Use of auxiliary data of topography, snow and ice to improve model performance in a glacier-dominated catchment in Central Asia
Hongkai Gao, Tianding Han, Youcun Liu and Qiudong Zhao

ABSTRACT
Whether coupling auxiliary information (except for conventional rainfall–runoff and temperature data) into hydrological models can improve model performance and transferability is still an open question. In this study, we chose a glacier catchment to test the effect of auxiliary information, i.e., distributed forcing input, topography, snow-ice accumulation and melting on model calibration–validation and transferability. First, we applied the point observed precipitation and temperature as forcing data, to test the model performance in calibration–validation and transferability. Second, we took spatial distribution of forcing data into account, and did the same test. Third, the aspect was involved to do an identical experiment. Finally, the snow–ice simulation was used as part of the objective function in calibration, and to conduct the same experiment. Through stepwisely accounting these three pieces of auxiliary information, we found that a model without involving forcing data distribution, local relief, or snow–ice data can also perform well in calibration, but adding forcing data distribution and topography can dramatically increase model validation and transferability. It is also remarkable that including the snow–ice simulation into objective function did not improve model performance and transferability in this study. This may be because the well-gauged hydro-meteorological data are sufficient to constrain a well-designed hydrological model.

Key words | auxiliary information, glacier hydrology, model transferability, Urumqi No.1 Glacier

INTRODUCTION
‘Auxiliary’ data in rainfall–runoff modelling are here defined as all data except conventional meteorological data (i.e., precipitation and temperature) and streamflow data which are indispensable information to force and calibrate hydrological models. The auxiliary data include, but are not limited to, observed evaporation (Baldocchi et al. 2001; Xiao et al. 2012), isotopic data to separate hydrographs (Uhlenbrook & Hoeg 2003; Weiler et al. 2003; Klaus & McDonnell 2013), saturated area fraction (Troch et al. 2001), groundwater level (Seibert 2003; Fenicia et al. 2008; Li et al. 2015a), lake water level (Duan & Bastiaanssen 2013; Lindström 2016), snow and ice depth (Pomeroy et al. 2007; Singh et al. 2008; Moore et al. 2009; Gao et al. 2012; Li et al. 2015b), and topography (Beven & Kirkby 1979; Gharari et al. 2011; Sun et al. 2017), etc. Vast amounts of in situ and remote sensing data are explosively accumulating and easier to access. Unfortunately, at least in practical implementation, it is still not uncommon that the only input for hydrological models are the conventional meteorological data as forcing, and streamflow data are utilized as an objective to calibrate and validate parameter sets with fixed model structures.

With auxiliary information, whether we can improve model performance or merely increase model redundancy is still an unaddressed question. Some hydrologists have
found that auxiliary information was useful to improve model performance, and some drew the opposite conclusion. For example, Winsemius et al. (2009) used satellite data, i.e., remote sensed evaporation, as auxiliary information to calibrate a model and gain deeper understanding of the hydrological processes in a real ungauged basin in Africa. Fenicia et al. (2008) used groundwater and water isotope auxiliary data in addition to streamflow to understand how different sources of data can motivate model development and help to understand catchment behavior. They found that both groundwater and isotope data could be used to understand threshold processes and mixing processes in the catchment, respectively. Tangdamrongsub et al. (2015) found that data assimilation of GRACE terrestrial water storage into a hydrological model had improved groundwater estimation, but not in streamflow estimation. Sometimes, due to observation artifacts, auxiliary data do not have undoubted superiority over simulated results. For example, Winsemius et al. (2006) found the discrepancy of GRACE and a hydrological model is probably caused by the data quality of GRACE; Matgen et al. (2012) found that assimilating the remote sensing soil moisture even resulted in a negative impact on discharge simulation.

In spite of the availability of auxiliary information, most studies have focused on the impact of model performance in either internal fluxes or streamflow (Hailegeorgis & Alfredsen 2015; Li et al. 2015a), and the impact of auxiliary information on model transferability is rarely tested. Therefore, hydrological model transferability is still a great challenge in the hydrology community (Hrachowitz et al. 2013; Biondi & De Luca 2017). Beyond calibration–validation for one catchment, model transferability can serve as an indicator to more rigorous testing (Refsgaard et al. 2014) of their physical realism, and whether models get the right answer for the right reasons (Kirchner 2006). Model transferability is also related to model upscaling (Hrachowitz et al. 2013; Gao et al. 2014a), which is essential in hydrological modeling and water resources management. Ignoring catchment landscape heterogeneity is one reason to ruin model transferability (Gao et al. 2014a). Theoretically, additional data are helpful for us to gain a deeper understanding of hydrological processes, to improve model performance and benefit model transferability; however, which information will practically be valuable is still an open question, requires more stringent hypothesis tests and more case studies.

The forcing data spatial heterogeneity has been well documented (Barry 1992; Willmott & Matsuura 1995; Daly et al. 2008; Yu et al. 2012). Since most meteorological stations in mountainous regions are located in valleys and impacted by local relief on precipitation and energy distribution, the observed meteorological data may not be able to represent the spatial distribution pattern (Klemes 1990; Hrachowitz & Weiler 2011). Immerzeel et al. (2013) found that the observed precipitation severely underestimates the actual precipitation in the upper Indus basin, which was essential to estimate the water balance. Regarding its impact on model performance, most studies have shown that involving distributed forcing data has improved model performance (Boyle et al. 2001; Andréassian et al. 2004; Ajami et al. 2006) and consistency (Euser et al. 2015), with the exception of Kling & Gupta (2009), who found that spatial distributed input is of less importance than physical properties. However, how the spatial distribution of forcing data impacts on snow and ice melting simulation and model transferability is still unclear.

Concerning the impact of topography on snow and ice melt, the physical process has been intensively studied (Luce et al. 1998; Jost et al. 2007; Ménard et al. 2014), and well coupled into hydrological models (Bloschl et al. 1991; Seibert 1997; Pomeroy et al. 2007). However, few studies have investigated the impact of topography on snow/ice hydrologic model performance to simulate hydrography and transferability. Therefore, whether considering topography merely increases model complexity or truly improves model performance and its transferability still needs to be investigated.

Snow and ice melt in cold and mountainous catchments in Central Asia (Immerzeel & Bierkens 2012; Li et al. 2016) plays an essential role to support the economic sustainable development in the middle stream, and maintain the health of ecosystems in the downstream surrounded by deserts (Shi et al. 2000; Yao et al. 2007; Qin & Ding 2010; Cheng et al. 2014). For example, snow and ice melt accounts for half of surface runoff for the entire Urumqi River basin (Ma 1999). In Tarim River, a neighbor catchment of the Urumqi River basin, glacier melt accounts for over 40% of
the total surface runoff (Liu et al. 2006). The snow and ice cover are monitored both by field survey and remote sensing, and a large amount of data has been collected (Liu et al. 2014). However, how to use this information to aid hydrological modeling and whether this type of information will improve model performance are still unaddressed questions.

In this study, we tested the hypothesis that adding forcing data spatial distribution, topography, and snow and ice information will improve model performance and transferability in a glacier catchment in Central Asia. Compared with other auxiliary information, i.e., groundwater storage and fluctuation, saturated area fraction, or isotopic data, topographic information and snow and ice data are more easily observable, more reliable and with less uncertainty. Therefore, it is worthwhile to test the benefit of this type of auxiliary information to improve model performance and transferability. Particularly, we selected a well-gauged catchment – the Urumqi Glacier No.1 catchment – as a case study to conduct the research. A stepwise modeling framework was implemented. First, we used the meteorological station observed precipitation and temperature as forcing data, to test the results of model simulation and transferability. Second, we took forcing data spatial distribution into account, and did the same test. Third, the local relief, i.e., aspect, was considered. Finally, the simulation of snow and ice was incorporated as part of the objective function to do calibration and then transferability tests.

**STUDY SITE AND DATA**

**Study site**

The Urumqi No.1 Glacier catchment is located in northwest China, Central Asia. It is the headwater of Urumqi River which sustains five million residents in the downstream. The elevation ranges from 3,740 to 4,490 m a.s.l. The Glacier No.1 runoff gauge station (No.1) controls an area of 3.34 km², with 55% covered by ice. Another gauge station in the downstream, Zong Kong (ZK), controls an area of 28.9 km², with 21% covered by glaciers. The non-glacierized areas are mainly bare soil/rock with sparse grass (Li et al. 2010), with shallow root zone storage capacity (Gao et al. 2014b).

**Datasets**

The No.1 Glacier has the longest glaciology measurement record in China. The observation program started in 1959 (Xie & Ge 1965) and has continued up to the present day. Field observations include yearly glacier accumulation and ablation, daily meteorological and hydrological data collection. Streamflow is observed at two runoff gauging stations, the No.1 and ZK. Daily streamflow is available during the main snow/glacier melting season (June–August) from 1985 to 2006 (Table 1). Daily meteorological data are available from the Da Xi Gou (DXG) meteorological station located at 3,539 m a.s.l., about 3 km downstream of the glacier (Figure 1) for the period 1958–2006. Between March 1987 and February 1988, Yang et al. (1992) conducted an intensive snow survey close to the DXG meteorological station. Daily snow depth and snow density were measured, from which the daily snow water equivalent (SWE) was derived.

The variations of glacier mass balance (GMB) and equilibrium line altitude (ELA) sensitively indicates and quantifies the glacier change with climate change (Cuffey & Paterson 2010). Both the GMB and ELA were observed by stake method, with a permanent stake network, properly distributed across different elevation zones (about 45–80 stakes in 8–9 rows) and additional snow pits (Ye et al. 2005). From the monthly change of stakes’ height above the ice surface in hydrological years (from the beginning

<table>
<thead>
<tr>
<th>Data</th>
<th>Period of observation</th>
<th>Time step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air temperature</td>
<td>1959–2006</td>
<td>Daily</td>
</tr>
<tr>
<td>Precipitation</td>
<td>1959–2006</td>
<td>Daily</td>
</tr>
<tr>
<td>SWE</td>
<td>03/1987–02/1988</td>
<td>Daily</td>
</tr>
</tbody>
</table>
of October to the end of the next September), we can calculate the ice mass balance of each observation point in that year. Based on point measurement, the GMB of the entire glacier can be calculated by the contour map (Elder et al. 1992). The annual ELA is the altitude at where the ice accumulation and ablation are equal over a hydrological year, in other words, it is where the mass balance was zero of that year (Dong et al. 2012). The annual GMB and ELA are available from 1959 to 1966, and from 1980 to 2006. From 1967 to 1979, the observation was stopped, and the GMB and ELA data for the (1967–1979) period were reconstructed based on the relationship between air temperature and observed GMB (Zhang 1981).

Topography discretization

Topography influences forcing data spatial distribution and the energy budget (e.g., Barry 1992), thus also glacier and snow distribution and melting. This is particularly true for elevation and aspect, which directly impact solar radiation allocation as the first order control on snow/ice melt (Hock 2005). To account for these influences, balancing accuracy with computational cost, the catchments in this study were discretized into 16 elevation zones in the Glacier No.1 and 21 in the ZK catchment, with 50 m intervals to do elevation classification. Subsequently, each elevation zone was further divided into three aspect zones, including the north (315–45°), south (135–225°), and the east/west (45–135° and 225–315°) facing aspects. In summary, considering elevations, aspects as well as glaciered and non-glaciered areas, the Glacier No.1 catchment was classified into 96 classes, and ZK was classified into 126 classes.

Forcing data and their interpolation

The long-term mean annual temperature is −5.1 °C, with −20 °C in winter and about 0 °C from June to August. The annual precipitation is around 450 mm a⁻¹, and snow is the main phase of precipitation in the glacier area and the non-glacier area in non-summer seasons. Over 90% of precipitation occurs between April and September. Potential evaporation was calculated by the Hamon equation (Hamon 1961) and reaches a long-term average of about 200 mm a⁻¹. In the Hamon equation, only temperature is required as input, with no any free parameters to be calibrated. Oudin et al. (2005) found that the performance of rainfall–runoff simulation is not very sensitive with different approaches to estimate potential evaporation, therefore we chose this parsimonious method in this study.

The DXG meteorological station is located in low elevated valleys to allow easier access for maintenance, which typically reduces the representativeness of the observed variables. To offset these biases, temperature in the individual elevation zones was corrected with a lapse rate of −0.007°C m⁻¹ (Li et al. 2013), while precipitation was adjusted with a lapse rate of 0.05% m⁻¹ (Yang et al. 1988).

METHODS

Model

Snow model

Separate snowfall and rainfall. Precipitation is simulated to be either snow (Pₛ) or rain (Pᵢ) depending on whether the daily average air temperature (T) is above or below a threshold temperature, Tₛ [°C] (Equations (1) and (2)) (Han et al. 2010). It is worthwhile to note that with more detailed
auxiliary information, a dynamic scheme can be a competent alternative to estimate the snowfall (Ding et al. 2014).

\[
P_s = \begin{cases} 
  P; & T \leq T_t \\
  0; & T > T_t 
\end{cases}
\]  
\[
P_l = \begin{cases} 
  P; & T > T_t \\
  0; & T \leq T_t 
\end{cases}
\]  
\[
\frac{dS_w}{dt} = P_s + R_{\text{ref}} - M_s
\]  
\[
\frac{dS_w}{dt} = P_l + M_s - R_{\text{ref}} - P_e
\]  
\[
M_s = \begin{cases} 
  F_{\text{dd}}(T - T_t); & T > T_t \\
  0; & T \leq T_t 
\end{cases}
\]  
\[
P_e = \begin{cases} 
  \frac{dS_w}{dt} - C_{\text{wh}} \frac{dS_w}{dt}; & S_{\text{wl}} > C_{\text{wh}} S_w \\
  0; & S_{\text{wl}} \leq C_{\text{wh}} S_w 
\end{cases}
\]  
\[
R_{\text{ref}} = \begin{cases} 
  F_{\text{dd}} F_{\text{rr}}(T_t - T); & T_t > T \\
  0; & T_t \leq T 
\end{cases}
\]

\textbf{Snowfall correction.} Caused by systematic errors in measurement, such as wind wetting and evaporative losses, snowfall is always being underestimated (Goodison et al. 1998; Yang et al. 2001). According to field observation in this study site, Yang et al. (1988) concluded that only 76.5% snowfall is captured by observation in this study site. Therefore, the amount of observed snowfall should be multiplied by 1.3 to correct the biased observation.

\textbf{Snowmelt simulation.} The snow pack was regarded as porous media which can hold the liquid melting/rainfall water and the liquid water could be refrozen into the snow pack. Therefore, the solid snow pack (S\textsubscript{w}) and the liquid water inside the snow pack (S\textsubscript{wl}) were conceptualized as two separate reservoirs. The water balance of the S\textsubscript{w} reservoir is shown in Equation (3), where R\textsubscript{ref} (mm d\textsuperscript{-1}) is the refreezing water from S\textsubscript{wl} to S\textsubscript{w}. M\textsubscript{s} (mm d\textsuperscript{-1}) indicates the melted snow. Equation (4) shows the water balance of S\textsubscript{wl} reservoir, where the P\textsubscript{e} (mm d\textsuperscript{-1}) means the effective precipitation from snow pack to soil and the SWE is the sum of solid and liquid water of snow pack. Snowmelt (M\textsubscript{s}) is calculated with the widely used temperature-index approach (Equation (5)) (Braithwaite & Olesen 1989; Hock 2003), which uses a degree-day factor F\textsubscript{dd} (mm (°C d\textsuperscript{-1}) to calculate melt water by the temperature above the threshold temperature T\textsubscript{t} (°C). The liquid water in the S\textsubscript{wl} from meltwater and rainfall is retained within the snowpack until it exceeds a certain fraction, C\textsubscript{wh} (°), of the solid SWE (S\textsubscript{w}) (Equation (6)) (Seibert 1997). Liquid water within the snowpack refreezes according to Equation (7). F\textsubscript{rr} (-) is the correct factor to simulate liquid water refreezing, while temperature is below T\textsubscript{t} (Seibert 1997).

\[
\frac{dS_w}{dt} = P_s + R_{\text{ref}} - M_s
\]  
\[
\frac{dS_s}{dt} = P_l + M_s - R_{\text{ref}} - P_e
\]  
\[
M_s = \begin{cases} 
  F_{\text{dd}}(T - T_t); & T > T_t \\
  0; & T \leq T_t 
\end{cases}
\]  
\[
P_e = \begin{cases} 
  \frac{dS_w}{dt} - C_{\text{wh}} \frac{dS_w}{dt}; & S_{\text{wl}} > C_{\text{wh}} S_w \\
  0; & S_{\text{wl}} \leq C_{\text{wh}} S_w 
\end{cases}
\]  
\[
R_{\text{ref}} = \begin{cases} 
  F_{\text{dd}} F_{\text{rr}}(T_t - T); & T_t > T \\
  0; & T_t \leq T 
\end{cases}
\]

\textbf{Model for non-glacier area}

\textbf{Unsaturated reservoir.} The water balance of the unsaturated reservoir (S\textsubscript{u}) is

\[
\frac{dS_u}{dt} = P_e - E_a - R_u
\]  

where P\textsubscript{e} (mm d\textsuperscript{-1}) is the effective rainfall to soil; E\textsubscript{a} (mm d\textsuperscript{-1}) is the actual evaporation, which was assumed to equal to potential evaporation, since energy is not the constraint factor for evaporation in this region (Kang et al. 2002); R\textsubscript{u} (mm d\textsuperscript{-1}) is the streamflow generated from the unsaturated reservoir (Equation (8)). Water retention curve of the Xinjiang model (Equation (9)) (Zhao 1992) was used to separate P\textsubscript{e} into retained water in S\textsubscript{u} and R\textsubscript{u}, and S\textsubscript{u,max} (mm) is the root zone storage capacity and β (-) is the shape parameter.

\[
R_u = \frac{P_e}{P_e} = 1 - \left(1 - \frac{S_u}{(1 + \beta)S_{u,max}}\right)^\beta
\]  

\textbf{Response reservoir in non-glacier area.} A splitter D (-) was applied to divide the R\textsubscript{u} into two fluxes (R\textsubscript{t} and R\textsubscript{e}) and into two response reservoirs (S\textsubscript{t} and S\textsubscript{e}). We used two linear reservoirs (S\textsubscript{t} and S\textsubscript{e}) to represent the response process of subsurface storm flow Q\textsubscript{r} (mm d\textsuperscript{-1}) and groundwater streamflow Q\textsubscript{g} (mm d\textsuperscript{-1}).

\[
\frac{dS_t}{dt} = R_t - Q_t
\]  
\[
\frac{dS_e}{dt} = R_e - Q_g
\]
\[
\frac{dS_f}{dt} = R_f - Q_f
\]

(11)

\[
Q_f = \frac{S_f}{K_f}
\]

(12)

\[
Q_s = \frac{S_s}{K_s}
\]

(13)

where \(R_f\) (mm d\(^{-1}\)) is the recharge into fast response reservoir \((S_f)\); and \(R_s\) (mm d\(^{-1}\)) is the recharge into slow response reservoir \((S_s)\); \(K_f\) (d) is the recession parameter of \(S_f\); and \(K_s\) (d) is the recession parameter of \(S_s\).

### Glacier melting and mass balance

If the ice is covered by snow, the energy is first provided to melt snow. The ice only starts to melt without snow cover. The temperature-index method is used to simulate glacier melt \(M_g\) (mm d\(^{-1}\)) (Equation (14)).

Mainly due to the lesser albedo of ice cover (Fujita & Sakai 2014), the degree-day factor of glaciers is larger than snow degree-day factor in the same region (Braithwaite & Olesen 1993; Seibert et al. 2014). Therefore, we use a multiplier \((C_g)\) to get the glacier degree-day factor by \(F_{dd}\).

\[
M_g = \begin{cases} 
F_{dd}C_g(T - T_t); & T > T_t \text{ and } S_w = 0 \\
0; & T \leq T_t \text{ or } S_w > 0 
\end{cases}
\]

(14)

The response routine on ice is calculated by an independent linear reservoir \(S_g\) (Equations (15) and (16)), with a recession parameter \(K_g\) (d).

\[
\frac{dS_g}{dt} = P_e + M_g - Q_g
\]

(15)

\[
Q_g = \frac{S_g}{K_g}
\]

(16)

The GMB of each elevation band \((S_g)\) can be derived from precipitation \((P)\) on glaciers and simulated glacier streamflow \((Q_g)\).

\[
\frac{dS_g}{dt} = P - Q_g
\]

(17)

The sum of the \(S_g\) weighted by their area proportion is the GMB of the entire glacier. It is worthwhile to note that the calculated annual GMB is the water equivalent, which should be transformed into the ice thickness before comparing with measured GMB, divided by the ice density (0.91 g/cm\(^3\)). The ELA is the altitude where accumulation and ablation are equal at a given period.

### Snow/ice melting on different aspects

With the same air temperature, the south facing aspects get more direct solar radiation, which provides the most critical energy source for snow/ice melting (Hock 2005), and resulting in more melting water; while the north facing aspects get less direct solar radiation due to the topography shadow impact. The east/west facing aspects receive the intermediate solar radiation and then melting water with the same air temperature. The influence of aspect is taken into account by a multiplier \(C_a\) \((-\)), which is larger than 1. Specifically, the \(F_{dd}\) in south facing aspects are multiplied by \(C_a\), the north facing aspects are multiplied by \(1/C_a\), and the east/west facing aspects are kept as \(F_{dd}\).

### Model calibration and evaluation approach

#### Objective functions

The Kling-Gupta efficiency (Gupta et al. 2009) \((I_{KGE})\) was used as objective function for calibration and the criteria to evaluate model performance. The equation is:

\[
I_{KGE} = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}
\]

(18)

where \(r\) is the linear correlation coefficient between simulation and observation; \(\alpha\) \((\alpha = \sigma_m/\sigma_o)\) is a measure of relative variability in the simulated and observed values, where \(\sigma_m\) is the standard deviation of simulated streamflow, and \(\sigma_o\) is the standard deviation of observed streamflow; \(\beta\) is...
the ratio between the average value of simulated and observed data.

Model evaluation

Some parameters are obtained from observation or the literature, such as the temperature lapse rate (Li et al. 2015), the precipitation lapse rate (Yang et al. 1988), and the snowfall correction factor (Yang et al. 1988) (Table 2).

Additionally, there are 13 free parameters to be calibrated. In order to calibrate the model and analyze the model uncertainty, the generalized likelihood uncertainty estimation (GLUE) (Beven & Binley 1992) was applied. The $I_{KGE}$ is set as the objective function. The prior ranges of parameters are mostly determined by the literature, and are shown in Table 2. Monte Carlo was applied to sample 50,000 sets of parameters within prior ranges, and then the best 1% (500 parameter sets) was selected as behavioral parameter sets to do further analysis. The daily streamflow from 1985 to 1996 was used to do calibration, while the rest of daily streamflow data were severed to validate the models. All the models were warmed-up by one year spin-up period.

Table 2 | Model parameters and their prior ranges for Monte Carlo sampling in GLUE method

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Unit</th>
<th>Prior range</th>
<th>Method to estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_t$</td>
<td>Temperature lapse rate</td>
<td>Cm$^{-1}$</td>
<td>0.007</td>
<td>Li et al. (2015)</td>
</tr>
<tr>
<td>$L_p$</td>
<td>Precipitation lapse rate</td>
<td>%m$^{-1}$</td>
<td>0.05</td>
<td>Yang et al. (1988)</td>
</tr>
<tr>
<td>$C_s$</td>
<td>Snowfall correction factor</td>
<td>–</td>
<td>1.3</td>
<td>Yang et al. (1988)</td>
</tr>
<tr>
<td>$T_t$</td>
<td>Threshold temperature to split snowfall and rainfall</td>
<td>°C</td>
<td>(0, 4)</td>
<td>Han et al. (2010)</td>
</tr>
<tr>
<td>$F_{dd}$</td>
<td>Degree-day factor of snow</td>
<td>mm (C d)$^{-1}$</td>
<td>(2, 9)</td>
<td>Zhang et al. (2006); Yang et al. (2012)</td>
</tr>
<tr>
<td>$C_g$</td>
<td>Factor for ice melt</td>
<td>–</td>
<td>(1, 2)</td>
<td>Gao et al. (2012)</td>
</tr>
<tr>
<td>$C_a$</td>
<td>Factor for the influence of aspect on melt</td>
<td>–</td>
<td>(1, 2)</td>
<td>Gao et al. (2012)</td>
</tr>
<tr>
<td>$C_{wh}$</td>
<td>Snow water holding capacity</td>
<td>–</td>
<td>(0, 1)</td>
<td>Gao et al. (2012)</td>
</tr>
<tr>
<td>$F_r$</td>
<td>Refreezing factor</td>
<td>–</td>
<td>(0, 1)</td>
<td>Gao et al. (2012)</td>
</tr>
<tr>
<td>$K_{fg}$</td>
<td>Recession coefficient of glacier streamflow</td>
<td>D</td>
<td>(1, 10)</td>
<td>Gao et al. (2014a)</td>
</tr>
<tr>
<td>$S_{u,m,x}$</td>
<td>Root zone storage capacity</td>
<td>mm</td>
<td>(30, 100)</td>
<td>Gao et al. (2012, 2014b)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Shape parameter</td>
<td>–</td>
<td>(0.1, 1)</td>
<td>Gao et al. (2012)</td>
</tr>
<tr>
<td>$D$</td>
<td>The splitter</td>
<td>–</td>
<td>(0.2, 0.8)</td>
<td>Gao et al. (2014a)</td>
</tr>
<tr>
<td>$K_f$</td>
<td>Recession coefficient of fast response reservoir</td>
<td>d</td>
<td>(2, 30)</td>
<td>Gao et al. (2014a)</td>
</tr>
<tr>
<td>$K_s$</td>
<td>Recession coefficient of slow response reservoir</td>
<td>d</td>
<td>(50, 200)</td>
<td>Gao et al. (2014a)</td>
</tr>
</tbody>
</table>

Experiments

We designed four model setups to conduct four virtual experiments. Forcing data, model structure, and method to calculate the objective function were modified step by step.

Experiment 1 ($M_{nAFOH}$)

Develop a glacier hydrological model, whose detailed information was described in the section ‘Model’. The impact of elevation on forcing data distribution and the influence of aspect on melting are not taken into account in this experimental scenario. The in situ observed meteorological data were used as input. Measured hydrograph was utilized to calibrate parameters and evaluate model performance with $I_{KGE}$ of hydrograph as the objective function. We named this model setup ‘$M_{nAFOH}$’, indicating the model not accounting for aspect, forcing by point meteorological observation, and using only the hydrograph simulation as the objective function. Subsequently, test the model transferability by transferring both model and behavioral parameter sets from No.1 catchment to ZK catchment.
Experiment 2 (M_{nA}F_{D}O_{H})

Keep model structure and objective function the same as M_{nA}F_{D}O_{H} while changing the input forcing data from \textit{in situ} observed data to the spatial distributed forcing data, considering the lapse rates of precipitation and temperature. Calibrate and validate the glacier hydrological model, and then test its capability to be transferred. We named this model setup ‘M_{nA}F_{D}O_{H}’, indicating the model does not account for aspect, forced by distributed precipitation and temperature, and using hydrograph simulation as the objective function.

Experiment 3 (M_{A}F_{D}O_{H})

In this virtual experiment, the impact of aspect on snow/ice melting was taken into account with the approach described in the section ‘Snow/ice melting on different aspects’. The spatial distributed forcing data and the calibration approach are kept the same as M_{nA}F_{D}O_{H}, while the effect of different proportions of aspects at distinct elevation bands was considered in the Glacier No.1 catchment while doing calibration and validation, and the ZK catchment in the model transferability test. This experimental setup was named ‘M_{A}F_{D}O_{H}’, indicating the model accounts for aspect, forced by distributed observation data, and using hydrograph simulation as the objective function.

Experiment 4 (M_{A}F_{D}O_{HGS})

Snow and ice accumulation and ablation is an essential sub-routine in this landscape-based hydrological model. Using the model structure and distributed forcing proposed in Experiment 3, we attempted to further test if cooperating the snow and ice auxiliary information in calibration could improve model performance on reproducing hydrograph and snow/ice, and the ability to be transferred and upscaled. Technically, we utilized not only the hydrograph simulation as the objective function to evaluate the model performance, but also took the snow and ice sub-routine simulation into account, by quantifying the simulation of SWE, GMB, and ELA into the objective function. The objective function \( I_{\text{KGE, HGS}} \) was applied to evaluate model performance of hydrograph and the snow/ice sub-routine simulation, by giving different weights to hydrograph \( I_{\text{KGE, H}} \), SWE \( I_{\text{KGE, SWE}} \), GMB \( I_{\text{KGE, GMB}} \), and ELA \( I_{\text{KGE, ELA}} \) (Equation (19)). This new objective function allows us to restrict the behavioral parameters’ distribution by involving snow and ice information in calibration. Then, test whether the model transferability will be improved by this auxiliary information. This experimental setup was named ‘M_{A}F_{D}O_{HGS}’.

\[
I_{\text{KGE, HGS}} = 0.7I_{\text{KGE, H}} + 0.1I_{\text{KGE, SWE}} + 0.1I_{\text{KGE, GMB}} + 0.1I_{\text{KGE, ELA}}
\] (19)

RESULTS AND DISCUSSION

Models’ calibration and temporal validation

The summary of four model setups (M_{nA}F_{p}O_{H}, M_{nA}F_{D}O_{H}, M_{A}F_{D}O_{H}, and M_{A}F_{D}O_{HGS}) in calibration and validation are shown in Figure 2. The results show that the first three model setups (M_{nA}F_{p}O_{H}, M_{nA}F_{D}O_{H}, and M_{A}F_{D}O_{H}) can reproduce hydrographs quite well in calibration, all the median values of \( I_{\text{KGE}} \) are above 0.7, with the highest \( I_{\text{KGE}} \) values around 0.8. Compared with the benchmark model setup (M_{nA}F_{p}O_{H}), there is no obvious improvement in calibration when involving spatial distribution of forcing data (M_{nA}F_{D}O_{H}). Taking into account the aspect information has slightly improved model calibration when comparing M_{A}F_{D}O_{H} with M_{nA}F_{D}O_{H}. It is not a surprise that the M_{A}F_{D}O_{HGS} does not perform well in calibration, compared with the other three model setups, although its median value of the \( I_{\text{KGE}} \) is above 0.7 as well. Since accounting for snow and ice simulation in the objective function will filter out the parameters fitting both hydrograph and snow and ice simultaneously, it may reduce the model performance if hydrograph is the only evaluation criterion. This is in line with other research on multi-objective calibration (Fenicia et al. 2007). In summary, all models can be used to fit the observed hydrographs by calibration, even neglecting the impacts of forcing data distribution and topography.

Furthermore, Figure 2 demonstrates that the four model setups all perform satisfactorily in temporal validation, with
median values of $I_{KGE}$ above 0.7. Not surprisingly, both the boxes of the 25/75th quantiles and whiskers of the 5/95th quantiles for all model setups have wider ranges compared with the calibrated results. Additionally, there are more outliers below 5% quantiles, indicating the validation results have larger uncertainty than for calibration. While comparing the four model setups in validation, both forcing data spatial distribution and the impact of aspect on snow/ice melting have improved model validation. This means these two pieces of auxiliary information can increase model simulation consistence (Euser et al. 2015). Interestingly, the median value of $M_{AFDoHGS}$ in validation is even better than its performance in calibration. It is probably caused by the better data quality in 1995–2005 than the calibration period (1985–1994).

Figure 3 shows the comparison between observed and calibrated hydrographs of the four model setups in the Glacier No.1 catchment, in 1986. The simulated hydrographs do not exhibit distinctive differences among these four model setups in calibration. In all these four setups, no matter if point or distributed forcing input, regardless of whether taking account of the aspect, or whether involving snow and ice into the objective function, all model setups can well reproduce the hydrographs. This means all of them have the ability to fit hydrograph by calibration. Interestingly, the hydrograph components (glacier and non-glacier runoff) simulated by the four model setups are also surprisingly comparable. This illustrates that the most part of the streamflow is contributed from the glacier area, and further confirms the reliability and robustness of hydrograph components’ simulation. Noticeably, both the observed and simulated hydrographs show similar variation with temperature, but are distinct regarding the fluctuation of rainfall. This illustrates the sensitivity of glacier melt and hydrograph with temperature change in this highly glaciered catchment. Quantitatively, with one unit area, the glacier area generates four to five times more streamflow than the non-glacier area.

Parameters’ uncertainty

Figure 4 shows the dotty plot of the parameters of the four model setups and their averaged values, generated by the GLUE parameter uncertainty estimation method. It is worthwhile noting that the parameters related to snow and ice accumulation and ablation are well identified in all the four model setups, which is in line with other research (van den Broeke et al. 2010; Hegdahl et al. 2016). Particularly well identifiable are the parameter controlling rainfall/snowfall split threshold temperature ($T_t$), degree-day factor ($F_{dd}$), glacier melt multiply factor ($C_g$), hold capacity of snow pack ($C_w$), recession parameter of glacier zone ($K_{c8}$). This provides further evidence supporting the fact that the hydrological process in this catchment is mostly influenced by snow and glacier melt. Less identifiability of the correct
factor to simulate liquid water refreezing in snow cover $F_r$ may indicate the lesser importance of the refreezing process in the melting season in this study site, due to the relative thin snow cover (Qin et al. 2006) compared with a humid area (Pulliainen 2006). The aspect multiple factor ($C_a$) for $M_{AFD}O_H$ and $M_{AFD}O_{HGS}$ is also not as identifiable as other snow- and ice-related parameters, which probably indicates the aspect information, to some extent, has been implicitly considered in the glacier distribution data due to the influence of aspect on spatial distribution of glaciers (Figure 1). Simultaneously, the parameters intended to simulate the non-glacierized areas, are not well identifiable, such as $S_{u, max}$, $\beta$, $C_v$, $K_f$, $K_s$. Therefore, the signal of non-glacier hydrograph components are harder to identify from hydrograph.

In Figure 5, the cumulative distribution of behavioral parameters is illustrated. If the accumulative values are close to diagonal, this indicates the behavioral parameters are close to uniform distribution. The farther the accumulative distribution to the diagonal is, the better identifiability of the parameter is. We can find that glacier- and snow-related parameters are distributed farther to the diagonal, indicating their better identifiability, while the non-glacier-related parameters have the opposite pattern. It is worthwhile to note that the distribution of parameter $D$ in $M_{AFD}O_{HGS}$ is different from its distribution pattern in other model setups. This parameter is a splitter to separate the generated runoff into the fast and slow response reservoirs. Larger $D$ value indicates more water will go to the fast response reservoir, and less water to the slow one. Therefore, different $D$ values will impact the shape of the hydrograph. Since $M_{AFD}O_{HGS}$ involves both hydrograph and snow/ice simulation to estimate the objective function, the trade-off between hydrograph and snow/ice simulation may impact on the distribution of this parameter. This trade-off may cause the parameters controlling the shape of hydrographs to be not well represented.

Remarkably, both Figures 4 and 5 illustrate that the $M_{AFPOH}$ model setup has the most identifiable parameter sets for snow and ice accumulation and melt. However, its simulation consistence indicated by validation is not as good as other model scenarios. If we use the parameter identifiability to judge the model uncertainty, we may draw the conclusion that $M_{AFPOH}$ model setup performs better with less uncertainty compared with other model setups, which is obviously not true. This result shows the parameter distribution is not a good indicator to judge the model performance or realism. We can only safely address that if the parameters are well
identified, which indicates the importance of certain hydrological processes represented by the parameters. There is no direct linkage between model reliability, model performance, and the parameter identifiability.

Moreover, we can also find the trade-off between related parameters, such as $T_t$ and $F_{dd}$. For example, the MnAFPOH fits hydrography with larger $T_t$ than the other model setups, which indicates snowfall occurs and starts to melt with higher temperature. On the one hand, this increases the proportion of snowfall, and simultaneously decreases the positive degree-days for snow and ice melting, therefore larger degree-day factor ($F_{dd}$) is needed to compensate the change. This parameter’s trade-off phenomenon might be hidden in calibration by compensation, but could be amplified if we do model validation and parameter transfer.

**Snow and ice simulation**

The observed and simulated SWE from March 1st, 1987 to February 29th, 1988 are shown in Figure 6. Interestingly,
Figure 5 | Identifiability of model parameters toward the related objectives. The performance measures based on which the cumulative performance $C(\cdot)$ is calculated are determined from the values of $I_{KGE}$.

Figure 6 | Comparison between the observed and the modelled daily snow water equivalent (SWE) from March 1987 to February 1988.
all four model setups perform well from October 1987 to February 1988. This indicates the limited uncertainty in this period, but from March 1987 to September 1987, the four models perform quite distinctively. The simulated SWE by M_nAFPOH is quite different from observation and other simulation, with large overestimation of the snow pack. From the parameter distribution in Figures 4 and 5, we can find M_nAFPOH has larger T_t than other model setups. This means that with higher threshold temperature for snowfall, the M_nAFPOH is prone to calculate more snowfall than other scenarios with the same temperature data. This overestimation is caused by neglecting the spatial distribution of forcing data. This result also illustrates that a model can reproduce hydrograph excellently, but it does not guarantee that it will satisfy the internal fluxes inspection. Once the forcing data distribution was taken into account, M_nAFDOH improves the snow pack simulation conspicuously. Comparing with M_AFDOH and M_AFDOHGS, M_nAFDOH starts to melt almost simultaneously but with less amount, which is caused by the comparable T_t value but smaller degree-day factor. The simulated SWEs by M_AFDOH and M_AFDOHGS are quite similar. It is almost impossible to separate these two lines apart in most time series. This supports the model structure, algorithm, and forcing data of M_AFDOH, allowing the outstanding simulation of SWE even without involving snow information in calibration.

Model transferability

Figure 8 shows the performance of the four model setups in transferability test from Glacier No.1 catchment to ZK catchment. M_nAFPOH is the last option for model transferability, with the lowest median value, and the widest range of the 25/75th quantiles boxes and 5/95th quantiles whiskers which indicate the largest uncertainty. Figure 9 shows the simulated hydrographs and hydrograph components from glacier and non-glacier areas of four model setups in the ZK catchment, while transferring both the model setup and the behavioral parameters from the donor catchment (Glacier No.1). The result obtained by M_nAFPOH performs worst among these four model setups. Especially in the beginning of the melting season, M_nAFPOH underestimates the amount of streamflow, and the estimated start time to melt was later than observation. When melting starts, snow melt first begins from lower elevations of the catchment, but the lumped forcing data did not consider this heterogeneity. Melting starts only when the lumped temperature is above the threshold temperature (T_t), which is later than the real start time and does not fit the physical reality.

Impact of forcing data distribution

The improvement of the median values of I_KGE while involving forcing data distribution (M_nAFDOH) is exhibited in Figure 8. The 25/75th quantiles boxes and 5/95th quantiles whiskers also become narrower, demonstrating the declining of uncertainty. The results support the hypotheses that involving the forcing data distribution will increase the model realism to reproduce catchment hydrological processes. The simulated hydrograph (Figure 9) by M_nAFDOH is also closer to the observed one, both in the perspective of the amount of streamflow and the start time of melting. The hydrograph components in Figure 9 show that the peak flow generated from the glacier area in the M_nAFDOH is larger than the M_nAFPOH, while the streamflow from the non-glacier area in the M_nAFDOH is smaller than the M_nAFPOH. Different from M_nAFPOH, in which glacier and non-glacier areas almost simultaneously contribute to streamflow, non-glacier areas melt earlier than glacier areas in M_nAFDOH due to the lower elevation of non-glacier areas.
Impact of aspect

Model transferability is also beneficial by considering the impact of aspect on melting ($M_{AFDOH}$) (Figure 8). The simulated hydrograph by $M_{AFDOH}$ in model transfer is closer to the observed one compared with $M_{MnDOH}$ and the other two model setups ($M_{MnDOH}$ and $M_{AFDOHGS}$), with less uncertainty and higher median value. Figure 9 shows the improved hydrograph simulation by $M_{AFDOH}$ when involving aspect as auxiliary information, although not as significantly as taking account of the spatial distributed forcing data.

Given that glacier melt is the dominant hydrological process in these two catchments, it is worthwhile to analyze the glacier distribution for different aspects to understand the influence of aspect on model transferability. By map algebra, we analyzed the aspect map together with the glacier map, and found that 55% of the Glacier No.1 catchment is covered by glacier: 52% is covered by east/west facing glacier, 46% with north facing glacier, and less

Figure 7 | Upper panel shows the observed and modelled (by $M_{MnDOH}$, $M_{AFDOH}$, and $M_{AFDOHGS}$ model setups) glacier mass balance (GMB) of the Urumqi No.1 Glacier from 1959 to 2008. Lower panel shows modelled and observed equilibrium line altitude (ELA) of the Urumqi No.1 Glacier from 1958 to 2006. Dashed lines represent the GMB and ELA reconstructed by temperature data from 1967 to 1979.
than 2% covered with south facing ice. While in the ZK catchment, 21% is covered by glaciers: 43% is covered by east/west facing glacier, 55% is covered by north facing glacier, and 2% is covered by south facing ice. The results interestingly showed the clear pattern of glacier distribution. In both catchments, east/west and north facing aspects shared around 50% of the glaciers, and the area of south facing glaciers is limited. This may be caused by the fact that the location of glacier is strongly impacted by aspect, due to the spatial distribution pattern of solar radiation caused by topography, and eventually the snow and ice accumulation and melt. The transferability test illustrates that since the aspect information is implicitly contained in the glacier distribution pattern, further involving aspect could improve model transferability, but not as remarkably as considering the impact of elevation on forcing data.

Impact of snow and ice information

The first three model scenarios (M_nAFpO_H, M_nAFDOH, and M_AFDOH) only employ hydrograph to do calibration validation and transferability test. While M_AFDOSH involved snow and ice simulation as part of the objective function for calibration. The snow and ice auxiliary information, used to constrain the model parameters, includes GMB and ELA of glaciers, and SWE of snow pack. Figures 8 and 9 show that after adding the snow and ice auxiliary information in calibration, there is no improvement of model transferability. This indicates involving the auxiliary information does not guarantee the improvement of model performance in well-gauged catchment.

The reason is probably caused by the fact that the dominant hydrological processes have been fairly reflected in the observed hydrograph and the well-designed model. Moreover, the key processes of submodels have been

Figure 8 | Model transferability results of four model scenarios.

Figure 9 | Observed daily average air temperature and daily precipitation; and the comparison between observed and the modelled daily streamflow. The right figure shows the streamflow generated by four model scenarios of glacierized and non-glacierized area in the ZK catchment.
well-constrained by hydrograph while doing parameter calibration. Precisely, Figures 6 and 7 show that snow and ice accumulation and ablation processes have been well represented in the M_AFOH model even though there is no snow and ice information applied to constrain the parameter calibration. Another interpretation might be the data quality of auxiliary information. In many cases, the auxiliary information is more difficult to access than conventional data. Due to the difficulty of doing measurements, the data quality of auxiliary information is probably not as reliable as hydrograph observation in well-gauged runoff stations. Moreover, the M_AFOH model setup has taken catchment heterogeneity into account properly, including the elevation zones, aspects, and landscape classification (glacier/non-glacierized areas), which probably makes the auxiliary snow and ice data redundant.

This study also guides us to improve model transferability by more reliable spatial distributed forcing data and more realistic model structure. Besides, in some cases, auxiliary information will not guarantee the improvement of model realism and model transferability. Moreover, involving the snow and ice information in objective functions may lead the calibration to put extra effort into snow and ice simulation, which weakens the impact of hydrograph simulation in objective functions and causes deterioration in the performance of calibration and validation in the criteria of $I_{KGE}$. This may also result in the slight deterioration of model transferability.

It is not uncommon that hydrological models, especially the landscape-based models, have been criticized by equifinality (Beven & Binley 1992), mainly caused by more complicated model structure and larger amount of parameters, compared with lumped models. However, this study shows that with more realistic model structure, the equifinality can be well restricted, even with more complex model structure and larger amounts of parameters.

CONCLUSIONS

This study tested the impact of auxiliary information, i.e., forcing data spatial distribution, topography, and snow and ice data, on model performance and transferability. We started from a model setup (M_nAFOH) without this auxiliary information. Subsequently, involving forcing data spatial distribution and the impact of aspect on snow and ice melt step by step, M_nAFOH and M_AFOH were developed to test the impact of these two pieces of information. The results indicate that the forcing data spatial distribution and taking account of topography to calculate snow/ice melting had a marginal effect on model calibration, but improved model validation and, more importantly, the model transferability. Interestingly, when we add snow and ice information in the objective function to do calibration, the model performance in validation and transferability is not improved; on the contrary, it slightly deteriorates model performance for streamflow simulation in this study site. The results demonstrate that cooperating auxiliary information will not guarantee a better model performance. From this study, we can draw the following conclusions:

1. Forcing data spatial distribution, including precipitation and temperature, is essential in a snow and ice melt dominant catchment, and helpful to improve model performance in validation and model transferability.
2. Accounting for topography, i.e., aspect, in glacier and snow melting model can increase model realism and improve model transferability, but not as obvious as the forcing data spatial distribution, because the aspect information, to some extent, has been implicitly involved while coupling the glacier distribution information.
3. Well-gauged hydro-meteorological data might be sufficient to constrain a well-designed hydrological model. Involving the snow and ice data to constrain model parameters does not guarantee the improvement of model performance for streamflow simulation and model transferability.

ACKNOWLEDGEMENTS

This research is supported by State Key Laboratory of Cryospheric Sciences, Cold and Arid Regions Environment and Engineering Research Institute, Chinese Academy Sciences (SKLCS-OP-2016-04), and National Natural Science Foundation of China (41471001). We thank two anonymous reviewers for their constructive comments and valuable suggestion which dramatically
improved the quality of this paper. Thanks also go to Fabrizio Fenicia for his review and corrections of the draft manuscript.

REFERENCES


Li, H., Xu, C.-Y. & Beldring, S. 2015a How much can we gain with increasing model complexity with the same model concepts? *Journal of Hydrology* **527**, 858–871. DOI: http://dx.doi.org/10.1016/j.jhydrol.2015.05.044.


First received 9 May 2016; accepted in revised form 11 September 2016. Available online 2 December 2016