Originally published as:


DOI: https://doi.org/10.1175/JHM-D-17-0106.1
Comparison and Correction of High-Mountain Precipitation Data Based on Glacio-Hydrological Modeling in the Tarim River Headwaters (High Asia)

MICHEL WORTMANN  
Potsdam Institute for Climate Impact Research, Potsdam, Germany  

TOBIAS BOLCH  
Department of Geography, University of Zurich, Zurich, Switzerland  

CHRISTOPH MENZ  
Potsdam Institute for Climate Impact Research, Potsdam, Germany  

JIANG TONG  
National Climate Centre, Chinese Meteorological Administration, Beijing, China  

VALENTINA KRYSANOVA  
Potsdam Institute for Climate Impact Research, Potsdam, Germany  

(Manuscript received 20 June 2017, in final form 29 March 2018)

ABSTRACT

Mountain precipitation is often strongly underestimated as observations are scarce, biased toward lower-lying locations and prone to wind-induced undercatch, while topographical heterogeneity is large. This presents serious challenges to hydrological modeling for water resource management and climate change impact assessments in mountainous regions of the world, where a large population depends on water supply from the mountains. The headwaters of the Tarim River, covering four remote and highly glacierized Asian mountain ranges, are vital water suppliers to large agricultural communities along the Taklamakan Desert, northwest China. Assessments of future changes to these water towers have been hampered because of the large precipitation uncertainties. In this study, six existing precipitation datasets (observation-based reanalysis datasets, satellite observation datasets, and the output of high-resolution regional climate models) were compared over five headwaters of the Tarim River. The dataset incorporating the highest observation density (APHRODITE) is then corrected by calibrating the glacio-hydrological model Soil and Water Integrated Model–Glacier Dynamics (SWIM-G) to observed discharge, glacier hypsometry, and modeled glacier mass balance. Results show that this form of inverse modeling is able to inform the precipitation correction in such data-scarce conditions. Substantial disagreement of annual mean precipitation between the analyzed datasets, with coefficients of variation in catchment mean precipitation of 68% on average, was found. The model-based precipitation estimates are on average 1.5–4.3 times higher than the APHRODITE data, but fall between satellite-based and regional climate model results.

1. Introduction

Many arid and desert regions in Asia are highly dependent on water resources generated in the mountainous headwaters of High Mountain Asia (High Asia), such as the Indus, Syr Darya, Amu Darja, and the Tarim River (Kaser et al. 2010; Pritchard 2017; Bolch 2017). The latter has China’s largest endorheic river basin encircling the Taklamakan Desert. Approximately 10 million people depend on a vast irrigation agriculture along the rivers that provides subsistence farming and produces large shares of China’s cotton and fruits (Thews 2011; Tao et al. 2011; Feike et al. 2015; Rumbaur et al. 2015). The mountain streams originate as either summer rain, snow, or glacier melt in the Karakoram,
Pamir, Kunlun Shan, and Tien Shan ranges. Despite their importance to downstream communities, the water cycle of the Tarim River and in High Asia in general is poorly gauged and understood (Hewitt 2005; Immerzeel et al. 2012a; Ragettli et al. 2013; Winiger et al. 2005).

The region’s high-mountain precipitation is determined by a combination of synoptic climate patterns and small-scale orographic precipitation, resulting in highly heterogeneous and localized precipitation fields (Barry 2008). Most of the moisture arriving in the highly continental Tarim basin is driven by the westerlies and the southwestern Indian monsoon creeping over the Karakoram range in the most southwestern part (Aizen et al. 1995; Mölg et al. 2014; Yao et al. 2012). The combination of a sparse meteorological station density and topographical heterogeneity spell an unrelenting data scarcity for any kind of water resource assessment. The few meteorological stations that exist are mostly located at more accessible, low-lying locations in valley bottoms and at the populated edge of the Taklamakan Desert (Tao et al. 2011). This introduces a measurement bias with a tendency to much drier conditions, a situation common in mountain regions around the world (De Jong et al. 2005).

A further precipitation bias is introduced by gauge undercatch, which is especially important in mountainous regions because of strong winds and high proportions of solid precipitation (Yang et al. 1999; Adam et al. 2006; Groisman et al. 1991). For example, unshielded precipitation gauges may collect less than 50% of actual solid precipitation at high wind speeds (Kochendorfer et al. 2017; Yang and Ohata 2001). Overcoming these sources of negative bias for accurate water resource assessments is a challenge, but promising recent advances such as high-resolution satellite-based precipitation measurements and better glacier inventories need to be utilized.

There have been numerous studies on observed changes in the Tarim basin’s climate, but they largely fail to address mountain precipitation. A general increasing trend of both precipitation and temperature has been observed in the last decades (Changchun et al. 2008; Chen et al. 2006; Tao et al. 2011; Fan et al. 2011) with especially the Tarim River’s most important tributary, the Aksu River, experiencing an increase in discharge (Wang et al. 2008; Krysanova et al. 2015). AVHRR and MODIS remotely sensed snow cover area at high elevations has been investigated and was shown to have undergone only mild changes over the past three decades (Peters et al. 2015). Changchun et al. (2008) detected slight increases in snow cover area and significant correlation with precipitation variability, but none with temperature. None of these observation-based studies have addressed precipitation biases in the Tarim River’s headwaters, where the vast majority of the region’s freshwater is generated and estimates of annual total catchment precipitation are absent.

The recent availability of spatially distributed precipitation products, observation-based interpolations, and reanalysis datasets has provided hydrological modelers with convenient input data, but they largely inherit the low-elevation bias of their boundary conditions, and differences between station densities exist (Lorenz and Kunstmann 2012). Prominent examples of global coverage at 0.5° resolution are the ERA-40/ERA-Interim (Dee et al. 2011) and Global Precipitation Climatology Centre (GPCC; Schneider et al. 2015) datasets, which have themselves been used for reanalysis (Weedon et al. 2011). A regional dataset for Asia with the highest station density is APHRODITE at 0.25°–0.5° resolution (Yatagai et al. 2012). Tong et al. (2014) compared some of these precipitation datasets over the Tibetan Plateau and several southeast draining headwaters to station observations and found considerable variability. A major problem in mountainous regions with these datasets is that their relatively coarse grids (0.25°–2°), which are a poor match for the complex and poorly represented process of orographic precipitation in those reanalysis models (You et al. 2012; Böhner 2006).

An improvement comes from recent satellite-based products of the Tropical Rainfall Measurement Mission (TRMM) that provide indirect measurements for each grid cell and are extended through modeling. Although it relies on algorithms that are mostly ill suited for small-scale, solid, or short precipitation events (Bookhagen and Burbank 2006; Huffman 2006), it at least performs a spatially homogeneous sampling, and improvements are being made in resolution, algorithms, and sensors (Smith et al. 2007). Another type of precipitation product comes from dynamical, regional climate models that downscale global products to finer resolutions. They are relatively rare, but one exists for the Tibetan Plateau and surrounding mountain ranges with simulations since 2000. The High Asia Refined analysis (HAR) dataset provides detailed and process-based precipitation fields at a 10-km resolution (and 30 km for greater High Asia), but only for 2000–14 (Maussion et al. 2014). It provides valuable estimates of high-altitude water fluxes that are independent from observations. Although a variety of both simulated and observation-based datasets is available, no assessments exist for the data-scarce Tarim River headwaters to determine catchment-wide precipitation relevant for modeling efforts.

While hydrological models use precipitation as a driving variable, they have also been instrumental in verifying catchment-wide precipitation estimates by...
calibrating them to available discharge observations. Orographic precipitation is traditionally estimated using linear elevation gradients, similar to temperature. However, they can be highly local and variable with elevation zone and thus rely on an acceptable station network (Sevruk and Mieglitz 2002). Inverse modeling has been used to utilize discharge observations and correct precipitation accordingly (Valéry et al. 2009; Duethmann et al. 2013; Henn et al. 2016). Although it relies heavily on the accurate modeling of other water balance components (most notably evapotranspiration), discharge records often provide the best available evidence in data-scarce regions for catchment precipitation that is higher than meteorological observations suggest. Using information on snow cover has been shown to improve the estimates significantly (Henn et al. 2016; Duethmann et al. 2014). For example, Henn et al. (2015) estimate basinwide precipitation totals using a Bayesian approach and several hydrological models. It was shown in a follow-up study that uncertainties are reduced by one-third if snow water equivalent information is included (Henn et al. 2016).

An important drawback of this method in glacierized regions is the added or withheld discharge from potentially negative or positive glacier mass balances, so those must be known or reasonably constrained. If that is possible, glaciers may even function as gauges of high-altitude precipitation by inverse glacier modeling, which is shown by studies in the Indus basin (Immerzeel et al. 2012b, 2015). By estimating glacier melt, the water balance of individual glacier catchments is closed and local precipitation gradients between glacier accumulation zone and nearby meteorological stations as inferred. To accurately estimate catchment precipitation in the poorly gauged Tarim headwaters, these are promising approaches that have not been tested in existing models.

A similar modeling approach is proposed here to untangle the complexities of underestimated precipitation and glacier mass balances and to provide precipitation estimates that are based on a reasonable glacier and water balance.

2. Study area, data, and methods

a. The Tarim River headwaters

The Tarim River has one of the largest endorheic catchments of the world with a size of about $800 \times 10^3 \text{km}^2$. As the catchment encompasses the Taklamakan Desert and various salt lakes, the hydrologically active catchment is less than half the area. The river is fed by three large tributaries with their confluence at the northern edge of the desert: the Aksu River originating in the central Tien Shan to the north, the Hotan River originating in the Kunlun Shan to the south, and the Yarkant River originating in the Karakoram and the eastern Pamir (see Fig. 1). As the desert climate in the lower parts of these tributaries produces virtually no river runoff (except for rare extreme rain events; annual potential evapotranspiration exceeds precipitation by a factor of 30-50), the vast majority of river discharge is generated in the glacierized, mountainous headwaters. This study focuses on the five catchments of the gauging stations that are situated at the boundary of the Taklamakan Desert, that is, before river abstractions and significant transmission losses occur (Fig. 1, Table 1).

The climate of the headwaters is highly continental with a strong seasonality governed by the westerlies (Aizen et al. 1995; Maussion et al. 2014). River regimes exhibit a strong peak in summer when snow and ice melt coincides with the precipitation peak. In the Aksu catchment, about 80% of precipitation is measured between the months of April and September. Yarkant and Hotan catchments receive about 60%-70% of precipitation over the same period, as the westerlies remain active in winter in the southwestern parts of the domain (Karakoram). Mean summer (winter) temperatures fall between 3°C and 16°C (between -22°C and -9°C; Weedon et al. 2011). Precipitation gauge stations in the catchments are sparse (further discussed below). The reanalysis dataset APHRODITE (Yatagai et al. 2012) that incorporates most of them indicates annual totals of 230–314 mm a$^{-1}$ in the Aksu catchments and 62–113 mm a$^{-1}$ in the Hotan and Yarkant catchments.

The mountain ranges surrounding the Taklamakan Desert comprise steep, high-altitude terrain that has given rise to an extensive glacier cover accounting for significant proportions of the catchment areas considered here (Table 1). The mean elevation of the Aksu catchments is 3500 m MSL and includes the highest peak of the Tian Shan, the Jengish Chokuso (Pik Pobedy in Russian or Tömür in Uyghur) at 7439 m MSL. The two catchments cover a total glacier area of about 3350 km$^2$ (≈2008; Pieczonka and Bolch 2015; Osmonev et al. 2013). A unique glaciological feature of the Aksu headwaters is the ice-dammed Merzbacher Lake that sends near-annually reoccurring subglacial outburst floods (jökulhlaups) downstream with consequences for communities and modeling efforts (Glazirin 2010; Wortmann et al. 2014). The two catchments of the Hotan River span the northwestern edge of the Tibetan Plateau at a mean elevation of 4695 m MSL, including the Liushi Shan at 7167 m MSL, the highest summit of the Kunlun Shan range. Glaciers cover some 5880 km$^2$ (≈2010) and are mostly cold-based at very high altitudes (Shangguan et al. 2007). The
Yarkant headwater catchment has an average elevation of 4425 m MSL with the Karakoram in the south, including the K2 at 8614 m MSL, and the East Pamir in the north. A total glacier area of about 5610 km² (2010) is found within the catchment at extremely high and steep altitudes (Frey et al. 2014).
The Tarim headwaters suffer from severe data scarcity due to the high-altitude, heterogeneous terrain, the associated inaccessibility, and very sparse population. Long-term meteorological stations are therefore extremely difficult to maintain and are scarce. Moreover, the mountain climate exhibits strong vertical gradients (e.g., temperature stratification and orographic precipitation), leading to microclimates and short correlation distances (Beniston 2006). The few meteorological stations are disproportionately located in valley or low-altitude locations (or worse still, at the edge of the Taklamakan Desert; Tao et al. 2011). This introduces a low-altitude bias in the climate data, which leads especially to underestimations of precipitation, making some form of correction and extrapolation necessary (Immerzeel et al. 2012b).

b. Precipitation datasets

Six precipitation datasets were selected that cover all five catchments (Table 2). They may be broadly distinguished by the manner in which they assimilate observations and produce distributed precipitation fields. APHRODITE and GPCC use direct precipitation ground observations and are thus referred to as direct precipitation datasets (DPD). The satellite-based datasets TRMM Multisatellite Precipitation Analysis (TMPA) 3B42v7 and Global Precipitation Measurement (GPM) 3IMERGM, as well as the ERA-Interim reanalysis dataset, assimilate mainly indirect observations from space (radiance, sea surface temperature) and some ground-based datasets. We refer to them as satellite-based precipitation datasets (SPD). ERA-Interim falls into this category because it is mainly driven by indirect satellite measurements and does not assimilate direct precipitation observations. A third category includes the COSMO Community Land Model (CCLM) and HAR datasets that are results of regional climate models. They only use global climate model results as boundary conditions. We refer to them as model-based precipitation datasets (MPD).

The six datasets were chosen to assess the region’s precipitation by means of all common sources (direct observations, satellites, and climate models). This range allows an adequate evaluation of uncertainties associated with precipitation data available to hydrological modelers for such data-scarce, mountainous catchments, as has been done for other high mountain regions (Tong et al. 2014; Henn et al. 2018). The gridded precipitation dataset of the Chinese Meteorological Administration (CMA) was not considered because it only covers the Chinese territory and is similar in quality to the APHRODITE dataset. We only describe the APHRODITE dataset here because it was used in the calibration of the hydrological model and is generally considered the best dataset available for the region (Tong et al. 2014; Lutz et al. 2014). The others are described in the online supplemental material.

The APHRODITE dataset is a gridded, daily precipitation interpolation of the densest gauge network in

<table>
<thead>
<tr>
<th>Name</th>
<th>Period</th>
<th>Resolution</th>
<th>Boundary conditions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>APHRODITE</td>
<td>1951–2007</td>
<td>0.25°/0.5°</td>
<td>Point observations</td>
<td>Yatagai et al. (2012)</td>
</tr>
<tr>
<td>GPCC</td>
<td>1901–2010</td>
<td>0.5°</td>
<td>Point observations</td>
<td>Schneider et al. (2015)</td>
</tr>
<tr>
<td>TMPA/GPM</td>
<td>1998–present</td>
<td>0.25°/0.1°</td>
<td>Radar data from multiple satellites</td>
<td>Liu (2015); Huffman et al. (2014)</td>
</tr>
<tr>
<td>ERA-Interim</td>
<td>1979–present</td>
<td>0.5°</td>
<td>Satellite and point observations</td>
<td>Dee et al. (2011)</td>
</tr>
<tr>
<td>CCLM</td>
<td>1959–2001</td>
<td>0.44°</td>
<td>ERA-40</td>
<td>Rockel et al. (2008)</td>
</tr>
<tr>
<td>HAR</td>
<td>2001–14</td>
<td>10 km/30 km</td>
<td>GFS-FNL</td>
<td>Maussion et al. (2014)</td>
</tr>
</tbody>
</table>
Asia (Yatagai et al. 2012). A Sheremap-type interpolation scheme was employed to construct daily 0.25° and 0.5° precipitation fields, considering dominant wind exposition and monthly varying correlation distances. While they are the best data available, the underlying station density is still extremely poor, as illustrated in Fig. 2. The Hotan catchments are devoid of any gauging stations, with the closest stations located at the edge of the Taklamakan Desert (see also Tao et al. 2011). Similarly, the Kaqun (Yarkant) catchment has only one station within the catchment boundary, with most surrounding stations at significantly lower-lying locations. The Karakoram range in the southwest of the catchment is particularly remote from any stations.

Despite the drawbacks of APHRODITE, it is the best dataset available for longer-term hydrological modeling, as it covers more than five decades and has a high correlation with observed discharge (in comparison, e.g., to the MPD). The comparison is focused on annual mean precipitation as it summarizes gauging biases well in a highly seasonal climate and is available in all datasets.

**c. SWIM-G**

The glacio-hydrological model Soil and Water Integrated Model–Glacier Dynamics (SWIM-G) was used to simulate both glacier and the catchment hydrology in a tightly integrated approach (Wortmann et al. 2016). SWIM-G was developed to overcome the conflict of scale between individual glaciers and the catchment hydrology when modeling larger, highly glacierized, typically data-scarce catchments. It relies on computational units that disaggregate complex terrain into parts of similar elevation, aspect, and hydrological subbasin as opposed to gridded (fully distributed) or empirical approaches (e.g., Immerzeel et al. 2012a; Huss et al. 2010). The units are conceptually similar and computationally equal to the hydrological response units (HRUs), enabling a tight integration between the glaciological and hydrological model over large catchments. Typical topographical properties govern the disaggregation considering the catchment characteristics and the desired level of detail. These properties are elevation zones with intervals of 100 m, four aspect classes, and the subbasin boundaries.

The mass balance is evaluated through snow accumulation based on a threshold temperature and an enhanced degree-day melting approach. The latter accounts aspect and terrain shading, debris melt-out, and sublimation. Like most hydrological processes, the glacier processes are computed daily for each unit, and ice is routed laterally between them according to an adapted Glen’s Flow Law approach (Marshall et al. 2011). More detail on the model and the specific input data for the Tarim catchments is provided in the supplemental material.

**d. Precipitation correction**

Where meteorological information has to be extrapolated over great distances both horizontally and/or vertically, a method to correct for orographic precipitation is paramount to the accurate modeling of both the glaciers and the catchment hydrology (Immerzeel et al. 2012b, 2015). Most studies use linear gradients to vary precipitation with elevation over a complex terrain with typical ranges of 0.05%–0.5% m⁻¹ (Sevruk and Mieglitz 2002; Hock and Holmgren 2005; Immerzeel et al. 2014). However, representing variance of precipitation as a linear function of elevation is inherently local, highly variable over varying altitude ranges, and generally unsuitable for elevations below the reference altitude, as precipitation would diminish to 0.

In this study, precipitation was corrected by a nonlinear function of altitude taking account of varying
gradients and an eventual decrease at very high elevations. The correction factor $f_c$ in Eq. (1) remains 1 over lower elevations for which observations are available, but increases exponentially up to a maximum gradient $a$ ($\% \text{ m}^{-1}$). The gradient is then reduced until a maximum correction $c$ (dimensionless) at altitude $m$ (m MSL) is reached and decreases again at higher elevations thereafter:

$$f_c(z) = (c - 1) \times \exp \left\{ -\frac{a}{(c - 1) \times 100} \right\} \times (z - m)^2 + 1. \quad (1)$$

This is a more continuous approach than the combination of two linear functions proposed by Immerzeel et al. (2012b). Parameter $c$ is effectively the greatest correction applied, while the altitude $m$ is the physical limit of the atmosphere to lose more moisture. For the Tien Shan, Aizen et al. (1995) described this exponential change in precipitation and provide valuable observations to constrain the parameters of the correction function for the multiobjective calibration. The investigations by Immerzeel et al. (2012b, 2015) helped to constrain the parameter ranges for the Karakoram region.

To account for local differences in precipitation underestimations across a catchment, the correction term $c$ was distributed over selected glacier groups, totaling 6–11 in each catchment (Fig. S1 in the supplemental material). They were identified according to major glacier clusters at mountain massifs and their exposition. For subbasins without any glaciers, the values were interpolated between the groups using an inverse distance weighting (IDW) interpolation.

e. Model initialization, calibration, and validation

The model was calibrated using a multiobjective automatic optimization, considering observed discharge at the five outlet gauge stations, observed glacier area, and mass balances, where those were available. A calibration run consisted of the glacier initialization and a full hydrological run starting in the year with available discharge data (1964) and covering the reference period 1971–2000. The reference period was chosen because daily discharge data were available for the 1970s and 1980s for most stations (with some years missing, as listed in Table S1), and the geodetic mass balance analysis for the Aksu catchment by Pieczonka and Bolch (2015) is for the period 1975–2000. Following the split-sample calibration–validation approach, the available discharge data were split into 4-yr periods (the calendar years 1964–87, yielding six periods). The model was calibrated to all odd-numbered periods (1964–67, 1972–75, 1980–83) and validated to the even-numbered periods (1968–71, 1976–79, 1984–87). This method was preferred, because the day-to-day correlation of the precipitation data with discharge was found to be worse in the 1980s than in the 1970s (most likely caused by changes in meteorological station density). This meant that choosing either decade as the calibration or validation period would distort the results of the validation.

The calibration was supported by the recent findings of catchment-wide glacier mass balances for the Aksu catchments by Pieczonka and Bolch (2015). Since catchment-wide mass balances or representative local mass balance records for the Hotan and Yarkant catchments were not available, we used three initial mass balance scenarios and compared those to regional research findings (e.g., Holzer et al. 2015), as well as accounting for those scenarios in the uncertainty analysis. The initialization and calibration steps are further described below, and all calibration parameters with their ranges are provided in Table 3.

Glacier dynamics models need to be initialized over typical equilibrium time frames (200–1000 years) in order to ensure consistency between driving data, model structure, and ice cover (Marshall et al. 2011; Clarke et al. 2015). We initialized the SWIM-G model over 300 years and compared the resultant glacier cover to the observed glacier inventory. The fit of the ice hypsometry in each catchment was used to match the initialized glacier area to the observed, through the calibration of all parameters affecting the glaciers. It is quantified through the $\chi^2$ objective function for histogram comparison between the two hypsometries as follows:

$$\chi^2 = \sum \frac{(A_o,i - A_s,i)^2}{A_o,i + A_s,i} \times 100, \quad (2)$$

with observed $A_o$ and simulated $A_s$ glacier area for the 50-m elevation band $i$, expressed as percent. It ensures good fit also at the tail ends of the hypsometries due to the sum of observed and simulated areas in the denominator. The $\chi^2$ is calculated for each glacier group of the catchment and weighted by the observed glacier area.

The initialization requires a climate period with known, quasi-stable mass balance conditions that is representative for the observed glacier cover. Although glaciers are never in a perfect equilibrium state, glaciers are globally identified as having stable conditions between the 1960s and the mid-1970s (Dyurgerov 2010). This is also true for the Tien Shan (Sorg et al. 2012; Farinotti et al. 2015). However, historic observations are limited for Kunlun Shan and Karakoram. Recent findings based on declassified satellite imagery indicate that glaciers in the Karakoram have been stable or showed only slight mass loss since the 1970s until today (Bolch et al. 2017;
Zhou et al. (2017). Dyurgerov (2010) provides “best guess” regional glacier mass balances since the 1960s for all major glacier regions. For the initialization period between 1960 and 1975, mean mass balances of $-282 \text{ mm w.e. a}^{-1}$ are given for the Tian Shan, $-237 \text{ mm w.e. a}^{-1}$ for the Tibetan Plateau, $176 \text{ mm w.e. a}^{-1}$ for the Karakoram, and $-252 \text{ mm w.e. a}^{-1}$ for the Pamir. While these values should be used as guidelines only, as they mostly originate from single glaciers, they show mass balances less negative than the global mean of $-2330 \text{ mm w.e. a}^{-1}$ in the period 1976–2005 (Zemp et al. 2009). This imbalance was corrected by a residual mass balance parameter during the initialization as was previously done by Clarke et al. (2015). A negative mass balance means ice is evenly added to all glacier surfaces at the end of the year and vice versa for a positive mass balance. Although this parameter is only used in the initialization run, it has a great impact on the mass balance simulations, as it changes the imbalance of the glacier cover at the beginning of the simulation. The residual mass balance was calibrated in the Aksu catchment, where mass balances estimates exist, and prescribed in a scenario approach in the Hotan and Yarkant catchment where they are absent.

Pieczonka and Bolch (2015) provide a catchment-wide mean mass balance including uncertainty range for the Xiehela (Aksu) catchment for the period 1975–2000. Because of their proximity and topographical similarities, we used these average values also in the neighboring Shaliguilanke catchment. The objective function for the automatic calibration assigned a score between 1 (worst) and 0 (exact match) based on the normal distribution probability function using the mass balance (MB) as the mean value and the uncertainty range as the standard deviation. It is defined as follows:

$$MB_c = 1 - \exp \left[ -\frac{(\bar{B}_o - \hat{B}_o)^2}{\epsilon^2} \right],$$

where $\bar{B}_o$ is the mean annual mass balance over the study period by Pieczonka and Bolch (2015) ($0.35 \text{ m w.e. a}^{-1}$) with an uncertainty range of $\epsilon (= 0.34 \text{ m w.e. a}^{-1})$ and the simulated mass balance $\hat{B}_o$ over the same period. In the Hotan and Yarkant catchments, representative mass balance records are absent, but some regional indications and recent findings exist (they are discussed in section 4). We therefore calibrated these catchments for three initial mass balance scenarios of $+0.1$, $0$, $-0.1 \text{ m w.e. a}^{-1}$, with the simulated mass balances discussed in the results. These values are guided by the best-guess values mentioned above. The $0.1 \text{ m w.e. a}^{-1}$ max/min values represent about 30% of the expected catchment precipitation and nearly twice as much as is recorded in the APHRODITE data. A mass balance outside of that range would mean a strong glacier imbalance that has not been observed elsewhere in High Asia during the 1960s and 1970s and is thus considered unlikely.

<table>
<thead>
<tr>
<th>Parameter/description</th>
<th>Range</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$ Maximum correction factor</td>
<td>1–6$^a$</td>
<td>1–15$^b$</td>
</tr>
<tr>
<td>$a$ Maximum precipitation gradient</td>
<td>0.1–0.9</td>
<td>% m$^{-1}$</td>
</tr>
<tr>
<td>$m$ Maximum precipitation altitude</td>
<td>4000–7000</td>
<td>m MSL</td>
</tr>
<tr>
<td>$\delta_s$ Snow degree-day factor</td>
<td>2–7$^a$</td>
<td>mm °C$^{-1}$ day$^{-1}$</td>
</tr>
<tr>
<td>$T_m$ Snowmelt threshold temperature</td>
<td>0 ± 3</td>
<td>°C</td>
</tr>
<tr>
<td>$t_e$ Temperature lapse rate</td>
<td>$-0.80 \text{ to } -0.55$</td>
<td>°C (100 m)$^{-1}$</td>
</tr>
<tr>
<td>$\delta_g$ Ice degree-day factor</td>
<td>6–14$^a$</td>
<td>mm °C$^{-1}$ day$^{-1}$</td>
</tr>
<tr>
<td>$\Gamma$ Proportion of energy consumed by sublimation</td>
<td>0.2–0.6$^a$</td>
<td></td>
</tr>
<tr>
<td>$b_r$ Residual mass balance during initialization</td>
<td>0 ± 0.34$^a$</td>
<td>m w.e. a$^{-1}$</td>
</tr>
<tr>
<td>$E_c$ Potential evaporation correction</td>
<td>0.7–1.3</td>
<td></td>
</tr>
<tr>
<td>$R_{2,4}$ Routing coefficients</td>
<td>1–20</td>
<td></td>
</tr>
<tr>
<td>$S_c$ Saturated conductivity correction</td>
<td>0.1–2</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Aksu catchments
$^b$ Hotan and Yarkant catchments
In addition to the simulated glacier dynamics, the catchment hydrology was calibrated to the available discharge records from the five outlet gauge stations (Table 1). The hydrological performance was measured by one objective function for the daily discharge and a second objective function for annual discharge. The first objective function applied to daily discharge is a combination of the standard and logarithmic Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe 1970). Because of the strong seasonality in the catchments, the inclusion of the logarithmic NSE prevents the strong summer peaks from receiving the greatest weight in the common NSE measure, as was observed by Duethmann et al. (2014, 2015). The objective function is given by the following equation:

$$\text{NSE}_c = 1 - \frac{1}{2} \left[ \frac{\sum (Q_o - Q_s)^2}{\sum (Q_o - \bar{Q}_c)^2} - \frac{\sum (\log Q_o - \log \bar{Q}_c)^2}{\sum (\log Q_o - \log \bar{Q}_c)^2} \right],$$

(4)

with observed $Q_o$ and simulated $Q_s$, daily mean discharge and the observed long-term average $\bar{Q}_c$. For the model validation, this objective function was also evaluated for the validation years (odd-numbered periods).

A second objective function using the annual discharge was used to also optimize the model to the annual water balance and the interannual variability. This was done through a simple root-mean-square function as described by

$$\text{RMSE}_{QA} = \frac{\sqrt{(QA_o - QA_s)^2}}{QA_o} \times 100,$$

(5)

with squared residuals between observed $QA_o$ and simulated $QA_s$, annual mean discharge weighted by the observed long-term average $QA_o$ expressed in percent.

We employed the S-metric selection evolutionary multiobjective optimization algorithm (SMS-EMOA) to optimize the four objective functions using the 12 calibration parameters listed in Table 3. The algorithm successively finds Pareto optimal solutions out of a parameter set population of 100 “individuals” by optimizing the hypervolume enclosing the space of dominated (not Pareto optimal) solutions (Beume et al. 2007; Emmerich et al. 2005). Its advantage over other commonly used multiobjective optimization algorithms [e.g., Nondominated Sorting Genetic Algorithm (NSGA-II)] is the preference of solutions well-spaced out on the Pareto front, rather than clustering many solutions over smaller parts of the front. It has been used successfully in hydrological modeling studies with similar multicriteria problems (Ficklin et al. 2014; Stagge and Moglen 2014).

The model was optimized for each of the five catchments individually and run for a population of 100 in parallel on a large computer cluster over 100 generations. Satisfactory convergence results were already attained after 50–70 generations, that is, 5000–7000 model runs. The remaining “Pareto optimal” solutions were inspected and the following minimum criteria were defined: 1) an NSE$_c$ of at least 0.6, 2) an RMSE$_{QA}$ less than 25%, 3) a $\chi^2$ less than 20%, and 4) an MB$_c$ error probability of lower than 0.05 (only in the Aksu catchments). These criteria were used to filter out solutions at the edges of the Pareto front. The remaining count out of 100 will be reported in the results along with median values over all remaining ones.

3. Results

a. Comparison of precipitation datasets

Mean annual precipitation varies widely between the selected datasets in both space and time. The discrepancies are evident in catchment mean values (Table 4) and mean annual and monthly values (Fig. 3, Fig. S2), as well as horizontal (Fig. 4) and vertical distributions of precipitation (Fig. 5). Some datasets refer to different periods than 1971–2000 (our reference period); these periods are clearly indicated in the figures, where they are shown together. Although care must be taken when comparing values of different periods, the annual variability is below the differences between most of the precipitation datasets (Fig. 3). As described in section 2b, we are referring to the three categories of datasets as direct, satellite-based, and model-based precipitation datasets (DPD, SPD, MPD).

In the Aksu catchments (Xiehela and Shaliguilanke), the DPD show mean annual values of 230–320 mm a$^{-1}$, while ERA-Interim and CCLM are higher at 482–838 mm a$^{-1}$ over the reference period (Table 4). These differences are also reflected in much higher summer precipitation for ERA-Interim and CCLM compared to the DPD and higher precipitation at higher altitudes for HAR, ERA-Interim, and CCLM (Fig. 5, Fig. S2). Although slightly lower, the DPD are in line with observations from the single high-altitude Tian Shan meteorological station ($=3600$ m MSL), ranging between 200 and 400 mm a$^{-1}$. However, the altitude dependence is only vaguely visible, with most of the Tian Shan showing values between 200 and 300 mm a$^{-1}$. At similar spatial resolutions, ERA-Interim, HAR, and CCLM show much higher values, especially at the north eastern fringe of the Xiehela catchment.
In the Hotan and Yarkant catchments, catchment mean values of the DPD are in the range of 44–113 mm a\(^{-1}\), values similar to the Taklamakan Desert. The region’s typical summer precipitation peak is not evident in the Wuluwati and Kaqun catchments, while there is virtually no precipitation recorded between October and March in the Tongguiziluoke catchment (see Fig. S2). This leads us to conclude that the DPD are strongly biased by the observations in the desert with annual mean values below 100 mm a\(^{-1}\) in the Hotan and around 100–150 mm a\(^{-1}\) in the Yarkant catchment. The SPD and MPD, on the other hand, have mean values of 147–609 mm a\(^{-1}\) and show a greater correlation with the major mountain ranges in the region. The high-resolution GPM and HAR datasets show the heterogeneities in precipitation well. For example, in the HAR dataset, the strong differences between the north of the Yarkant catchment and the much wetter Karakoram in the south are striking (Fig. 4). This elevational dependence in precipitation over the catchments is barely noticeable in the DPD (Fig. 5).

A first-order indication of precipitation underestimation is also provided by runoff coefficients, the ratio of catchment discharge to precipitation. For four of the five catchments, runoff coefficients are greater than 1 with precipitation from the DPD (Table 5). The Hotan and Yarkant catchments have exceptionally high coefficients above 1 and up to 3.9. Runoff ratios for the ERA-Interim and CCLM precipitation, on the other hand, are much lower at 0.19–0.60. Coefficients roughly vary with glacier cover and aridity, with the highly glacierized catchments Xiehela and Tongguiziluoke showing significantly higher values than their neighboring less glacierized catchments, Shaliguilanke and Wuluwati.

The extensive glacier cover of the catchments has the potential to alter runoff without a change in precipitation due to its dependence on temperature. Negative mass balances drive the ratio up and vice versa, but only within the limits of reasonable mass balance assumptions. Simultaneously, mountainous catchments generally exhibit high runoff coefficients, as the existence of permafrost, shallow soils, and steep slopes favor fast surface runoff and low temperatures coupled with sparse vegetation keep evapotranspiration low (Goel 2011). However, in the Tarim headwaters this is counterbalanced by an arid to hyperarid climate. Numerous high-altitude saltwater (endorheic) lakes are close to the catchments on the Tibetan Plateau and in the western central Tien Shan (e.g., Aksai Chin Lake, Chatyr Kol), providing evidence for evaporation-driven water cycles. These opposing conditions and uncertain mass balances in the Yarkant and Hotan catchments for this period make it difficult to estimate reasonable runoff coefficients. But even with—

<table>
<thead>
<tr>
<th>Catchment</th>
<th>Aksu Xiehela</th>
<th>Shaliguilanke</th>
<th>Hotan Wuluwati</th>
<th>Tongguiziluoke</th>
<th>Yarkant Kaqun</th>
</tr>
</thead>
<tbody>
<tr>
<td>APHRODITE</td>
<td>314 ± 63</td>
<td>230 ± 20</td>
<td>98 ± 28</td>
<td>62 ± 11</td>
<td>113 ± 24</td>
</tr>
<tr>
<td>GPCC</td>
<td>320 ± 90</td>
<td>238 ± 25</td>
<td>63 ± 27</td>
<td>44 ± 3</td>
<td>111 ± 39</td>
</tr>
<tr>
<td>TMPA</td>
<td>390 ± 83</td>
<td>300 ± 26</td>
<td>147 ± 38</td>
<td>148 ± 44</td>
<td>211 ± 54</td>
</tr>
<tr>
<td>ERA-Interim</td>
<td>838 ± 123</td>
<td>486 ± 80</td>
<td>566 ± 137</td>
<td>572 ± 138</td>
<td>566 ± 195</td>
</tr>
<tr>
<td>CCLM</td>
<td>617 ± 150</td>
<td>482 ± 160</td>
<td>548 ± 167</td>
<td>567 ± 160</td>
<td>609 ± 134</td>
</tr>
</tbody>
</table>

In summary, the comparison of the six precipitation datasets in the Tarim headwater catchments revealed the following:

1) The SPD and MPD give significantly higher values than the DPD (catchment means 3–4 times higher, locally up to 12 times).
2) The mountain ranges are not well reproduced by the DPD, especially where the closest meteorological stations are at the desert fringes.
3) The DPD do not capture the heterogeneity and elevation dependence of precipitation in the catchment as it is, for example, measured by the satellite-based datasets or modeled by the MPD.
4) All datasets vary significantly from each other, especially in the elevation distribution, with the exception of the strong similarity between the DPD.
model-based datasets. But using the findings of the above comparison allowed us to constrain the correction with the help of the glacio-hydrological model SWIM-G.

b. Model calibration

The automatic calibration of the SWIM-G yielded good results, but performances vary with meteorological station density (Table 6). The population size of 100 parameter sets was subset using acceptable threshold values (see section 2e and top of Table 6) for the four objective functions (NSE, RMSE of annual discharge, $\chi^2$ of glacier hypsometry, and MB). The subsets include all of the runs in the Xiehela (Aksu) catchment, 49 runs for the Shaligualanke (Aksu catchment), 51–83 in the Hotan catchments, and 84–97 in the Yarkant catchment. The results for the Hotan and Yarkant catchments are given with uncertainty ranges over the three mass balance assumptions in the initialization period. Calibrated parameter values are listed in Tables S2–S4 in the supplemental material.

The glaciological calibration is mainly characterized by the match between the observed and simulated (initialized) ice hypsometries (Fig. 6), evaluated by the objective function $\chi^2$. The overall hypsometry is generally in good agreement with the observed one, with median $\chi^2$ deviations of 4.0%–14.7%. Departures from the observations mainly occur at overestimated hypsometry peaks, as a result of the $\chi^2$ objective function that weighs errors relative to the observed magnitude, compromising overestimations in the peaks against better fits at the lower ends. Differences between the initial mass balance assumptions for the Hotan and Yarkant catchments are small (0.6%–1.3% in the median $\chi^2$), giving evidence that the precipitation correction has compensated for the loss/gain in ice for positive/negative mass balances in the initialization period.

The hydrological calibration, gauged by the combined NSE and the RMSE of annual mean discharge, yielded good results, but differences with meteorological station
Fig. 4. Mean annual precipitation for the selected datasets over the periods stated in the title. The APHRODITE, GPCC, and CCLM datasets all refer to the reference period 1971–2000. All datasets are shown at their original resolution but in UTM projection. Elevation from the SRTMv3 DEM for reference; refer to Fig. 1 for elevations and scales. The horizontal line in the HAR dataset indicates the boundary of the 10-km-resolution domain; the northern part is the 30-km-resolution domain. Note the nonlinear color scaling.
densities are apparent (Fig. 7). Whereas the two Aksu catchments and the Yarkant catchment have NSE\textsubscript{c} values of above 0.82 and an annual discharge error of less than 16%, these objective functions decline to around 0.74% and 21% in the Hotan catchments. The validation NSE\textsubscript{c} value is 0.035–0.045 lower than the calibration value. The day-to-day variability as well as the day of year mean seasonality are reproduced well, but the erratic discharge peaks in summer are often missed by the simulations, especially in the Hotan catchments. The range in discharge over the calibration ensemble is relatively small but is usually greatest in autumn, with some exceptions in the Yarkant catchment. While catchment discharge was used for the automatic calibration, it is only one component of the catchment hydrology. Results of the other components (snowmelt, evapotranspiration, etc.) will be presented in the next section and should also be considered in the model calibration and its hydrological plausibility. Most discharge

![Fig. 5. Mean annual precipitation (1971–2000, unless otherwise stated) hypsometries over the five headwater catchments. The 50-m elevation intervals are sampled from the original values at a horizontal 90-m resolution. The shaded area indicates the catchment hypsometry.](image)

### TABLE 5. Runoff coefficients for the datasets with overlap with discharge observations (1971–87).

<table>
<thead>
<tr>
<th></th>
<th>Aksu</th>
<th>Hotan</th>
<th>Yarkant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Xiehela</td>
<td>Shaligulanke</td>
<td>Wuluwati</td>
</tr>
<tr>
<td>APHRODITE</td>
<td>1.10</td>
<td>0.62</td>
<td>1.26</td>
</tr>
<tr>
<td>GPCC</td>
<td>1.08</td>
<td>0.61</td>
<td>1.99</td>
</tr>
<tr>
<td>ERA-Interim</td>
<td>0.43</td>
<td>0.30</td>
<td>0.19</td>
</tr>
<tr>
<td>CCLM</td>
<td>0.60</td>
<td>0.31</td>
<td>0.22</td>
</tr>
</tbody>
</table>
is generated by surface runoff (in the Aksu catchments) or interflow (Hotan and Yarkant catchments), while groundwater contributes about a third of total annual runoff (Fig. S3).

The simulated glacier mass balance in the two Aksu catchments was accurately calibrated to the catchment-wide values found by Pieczonka and Bolch (2015) for the period 1975–2000 (Fig. 8, top). While the simulated annual variability is close to the in situ mass balance record at Karabatkak Glacier (scaled in Fig. 8, top), it is more positive in the early 1980s and slightly more negative in the late 1990s. The calibration ensemble range ($-0.35 \pm 0.06 \text{ m w.e. a}^{-1}$) is lower than the accumulated uncertainty found by Pieczonka and Bolch (2015) ($-0.35 \pm 0.34 \text{ m w.e. a}^{-1}$). Mass balance is mainly calibrated via snow and ice degree-day melt rates as well as the residual mass balance during initialization. The calibrated residual mass balance median (± standard deviation) is $-0.19 \pm 0.02 \text{ m w.e. a}^{-1}$.

### Table 6. Model performance according to the median values of the four objective functions over the calibration ensemble (NSE$_c$, RMSE$_{Q_A}$, $\chi^2$, and MB$_c$). By applying the thresholds shown in the table header, the 100 Pareto-optimal calibrations are reduced to number $N$. The NSE$_c$ is given for the calibration (even years) and the validation years (odd years). In the Hotan and Yarkant catchments, mean values are provided for three mass balance assumptions ($b_r = +0.1, 0, -0.1 \text{ m w.e. a}^{-1}$), instead of MB$_c$.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>$N$</th>
<th>Calibration</th>
<th>Validation</th>
<th>RMSE$_{Q_A}$ (%)</th>
<th>$\chi^2$ (%)</th>
<th>MB$_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aksu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xiehela</td>
<td>100</td>
<td>0.936</td>
<td>0.889</td>
<td>13.4</td>
<td>8.4</td>
<td>0.001</td>
</tr>
<tr>
<td>Shaliguilanke</td>
<td>49</td>
<td>0.827</td>
<td>0.787</td>
<td>16.3</td>
<td>14.7</td>
<td>0.023</td>
</tr>
<tr>
<td>Hotan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wuluwati</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+0.1</td>
<td>51</td>
<td>0.739</td>
<td>0.701</td>
<td>20.8</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>73</td>
<td>0.739</td>
<td>0.700</td>
<td>20.9</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>-0.1</td>
<td>67</td>
<td>0.734</td>
<td>0.696</td>
<td>20.1</td>
<td>11.8</td>
<td></td>
</tr>
<tr>
<td>Tongguziluoke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+0.1</td>
<td>75</td>
<td>0.704</td>
<td>0.669</td>
<td>23.0</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>83</td>
<td>0.742</td>
<td>0.706</td>
<td>21.9</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>-0.1</td>
<td>83</td>
<td>0.765</td>
<td>0.728</td>
<td>21.1</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>Yarkant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaqun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+0.1</td>
<td>93</td>
<td>0.827</td>
<td>0.785</td>
<td>16.9</td>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>84</td>
<td>0.827</td>
<td>0.786</td>
<td>15.8</td>
<td>10.2</td>
<td></td>
</tr>
<tr>
<td>-0.1</td>
<td>97</td>
<td>0.856</td>
<td>0.813</td>
<td>16.6</td>
<td>11.1</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 6.** Initialized and observed glacier hypsometry of the five headwater catchments. Hypsometry intervals are 50 m. Note the variable horizontal scales.
for the Xiehela catchment and $-0.16 \pm 0.04$ m w.e. a$^{-1}$ for the Shaligualanke catchment, which is within the values found in a modeling study for the Tien Shan for the same period (Farinotti et al. 2015).

The simulated mass balances in the Hotan and Yarkant catchments are stable or increasing under all three initial mass balance assumptions (Fig. 8, bottom two panels). Increasing with the residual mass balance assumption,
median values range between 0.03 and 0.20 m w.e. a\(^{-1}\) in the Hotan and between 0.05 and 0.27 m w.e. a\(^{-1}\) in the Yarkant catchment, with uncertainty ranges (half-range over all runs) of 0.03–0.07 m w.e. a\(^{-1}\). The annual variability is similar in those catchments; it is stable to slightly negative between 1975 and 1985 and increasing thereafter. These simulations are in line with several, mostly local, glaciological studies from the Kunlun Shan and the Karakoram, as is discussed in more detail in section 4. The increasing trend, even with a negative initialization mass balance, is driven by increases in precipitation in comparison to the initialization period (1961–75), while temperature changes are moderate.

c. Corrected precipitation

Corrected catchment mean precipitation as well as other water balance components are shown in Fig. 9 and the distribution of mean annual precipitation over the five catchments in Fig. 10. The simulated precipitation is far higher than the APHRODITE dataset suggests: at the catchment average by a factor of about 1.5 in the Aksu catchments, 2.2 in the Wuluwati (Hotan) and the Kaqun (Yarkant) catchments, and by a factor of about 4.3 in the Tongguziluoke (Hotan) catchment (Fig. 11). In the calibration ensemble median, the highly glaciated Xiehela (Aksu) catchment receives 487 mm a\(^{-1}\) (corrected from 314 mm a\(^{-1}\)) on average over the reference period 1971–2000 and the less glacierized, lower-lying Shaliguilanke (Aksu) catchment to the southwest receives 327 mm a\(^{-1}\) (corrected from 230 mm a\(^{-1}\)). Correction factors over the large massifs, the Kokshal–Too in the west and the Inylcheck–Tomur regions, are between 2 and 5 with precipitation totals of up to 1400 mm a\(^{-1}\). The two Hotan catchments south of the Taklamakan Desert were strongly corrected due to the
absence of any meteorological station to 285 mm a\textsuperscript{−1} (from 62 mm a\textsuperscript{−1}) in the highly glacierized Tongguziluoke catchment and 230 mm a\textsuperscript{−1} (corrected from 98 mm a\textsuperscript{−1}) in the drier Wuluwati catchment with large parts on the Tibetan Plateau. The West Kunlun Shan massif in the southwest of Tongguziluoke catchment requires factors of 5–10 to sustain its vast glacier cover. The precipitation patterns are more heterogeneous in the much larger Yarkant catchment; the catchment-wide average was corrected to 267 mm a\textsuperscript{−1} (from 113 mm a\textsuperscript{−1}). The Karakoram in the south of the catchment, however, receives a range of 500–1200 mm a\textsuperscript{−1}, much more than the rest of the catchment, with high correction factors of 5–9.

It is apparent that regions with very large correction factors (i.e., >5) are also very far away from any meteorological station (Fig. 2), both horizontally and vertically. For example, the West Kunlun Shan massif and the northeastern Karakoram are 50–100 km away from the nearest precipitation gauge as well as several thousand meters higher. The large correction factors clearly stem from entirely different precipitation regimes and not only orographic precipitation. In the Aksu catchments, on the other hand, meteorological stations are closer but mainly located at lower altitudes, leading to lower correction factors that mainly correspond to orographic precipitation.

The simulated runoff coefficients are more reasonable than those calculated with the DPD (Fig. 9). Evaporation ranges between 100 and 200 mm a\textsuperscript{−1}, comparing well with global modeling results (Table S5). Low relative humidity, high elevations, and windy conditions favor sublimation, consuming large proportions of a glacier’s energy balance, as studies from tropical glaciers have shown (Winkler et al. 2009; Gascoin et al. 2011). In comparison to glacier melt, however, sublimation from glaciers is a relatively small factor in the catchments’ water balance because it consumes approximately 8.5 times more energy than melting (Mölg et al. 2009). The fraction of energy consumed by sublimation was calibrated to 0.2–0.3 in the Aksu catchments and 0.5–0.6 in the drier Hotan and Yarkant catchments.

We use the coefficient of variation (standard deviation as percent of mean) over the calibration ensemble to describe the spatial pattern of the uncertainty, referred to as just uncertainty hereafter (Fig. 12). Since the mass balance of the Aksu catchments is calibrated, the uncertainty is at around 10%–20% smaller than in the Hotan and Yarkant catchments, where most parts have uncertainties of about 20% and in some locations of up to 45%. The mean catchment uncertainty, however, is much smaller at 5%–7% in the Aksu catchments and 12%–15% in the Hotan and Yarkant catchments. Uncertainties are generally greater at higher altitudes, where they are mainly distributed according to the glacier area that the calibration to ice hypsometry depends on. The highly glacierized areas mostly have the highest correction factors but also lower uncertainties than other high-altitude parts of the catchment. These latter areas typically have small glaciers covering only small
parts of the terrain’s hypsometry. The parameters of the precipitation correction are free to vary more than in areas where glaciers cover larger parts of the altitude range. For example, the precipitation estimations in the central east of the Wuluwati (Hotan) and the southeast of the Yarkant catchment with uncertainty values of around 40% are based on a relatively small glacier group with predominantly smaller glaciers.

In comparison to the six selected precipitation datasets, the corrected values fall between the SPD and the MPD. In the Aksu catchments, they are closest to the TMPA and CCLM catchment averages, and in the Hotan and
Yarkant catchment they are between the TMPA and HAR values. The spatial distribution of the corrected precipitation corresponds closer to higher-resolution datasets than the original APHRODITE data does (cf. Fig. 4 and Fig. 10), that is, the HAR dataset in particular. At the catchment average, the two SPD and MPD covering the simulation period, CCLM and ERA-Interim (only from 1979), show larger values with catchment means greater by a factor of 1.2–2.1 and 1.5–2.3, respectively. Although it is covering a different period, the high-resolution HAR dataset is closer in the Hotan and Yarkant catchment with deviations of 1.2–1.5 times greater. The precipitation patterns around the large massifs are similar to the high-resolution HAR and the GPM datasets. Although they are only available for more recent periods, they provide evidence for high orographic precipitation of up to 1000–1500 mm a\(^{-1}\), for example, on the northern ridge of the Karakoram or at the crests of the Kunlun Shan, corroborating the large correction factors found with our simulations. Our corrected values, however, deviate from most other SPD and MPD in the north of the Yarkant catchment: corrected values range between 80 and 200 mm a\(^{-1}\), but most SPD and MPD indicate higher values of around 200–500 mm a\(^{-1}\). This is also reflected in relatively high uncertainties (Fig. 12) that are probably a result of the contrast between sparse and relatively small glaciers in the north and an extensive glacier cover in the south in this vast catchment.

4. Discussion

a. Uncertainties in precipitation datasets

Strong precipitation underestimations and necessary corrections in remote mountainous regions associated with strong elevation gradients have been studied before, but not in the Tarim headwaters (You et al. 2012; Henn et al. 2016). Deficiencies in various reanalysis datasets were corrected in the Syr Darya headwaters in Kyrgyzstan using a hydrological model under data-scarce conditions (Duethmann et al. 2013). Also, observations from the better-gauged Alps confirm large gradients between valley and high altitudes, providing further evidence for the necessity for strong corrections in regions with a low elevation sampling bias (Frei and Schär 1998; Sevruk 1985). As the great variability between the selected datasets and the high simulated correction factors in our results show, the Tarim headwaters are another example of such deficiencies.

The comparison of the six precipitation datasets has exposed considerable disagreement. The DPDs seem to be ineffective in representing orographic precipitation, although some effort is made to give greater weight to observations on windward facing and higher-elevation slopes [e.g., Yatagai et al. (2012) use the Mountain Mapper approach]. However, they do not extrapolate into ungauged elevation ranges, leading to a low elevation–low precipitation bias. But station distribution may not be the only reason for bias, as
measurement errors such as undercatch lead to further underestimations (Kochendorfer et al. 2017; Yang and Ohata 2001). Strong winds and large shares of solid precipitation, as found in the Tarim headwaters, make the observations especially susceptible. A recent study of high-resolution datasets over the western United States has shown biases of 5%–60% in long-term mean annual precipitation in a region with a much denser observation network compared to the Tarim headwaters (Henn et al. 2018).

The precipitation correction presented here is able to overcome these biases by incorporating available hydrological and glaciological observations, bringing it closer to the satellite and model-based datasets. It is, however, lower than the coarser-resolution datasets, ERA-Interim (0.5°) and CCLM (0.44°) as well as HAR (30 km) in the Aksu catchments, agreeing with previous findings over the southeastern Tibetan Plateau (Tong et al. 2014). Runoff coefficients and the associated water balance components are more plausible with the corrected precipitation than those for the DPD (Fig. 9). For the arid yet highly mountainous catchments, the simulations result in runoff coefficients between 0.4 and 0.6. An exception is the relatively wet and highly glacierized Xiehela (Aksu) catchment, where higher runoff paired with a negative mass balance results in a median value of 0.76. As a result, glacier melt accounts for 40%–50% of river runoff in this catchment [as has been found before by Duethmann et al. (2015) and also by older Soviet literature as described in Sorg et al. (2012)]. In the other catchments it is between 25% and 35%.

Assessing the uncertainty of the simulated evapotranspiration in comparison to observations is difficult. Observations of evapotranspiration in mountainous regions are extremely scarce, and the same applies to the headwaters of the Tarim River. To our best knowledge there are no direct observations available. Remote sensing–based products are insufficient for mountainous regions as they rely on vegetation indices and are thus often not available for desert and mountainous regions. The most widely available MODIS-based product (MOD16; Mu et al. 2007) is not available for approximately 90% of our model domain. However, we compared our simulations to three global hydrological model runs (see Table S5) from an ensemble presented in Haddeland et al. (2011) and provide a parameter uncertainty range through the ensemble calibration approach. Our simulations compare well with simulations of the Global Water Availability Assessment (GWAVA) and the Max Planck Institute Hydrological Model (MPI-HM) when they are driven by reference runs of a global climate model (ECHAM5). The values fall within ±20% of the SWIM-G simulations. The MPI-HM model was also driven by the EU Water and Global Change (WATCH) reanalysis data, which is based on the GPCC precipitation data also analyzed here. Although slightly lower, the results are comparable in the Aksu catchments (~7%) and are much lower in the Hotan and Yarkant catchments (up to ~75%). This is in-line with the extremely low precipitation found in these catchments for the GPCC (cf. Table 4) that clearly constrain the evapotranspiration in this model run. The SWIM-G internal parameter uncertainty is relatively small (below ±5%). Considering the precipitation correction of 1.5–4.3 times the observational dataset, these uncertainty ranges are acceptable.

To reduce uncertainties further and corroborate existing estimates, the existing meteorological stations network at higher elevations in the Tarim headwaters as well as all of High Asia needs to be extended (Bolch et al. 2012; Mountain Research Initiative EDW Working Group 2015; Immerzeel et al. 2014). Better incorporation of satellite observations of precipitation is also promising, but suitable datasets have only been developed recently and do not extend as far into the past. For example, the high-resolution precipitation dataset from the GPM network, which we demonstrated the potential of in our comparison with other datasets, has only been operational since 2014 (Hou et al. 2014), but promises to be useful especially for the spatial distribution of precipitation (and probably less for absolute magnitudes). Synthesizing different types of observations has already shown to be useful, for example, using gauge, radar, and satellite data (Nelson et al. 2010). Similarly, information about glaciers combined with hydrological records should be considered in mountainous regions.

b. Precipitation correction by means of glacio-hydrological models

Our results compare well with similar studies. For example, Immerzeel et al. (2015) find similar discrepancies with existing datasets, such as APHRODITE or TRMM-based products in the neighboring upper Indus basin, southwest of the Yarkant and Hotan catchments. Similar to our results, they obtain corrected precipitation of around 800–1200 mm a⁻¹ along the Karakoram range, with similar correction factors of up to 10 for the APHRODITE dataset. They also find much drier conditions at the southeastern edge of the Yarkant catchment and south of the Wuluwati (Hotan) catchment with similar precipitation totals of 200–500 mm a⁻¹. Parameter uncertainties also compare well around the Karakoram between 10% and 50% (35% mean).

Similarly high correction factors were found by Sakai et al. (2015), who use a glacier mass balance model of
High Asian glaciers and correct the APHRODITE by tuning the model to independently estimated equilibrium line altitudes. But the relatively coarse model resolution (0.5°) fails to represent the heterogeneity of the terrain and is based on the assumption that glacier median elevation represents the equilibrium line altitude.

Although our and the two above approaches differ in time covered, model, and calibration strategy, the results provide new insights into poorly gauged and investigated catchments and reduce the uncertainty for future modeling studies. They highlight the importance of in situ and geodetic glaciological measurements of both glacier outlines and mass balance. While not immediately obvious, they are vital to constrain models that provide a greater understanding and management tools for downstream communities. Our approach has particular merit in other catchments with only discharge observations and preferably geodetic mass balance estimates available.

Some uncertainties remain in the calibration strategy and model structure, but we attempt to quantify them through a calibration ensemble (Fig. 12). We addressed parameter conflicts due to better correlation of discharge with snow/ice melt than precipitation by using agreement in annual discharge as one objective function. Incorporating snow cover maps in the automatic calibration, as was done by Duethmann et al. (2014), could further reduce the parameter uncertainty, but was outside the scope of this study. Temporal changes in correction factors as well as valley and slope specific precipitation patterns were also not considered. This would likely require a greater coupling of a high-resolution climate model (such as the HAR dataset) with the glaciological model.

c. Mass balances in the Hotan and Yarkant catchments

Despite the uncertainty about mass balances in the Hotan and Yarkant catchments, the parameter uncertainty of our precipitation correction with three reasonable mass balance assumptions is lower than the overall parameter uncertainty. Differences over the simulation period between the three scenarios are most significant in simulated mass balance, precipitation, and snowfall. Under all scenarios glacier mass balances are positive over the simulation period, although not in every year.

Despite a general trend of glacier recession in High Asia, glaciers in the Karakoram (Yarkant) and Kunlun Shan (Hotan) have shown strong signs of opposing this trend (sometimes referred to as the Karakoram anomaly). Hewitt (2005) show many advancing and thickening glacier tongues in the Karakoram in the late 1990s, and Liu et al. (2006) find evidence of advancing glaciers in the Kunlun Shan between 1968 and 1999. Recent investigations in the Karakoram based on declassified satellite imagery (Hexagon KH-9) from the 1970s indicate stable conditions or only slight mass loss: Bolch et al. (2017) find mass balance of $-0.01 \pm 0.09$ m w.e. a$^{-1}$ over 1973–99 and $-0.08 \pm 0.21$ m w.e. a$^{-1}$ over 1999–2009 for the Hunza catchment. For the central Karakoram, Zhou et al. (2017) give a mass balance of $-0.04 \pm 0.05$ m w.e. a$^{-1}$ (1973–2000). Similar findings were made for the Mutuztagh Ata massif (eastern Pamir) in the north of the Yarkant catchment (Holzer et al. 2015). Studies of changes in the 2000s using ICESat satellite measurements show a positive trend in the Kunlun Shan and Tarim part of the Karakoram with mass balances of $0.03 \pm 0.25$ m w.e. a$^{-1}$ (2003–09; Neckel et al. 2014), $0.05 \pm 0.07$ m w.e. a$^{-1}$ (2003–08; Kääb et al. 2015), and $0.17 \pm 0.15$ m w.e. a$^{-1}$ (2000–10; Gardelle et al. 2013). Brun et al. (2017) report a change of $+0.14 \pm 0.08$ m w.e. a$^{-1}$ for Kunlun Shan (2000–16). The strengthening of the westerlies is assumed to be the cause of these stable conditions (Yao et al. 2012). In combination with the extremely high elevations, increases in precipitation lead to greater glacier accumulation even with slightly increasing temperatures (Hewitt 2005).

In light of these studies, the stable or increasing mass trends simulated here (Fig. 8) are reasonable. Although spatial and temporal domains vary, the lower initial mass balance assumption of $-0.1$ m w.e. a$^{-1}$ is the most likely scenario with $0.03 \pm 0.04$ m w.e. a$^{-1}$ and $0.05 \pm 0.06$ m w.e. a$^{-1}$ for the Hotan and Yarkant catchments, respectively, over the period 1975–2000.

5. Conclusions

We compare six available precipitation datasets for the five major, highly glacierized headwaters of the Tarim River, northwest China. The analysis of six datasets exposes large disparities between different types, with direct observation-based datasets much lower than satellite-based or climate model-based estimates. Runoff coefficients above one support the assumption that the APHRODITE precipitation is strongly underestimating catchment-wide precipitation, especially in the data-scarce catchments of the Hotan and Yarkant rivers.

We provide a precipitation correction for the five catchments by optimizing an integrated glacier dynamics and hydrological model to observed daily and annual discharge as well as glacier area and mass balances. Satisfactory model performance suggests actual precipitation must be higher by factors of 1.5–4.3 on average to support the observed glacier area and attaining
reasonable mass balances for the reference period 1971–2000. Although high uncertainty ranges are still present (especially in areas with sparse glacier cover), the calibration ensemble represents a hydrologically and glaciologically consistent correction. The results are indicative of data discrepancies across High Asia and other sparsely gauged mountainous regions.

Acknowledgments. This study was conducted within the project SuMaRiO (Sustainable Management of River Oases along the Tarim River; http://www.sumario.de/), funded by the German Federal Ministry of Education and Research (BMBF Grants 01LL0918J and 01LL0918B). T. Bolch acknowledges funding by Deutsche Forschungsgemeinschaft (DFG, Code BO3199/2-1).

REFERENCES


Winkler, M., I. Juen, T. Mölg, P. Waggon, J. Gómez, and G. Kaser, 2009: Measured and modelled sublimation on the


