (Fig. 9f), 15 - 21.5% from glacier melt and 35.8 - 53.8% from groundwater (Figs. 9k and 9p). Rainfall mostly contributed to streamflow in spring and summer (Figs. 9b-c), while contributing in autumn and winter remained below 8% (Figs. 9d-e). Groundwater was the largest runoff component in all seasons, with the highest contributions of 90% during winter and the smallest contributions of 35.7% in summer (Figs. 9t and 9r, from median values of the four-dataset method). Snowmelt was estimated to be the second largest runoff component in spring (Fig. 9g, 33.8%) and autumn (Fig. 9i, 12.9%). Glacier melt was the second largest runoff component in summer with contribution of 28.5% (Fig. 9m).

The most obvious differences between the different calibration methods were in the uncertainty ranges. The single-dataset calibration method generally generated the largest uncertainty ranges, which partly resulted in estimates that were not reliable due to their large uncertainty range (e.g. estimated contribution of glacier melt of 6-44.8% at the annual scale in Fig. 9k, and estimated winter snowmelt contribution of 0.1-79% in Fig. 9j). The four-dataset calibration method resulted in the smallest uncertainty ranges. The difference in the seasonal contributions of snowmelt and groundwater are also visible (Figs. 9g-j and 9q-t). The single-dataset calibration tended to produce the lowest (highest) contributions of snowmelt (groundwater), while the four-dataset calibration tended to produce the highest (lowest) contributions of snowmelt (groundwater). It becomes clear that the poorly constrained model tries to compensate the lack of meltwater by the higher groundwater contribution.

4.5 Comparison of the modeled contributions of runoff component to estimates from a tracer-based mixing method for two summer peak flows

Fig. 10 shows the $\delta^{18}\text{O}$ and EC values analyzed from the water samples during the two selected summer peak flows. As expected, the $\delta^{18}\text{O}$ of snow and glacier meltwater are the most depleted, followed by the $\delta^{18}\text{O}$ of groundwater. The groundwater presents the highest EC value followed by the rainfall samples. Snowmelt samples present similar $\delta^{18}\text{O}$ and EC characteristics to the glacier melt samples. It is noted that the $\delta^{18}\text{O}$ of snow and glacier melt samples show a high variability. The variability in snowmelt $\delta^{18}\text{O}$ can be attributed to the fact that the snow samples were collected at different depths from the snowpack. Samples collected from deeper snow layers tend to have more depleted $\delta^{18}\text{O}$ values. The variability in glacier melt $\delta^{18}\text{O}$ can be explained by the mixing of glacier melt water and potential snowmelt on glacier surface. The streamflow samples were perfectly located within the triangle areas formed by the runoff component samples. This provides a good basis for the tracer-based method to estimate the contributions of runoff components. The average values and the 95% confidence intervals for the contributions of individual runoff components estimated by the tracer-based method are presented in Table 5 and Fig. 11. The contributions show obvious uncertainty, especially for rainfall (Figs. 11b and 11e). Meltwater dominated the peak flow 1 with an average contribution of 54% (Fig. 11c), while groundwater dominated the peak flow 2 with an average contribution of 50% (Fig. 11d). Average contributions of rainfall during the two peak flows are both around 21%.

To estimate the contributions of runoff components based on the hydrological model output, we defined the behavioral parameter sets using the same threshold values for the performance metrics in section 4.1, except the threshold value of *avNSE* for discharge which was set to 0.87 in this section to assure good performance for the reproductions of summer peak flows. The minimum/maximum and mean contributions of runoff components modeled by the

four calibration methods are compared in Fig. 11 and Table 5. Generally, the four-dataset method produced the smallest uncertainties in the runoff components, while the single-dataset method produced the largest uncertainties. The contribution ranges produced by the four-dataset method are generally located within the contribution ranges estimated by the tracer-based method. The single-dataset calibration method presents the largest deviations from the tracer-based method in both two peak flows. The different calibration methods resulted in significant differences in the mean contributions. For example, the single-dataset calibration method produced the highest contribution of groundwater and the lowest contribution of meltwater in peak flow 1. The four-dataset method produced the lowest contribution of groundwater (Fig. 11a). The bi-dataset method yielded the highest contribution of meltwater (Fig. 11c). The tri-dataset method estimated the meltwater contributions similarly to the four-dataset method, due to the dominance of snow and glacier melt runoff in this period. Snow and glacier observation data used in the tri-dataset method provide important information for the optimization of melt parameters. For peak flow 2 (Figs. 11d-f), the bi-dataset method estimated the contributions similarly to the four-dataset method, partly due to the dominance of groundwater in this period. The $\delta^{18}\text{O}$ data used in the bi-dataset method provided additional information on the partitioning between direct runoff and groundwater.

In peak flow period 1, the single-dataset calibration overestimated (underestimated) the contribution of groundwater (meltwater) compared to the tracer-based method, which could be partly attributed to the overestimated (underestimated) precipitation (temperature) input (Figs. 12a and 12c). The single-dataset calibration estimated a negative C_T for the peak flow period 1 (Fig. 12c), and subsequently estimated a large C_P to compensate the underestimated temperature input (Fig. 12a). The four-dataset calibration method estimated the C_T around 0 °C (Figs. 12c-d) in both two peak flow periods, indicating the temperature in high-elevation areas has likely been appropriately captured by the gauged temperature and lapse rate. In peak flow

period 2, the single-dataset calibration method underestimated the contribution of rainfall compared to the tracer-based method (Fig. 11e), partly due to the underestimated precipitation input (Fig. 12b); and overestimated the contribution of meltwater (Fig. 11f), partly caused by the overestimated temperature input (Fig. 12d). The contributions of runoff components quantified by the bi- and tri-dataset calibration methods in Fig. 11 are generally consistent to the magnitudes of the precipitation and temperature inputs estimated by correction factors in Fig. 12.

As expected, the four-dataset method helped to reduce the compensation between glacier melt (typically driven by temperature) and precipitation-triggered runoff (i.e., sum of the rainfall and snowmelt runoff) during the peak flows as shown in Fig. 13. The single-dataset method generally presents the most obvious compensation between the glacier melt and precipitation-triggered runoff indicated by the widest ranges for the quantifications of the runoff components, followed by the bi- and tri-dataset calibration methods. The simulated precipitation-triggered runoff and glacier melt runoff in peak flow 1 are larger than those in peak flow 2, due to that precipitation and temperature in the period of peak flow 1 were both higher than those in the period of peak flow 2. Results in Figs. 12-13 demonstrate that the four-dataset calibration method performed well in correcting the model climate inputs and reducing the compensation between glacier melt and precipitation-triggered runoff.

5. Discussions

5.1 Benefits of the water isotope data for hydrological modeling in glacierized basins

Comparisons between the single- and bi-dataset calibration methods indicate that involving the water isotope data for model calibration helped to improve the performance for the simulations of snow and glacier dynamics (Figs. 4-5). In glacierized basins where ground measured glacier mass balance data are not available, water isotope data bear the potential to

provide constraint on the simulation of glacier melt runoff. The four-dataset calibration demonstrated the benefit of streamflow $\delta^{18}\text{O}$ to reduce the simulation uncertainty ranges in comparison to the tri-dataset calibration. The performance of the tri-dataset method indicates that SCA and GMB data helped to constrain the simulation of $\delta^{18}\text{O}$. This further indicates that the water isotopic composition of streamflow is capable of representing the dominance of snow and glacier melt during the summer months.

Compared to the other multi-dataset calibration runs, involving additional water isotopic compositions in the hydrological model calibration procedure further narrowed the uncertainty ranges in the quantification of runoff components. The compensations between runoff components were obviously reduced using the water isotopic compositions. The model parameters controlling the generation of groundwater can be identified well by the isotopic compositions, thus improving the constraining for the other remaining model parameters controlling the generations of rainfall and melt runoff. The water isotopic compositions bear the potential to provide insights into the internal apportionments of runoff components.

Our modeling results could also provide insights into the field sampling in glacierized basins. Weekly sampling from streamflow is prerequisite for the calibration of the isotope-hydrological integrated model. Monthly sampling for precipitation at two meteorological stations seems appropriate to capture the spatial and seasonal variability of the isotopic composition of precipitation. Samples of glacier meltwater in summer from the ablation zone at different elevations are necessary for the initialization of isotopic composition of glacier melt. Taking stream water samples from sites with different elevations in January are required for the initialization of isotopic composition of groundwater in January. To apply the integrated modeling approach, water samples from snowmelt can be appropriately reduced, since the measurement of isotopic composition of snowmelt sample is not required for the model initialization and calibration. To run the integrated modeling approach in a short period, the

sampling of glacier melt water in the ablation area can be also appropriately reduced once the elevation-gradient of glacier melt isotope has been defined using sampling data during field campaigns, assuming the glacier melt isotopic compositions are constant at specific locations in a short period. To improve the model performance, more sampling work could be spent on increasing sampling intervals for the streamflow and precipitation. For example, more frequent sampling from streamflow over the day during the summer melt period could help to capture the diurnal cycle of the melt contribution. Using volume-weighted isotopic composition could provide more reliable estimates of the daily isotopic composition of streamflow. The sampling cost in high-elevation areas can be appropriately reduced. Our results show that the integrated modeling approach quantified the contributions of runoff components comparably to the tracer-based mixing method, and narrowed the uncertainty in the quantifications. Considering the requirement for larger sample sizes from various water sources in the end-member tracer-mixing approaches, the integrated modeling approach presents the superiority on quantifying runoff components based on less water tracer data.

5.2 Comparison with previous studies on tracer-aided hydrological modeling

Some previous studies developed tracer-aided hydrological models, including applications in snow-dominated basins. The performance of the IsoWASA model is comparable to the performance of these hydrological models, such as from Delavau et al. (2017) and Ala-aho et al. (2017b). The finding that water isotopic compositions helped to reduce model parameter uncertainty is consistent with that in Birkel et al. (2010) and Capell et al. (2012), who applied the tracer-aided hydrological model in two lowland catchments in Scotland.

For the simulation of the isotopic composition of snowmelt, Ala-aho et al. (2017a; 2017b) emphasized that sublimation from interception and ground snow storage, as well as the snowmelt fractionation could enrich the heavy isotopes in snowpack. In our study basin, the air humidity is typically high in the mountainous areas during the snow accumulation period, and

the snow interception by vegetation canopy is small due to the low vegetation coverage in winter; thus the sublimation effect on snow isotopes was not considered. We took into account the snowmelt fractionation effect to enrich the heavy isotopes in snowpack. Birkel et al. (2010) and Stadnyk et al. (2013) used the model proposed by Craig and Gordon (1965) to describe the effect of evaporation fractionation on water isotopes. Considering the complex runoff mechanism in our glacierized basin, we simplified the evaporation fractionation module to reduce the number of parameters, and focused on the calculation of water isotopes of runoff components. Our results indicated that the proposed water stable isotope module is capable of reproducing the isotopic composition of streamflow in the glacierized study basin.

5.3 Limitations

We calculated the elevation-gradients for isotopic compositions of precipitation and glacier melt from field data to estimate the spatial variability of isotopic compositions in high-elevation areas. To provide the initial isotopic composition of groundwater, we used the elevation-gradient of isotopic composition estimated from the streamflow isotopic measurements in January. The elevation effects on isotopic compositions of runoff components found in our study basin are similar to those reported in previous studies in mountain areas (e.g., Payne, 1978; Dalai et al., 2002; Maurya et al., 2011; Penna et al., 2017). In particular, Ohlanders et al. (2013) estimated the elevation-gradient of precipitation $\delta^{18}\text{O}$ as $-0.56 (\text{‰}) \sim -0.44(\text{‰})/100\text{m}$ in a Central Andean glacierized catchment. Engel et al. (2016) used an elevation-gradient of around $-0.77(\text{‰})/100\text{m}$ to estimate the $\delta^{18}\text{O}$ of snowmelt in a glacierized basin in Italy. The estimation of elevation-gradients for isotopic compositions of runoff components in the study basin can be further improved by taking more water samples from different locations.

The representativeness of water samples for runoff components is another limitation of this study. We only found one spring in the basin to take groundwater samples. Our sampled

spring is near the river channel and located at the elevation of 2400 m a.s.l., which is close to the mean basin elevation (2530 m a.s.l.). We assumed the isotopic composition measured from this spring could represent the average isotopic composition of groundwater draining to the river channel. Since groundwater wells are commonly not available in glacierized basins, taking water samples from springs to represent groundwater could be a viable alternative as reported in previous studies (e.g., Ohlanders et al., 2013 took samples from one spring for a study area of 256 km², and Mark and McKenzie, 2007 took samples from two springs in a basin with area of around 5000 km²).

The snow and glacier melt samples were only collected from the Golubin glacier area during the summer melt season. The elevation range (3232 - 4458 m a.s.l.) and the average elevation (3869 m a.s.l.) of the Golubin glacier are close to those of the entire glacierized area in our study basin (3218 – 4857 and 3945 m a.s.l., respectively). Considering that the spatial variability in isotopic composition of summer meltwater is mainly caused by the elevation effect, we used the measurements from the samples on the Golubin glacier to represent the average isotopic composition of meltwater in the entire glacierized area during the melt season. Given the logistic limitations in high-elevation glacierized basins, taking snow and glacier melt samples from specific locations and assuming summer snow and glacier meltwater at the same elevation have similar isotopic compositions, have been adopted in previous studies (e.g., Maurya et al., 2011; Penna et al., 2017). We assumed that the samples collected from the Golubin glacier could provide appropriate isotopic compositions of meltwater for the hydrograph separation in the two summer peak flow periods. In our isotope-hydrological integrated modeling approach, measurements of the isotopic composition from snowmelt samples were not used to force and calibrate the model. The representativeness of water samples for snowmelt should have no effect on the calibration of model parameters.

We investigated the ability of the isotope-hydrological integrated modeling approach to quantify runoff components only during two summer peak flow periods, due to the limited availability of groundwater and meltwater samples. More frequent water sampling in the summer melt period would allow a more insightful comparison between the isotope-hydrological integrated modeling and the tracer-based mixing method. Currently, it is not possible to discriminate snowmelt and glacier melt with the three-component tracer-based mixing method due to very similar isotopic signatures. Therefore, the compensation between the snowmelt and glacier melt contributions could not be examined in the comparison between the modeling and tracer-based methods. Further water chemistry tracers might be useful for the quantification of the contributions of snowmelt and glacier melt. On the other hand, the strong variabilities in the measured water isotopes resulted in large uncertainties in the quantification of runoff components derived from the tracer-based mixing method (similar to Engel et al., 2016 and Penna et al., 2017). We assume that more frequent water sampling from different water sources could help to reduce this uncertainty.

6. Conclusions

In this study, we integrated the modeling of water stable isotopic compositions into a glacio-hydrological model, and calibrated the model parameters using additionally the isotopic composition of streamflow. The novelties of this study are two-fold: we investigated the value of the isotopic composition for model calibration in comparison to calibration methods using glacier mass balance, snow cover area data and discharge; and we examined the power of the isotope-hydrological integrated modeling approach to quantify the contributions of runoff components in comparison to a tracer-based mixing method. Our main findings are summarized as follows.

(1) The proposed isotope-hydrological integrated model was able to simultaneously simulate discharge, streamflow isotopic composition, and snow/glacier cover observations in a glacierized basin. Model parameters calibrated additionally with the isotopic composition produced higher performance in the evaluation period compared to the calibration without using isotopic signatures.

(2) Involving water stable isotopic composition for model calibration significantly reduced the parameter uncertainty. The isotopic composition showed significant ability on constraining the parameters controlling the generation of meltwater runoff and groundwater.

(3) The isotope-hydrological integrated modeling approach quantified the contributions of runoff components comparably to the tracer-based mixing method during summer peak flows, even using less water tracer data. The integrated modeling approach also showed the superiority on narrowing uncertainty in the quantification of contributions of runoff components compared to the tracer-based mixing method.

Acknowledgements

This work has been carried out in the frame of the GlaSCA (grant number 01DK15002A and B) and GlaSCA-V (grant number 88 501) projects funded by the German Federal Ministry for Science and Education (BMBF) and Volkswagen Foundation, respectively. We thank the Kyrgyz Hydrometeorological Service for providing the meteorological and discharge data for the Ala-Archa basin.

Reference

- Ala-aho, P., Tetzlaff, D., McNamara, J. P., Laudon, H., Soulsby, C., 2017a. Using isotopes to constrain water flux and age estimates in snow-influenced catchments using the STARR (Spatially distributed Tracer-Aided Rainfall-Runoff) model. *Hydrol. Earth Syst. Sci.* 21, 5089-5110.
- Ala-Aho, P., Tetzlaff, D., McNamara, J. P., Laudon, H., Kormos, P., Soulsby, C., 2017b. Modeling the isotopic evolution of snowpack and snowmelt: Testing a spatially distributed parsimonious approach. *Water Resour. Res.* 53, 5813-5830.
- Allen, S. T., Kirchner, J. W., Goldsmith, G. R., 2018. Predicting spatial patterns in precipitation isotope ($\delta^2\text{H}$ and $\delta^{18}\text{O}$) seasonality using sinusoidal isoscapes. *Geophys. Res. Lett.* 45, 4859–4868. <https://doi.org/10.1029/2018GL077458>.
- Aizen, V. B., Aizen, E., Glazirin, G., Loaiciga, H. A., 2000. Simulation of daily runoff in Central Asian alpine watersheds. *J. Hydrol.* 238, 15-34.
- Aizen, V. B., Aizen, E. M., Melack, J. M., 1995. Climate, Snow Cover, Glaciers, and Runoff in the Tien-Shan, Central-Asia. *Water Resour. Bull.* 31, 1113-1129.
- 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. *Glob. Planet. Change.* 56, 328-340.
- Barandun, M., Huss, M., Sold, L., Farinotti, D., Azisov, E., Salzmann, N., Usubaliev, R., Merkushkin, A., Hoelzle, M., 2015. Re-analysis of seasonal mass balance at Abramov glacier 1968–2014. *J. Glaciol.* 61(230), 1103-1117.
- Barnett, T. P., Adam, J. C., Lettenmaier, D. P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature.* 438, 303-309.

- Birkel, C., Dunn, S. M., Tetzlaff, D., Soulsby, C., 2010. Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall-runoff model. *Hydrol. Process.* 24, 2335-2348.
- Capell, R., Tetzlaff, D., Soulsby, C., 2012. Can time domain and source area tracers reduce uncertainty in rainfall-runoff models in larger heterogeneous catchments? *Water Resour. Res.* 48, W09544, doi:10.1029/2011WR011543.
- Craig H., Gordon L. I, 1965. Deuterium and oxygen 18 variations in the ocean and the marine atmosphere. In *Stable Isotopes in Oceanographic Studies and Paleotemperatures*. Tongioli E (ed.). Lab. De Geologica Nucleare: Pisa, Italy; 9–130.
- Dahlke, H. E., Lyon, S. W., Jansson, P., Karlin, T., Rosqvist, G., 2014. Isotopic investigation of runoff generation in a glacierized catchment in northern Sweden. *Hydrol. Process.* 28, 1383-1398.
- Dalai, T. K., Bhattacharya, S. K., Krishnaswami, S., 2002. Stable isotopes in the source waters of the Yamuna and its tributaries: seasonal and altitudinal variations and relation to major cations. *Hydrol. Process.* 16, 3345-3364.
- Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 6, 182-197.
- Delavau, C. J., Stadnyk, T., Holmes, T., 2017. Examining the impacts of precipitation isotope input (δ O-18(ppt)) on distributed, tracer-aided hydrological modelling. *Hydrol. Earth Syst. Sci.* 21, 2595-2614.
- Duethmann, D., Bolch, T., Farinotti, D., Kriegel, D., Vorogushyn, S., Merz, B., Pieczonka, T., Jiang, T., Su, B. D., Guntner, A., 2015. Attribution of streamflow trends in snow and glacier melt-dominated catchments of the Tarim River, Central Asia. *Water Resour. Res.* 51, 4727-4750.

- Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., Guntner, A., 2014. The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. *Water Resour. Res.* 50, 2002-2021.
- Duethmann, D., Zimmer, J., Gafurov, A., Guntner, A., Kriegel, D., Merz, B., Vorogushyn, S., 2013. Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling. *Hydrol. Earth Syst. Sci.* 17, 2415-2434.
- Dunn, S. M., Bacon, J. R., 2008. Assessing the value of Cl- and delta O-18 data in modelling the hydrological behaviour of a small upland catchment in northeast Scotland. *Hydrol. Res.* 39, 337-358.
- Engel, M., Penna, D., Bertoldi, G., Dell'Agnese, A., Soulsby, C., Comiti, F., 2016. Identifying run-off contributions during melt-induced run-off events in a glacierized alpine catchment. *Hydrol. Process.* 30, 343-364.
- Engelhardt, M., Schuler, T. V., Andreassen, L. M., 2014. Contribution of snow and glacier melt to discharge for highly glacierised catchments in Norway. *Hydrol. Earth Syst. Sci.* 18, 511-523.
- 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 Resour. Res.* 51, 1939-1958.
- Gafurov, A., Luedtke, S., Unger-Shayesteh, K., Vorogushyn, S., Schoene, T., Schmidt, S., Kalashnikova, O., Merz, B., 2016. MODSNOW-Tool: An operational tool for daily snow cover monitoring using MODIS data. *Environ. Earth Sci.* 75, 1078. <https://doi.org/10.1007/s12665-016-5869-x>.
- Genereux D., 1998. Quantifying uncertainty in tracer-based hydrograph separations. *Water Resour. Res.* 34, 915-919.

- Güntner, A., Bronstert, A., 2004. Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas. *J. Hydrol.* 297, 136-161.
- He, Z. H., Hu, H. C., Tian, F. Q., Ni, G. H., Hu, Q. F., 2017. Correcting the TRMM rainfall product for hydrological modelling in sparsely-gauged mountainous basins. *Hydrol. Sci. J.*, 62, 306–318, doi:10.1080/02626667.2016.1222532.
- He, Z. H., Parajka, J., Tian, F. Q., Blöschl, G., 2014. Estimating degree-day factors from MODIS for snowmelt runoff modeling, *Hydrol. Earth Syst. Sci.* 18, 4773-4789.
- He, Z. H., Tian, F. Q., Gupta, H. V., Hu, H. C., Hu, H. P., 2015. Diagnostic calibration of a hydrological model in a mountain area by hydrograph partitioning. *Hydrol. Earth Syst. Sci.* 19, 1807-1826.
- 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 Resour. Res.* 54, 2336-2361, doi.org/10.1002/2017WR021966.
- Hindshaw, R. S., Tipper, E. T., Reynolds, B. C., Lemarchand, E., Wiederhold, J. G., Magnusson, J., Bernasconi, S. M., Kretzschmar, R., Bourdon, B., 2011. Hydrological control of stream water chemistry in a glacial catchment (Damma Glacier, Switzerland). *Chem. Geol.* 285, 215-230.
- Hock, R., 2003. Temperature index melt modelling in mountain areas. *J. Hydrol.* 282, 104-115.
- Hoelzle, M., Azisov, E., Barandun, M., Huss, M., Farinotti, D., Gafurov, A., Hagg, W., Kenzhebaev, R., Kronenberg, M., Machguth, H., Merkushkin, A., Moldobekov, B., Petrov, M., Saks, T., Salzmann, N., Schöne, T., Tarasov, Y., Usubaliev, R., Vorogushyn, S., Yakovlev, A., Zemp, M., 2017. Re-establishing glacier monitoring in Kyrgyzstan and Uzbekistan, Central Asia. *Geosci. Instrum. Methods Data Sys.* 6(2), 397.

- Huss, M., Bauder, A., Funk, M., 2009. Homogenization of long-term mass-balance time series. *Ann. Glaciol.* 50(50), 198-206.
- Huss, M., Juvet, G., Farinotti, D., Bauder, A., 2010. Future high-mountain hydrology: a new parameterization of glacier retreat. *Hydrol. Earth Syst. Sci.* 14, 815-829.
- Immerzeel, W. W., Petersen, L., Ragetti, S., Pellicciotti, F., 2014. The importance of observed gradients of air temperature and precipitation for modeling runoff from a glacierized watershed in the Nepalese Himalayas. *Water Resour. Res.* 50, 2212-2226.
- 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. *Hydrol. Earth Syst. Sci.* 19, 4673-4687.
- Jasper, K., Gurtz, J., Herbert, L., 2002. Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. *J. Hydrol.* 267, 40-52.
- Joerin, C., Iorgulescu, I., Musy, A., Beven, K. J., 2002. Uncertainty in hydrograph separations based on geochemical mixing models. *J. Hydrol.* 255, 90-106.
- Jost, G., Moore, R. D., Menounos, B., Wheate, R., 2012. Quantifying the contribution of glacier runoff to streamflow in the upper Columbia River Basin, Canada. *Hydrol. Earth Syst. Sci.* 16, 849-860.
- Koboltschnig, G. R., Schoner, W., Zappa, M., Kroisleitner, C., Holzmann, H., 2008. Runoff modelling of the glacierized Alpine Upper Salzach basin (Austria): multi-criteria result validation. *Hydrol. Process.* 22, 3950-3964.
- Kollat, J. B., Reed, P. M., 2006. Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Adv. Water Resour.* 29, 792-807.

- Konz, M., Seibert, J., 2010. On the value of glacier mass balances for hydrological model calibration. *J. Hydrol.* 385, 238-246.
- Li, X. G., Williams, M. W., 2008. Snowmelt runoff modelling in an arid mountain watershed, Tarim Basin, China, *Hydrol. Process.* 22, 3931-3940.
- Lutz, A. F., Immerzeel, W. W., Shrestha, A. B., Bierkens, M. F. P., 2014. Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nat. Clim. Chang.* 4, 587-592.
- Majoube, M., 1971. Fractionnement en oxygene 18 et en deuterium entre l'eau et sa vapeu. *J. Chim. Phys.* 68, 1423-1436.
- Mark, B. G., McKenzie, J. M., 2007. Tracing increasing tropical Andean glacier melt with stable isotopes in water, *Environ. Sci. Technol.* 41, 6955-6960.
- Maurya, A. S., Shah, M., Deshpande, R. D., Bhardwaj, R. M., Prasad, A., Gupta, S. K., 2011. Hydrograph separation and precipitation source identification using stable water isotopes and conductivity: River Ganga at Himalayan foothills. *Hydrol. Process.* 25, 1521-1530.
- Nolin, A. W., Phillippe, J., Jefferson, A., Lewis, S. L., 2010. Present-day and future contributions of glacier runoff to summertime flows in a Pacific Northwest watershed: Implications for water resources. *Water Resour. Res.* 46, W12509, doi:10.1029/2009WR008968.
- Ohlanders, N., Rodriguez, M., McPhee, J., 2013. Stable water isotope variation in a Central Andean watershed dominated by glacier and snowmelt. *Hydrol. Earth Syst. Sci.* 17, 1035-1050.
- Parajka, J., Blöschl, G., 2008. The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. *J. Hydrol.* 358, 240-258.

- Payne, B. R., Leontiadis, J., Dimitroulas, Ch, 1978. A study of the Kalamos springs in Greece with environmental isotopes. *Water Resour. Res.* 14, 653-658.
- Penna, D., Engel, M., Bertoldi, G. and Comiti, F., 2017. Towards a tracer-based conceptualization of meltwater dynamics and streamflow response in a glacierized catchment. *Hydrol. Earth Syst. Sci.* 21, 23-41.
- Penna, D., Engel, M., Mao, L., Dell'Agnese, A., Bertoldi, G., Comiti, F., 2014. Tracer-based analysis of spatial and temporal variations of water sources in a glacierized catchment. *Hydrol. Earth Syst. Sci.* 18, 5271-5288.
- Phillips, D. L., Gregg, J. W., 2001. Uncertainty in source partitioning using stable isotopes. *Oecologia.* 127, 171-179.
- Pu, T., Qin, D. H., Kang, S. C., Niu, H. W., He, Y. Q., Wang, S. J., 2017. Water isotopes and hydrograph separation in different glacial catchments in the southeast margin of the Tibetan Plateau. *Hydrol. Process.* 31, 3810-3826.
- Prasch, M., Mauser, W., Weber, M., 2013. Quantifying present and future glacier melt-water contribution to runoff in a central Himalayan river basin. *Cryosphere.* 7, 889-904.
- Rahman, K., Besacier-Monbertrand, A. L., Castella, E., Lods-Crizet, B., Ilg, C., Beguin, O., 2015. Quantification of the daily dynamics of streamflow components in a small alpine watershed in Switzerland using end member mixing analysis. *Environ. Earth Sci.* 74: 4927. <https://doi.org/10.1007/s12665-015-4505-5>.
- Schaefli, B., Huss, M., 2011. Integrating point glacier mass balance observations into hydrologic model identification. *Hydrol. Earth Syst. Sci.* 15, 1227-1241.
- Seibert, J., Rodhe, A., Bishop, K., 2003. Simulating interactions between saturated and unsaturated storage in a conceptual runoff model. *Hydrol. Process.* 17, 379-390.
- Soulsby, C., Tetzlaff, D., 2008. Towards simple approaches for mean residence time estimation in ungauged basins using tracers and soil distributions. *J. Hydrol.* 363, 60-74.

- Stadnyk, T. A., Delavau, C., Kouwen, N., Edwards, T. W. D., 2013. Towards hydrological model calibration and validation: simulation of stable water isotopes using the isoWATFLOOD model. *Hydrol. Process.* 27, 3791-3810.
- Stadnyk, T., St Amour, N., Kouwen, N., Edwards, T. W. D., Pietroniro, A., Gibson, J. J., 2005. A groundwater separation study in boreal wetland terrain: The WATFLOOD hydrological model compared with stable isotope tracers. *Isotopes Environ. Health Stud.* 41, 49-68.
- Stahl, K., Moore, R. D., 2006. Influence of watershed glacier coverage on summer streamflow in British Columbia, Canada. *Water Resour. Res.* 42, W06201, doi:10.1029/2006WR005022.
- Stahl, K., Moore, R. D., Shea, J. M., Hutchinson, D., Cannon, A. J., 2008. Coupled modelling of glacier and streamflow response to future climate scenarios. *Water Resour. Res.* 44, W02422, doi:10.1029/2007WR005956.
- 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 Resour. Res.* 52, 4674-4699.
- Tetzlaff, D., Buttle, J., Carey, S. K., van Huijgevoort, M. H. J., Laudon, H., McNamara, J. P., Mitchell, C. P. J., Spence, C., Gabor, R. S., Soulsby, C., 2015. A preliminary assessment of water partitioning and ecohydrological coupling in northern headwaters using stable isotopes and conceptual runoff models. *Hydrol. Process.* 29, 5153-5173.
- van Huijgevoort, M. H. J., Tetzlaff, D., Sutanudjaja, E. H., Soulsby, C., 2016. Using high resolution tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-runoff model. *Hydrol. Process.* 30, 4761-4778.

- Verbunt, M., Gurtz, J., Jasper, K., Lang, H., Warmerdam, P., Zappa, M., 2003. The hydrological role of snow and glaciers in alpine river basins and their distributed modeling. *J. Hydrol.* 282, 36-55.
- Weiler, M., McGlynn, B. L., McGuire, K. J., McDonnell, J. J., 2003. How does rainfall become runoff? A combined tracer and runoff transfer function approach. *Water Resour. Res.* 39 (11), 1315, doi:10.1029/2003WR002331.
- Wolfe, B. B., Karst-Riddoch, T. L., Hall, R. I., Edwards, T. W. D., English, M. C., Palmini, R., McGowan, S., Leavitt, P. R., Vardy, S. R., 2007. Classification of hydrological regimes of northern floodplain basins (Peace -Athabasca Delta, Canada) from analysis of stable isotopes (δ O-18, δ H-2) and water chemistry. *Hydrol. Process.* 21, 151-168.
- Zhao, Q. D., Ye, B. S., Ding, Y. J., Zhang, S. Q., Yi, S. H., Wang, J., Shangguan, D. H., Zhao, C. C., Han, H. D., 2013. Coupling a glacier melt model to the Variable Infiltration Capacity (VIC) model for hydrological modeling in north-western China. *Environ. Earth Sci.* 68, 87-101.

LIST OF TABLES

Table 1. Summary of the characteristics of $\delta^{18}\text{O}$ and EC of various water samples	44
Table 2. Model parameters calibrated in all the four calibration methods.....	45
Table 3. Summary of the model performance produced by four calibration methods.....	46
Table 4. Summary of the sensitivity of the model performance for the simulation of streamflow $\delta^{18}\text{O}$ to the model parameters.....	47
Table 5. Contributions (%) of three runoff components to streamflow quantified by various methods during two selected summer peak flows.....	48

Table 1. Summary of the characteristics of $\delta^{18}\text{O}$ and EC of various water samples. The coefficient of variation (CV) refers to the ratio between the standard deviation and mean value.

The lapse rates for $\delta^{18}\text{O}$ of precipitation (LPI), glacier melt (LGI) and initial groundwater (LGWI) were derived from the precipitation, glacier melt and winter streamflow samples.

Tracer	Variable	Sample number	Mean	Range	CV
$\delta^{18}\text{O}$ (‰)	Streamflow at Baitik	158	-11.32	(-12.37, -10.82)	0.03
	Streamflow at Alplager	184	-11.73	(-12.90, -10.94)	0.03
	Precipitation at Baitik	36	-11.21	(-20.99, 1.51)	0.53
	Precipitation at Alplager	43	-11.41	(-22.82, -0.06)	0.51
	Groundwater	14	-11.17	(-11.70, -10.61)	0.03
	Snowmelt	45	-13.98	(-24.24, -10.53)	0.26
	Glacier melt	17	-13.46	(-15.66, -12.33)	0.08
	LPI (‰/100m)	-	-0.530	-	-
	LGI (‰/100m)	-	-0.226	-	-
	LGWI (‰/100m)	-	-0.210	-	-
Electrical Conductivity (EC, $\mu\text{s}/\text{cm}$)	Streamflow at Baitik	25	114.7	(81.0, 139.3)	0.13
	Streamflow at Alplager	78	108.7	(66.7, 137.1)	0.18
	Precipitation at Baitik	14	69.9	(26.6, 99.6)	0.30
	Precipitation at Alplager	23	68.3	(21.3, 102.0)	0.30
	Groundwater	12	126.8	(69.6, 167.2)	0.23
	Snowmelt	7	28.4	(11.0, 55.1)	0.55
	Glacier melt	3	32.1	(30.1, 33.4)	0.06

Table 2. Model parameters calibrated in all the four calibration methods.

Parameter	Unit	Value range	Description
snowmelt factor	mm/°C/day	0-5	Degree-day factor for snowmelt
glacier melt factor	mm/°C/day	5-20	Degree-day factor for glacier melt
<i>k_sat_f</i>	-	0.01-300	Soil hydraulic conductivities
<i>kf_corr_f</i>	-	0.01-300	Soil hydraulic conductivities correction factor
<i>frac2gw</i>	-	0-1	Recharge fraction factor from shallow subsurface flow to groundwater
interflow delay factor	days	0-100	Outflow delay factor for shallow subsurface flow
groundwater delay factor	days	30-400	Outflow delay factor for groundwater
<i>frac_riparian</i>	-	0-0.05	Fraction of saturated area
<i>sat_area_var</i>	-	0-0.3	Spatial variability of saturated areas within a model unit
<i>p_correct_f</i>	-	0.0-2.0	Precipitation correction factor

Table 3. Summary of the model performance produced by four calibration methods.

		Sing.	Bi.	Tri.	Four.
Calibration period	Max	0.8842	0.8736	0.8732	0.8541
	avNSE(Q) Median	0.8623	0.8542	0.8376	0.8489
	Min	0.8009	0.8041	0.8002	0.8035
	Max	0.8057	0.8174	0.8080	0.8130
	SE($\delta^{18}\text{O}$) Median	0.7839	0.8100	0.7838	0.8034
	Min	0.5993	0.8001	0.7663	0.8002
	Max	0.8164	0.8134	0.8199	0.8028
	SE(SCA) Median	0.7923	0.8013	0.8127	0.8010
	Min	0.6955	0.7600	0.8001	0.8000
	Max	0.9969	0.9884	0.9999	0.9993
	VE(GMB) Median	0.5913	0.6607	0.9804	0.9802
	Min	-0.9163	0.2606	0.9501	0.9523
Evaluation period	Max	0.6911	0.5217	0.7866	0.7683
	avNSE(Q) Median	0.0219	0.1345	0.4353	0.6888
	Min	-2.1203	-0.0716	0.0071	0.6202
	Max	0.7662	0.7337	0.7554	0.7576
	SE($\delta^{18}\text{O}$) Median	0.6609	0.6745	0.7023	0.7187
	Min	0.4385	0.5898	0.6542	0.7080
	Max	0.7634	0.7104	0.7493	0.7616
	SE(SCA) Median	0.6773	0.6753	0.7200	0.7519
	Min	0.6312	0.6450	0.6768	0.7490
	Max	0.9992	0.9980	0.9996	0.9879
	VE(GMB) Median	0.7285	0.5940	0.8231	0.9660
	Min	-1.3324	0.2606	0.5110	0.9237

Table 4. Summary of the sensitivity of the model performance for the simulation of streamflow $\delta^{18}\text{O}$ to the model parameters. Width of behavioral range refers to the difference between the maximum and minimum normalized parameter values that produced SE ($\delta^{18}\text{O}$) value higher than 0.8. SEmax and SEmin stand for the maximum and minimum SE ($\delta^{18}\text{O}$) values produced by the individual parameters. Po represents the optimal normalized parameter value that produced the highest SE ($\delta^{18}\text{O}$) value, and Pi represents the normalized parameter that produced a SE ($\delta^{18}\text{O}$) value as 0.8.

Parameter	Width of behavioral range	$ (SE_{\max}-SE_{\min})/(P_o-P_i) $
Snowmelt factor	0.060	0.100
glacier melt factor	0.220	0.029
<i>k_sat_f</i>	0.858	0.004
<i>kf_corr_f</i>	0.071	0.021
<i>frac2gw</i>	0.128	0.012
interflow delay factor	0.765	0.019
groundwater delay factor	0.557	0.066
<i>frac_riparian</i>	0.216	0.007
<i>sat_area_var</i>	0.784	0.030
<i>p_correct_f</i>	0.062	0.050

Table 5. Contributions (%) of three runoff components to streamflow quantified by various methods during two selected summer peak flows. Lower/upper limits refer to the 95% confidence intervals for the Tracer-based mixing method (Trac.), and refer to the minimum and maximum contributions produced by the behavioral parameter sets in the four calibration methods.

			Sing.	Bi.	Tri.	Four.	Trac.
Peak flow 1	Groundwater	upper limit	38.1	29.4	31.1	27.3	30.6
		mean	30.0	26.7	24.9	24.7	25.4
		lower limit	24.3	22.6	21.5	21.9	20.2
	Rainfall	upper limit	19.5	17.5	21.1	21.0	25.3
		mean	16.9	15.1	19.3	19.8	20.8
		lower limit	12.4	12.8	14.0	18.0	16.3
	Meltwater	upper limit	58.7	61.1	60.4	57.6	57.6
		mean	53.1	58.2	55.8	55.5	53.8
		lower limit	43.1	54.0	52.1	52.6	49.9
Peak flow 2	Groundwater	upper limit	63	53.4	56.1	51.1	52.1
		mean	51	52.2	53.5	50.6	50.1
		lower limit	43	51.1	52.9	50.4	48
	Rainfall	upper limit	19.9	22.1	19	22.3	27.8
		mean	9.5	21.3	17.9	22	22.3
		lower limit	8.7	19.4	16.5	21.4	16.8
	Meltwater	upper limit	47.1	28	30.4	28	31.9
		mean	39.5	26.5	28.6	27.4	27.6
		lower limit	27.3	25.4	25.4	27.1	23.4

LIST OF FIGURES

Fig. 1. Study area and water sampling points.....	50
Fig. 2. EC measurements from water samples collected during the study period.....	51
Fig. 3. Schematic representation of the isotope-hydrological integrated modeling approach..	52
Fig. 4. Values of objective functions for observation datasets produced by four calibration methods in the calibration (a-d) and evaluation (e-h) periods.....	53
Fig. 5. Simulation ranges produced by the behavioral parameter sets of the four calibration methods in calibration period.....	54
Fig. 6. Uncertainty ranges of the behavioral parameter sets produced by the four calibration methods.....	55
Fig. 7. Sensitivity of the simulation performance of streamflow $\delta^{18}\text{O}$ to individual parameter values.....	56
Fig. 8. Seasonality of simulated $\delta^{18}\text{O}$ of various runoff components produced by the four-dataset calibration.....	57
Fig. 9. Contributions of runoff components quantified by four calibration methods at annual and seasonal scales.....	58
Fig. 10. Variability in the values of $\delta^{18}\text{O}$ and EC of water samples during the two summer peak flow periods.....	59
Fig. 11. Contributions of runoff components for the selected peak flows estimated by the four calibration methods and the tracer-based mixing model	60
Fig. 12. Estimated correction factors for precipitation and temperature inputs during the two peak flow periods by four calibration methods.....	61
Fig. 13. Compensation between glacier melt and precipitation-triggered runoff during the two summer peak flow periods produced by different calibration methods	62

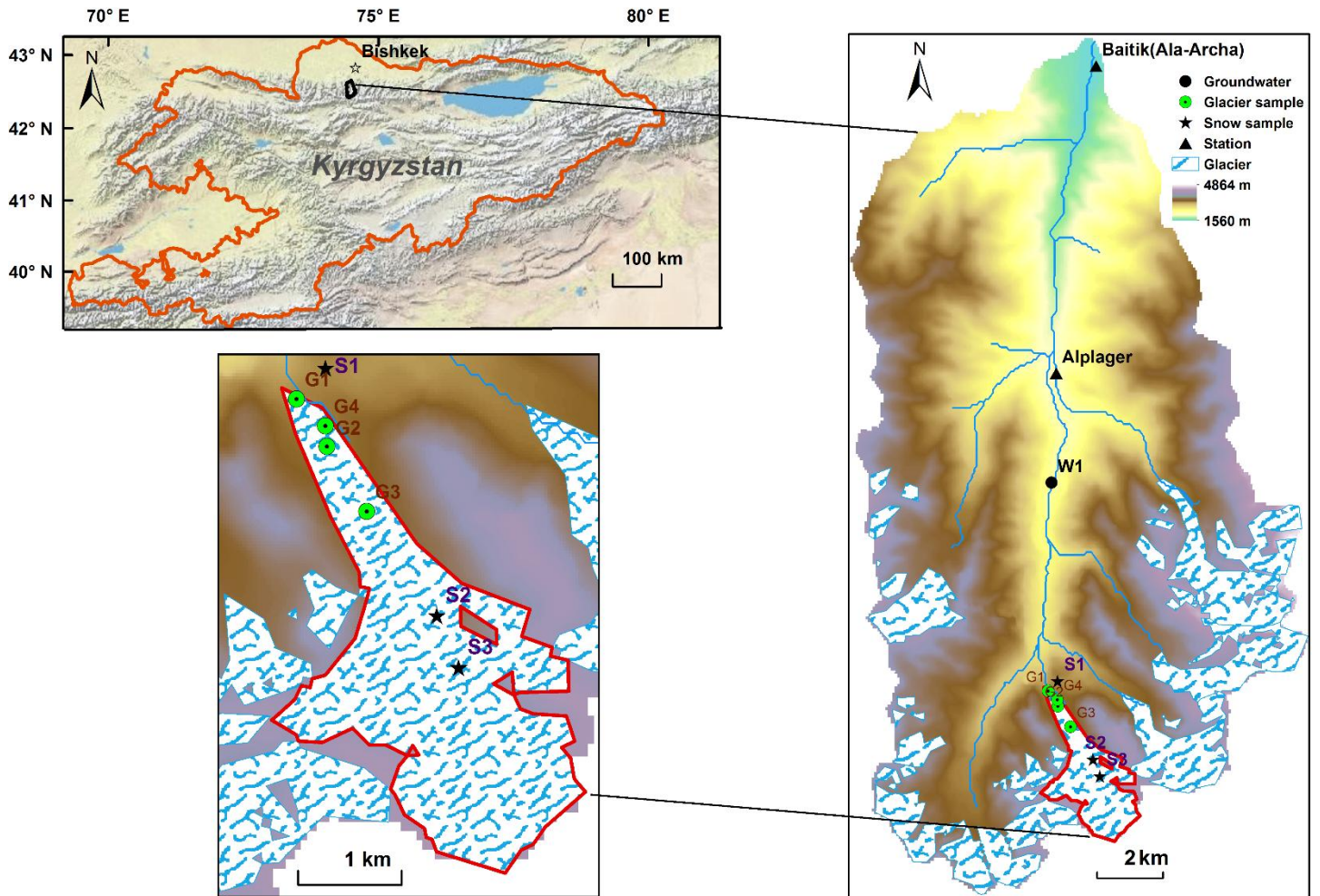


Fig. 1. Study area and water sampling sites. Red polygon refers to the Golubin glacier where annual glacier mass balance measurements were carried out. S1-3: snowmelt sampling sites; G1-4: glacier melt sampling sites; W1: groundwater sampling site.

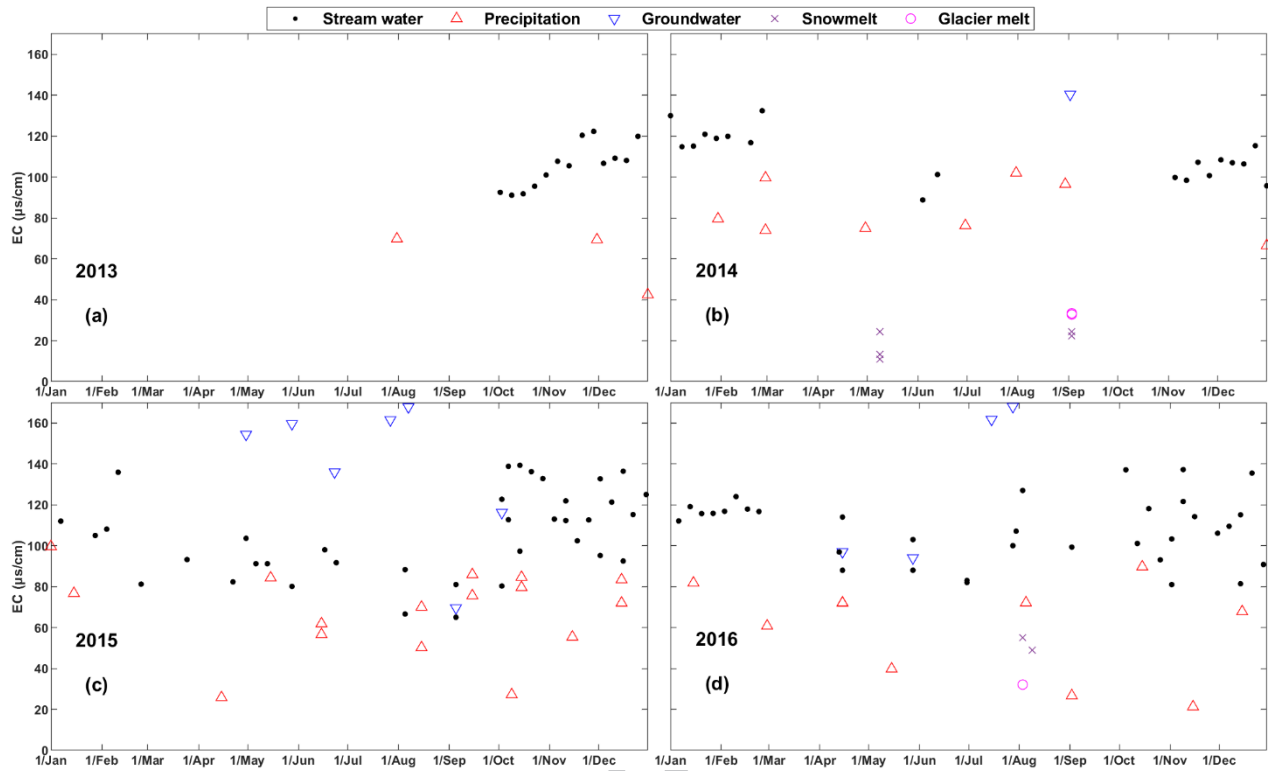


Figure 2. EC measurements from water samples collected during the study period. The period of 2013-2014 is the validation period and the period of 2015-2016 is the calibration period used in the hydrological model.

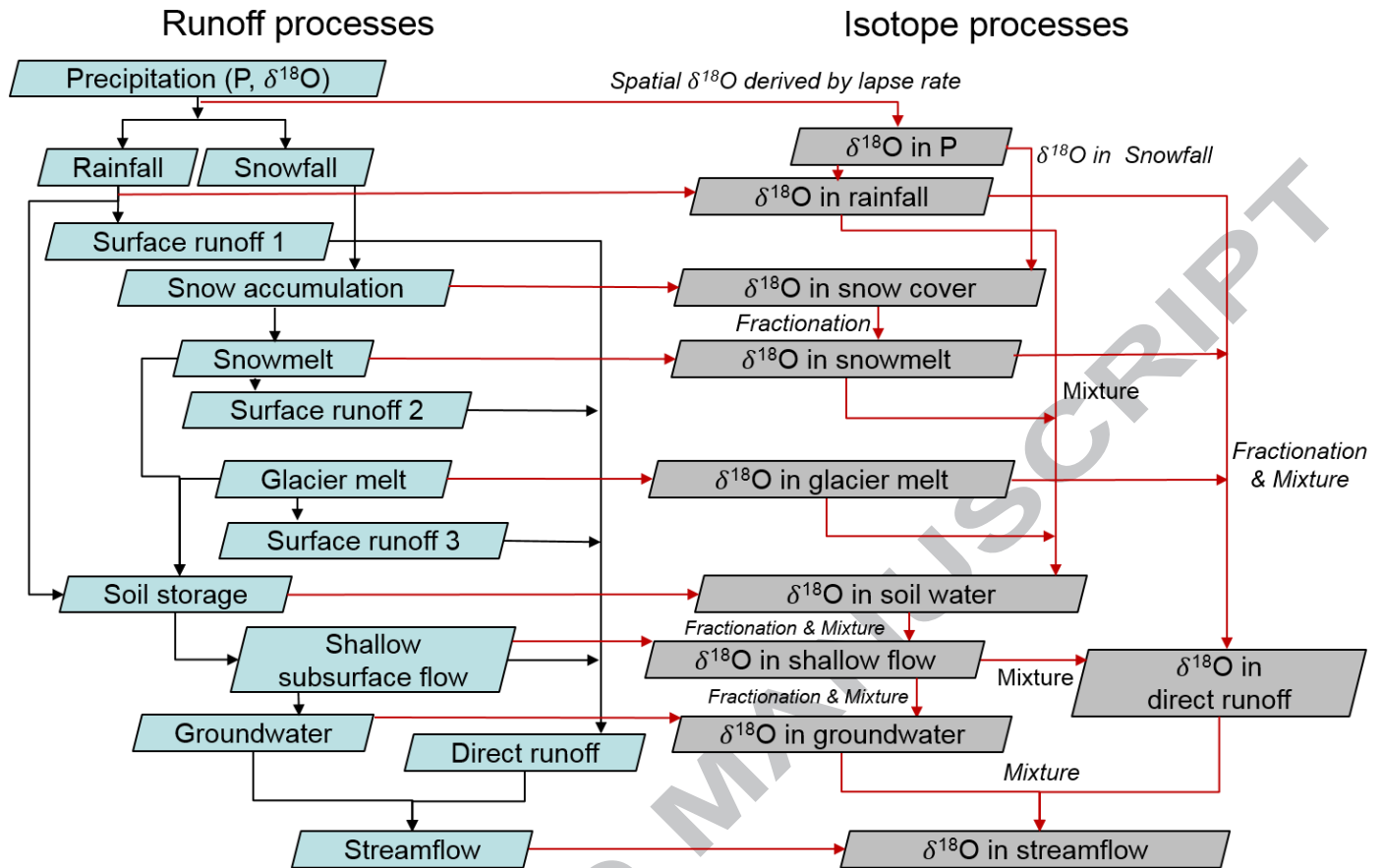


Fig. 3. Schematic representation of the isotope-hydrological integrated modeling approach in the IsoWASA model.

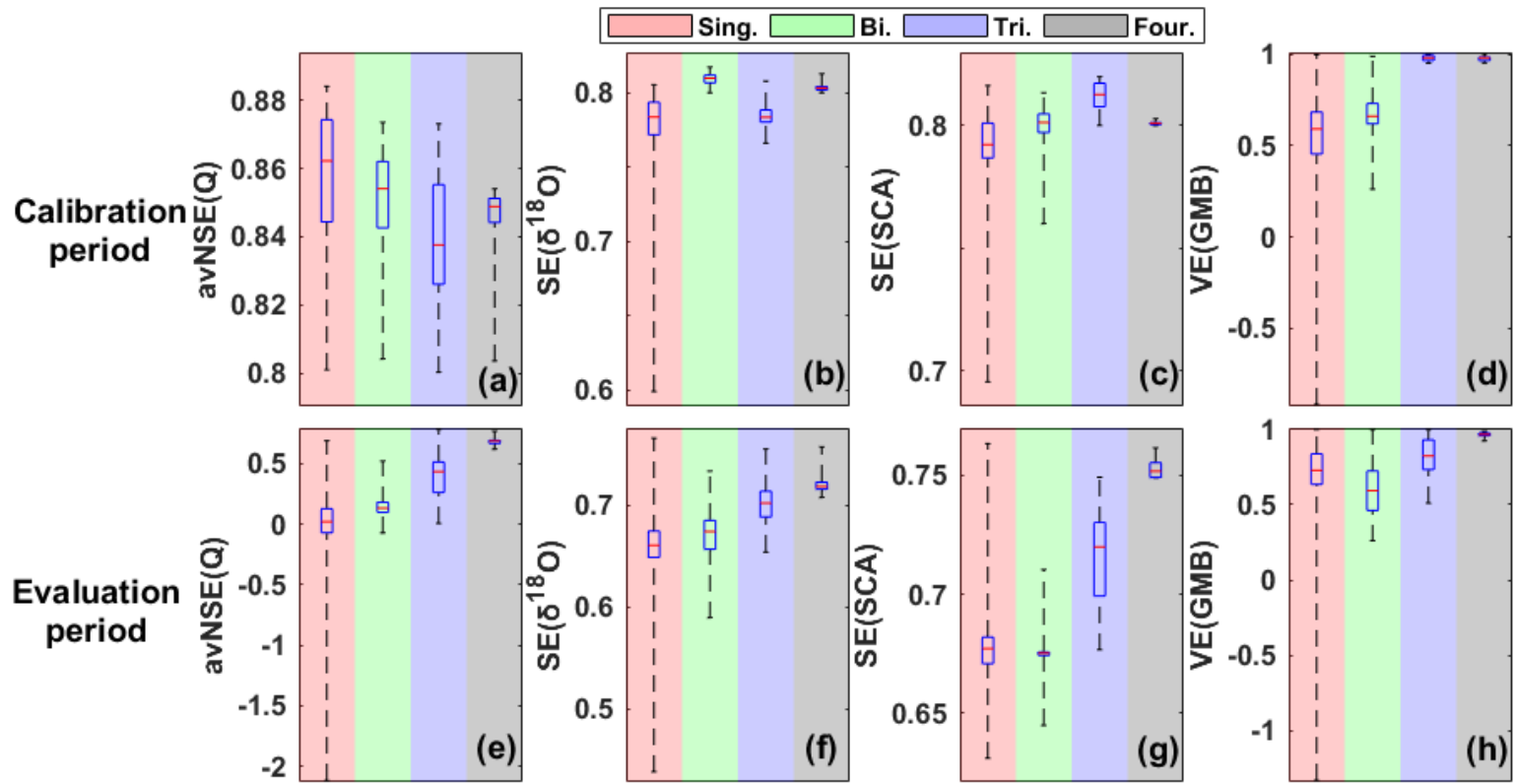


Fig. 4. Values of objective functions for observation datasets produced by four calibration methods in the calibration (a-d) and evaluation (e-h) periods. Sing.: single-dataset calibration using solely discharge; Bi.: bi-dataset calibration using discharge and isotopic composition of streamflow; Tri.: tri-dataset calibration using SCA, discharge and glacier mass balance; Four.: four-dataset calibration based on the tri-dataset calibration but additionally used the isotopic composition of streamflow.

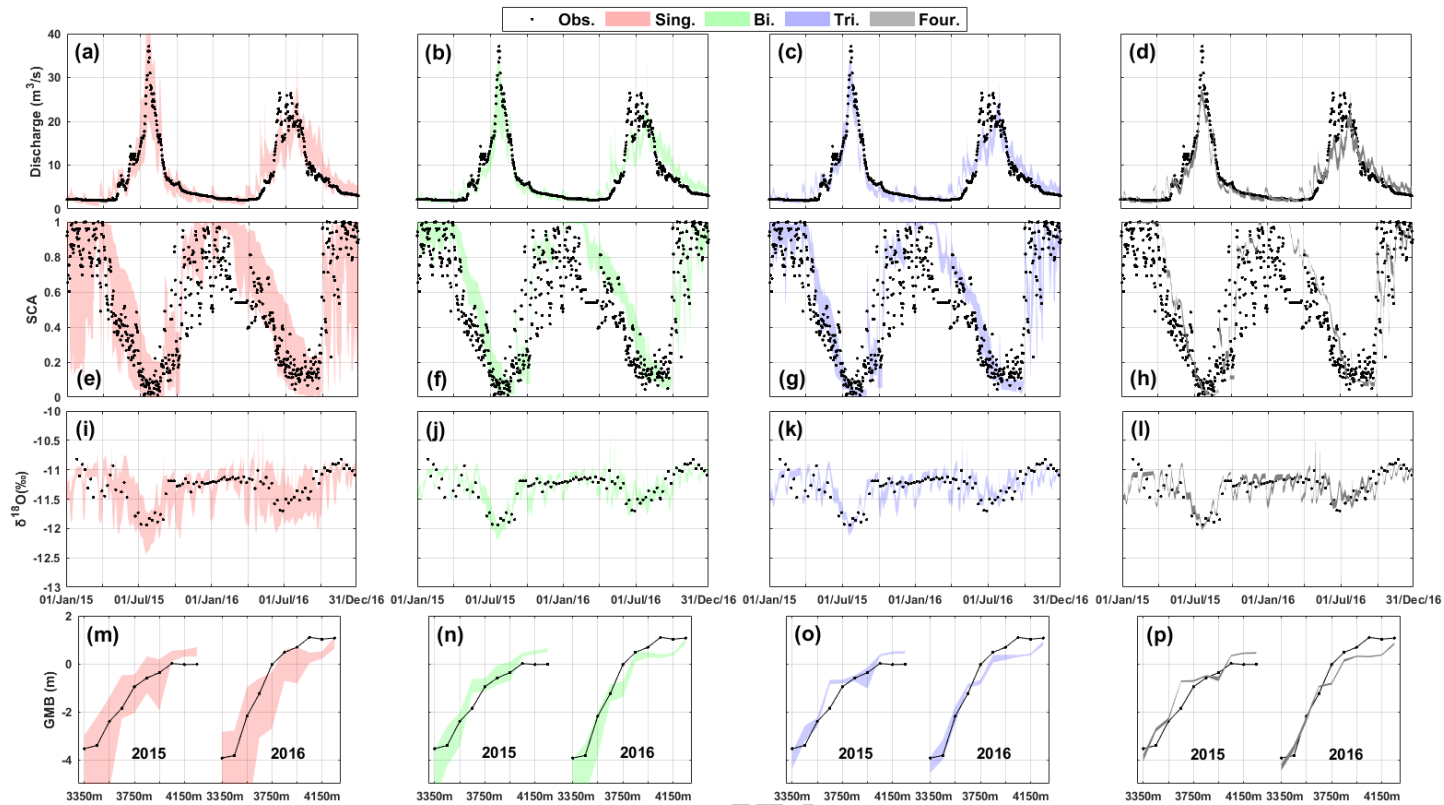


Fig. 5. Simulation ranges produced by the behavioral parameter sets of the four calibration methods in calibration period. (a-d) Daily discharge; (e-h) daily snow cover area (SCA); (i-l) daily $\delta^{18}\text{O}$ (‰) of streamflow; (m-p) annual glacier mass balance (GMB) at ten elevation bands. Sing.: single-dataset calibration using solely discharge; Bi.: bi-dataset calibration using discharge and isotopic composition of streamflow; Tri.: tri-dataset calibration using SCA, discharge and glacier mass balance; Four.: four-dataset calibration based on the tri-dataset calibration but additionally used the isotopic composition of streamflow.

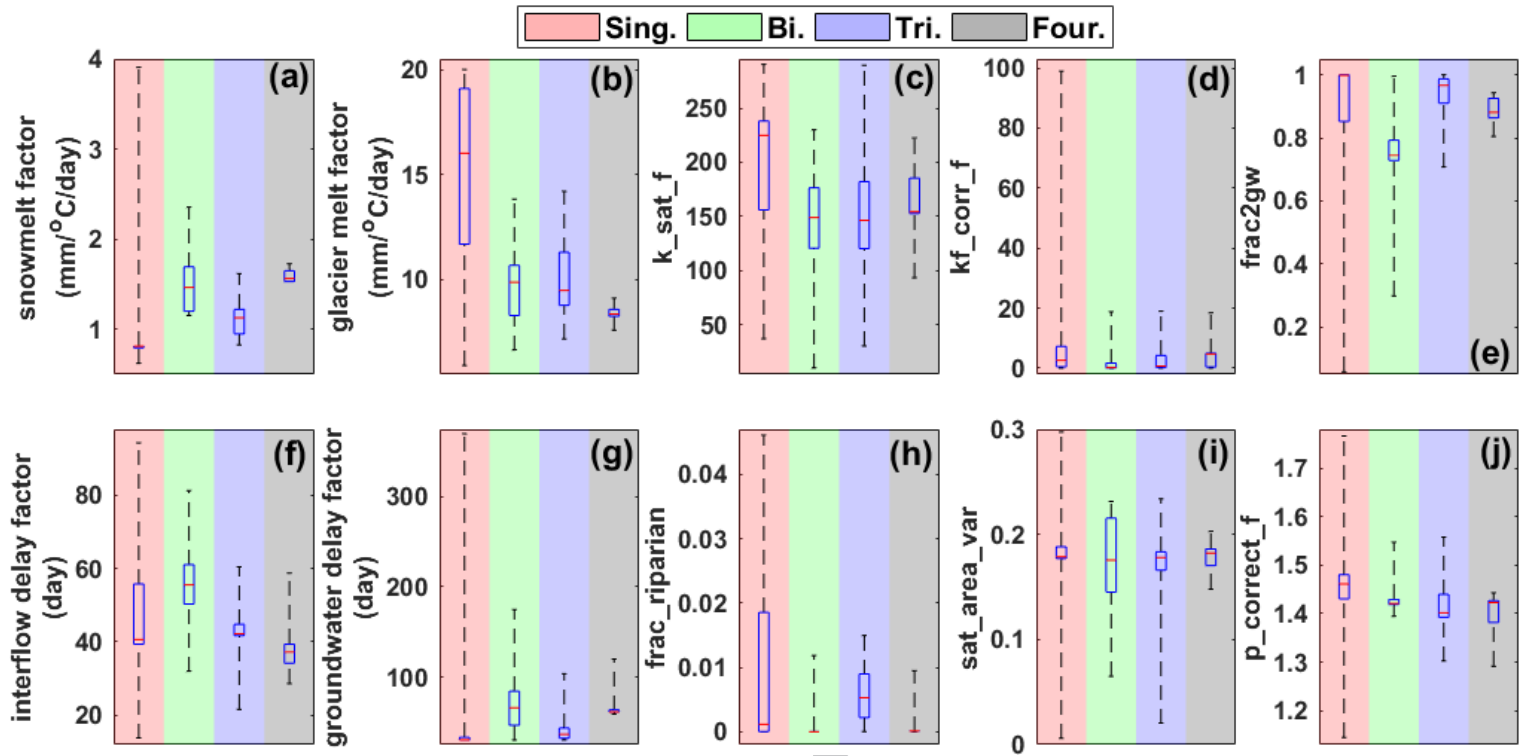


Fig. 6. Uncertainty ranges of the behavioral parameter sets produced by the four calibration methods.

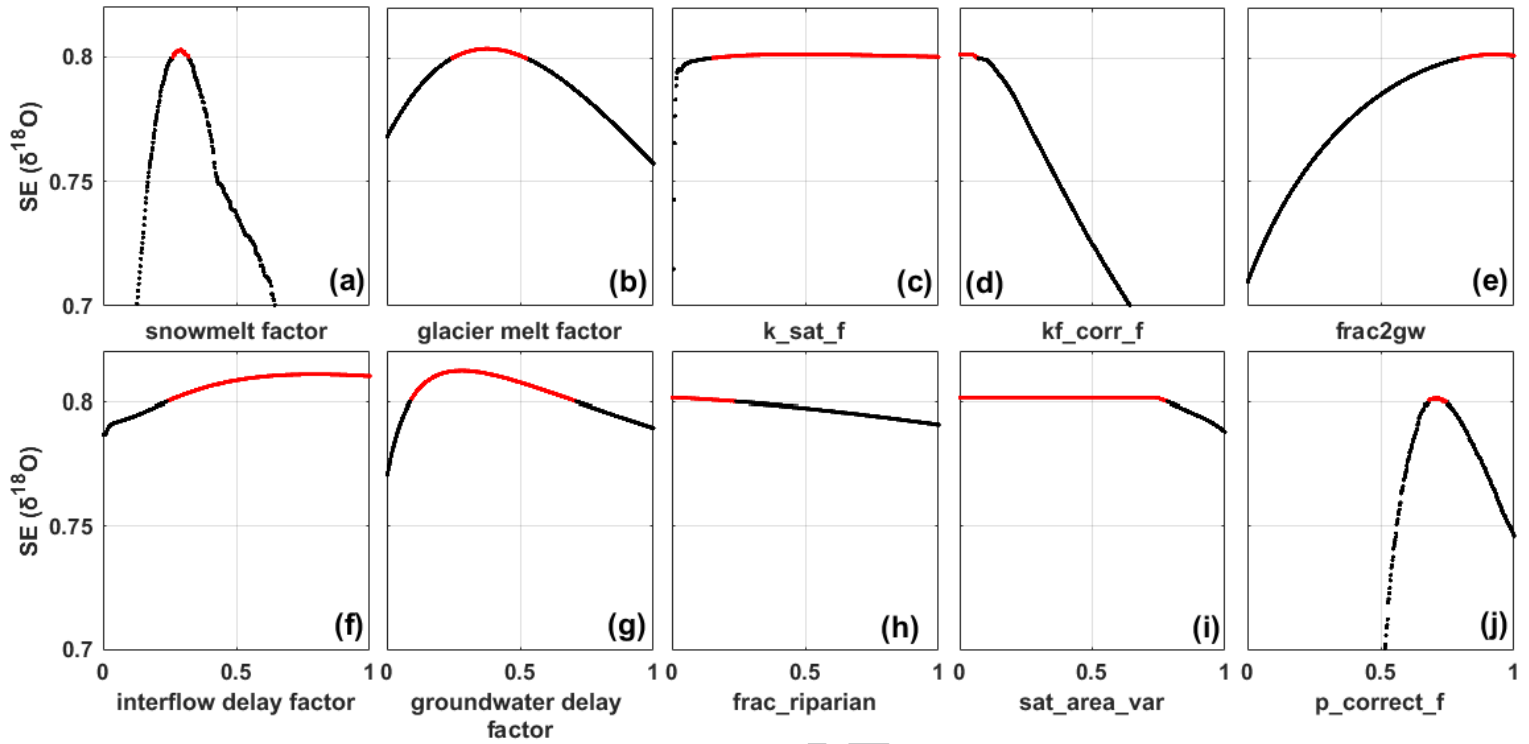


Fig. 7. Sensitivity of the performance for the simulation of streamflow $\delta^{18}\text{O}$ to individual parameter values. Parameter value ranges that produced SE value higher than 0.8 for the simulation of $\delta^{18}\text{O}$ are labeled in red. Parameter values range from their lower limits (normalized as 0) to the upper limits (normalized as 1) in Table 2.

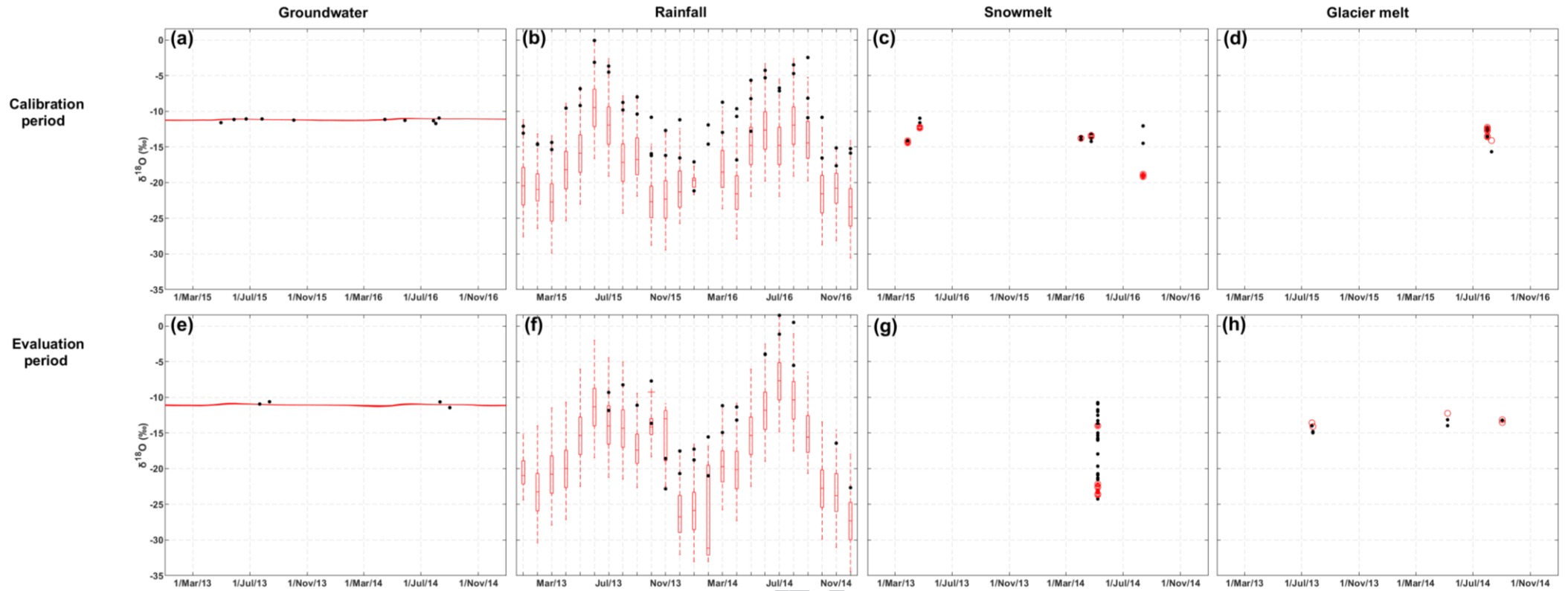


Fig. 8. Seasonality of simulated $\delta^{18}\text{O}$ of various runoff components produced by the four-dataset calibration in both the calibration and evaluation periods. Boxplots and red dots refer to simulated $\delta^{18}\text{O}$, and the black dots refer to measured $\delta^{18}\text{O}$ from available samples.

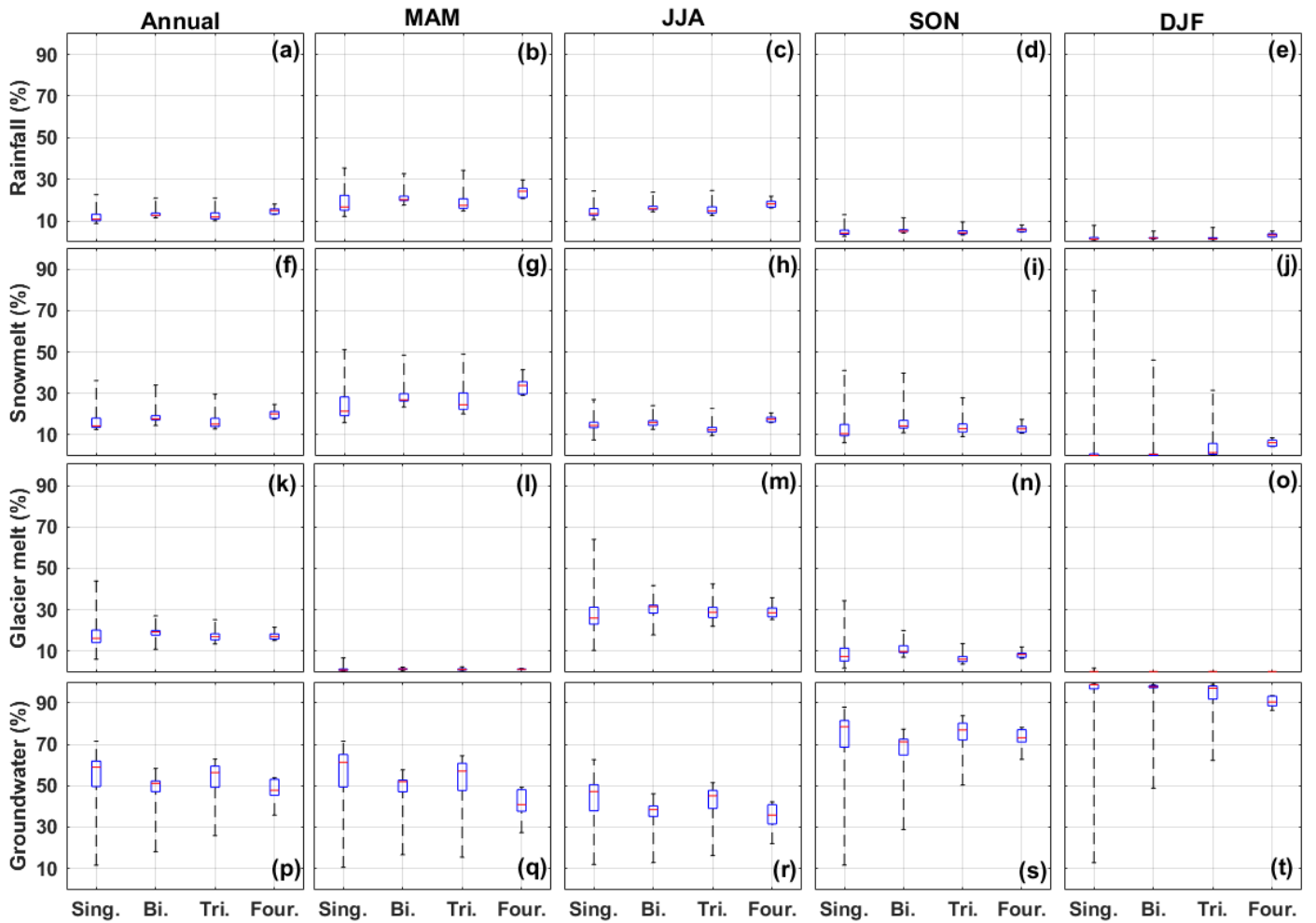


Fig. 9. Contributions of runoff components quantified by four calibration methods at annual and seasonal scales in the calibration period. The uncertainty ranges were produced by the individual behavioral parameter sets. ‘MAM’ is the spring months: March, April and May; ‘JJA’ is the summer months: June, July and August; ‘SON’ is the autumn months: September, October and November; ‘DJF’ is the winter months: December, January and February.

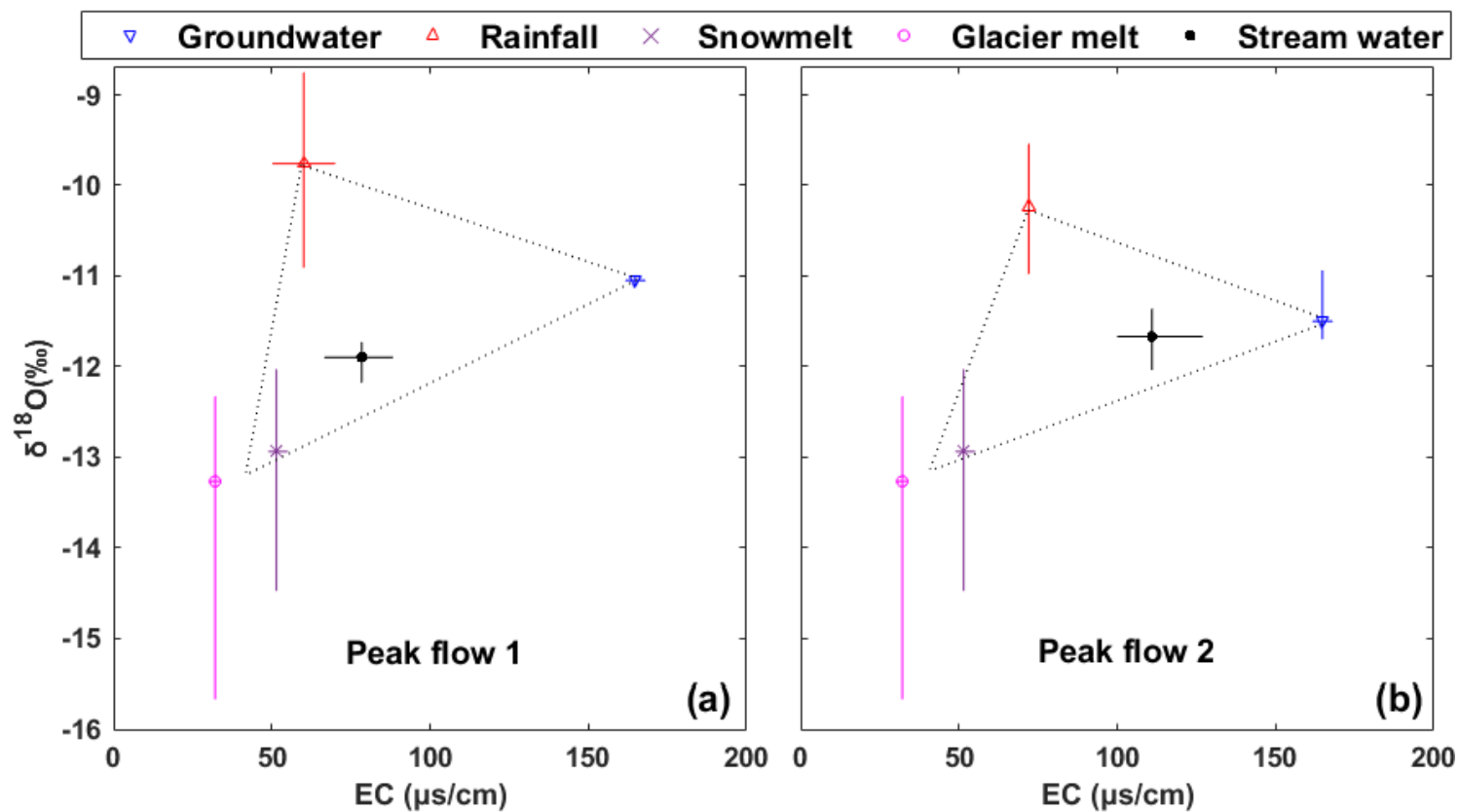


Fig. 10. Variability in the values of $\delta^{18}\text{O}$ and EC of water samples during two summer peak flow periods. Solid lines indicate the lower and upper limits in the measurements of the tracers for individual water sources. Markers indicate the average values. Isotope and EC values of stream water from both Baitik and Alplager stations are presented.

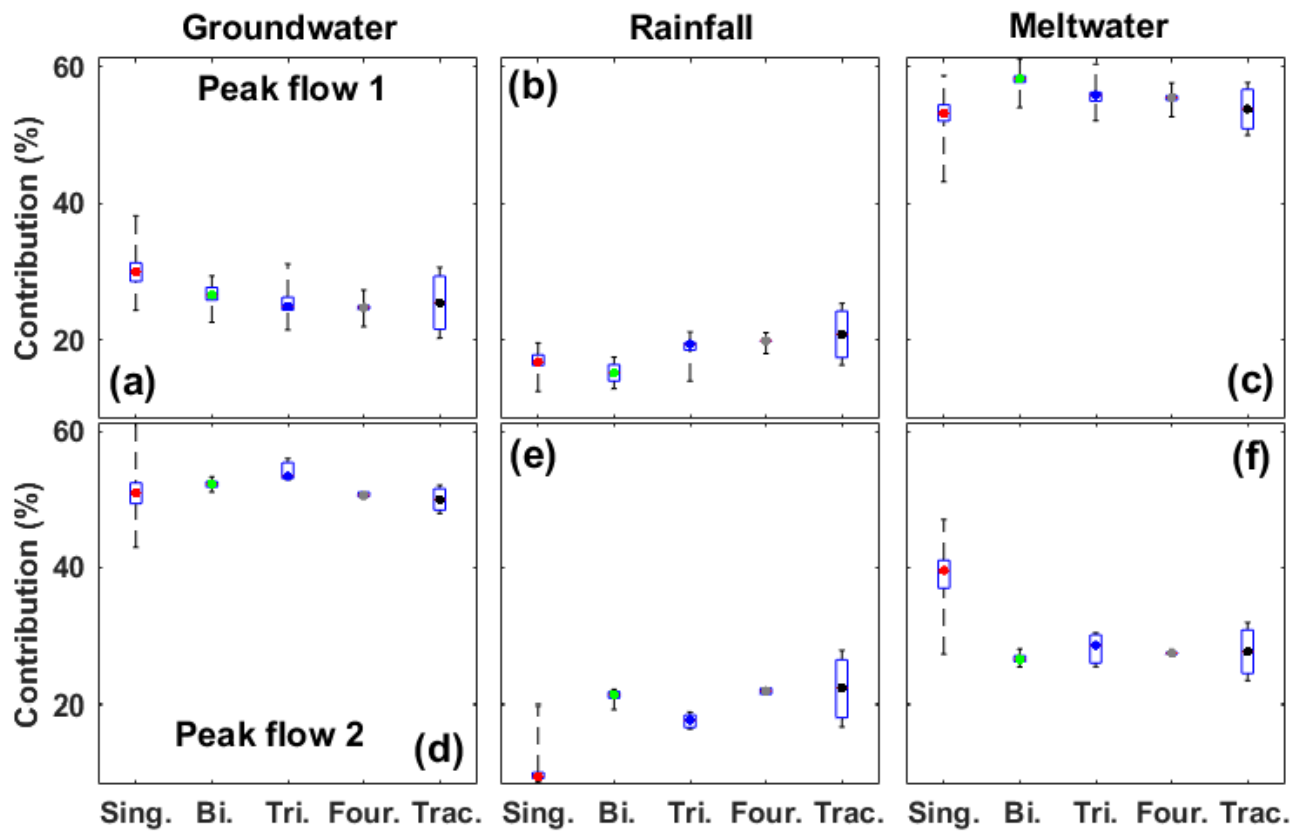


Fig 11. Contributions of runoff components for the two selected peak flows estimated by the four calibration methods and the tracer-based mixing model (Trac.). (a)-(c) for peak flow 1, and (d)-(f) for peak flow 2.

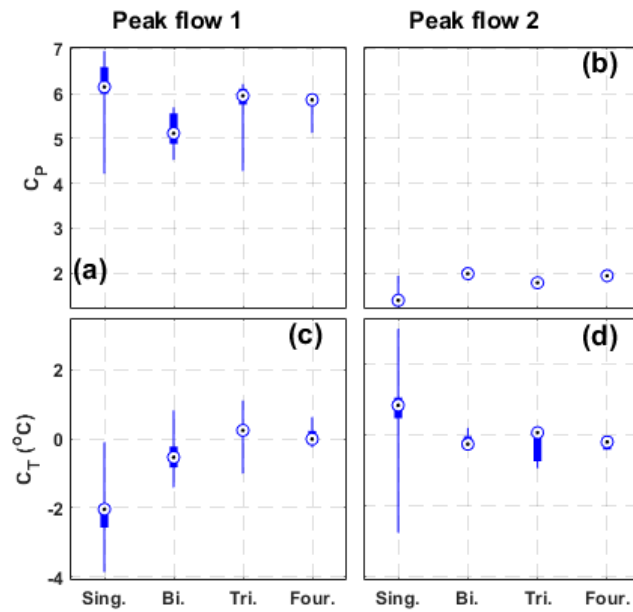


Fig. 12. Estimated correction factors for precipitation (C_p) and temperature (C_T) inputs during the two peak flow periods by the four calibration methods. ‘Sing.’ stands for the correction factors estimated by the single-dataset calibration method. ‘Bi.’, ‘Tri.’ and ‘Four.’ refer to the correction factors estimated by the bi-, tri- and four-dataset calibration methods.

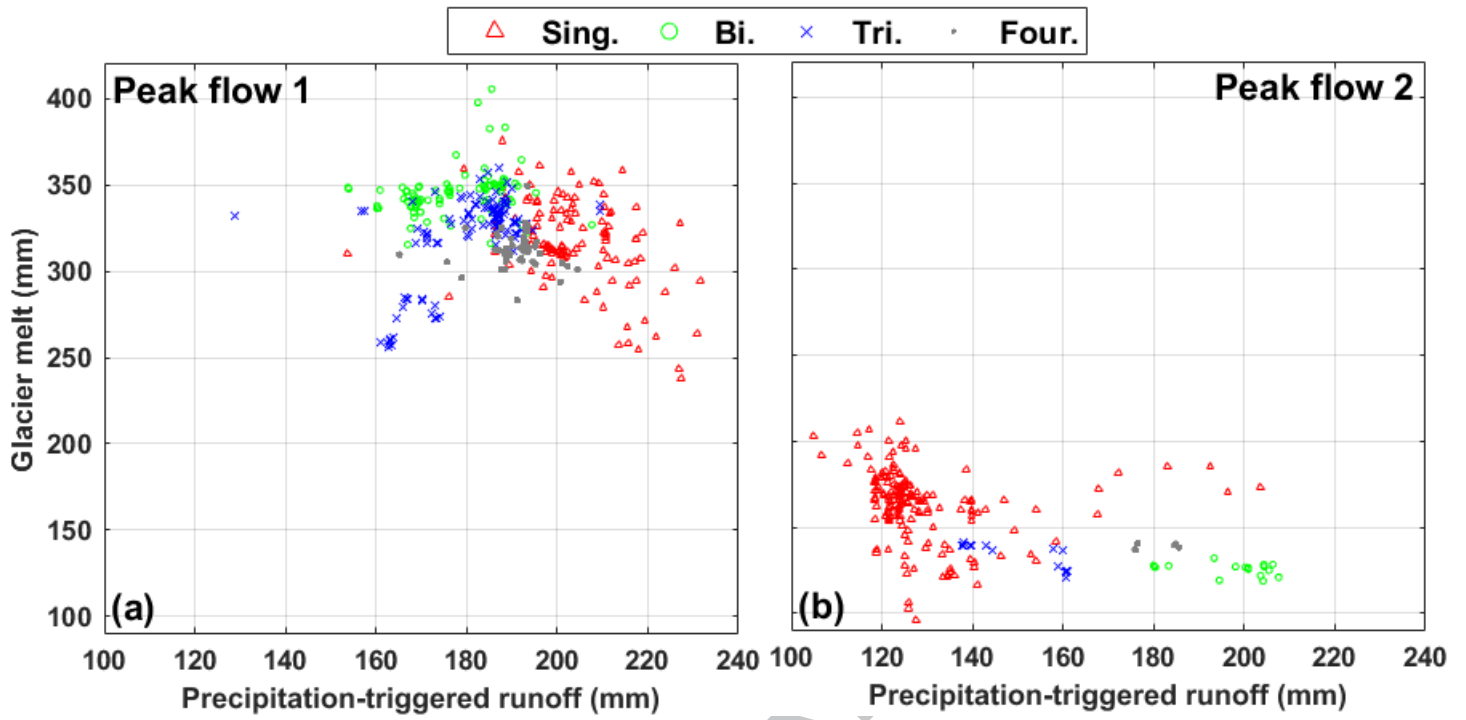


Fig. 13. Compensation between glacier melt and precipitation-triggered runoff during the two summer peak flow periods for different calibration methods.

Highlights

- 1) The proposed isotope-hydrological integrated modeling approach was able to reproduce the isotopic composition of streamflow, and improved the model performance in the evaluation period;
- 2) Involving water isotopic composition for model calibration reduced the model parameter uncertainty, and helped to reduce the uncertainty in the quantification of runoff components;
- 3) The isotope-hydrological integrated modeling approach quantified the contributions of runoff components comparably to a tracer-based mixing approach for summer peak flows, even using less water tracer data.