Evaluation of global forcing datasets for hydropower inflow simulation in Nepal

Bikas Chandra Bhattarai, John Faulkner Burkhart, Lena M. Tallaksen, Chong-Yu Xu and Felix Nikolaus Matt

ABSTRACT

Discharge over the Narayani river catchment of Nepal was simulated using Statkraft’s Hydrologic Forecasting Toolbox (Shyft) forced with observations and three global forcing datasets: (i) ERA-Interim (ERA-I), (ii) Water and Global Change (WATCH) Forcing Data ERA-I (WFDEI), and (iii) Coordinated Regional Climate Downscaling Experiment with the contributing institute Rossy Centre, Swedish Meteorological and Hydrological Institute (CORDEX-SMHI). Not only does this provide an opportunity to evaluate discharge variability and uncertainty resulting from different forcing data but also it demonstrates the capability and potential of using these global datasets in data-sparse regions. The fidelity of discharge simulation is the greatest when using observations combined with the WFDEI forcing dataset (hybrid datasets). These results demonstrate the successful application of global forcing datasets for regional catchment-scale modeling in remote regions. The results were also promising to provide insight of the interannual variability in discharge. This study showed that while large biases in precipitation can be reduced by applying a precipitation correction factor \( p_{\text{corr \_factor}} \), the best result is obtained using bias-corrected forcing data as input, i.e. the WFDEI outperformed other forcing datasets. Accordingly, the WFDEI forcing dataset holds great potential for improving our understanding of the hydrology of data-sparse Himalayan regions and providing the potential for prediction. The use of CORDEX-SMHI- and ERA-I-derived data requires further validation and bias correction, particularly over the high mountain regions.

Key words | discharge, global forcing dataset, Himalaya, hydrological modeling, hydropower inflow simulation

INTRODUCTION

The Himalayan and adjacent Tibetan Plateau water supply is intricately linked to the livelihoods, economic and social, to millions of people (Bookhagen & Burbank 2010; Immerzeel et al. 2010; Remesan et al. 2019; Zhang et al. 2019). This region provides for drinking water, hydropower generation, agricultural demands, as well as water-powered grain mills and other agro-economic activities (Ménégoz et al. 2013). Large Asian rivers, such as the Indus, Sutlej, Ganges, and Brahmaputra, are widely acclaimed for their great cultural, spiritual, economic, and ecological significance (Kumar 2017). The Ganges River and its tributaries alone fulfill significant water demands for more than 250 million people. Originating in the Himalayas, the river travels over 2,500 km, aggregating water from its tributaries...
to become the third largest freshwater delta (Chowdhury & Ward 2004).

Due to the high precipitation rates (1,500–2,500 mm/year; Dahal & Hasegawa 2008) and steep elevation gradients, hydropower has significant potential in the region and, if well managed, could provide a resource for economic growth. Hydropower potential depends on climatic conditions, particularly on precipitation, evaporation, temperature, and snow/ice in the catchment (Edenhofer et al. 2011). However, the barriers for sustainable water resource management in this region are manifold. First, there are numerous socio-political issues related to trans-border management issues (Biswas 2011). Despite efforts, a comprehensive framework to guide the development and international cooperation has yet to be ratified (Biswas 2011). Second, portions of the region are heavily glaciated and climate change is exacting an immediate and tangible impact (Immerzeel et al. 2010; Dehecq et al. 2019).

In this paper, we focus on a third significant barrier: providing robust and reliable analysis of the water resource potential and impacts from climate change. Data scarcity and the complexity of the terrain creating large spatial gradients and variability in weather and climate make such analyses a challenging task. The extremely heterogeneous topography in the region presents a great challenge for the accurate measurement of meteorological variables, giving rise to data scarcity (Pellicciotti et al. 2012). Although there are some meteorological stations, they are not well distributed in space. To overcome these challenges, regional climate models and reanalysis data offer gridded datasets of many meteorological variables, although with a rather coarse spatial resolution (Guo & Su 2019). The Coordinated Regional Climatic Downscaling Experiment (CORDEX) (Giorgi et al. 2009) provides extensively used regional datasets for past and future climate. For a historical perspective, the reanalysis dataset ERA-Interim (ERA-I) from the European Center for Medium-Range Weather Forecasts (ECMWF) (Dee et al. 2011) is widely used (Li et al. 2013, 2018; Xu et al. 2016). Bharti & Singh (2015) reported that the ERA-I precipitation is largely overestimated over the Indian Himalayan region. To address the bias inherited in ERA-I, bias- and elevation correction-based monthly observations were carried out when creating the Water and Global Change Forcing Data ERA-I (WFDEI) datasets (Weedon et al. 2014; Kim et al. 2019).

Various factors make hydrological modeling in the region challenging, including a large spatial variability in hydrometeorological variables, steep gradients, marked seasonality driven by the Indian Monsoon, and contrasting moisture regimes between the high elevation Tibetan Plateau and regions in the vicinity of the Indian Ocean. Changes in the climatic condition in the region may lead to changes in regional water balance components impacting hydrologic regimes (Koch et al. 2011). Moreover, snow and glacier storage and melt play an important role for the river discharge generation (Radić & Hock 2014; Li et al. 2015, 2016). The effective assessment and implementation of hydropower projects and other water resource management projects depend on a thorough analysis of the discharge and hydrological storages (Bakken et al. 2013).

Hydrologic simulation of discharge and other water balance components (evapotranspiration, snow, and groundwater storage) (Bhattacharyya & Regmi 2015; Matt & Burkhart 2018; Li et al. 2019) is used for the analysis of available water resources both in the past and future analyses, but the quality of discharge simulation remains a challenge (Rochester 2010; Engeland et al. 2016; Kauffeldt et al. 2016). The choice of a suitable hydrological model and appropriate forcing data are critical for any analysis and will greatly affect the outcome (Kauffeldt et al. 2016). Currently, no standardized or community modeling framework within hydrology and prior studies of the water resources of the Himalayan region have been conducted using different types of hydrologic models ranging from simple conceptual models (e.g. Pradhananga et al. 2014; Bhattacharyya & Regmi 2015; Skaugen & Weltzien 2016; Bhattacharyya et al. 2018) to more advanced, distributed models (e.g. Pellicciotti et al. 2012; Jain et al. 2017). Hydrological modeling and water balance studies in Himalayan regions have taken a range of approaches and used different models to focus on topics such as glacier melt and retreat, water balance, flooding, and the impact of climate change (Bookhagen & Burbank 2010; Pellicciotti et al. 2012; Wortmann et al. 2014; Khanal et al. 2015). Recent studies have addressed the performance of large-scale forcing datasets, including the WFDEI and ERA-I, for discharge simulation in various regions across the world (e.g. Li et al. 2013, 2016; Weedon et al. 2014; Essou et al. 2016; Nkiaka et al. 2017), but, to date, the use of these
global forcing datasets to simulate discharge in a Himalayan catchment is limited.

In this study, we use a distributed, conceptual hydrologic framework with an energy balance-based snow routine that has been demonstrated to perform well in the region (Xu et al. 2015; Hegdahl et al. 2016; Matt & Burkhart 2018). The main goal of this work is to evaluate the impact of four different forcing datasets on the simulated discharge and associated water balance assessment for the Narayani river catchment in Nepal. The forcing datasets include WFDEI, ERA-I, CORDEX, and a hybrid dataset referred to as Observed + WFDEI (all datasets are described in the Data and Methods section). Our specific objectives are to evaluate: (i) the discharge sensitivity to the forcing dataset for hydrologic modeling inflow, (ii) model performance using bias-corrected versus non-bias-corrected forcing as input to a (calibrated) hydrological model, and (iii) the sensitivity to forcing data for water balance assessment. Furthermore, we assess the overall ability of the Statkraft Hydrological Forecasting Toolbox (Shyft) (Burkhart et al. 2016) to simulate discharge in this complex and data-sparse regions.

**STUDY AREA**

The Narayani river catchment lies in the central part of Nepal. About 13% of the total area lies in China (Figure 1), and thus, it is a transboundary catchment. The main Narayani river gauging station is located in Narayanghat (27°42'30"N, 84°25'50"E) and is operated by the Department of Hydrology and Meteorology, Government of Nepal (DHM, GoN). The catchment area is 31,692 km² and is partly glacier covered (~8%; Omani et al. 2017) with elevation ranging from 175 m a.s.l. in the south to 8,148 m a.s.l. in the north. The hydrological regime is heavily influenced by the season. The seasons are defined as monsoon (June to September), post-monsoon (October to November), winter (December to February), and pre-monsoon (March to May) (Shrestha et al. 1999; Bhattarai et al. 2019). Almost 80% of total annual precipitation occurs during the monsoon period (Nayava 1974; Kripalani & Sontakke 1996). Tributaries to the Narayani river are either monsoon fed (those originating in middle and high mountain regions) or glacial and snow melt fed (those originating in a higher Himalayan region).

![Map of Nepal](http://iwaponline.com/hr/article-pdf/51/2/202/682207/nh0510202.pdf)
region). The Narayani river catchment lies also in the primary hydropower development region of Nepal, providing 44% of the total electric generation (Adhikari 2006).

DATA AND METHODS

Meteorological forcing data

Forcing variables used in the study include air temperature (T), precipitation (P), relative humidity (RH), wind speed (WS), and shortwave incoming solar radiation (S) at a daily resolution. Details of each dataset are described in the following sections.

Observed data

The DHM, GoN is responsible for collecting and disseminating hydrological and meteorological information for water resources, agriculture, energy, and other development activities in the country. Most of the measured data are based on the conventional manual observation. Measurements from the automatic weather station (AWS) are available only after the year 2000 (Karki 2010). The geographic location and installation date for each station are available from http://dhm.gov.np/meteorological-station/.

Observed temperature and precipitation station data which are located inside or in the vicinity of the catchment (Figure 1) were collected from the DHM. Data from the DHM were manually plotted for each station and screened for data quality. Clearly, erroneous departures from the historical pattern were removed manually from the datasets. The highest temperature station elevation is 3,870 m a.s.l., located in Chhoser (DHM st. no 633), and the highest precipitation station elevation is 3,705 m a.s.l., located in Mustang (DHM st. no 612).

To accommodate missing data, stations with less than 10 years of record or missing more than 15% of the observations were removed from the datasets. Ten precipitation stations (out of 82) did not meet this criterion and were removed from the datasets. Similarly, seven (out of 30) temperature stations were removed from datasets. This resulted in 72 precipitation stations and 23 temperature stations that were used for further analysis. Maximum numbers of stations with temperature and precipitation data were observed in the 2000–2009 period. Therefore, in this study, data for the period 2000–2009 were used.

All the stations are located below 4,000 m a.s.l., covering only 60% of the total catchment area (Figure 2(a)). Generally, temperature and precipitation data from higher elevations are missing, which add uncertainty to the current study. Daily mean temperature and mean annual precipitation for each station are plotted versus station elevation in Figure 2(b). Temperature shows as expected a consistent decrease with elevation, whereas precipitation shows a more mixed picture, with no clear trend below approximately 2,000 m a.s.l., but a notable decrease in precipitation above, i.e. from about 2,000 to 3,500 m a.s.l. Normally precipitation increases with elevation (Daly et al. 1994) in the mountainous region due to the orographic effect, but in the Himalayan region, an opposite pattern has been reported after a certain elevation level (e.g. Nayava 1980; Kansakar et al. 2004), in agreement with what is observed for the Narayani river catchment. From the analysis, station average daily temperature and mean annual precipitation for the period 2000–2009 were 10.7 °C and 1,292 mm/year, respectively.

Discharge observed at the Narayanghat station is available from the year 2000–2009 with no missing values, and this is also the period used for the hydrological model calibration and validation. Daily average discharge for the period 2000–2009 was 1,482.5 m³/s.

Reanalysis and regional climate model data

A summary of the available time periods and the resolution of the forcing datasets are provided in Table 1. All datasets are well documented and are freely available.

ERA-I is a reanalysis global forcing dataset available from 1979, produced by the ECMWF. ERA-I temperature results from the assimilated surface temperature (Essou et al. 2016), while precipitation data are based on a reanalysis of precipitation fields generated with a meteorological model (Berrisford et al. 2011; Dee et al. 2011). The obtained precipitation data are not scaled using observation data. ERA-I is continuously updated once per month, with a delay of 2 months, and is freely available from http://apps.ecmwf.int/datasets/. Monthly mean meteorological variables
averaged over the Narayani river catchment using the Shyft interpolation routines (see the section ‘Spatial-temporal interpolation of forcing data’) are shown in Figure 3. The catchment average mean temperature for the period (2000–2009) was 7.6°C, while catchment average annual precipitation was 4,660 mm/year.

The WFDEI is a global forcing dataset obtained by downscaling and bias-correcting ERA-I data (Weedon et al. 2014; Raimonet et al. 2017). The temporal and horizontal resolution of the dataset is shown in Table 1. WFDEI has two sets of rainfall generated by using either Climate Research Unit (CRU) or Global Precipitation Climatology Centre (GPCC) precipitation correction methods (Weedon et al. 2014). In this study, GPCC-corrected data were preferred to CRU because of their higher resolution and data quality (Weedon et al. 2014; Raimonet et al. 2017). The WFDEI dataset is freely available online from ftp.iiasa.ac.at. Catchment average mean daily temperature and annual precipitation (for the period 2000–2009) were found to be 7.9°C and 1,764 mm/year, respectively.

CORDEX is a program sponsored by the World Climate Research Programme (WCRP), to produce an improved generation of regional climate change projections (Giorgi et al. 2009). CORDEX has two datasets, referred to Evaluation and Historical. Evaluation is run within reanalysis and is used to ‘mimic’ observations (i.e. represent real weather), whereas Historical is run within a climate model, and the results can only be used in a climatological sense. Daily datasets of the Evaluation product over the study area are available for the South Asia CORDEX.

Table 1 | Summary of selected forcing datasets (n refers to the number of grid cells in the Narayani river catchment)

<table>
<thead>
<tr>
<th>Forcing dataset</th>
<th>Data period</th>
<th>Spatial resolution (degrees)</th>
<th>Temporal resolution</th>
<th>n</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERA-I</td>
<td>1979–2015</td>
<td>0.75 × 0.75</td>
<td>3 h, 6 h</td>
<td>12</td>
<td>Dee et al. (2011)</td>
</tr>
<tr>
<td>WFDEI</td>
<td>1979–2016</td>
<td>0.5 × 0.5</td>
<td>Daily</td>
<td>20</td>
<td>Weedon et al. (2014)</td>
</tr>
<tr>
<td>CORDEX-SMHI</td>
<td>1980–2010</td>
<td>0.44 × 0.44</td>
<td>Daily</td>
<td>25</td>
<td>Sanjay et al. (2017)</td>
</tr>
<tr>
<td>Observed + WFDEI</td>
<td>1999–2010</td>
<td>–</td>
<td>Daily</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Figure 2 | (a) Station altitude (temperature and precipitation) plotted against the hypsographic curve for the Narayani river catchment and (b) station average daily mean temperature and annual mean precipitation (for the period 2000–2009) plotted against altitude. Station name, number, and data availability are provided in the Supplementary Data.
domain (Region 6) from two institutional runs: Indian Institute of Tropical Meteorology (CORDEX-IITM) and Swedish Meteorological and Hydrological Institute (CORDEX-SMHI) (Sanjay et al. 2011). Based on a comparison of seasonal precipitation patterns over the study area, the CORDEX-SMHI was selected for this study. The horizontal resolution and available periods for the CORDEX-SMHI data are shown in Table 1. CORDEX-SMHI data can be freely downloaded on request from http://cccr.tropmet.res.in/home/index.jsp. Catchment average mean daily temperature and annual precipitation (for the period 2000–2009) are found to be 6.1 °C and 5,431 mm/year, respectively.

**Topographical and land cover datasets**

In this study, a digital elevation model (DEM) of 90 m spatial resolution from NASA’s Shuttle Radar Topography Mission (NASA-SRTM) was used. The NASA-SRTM DEM is freely available for download from https://eros.usgs.gov/. The hydrology tool available in the software package, System for Automated Geoscientific Analysis (SAGA) (Conrad et al. 2015) under Qgis-2.18 (QGIS Development Team 2016), was used for automatic catchment delineation. Catchment delineation was based on the gauge point at Narayanghat and NASA-SRTM DEM. Catchment slope
and aspect data for Shyft (see the Hydrological Model section) were generated during the catchment delineation subprocesses in Qgis. The calculated catchment domain was gridded into 2 km × 2 km cells for input to Shyft. The centroid of each vector grid cell was calculated using Qgis as further input to the model.

A land cover map (0.5° × 0.5°, resolution) providing forest, lake, glacier, and reservoir cover of the study area was extracted from the Moderate Resolution Imaging Spectroradiometer (MODIS) land cover data (Channan et al. 2014). MODIS land cover data are available free online at http://glcf.umd.edu/data/lc/. Land cover fractions of forest, lake, glacier, and reservoir cover area at the centroid point per each grid cell within the catchment were also calculated. The python code and algorithm for calculating land cover fraction per each vector grid cell at centroid are available online at https://github.com/felix-matt/shyft-gis.

Spatial interpolation of observed and gridded forcing data

Shyft is able to ingest daily spatially distributed meteorological data as input variables. Here, we interpolate both observations (station data) and gridded forcing data (WFDEI, ERA-I, and CORDEX-SMHI).

The gridded forcing data have a coarser resolution than the simulation domain in the hydrological modeling framework of Shyft (Table 1). Accordingly, Shyft provides interpolation routines used for downscaling the forcing data sets to the model grid cell scale. Data interpolation methods were selected based on prior studies by Sapkota (2016), Lombraña (2017), Matt et al. (2018), Matt & Burkhart (2018), and Teweldebrhan et al. (2018). Observed temperature and gridded data were extrapolated/interpolated by three-dimensional Kriging where elevation is the third dimension (Oliver & Webster 1990). Observed precipitation, gridded precipitation, global radiation, WS, and RH data were extrapolated/interpolated by inverse distance weighting (IDW) (Shepard 1968). A precipitation gradient of −0.07 mm/100 m (Table 2) and a temperature gradient of −0.6 °C/100 m are used during interpolation. The precipitation gradient was calculated from the observed precipitation data. The different interpolation methods along with their parameter values are presented in Table 2.

| Parameter values used for IDW and Kriging interpolation |
|---------------------------------|------------|
| **IDW**                        | **Kriging** |
| Parameter                      | Value      | Parameter | Value      |
| Maximum distance               | 400,000 m  | Temperature gradients | −0.6 °C/100 m |
| Maximum number                 | 10         | Range     | 20,000 m   |
| Distance measure factor        | 1          | z-scale   | 20         |
| Resolution                     | 2 km × 2 km| Resolution | 2 km × 2 km |
| Precipitation gradient         | −0.07 mm/100 m | Sill | 25 °C    |

Hydrological model

The Shyft modeling framework was used to simulate daily discharge from the catchment. The modeling framework has three main hydrologic modeling routines (https://gitlab.com/shyft-os). These three models are different in the way to calculate evapotranspiration and snow estimation and melt.

In this study, the PT_GS_K model was used for discharge simulation. The PT_GS_K model uses the Priestley–Taylor (PT) method (Priestley & Taylor 1972) for estimating potential evapotranspiration, the Gamma snow routine (GS; described in the Gamma Snow section) for snowmelt, sub-grid snow distribution and mass balance calculation, and a simple storage–discharge function (K) for catchment response calculation (Lambert 1972; Kirchner 2009). Actual evapotranspiration (AE) was assumed to take place from the snow-free area and was estimated as a function of potential evapotranspiration and a scaling factor.

The catchment response function ‘K’ is based on the storage–discharge relationship concept described in Kirchner (2009) and represents the sensitivity of discharge to storage changes as given by the following equation. Generally, discharge is nonlinear and varies by many orders of magnitude, and it is recommended to use log-transformed
discharge values to avoid numerical instability.

$$\frac{d(ln(Q))}{dt} = g(Q)(\frac{P - E}{Q} - 1)$$  \hspace{1cm} (1)

where $P$, $Q$, and $E$ are the rate of precipitation (input from either snow melt or rain), discharge, and AE (from the snow-free area), respectively, in units of depth per time.

The idea behind this method is that the discharge sensitivity to changes in storage, i.e. $g(Q)$, can be estimated from the time series of the discharge alone through fitting empirical functions to the data, such as the quadratic equation (Kirchner 2009), which is given by:

$$g(Q) = e^{c_1 + c_2 (ln(Q)) + c_3 (ln(Q))^2}$$  \hspace{1cm} (2)

where $c_1$, $c_2$, and $c_3$ are the catchment-specific outlet parameters (hereafter called Kirchner parameters) obtained during the model calibration. There is no routing function in the model, but the Kirchner response function represents a delayed outflow from storage within the catchment.

**Gamma snow**

In Shyft, the Gamma snow routine, an energy balance approach (Equation (3)) for snow ablation, and the snow depletion curve following a gamma distribution are combined into a single routine. The energy balance approach in the Gamma snow routine is based on DeWalle & Rango (2008) and also briefly explained in Hegdahl et al. (2016). The net energy flux ($\Delta E$) at the surface available for snow ablation is expressed as follows:

$$\Delta E = S(1 - a) + L_{in} + L_{out} + H_{SE} + H_L + E_G$$  \hspace{1cm} (3)

where $S$ is the net shortwave radiation, $L_{in}$ and $L_{out}$ are the incoming and outgoing longwave radiations, $H_{SE}$ and $H_L$ represent sensible and latent heat fluxes, and $E_G$ is the net ground heat flux calculated using a bulk-transfer approach. Two parameters defining the wind profile, intercept (wind constant), and slope (wind scale) are determined either by model calibration or as provided (Table 3). For a given time step ($t$), the snow albedo ($a$) at each cell depends on the minimum ($a_{min}$) and maximum albedo ($a_{max}$) as well as the albedo decay rate, temperature, and snowfall as described in Hegdahl et al. (2016):

$$a_t = \begin{cases} a_{min} + (a_{t-1} - a_{min}) \left(1 - \frac{1}{2 \text{FDR}}\right) & \text{if } T > 0 \text{°C} \\ a_{t-1} + (a_{max} - a_{min}) \left(1 - \frac{1}{2 \text{SDR}}\right) & \text{if } T \leq 0 \text{°C} \end{cases}$$  \hspace{1cm} (4)

In Equation (4), FDR and SDR denote fast and slow snow cover decay rates, respectively. In this study, $a_{max}$ and $a_{min}$ are prescribed (refer Table 3).

Within the Gamma snow routine, precipitation falling in each cell is classified as solid or liquid depending on a threshold temperature ($t_x$) and the actual cell temperature. Snow distribution within each cell is estimated by using a three-parameter gamma probability distribution. The third parameter in the gamma probability distribution represents the bare ground fraction in the cell. Finally, snowmelt depth (mm/day) is calculated by multiplying $\Delta E$ (available energy) with the latent heat of fusion for water.

A temperature index model which does not require glacier ice albedo was used to calculate glacier melt (see Hock 2003). The glacier reservoir was assumed to be infinite, and the glacier area was assumed to be constant throughout the simulation periods. Within a glacier-covered cell, glacier melt only happens from the snow-free fraction.

**Parameters and calibration**

Hydrological simulation from distributed models such as Shyft generally requires the estimation of model parameters through calibration with measured data (Madsen 2005). In the PT_GS_K routine, there are 14 parameters, out of which eight parameters (Table 3: top eight parameters) have been found to be significantly more sensitive than the rest (Teweldebrhan et al. 2018) and were selected for calibration in this study. The remaining six parameters were prescribed (Table 3: lower six parameters). The precipitation correction factor ($p_{corr\_factor}$), which is used to correct bias, was also set as a calibration parameter.

Manual calibration can be time-consuming and subjective; therefore, an automatic calibration was carried out.
Upper and lower limits for each parameter are shown in Table 3. CREST v.2.1, the Shufue Complex Evolution University of Arizona (SCE-UA) (Duan et al. 1992), is used as the kernel algorithm in the automatic calibration process. Typically, the procedure involves the selection of samples in the parameter space through the use of competitive evolution schemes, such as the simplex scheme, to reproduce better the observations. After several iterations, either due to convergence or when the maximum number of iteration is reached, the best set of parameter values based on the Nash–Sutcliffe coefficient of efficiency (NSE) is determined (Chu et al. 2010).

In this study, the model was calibrated and validated for each forcing dataset independently (hereafter named ‘independent calibration mode’). The model calibration was based on observed discharge data from 2000 to 2004, and validation was based on data from 2005 to 2009.

**Water balance estimation**

A water balance analysis is a useful tool to describe the principal components of water in and out of a catchment (Rochester 2010), where the volume of water inflow should be balanced with water outflow, assuming no changes in storage (ΔS/Δt). The following water balance components: mean annual precipitation, discharge, and AE were calculated. Water balance components were calculated using calendar years, so that the change in storage is mainly a

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Model calibration parameters with upper and lower limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Description and unit</td>
</tr>
<tr>
<td>c1</td>
<td>Outlet empirical coefficient 1 (–) Equation (4)</td>
</tr>
<tr>
<td>c2</td>
<td>Outlet empirical coefficient 2 (–) Equation (4)</td>
</tr>
<tr>
<td>c3</td>
<td>Outlet empirical coefficient 3 (–) Equation (4)</td>
</tr>
<tr>
<td>wind scale</td>
<td>Slope in turbulent wind function (m/s)</td>
</tr>
<tr>
<td>tx</td>
<td>Temperature threshold rain/snow (°C)</td>
</tr>
<tr>
<td>FDR</td>
<td>Fast albedo decay rate during melt (days)</td>
</tr>
<tr>
<td>SDR</td>
<td>Slow albedo decay rate during cold condition (days)</td>
</tr>
<tr>
<td>p_corr_factor</td>
<td>Scaling factor for precipitation (–)</td>
</tr>
<tr>
<td>ae scale factor</td>
<td>Scaling factor for AE (–)</td>
</tr>
<tr>
<td>Surface magnitude</td>
<td>Snow heat constant (mm SWE)</td>
</tr>
<tr>
<td>wind constant</td>
<td>Intercept in turbulent wind function (–)</td>
</tr>
<tr>
<td>min albedo</td>
<td>Minimum snow albedo used in snow routine</td>
</tr>
<tr>
<td>max albedo</td>
<td>Maximum snow albedo used in snow routine</td>
</tr>
<tr>
<td>max water</td>
<td>Frictional max water constant of snow (–)</td>
</tr>
</tbody>
</table>

In the table, ‘K’ is the catchment response function; ‘GS’ is the Gamma snow; ‘PT’ is the Priestley and Taylor; and ‘AE’ is the actual evapotranspiration.
difference in storage between first and last days of simulation.

The annual change in storage was calculated according to the following equation:

\[
\frac{\Delta S}{\Delta t} = (P + G) - (Q + ET)
\]

where \( P \) is precipitation, \( Q \) is the discharge, \( ET \) is actual evapotranspiration, \( G \) is the glacier melt, and the unit of measurement is mm/year. \( \Delta S/\Delta t \) is the change in storage per time unit.

### Model performance evaluation

To determine the agreements between observed and simulated discharges, NSE (Nash & Sutcliffe 1970), Kling–Gupta efficiency (KGE) (Gupta et al. 2009), square-root-transformed NSE (NSE\textsubscript{sqrt}) (Seiller et al. 2012), benchmark series (G\textsubscript{bench}) (Seibert 2001), and percentage volume difference (\( D_v \)) (Martinec et al. 1998) were used (Table 4). The NSE, NSE\textsubscript{sqrt}, and KGE efficiency measures are used to access the predictive power of the hydrological model and can range from \(-\infty \) to 1. An efficiency of 1 corresponds to a perfect match of simulated and observed discharges. To improve the estimation of the performance error, KGE considers three components: bias (\( \alpha \)), variability (\( \beta \)), and linear correlation coefficient (\( r \)) to overcome the problems associated with NSE, i.e. NSE results in the underestimation of the streamflow variability and the runoff peaks (Gupta et al. 2009). NSE\textsubscript{sqrt} gives more emphasis to the overall agreement between observed and simulated discharges (Seiller et al. 2012; Peña-Arencibia et al. 2015) as compared to NSE, so it is also included as an evaluation criterion. Percentage volume difference (\( D_v \)) gives the percentage bias between the simulated and observed series and can range from \(-\infty \) to \( +\infty \). \( D_v \) equals to zero indicates a perfect agreement between simulated and observed discharges. Since the observed discharge shows strong seasonal patterns, we also used a benchmark series (G\textsubscript{bench}) for model evaluation. The monthly average discharge was used as a benchmark series. G\textsubscript{bench} is negative if the model performance is poorer than the benchmark, zero if the model performs as well as the benchmark, and positive if the model is superior, with a highest value of one for a perfect fit. In this study, we aimed to achieve NSE > 0.7, KGE > 0.7, NSE\textsubscript{sqrt} > 0.7, G\textsubscript{bench} > 0.5, and \( D_v \) within \( \pm 15\% \), during both the model calibration and validation periods.

### Table 4: Definitions of the evaluation criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Mathematical expression</th>
<th>Description</th>
<th>Best value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_v )</td>
<td>( \frac{V_{\text{obs}} - V_{\text{sim}}}{V_{\text{obs}}} \times 100 )</td>
<td>Volume difference in percentage</td>
<td>0</td>
</tr>
<tr>
<td>NSE</td>
<td>( 1 - \frac{\sum_{i=1}^{n} (Q_{\text{sim},i} - Q_{\text{obs},i})^2}{\sum_{i=1}^{n} (Q_{\text{obs},i} - Q_{\text{obs}})^2} )</td>
<td>Nash–Sutcliffe efficiency</td>
<td>1</td>
</tr>
<tr>
<td>NSE\textsubscript{sqrt}</td>
<td>( 1 - \frac{\sum_{i=1}^{n} (\sqrt{Q_{\text{sim},i}} - \sqrt{Q_{\text{obs},i}})^2}{\sum_{i=1}^{n} (\sqrt{Q_{\text{obs},i}} - \sqrt{Q_{\text{obs}}})^2} )</td>
<td>Squared-root transformed Nash–Sutcliffe efficiency</td>
<td>1</td>
</tr>
<tr>
<td>KGE</td>
<td>( \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} )</td>
<td>Modified Kling–Gupta efficiency</td>
<td>1</td>
</tr>
<tr>
<td>Here, ( \beta = \frac{Q_{\text{min}}}{Q_{\text{obs}}}; \alpha = \frac{Q_{\text{sim, std}}}{Q_{\text{obs, std}}} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G\textsubscript{bench}</td>
<td>( \frac{\sum_{i=1}^{n} (Q_{\text{obs},i} - Q_{\text{bench}})^2}{\sum_{i=1}^{n} (Q_{\text{obs},i} - Q_{\text{sim},i})^2} )</td>
<td>Goodness of fit with respect to the benchmark series</td>
<td>1</td>
</tr>
</tbody>
</table>

\( n \) = total number of days in the evaluation period; \( Q_{\text{sim}} \) and \( Q_{\text{obs}} \); simulated and observed discharges; \( Q_{\text{sim, std}} \) and \( Q_{\text{obs, std}} \); standard deviation for simulated and observed discharges; \( V_{\text{obs}} \) and \( V_{\text{sim}} \); arithmetic mean for simulated and observed discharges; \( V_{\text{sim}} \); total discharge volume for simulated and observed discharges; \( Q_{\text{bench}} \); monthly long-term average observed discharge.
RESULTS

Meteorological forcing data analysis

As discussed in the section ‘Reanalysis and regional climate model data’, the resolutions of gridded datasets are different. For the sake of comparison, Shyftinterpolated meteorological variables (for the period 2000–2009) for each forcing dataset were used. Temperature shows an overall agreement where all three datasets demonstrate similar seasonality and distinct cool or negative bias as compared to the observations during the winter and pre-monsoon seasons (Figure 3). The bias was the strongest in the CORDEX-SMHI data during the September–December period. Long-term spatially averaged temperature for July (Figure 4) shows that the CORDEX-SMHI was not able to capture higher daily average temperatures in the river valley. Daily average temperatures during July over the lower elevation region from ERA-I and CORDEX-SMHI were lower than the temperatures from the Observed and WFDEI interpolated datasets. What was most interesting is that negative catchment average temperatures, which were not at all captured by the interpolated observations (Figure 3), will have a significant impact on any simulation. We attribute the lack of negative catchment average temperatures to the lack of stations above 4,000 m a.s.l. combined with a too low-temperature lapse rate. This is a typical challenge when observations lack representativity to topography.

Precipitation varies notably among the forcing datasets. WFDEI precipitation agrees well with the observation (Figures 3 and 4). The strongest positive bias was observed for CORDEX-SMHI and ERA-I datasets (Figure 5). Long-term daily averaged precipitation for July–September shows the strongest deviation among forcing. The catchment precipitation for July was more than 150 mm/day in CORDEX-SMHI and ERA-I datasets, and the highest precipitation was found in the lower elevations. A seasonal comparison shows that the WFDEI was slightly higher than the observed precipitation during the monsoon season (Figure 6). Similar findings are reported for the Indus catchment by Dahri et al. (2016).

Monthly averaged wind speeds and incoming short-wave radiation from WFDEI and ERA-I were fairly similar but different for CORDEX-SMHI. Station
observations for RH, WS, and shortwave incoming global radiation were not available for the catchment, but previous global studies by Iizumi et al. (2014) and Weedon et al. (2014) showed that the WFDEI is more representative for the observations.

Model parameters

Shyft was calibrated for each forcing dataset, and the calibrated parameters are shown in Table 5. Normal ranges for all calibrated parameters are given in Table 3. The
highest range in values (greater than 0.15 variance in the normalized scale of –1 to 1) was seen for the parameters $p_{\text{corr\_factor}}$ (0.4–1.34), wind scale (2.01–4.84), threshold temperature (tx) (–1.68 to 0.94), and slow albedo decay rate (29.04–39.98). A relatively small variation in parameter values (less than 0.05 variance in the normalized scale of –1 to 1) was seen for the Kirchner coefficients ($c_1$ (–6.64 to –5.51), $c_2$ (0.14 to 0.44), and $c_3$ (–0.05 to –0.07)) and fast albedo decay rate (9.94–12.03). The highest variance (0.27) in a normalized scale of –1 to 1 was found for $p_{\text{corr\_factor}}$. The lowest and highest $p_{\text{corr\_factor}}$ values were observed for CORDEX-SMHI (0.4) and Observed + WFDEI (1.34) forcing dataset, respectively. The higher $p_{\text{corr\_factor}}$ for Observed + WFDEI is interesting and suggests an underprediction of precipitation in the catchment when gridded precipitation was based on Shyft-interpolated station observations. Unique to the CORDEX-SMHI dataset, the Gamma snow threshold temperature (tx) was found to be positive (0.94). Less sensitive parameters as reported by Teweldebrhan et al. (2018) were not calibrated but are listed with given values in Table 3.

### Evaluation of discharge simulation using different forcing datasets

Figure 7 compares daily observed and simulated discharges using the different forcing datasets. Calibration and validation results in terms of error statistics are shown in Table 6. Simulations from Observed + WFDEI forcing give the best results in terms of error statistics during both calibration (NSE = 0.90, NSE$\sqrt{\text{sqrt}} = 0.91$, KGE = 0.94, $G_{\text{bench}} = 0.43$, and $D_v = -0.91\%$) and validation (NSE = 0.90, NSE$\sqrt{\text{sqrt}} = 0.95$, KGE = 0.92, $G_{\text{bench}} = 0.56$, and $D_v = -1.29\%$) periods. $D_v$ in both calibration and validation was also best for Observed + WFDEI and found to be less than –2%. The second best performance was achieved for WFDEI (Table 6). For NSE, NSE$\sqrt{\text{sqrt}}$, and KGE, the ERA-I and WFDEI datasets were similar, but the volume differences during calibration were comparatively higher for ERA-I than WFDEI. Comparatively poorer model performance and higher volume differences (Table 6) were found.

### Table 5 | Estimated parameter values from the independent calibration mode for the four forcing datasets

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Unit</th>
<th>Observed + WFDEI</th>
<th>WFDEI</th>
<th>ERA-I</th>
<th>CORDEX-SMHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>–</td>
<td>–5.51</td>
<td>–6.09</td>
<td>–6.64</td>
<td>–5.99</td>
</tr>
<tr>
<td>$c_2$</td>
<td>–</td>
<td>0.44</td>
<td>0.25</td>
<td>0.14</td>
<td>–0.20</td>
</tr>
<tr>
<td>$c_3$</td>
<td>–</td>
<td>–0.05</td>
<td>–0.05</td>
<td>–0.05</td>
<td>–0.07</td>
</tr>
<tr>
<td>tx</td>
<td>°C</td>
<td>–1.68</td>
<td>–1.38</td>
<td>–0.36</td>
<td>0.94</td>
</tr>
<tr>
<td>wind scale</td>
<td>m/s</td>
<td>2.01</td>
<td>4.70</td>
<td>4.84</td>
<td>3.40</td>
</tr>
<tr>
<td>FDR</td>
<td>days</td>
<td>11.59</td>
<td>12.03</td>
<td>9.94</td>
<td>10.75</td>
</tr>
<tr>
<td>SDR</td>
<td>days</td>
<td>39.98</td>
<td>32.66</td>
<td>29.04</td>
<td>31.55</td>
</tr>
<tr>
<td>$p_{\text{corr_factor}}$</td>
<td>–</td>
<td>1.34</td>
<td>1.09</td>
<td>20.41</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Figure 7 | Calibrated and simulated daily discharge for the Narayani river catchment for the period 2000–2009.
Goodness of fit with respect to the benchmark ($G_{bench}$) for ERA-I and CORDEX-SMHI during calibration periods was found negative, indicating poorer performance than the benchmark series. However, it should be noted the peak discharge was best simulated with CORDEX-SMHI data (Figure 7).

In Figure 8, the year 2004 was plotted to show in more details the representation of the annual cycle by the different forcing datasets. Overall, all datasets were able to simulate the cycle well; however, the CORDEX-SMHI simulation deviates somewhat from the rest and shows a higher peak flow early in the wet season.

Quantile–Quantile (QQ) plots for the different forcing datasets are shown in Figure 9. The QQ-plot for the calibration period is shown in Figure 9(a), for the validation period in Figure 9(b). To highlight low flows, Figures 9(c) and 9(d) show the QQ-plot on log-scale. The QQ-plot is a graphical technique to determine if two datasets come from the same population with a common distribution (Renard et al. 2013). Departures from the 1:1 reference line indicate discrepancies between the simulated and the observed discharge. Figure 9(a) shows that all simulations slightly overestimate discharge (with respect to the observed discharge) up to a value of around 4,000 m$^3$/s. Between 4,000 and 8,000 m$^3$/s, the CORDEX-SMHI simulations were higher than observed, whereas other simulations were lower than observed. After 8,000 m$^3$/s, CORDEX-SMHI simulations were able to capture peaks where others simulations were lower than observed. Although the

Table 6 | Performance statistics for different forcing datasets

<table>
<thead>
<tr>
<th>Error parameters</th>
<th>Observed - WFDEI</th>
<th>WFDEI</th>
<th>ERA-I</th>
<th>CORDEX-SMHI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calibration</td>
<td>Validation</td>
<td>Calibration</td>
<td>Validation</td>
</tr>
<tr>
<td>NSE</td>
<td>0.90</td>
<td>0.90</td>
<td>0.83</td>
<td>0.80</td>
</tr>
<tr>
<td>KGE</td>
<td>0.94</td>
<td>0.92</td>
<td>0.86</td>
<td>0.81</td>
</tr>
<tr>
<td>$D_v$</td>
<td>-0.91</td>
<td>-1.29</td>
<td>-1.72</td>
<td>-5.68</td>
</tr>
<tr>
<td>$NSE_{sqr}$</td>
<td>0.91</td>
<td>0.95</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>$G_{bench}$</td>
<td>0.43</td>
<td>0.56</td>
<td>0.01</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Figure 8 | Daily observed and simulated discharges from different forcing models for the year 2004.
Observed + WFDEI forcing dataset resulted in discharge simulations with the best performance, it was not able to capture the highest observed discharge (simulated discharge points after 95% quantile). The simulations from CORDEX-SMHI were able to capture relatively high peak-discharge events (above the 99% quantile); however, they have large deviations from the observations between 90% and 95% quantiles. For the low flows (i.e. 1–10% quantile), the simulation from the WFDEI forcing captures observation better than the Observed + WFDEI forcing (Figure 9(c)). However, low-flow simulation from ERA-I until 5% quantile was better than the rest.

Similarly, in the validation period (Figure 9(b)), simulated discharge and observed daily discharge for values less than 90% quantile from all forcing datasets fit well to the observations. Above 95% quantile, most forcing datasets fail to capture these highest discharges. Interestingly, the CORDEX-SMHI, which generally provides lower performance (Table 6), notable in the lower flow range, manages to capture the highest flows best. However, these
simulations were mostly higher than the reference line, indicating a general overestimation during the full range of simulations. Similar to the calibration periods, low-flow simulation from WFDEI was also found closest to the reference (observations) during the validation periods (Figure 9(d)).

**Water balance analysis**

Figure 10 compares the catchment average (2000–2009) water balance components for the four different forcing datasets (the exact values are given in Table 7). Annual averages over the 10-year simulation period, 25.1%, 24.8%, 30.7%, and 17.8% of precipitation, were lost to evapotranspiration for Observed + WFDEI, WFDEI, ERA-I, and CORDEX-SMHI datasets, respectively. Among all forcing datasets, Observed + WFDEI shows lower runoff than the others. Higher runoff from CORDEX-SMHI forcing was observed, and it is associated with higher average precipitation during the monsoon period (Figure 3). The lowest (72.8 mm) total glacier melt during the simulation period was observed for the WFDEI dataset, while the highest (297.3 mm) was observed from ERA-I. The change in storage (Equation (5)) shows that 73.1, 29.2, 68.7, and 32.7 mm of water were surplus (a positive change in storage) in the catchment for Observed + WFDEI, CORDEX-SMHI, ERA-I, and WFDEI forcing data, respectively. The largest glacier melt contribution (19%) to total runoff was observed for ERA-I forcing, while the lowest was observed for WFDEI (4%). Different water balance components and their associated changes in storage are shown in Table 7.

**DISCUSSION**

**Discussion on model parameters**

As with any hydrologic model, the model used in this study is sensitive to the parameters; particularly the Kirchner coefficients, precipitation correction factor, and the threshold temperature (Teweldebrhan et al. 2018). While calibrating the model, different values of the parameters were obtained for different forcing datasets (Table 5). An examination of the model parameters revealed that the calibrated values of the precipitation correction factor, wind scale, threshold temperature, and slow albedo decay rates differ a lot

---

### Table 7 | Water balance components for different forcing (numbers in mm/year)

<table>
<thead>
<tr>
<th>Water balance components</th>
<th>Observed + WFDEI</th>
<th>WFDEI</th>
<th>ERA-I</th>
<th>CORDEX-SMHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>1,771.1</td>
<td>1,965.7</td>
<td>1,901.4</td>
<td>2,234.1</td>
</tr>
<tr>
<td>Evapotranspiration</td>
<td>449.6</td>
<td>487.4</td>
<td>584.6</td>
<td>399.5</td>
</tr>
<tr>
<td>Glacier melt</td>
<td>242.2</td>
<td>72.8</td>
<td>297.3</td>
<td>132.5</td>
</tr>
<tr>
<td>Runoff</td>
<td>1,490.7</td>
<td>1,521.9</td>
<td>1,545.4</td>
<td>1,954.4</td>
</tr>
<tr>
<td>Change in storage (Equation (5))</td>
<td>73.0</td>
<td>29.2</td>
<td>68.7</td>
<td>32.7</td>
</tr>
</tbody>
</table>

---

**Figure 10 | Annual average (2000–2009) water balance components (Evap, actual evapotranspiration; Prec, precipitation calculated by using p_corr_factor) for different forcing datasets in the Narayani catchment.**
between the forcing datasets. Larson & Peck (1974) and Lakew et al. (2017) pointed out that the model simulations are highly sensitive to the precipitation correction factor, which helps adjusting systematic bias embedded in the precipitation forcing. The precipitation correction factor may account for both under catch and lack of representative stations (Lakew et al. 2017) and is directly related to the bias in precipitation data. Prior studies by Engeland et al. (2016) and Teweldebrhan et al. (2018) suggested that the Kirchner parameters are highly sensitive since they have a large influence on a given simulation. We found that the variance for Kirchner response parameters was less than 0.05. An explanation is that these parameters are not dependent on the forcing datasets but rather on the physical catchment characteristics and thereby the concentration time of the catchment. Kirchner response parameters are dependent on the physical catchment characteristics and thereby the concentration time of the catchment and are therefore less sensitive to the forcing.

**Potential factors controlling hydrological model efficiency during model calibration and validation**

Different forcing datasets lead to different levels of performance efficiency in terms of discharge simulations (Table 6). Explaining these differences is not easy, as each forcing dataset has different characteristics such as the spatial resolution and the assimilation methods. Figures 3–5 present the catchment average monthly mean forcing variables for the different forcing datasets. From the figures, we can see that in general, the spatio-temporal distribution of precipitation from WFDEI was quite similar to the interpolated observations. The extent of higher precipitation than observation during pre-monsoon and monsoon seasons was larger for ERA-I and CORDEX-SMHI. Overestimation of precipitation during the monsoon season identified for CORDEX-SMHI in the Himalayan region (Ghimire et al. 2018) and for the ERA-I over the Indus catchment (Dabhi et al. 2016). Moulin et al. (2009) and Zhu et al. (2014) indicate that a poor representativeness of precipitation over the catchment is a major source of uncertainty in discharge simulation.

Despite some of the biases in the forcing data, the performance of the model during calibration and validation showed that all datasets performed reasonably well. The good agreement between simulated and observed discharges during the model calibration and validation periods (Figure 7) is to a large extent due to the application of the precipitation correction factor. The multiplicative correction factor to the precipitation removes a substantial proportion bias in both precipitation and simulated discharges. It should be clarified here that such an adjustment of bias in discharge (through the input forcing correction) is an ‘ad hoc’ procedure to overcome the quality of forcing data. The ‘p_corr_factor’ shown in Table 5 indicates that the CORDEX-SMHI was reduced by a factor of 0.4, and ERA-I precipitation was reduced by a factor of 0.41 during the calibration period to minimize bias in discharge estimation. These numbers indicate that there were significant biases in these two forcing datasets and suggest that some kind of bias correction should be considered in any application. We also notice that the precipitation correction factor for WFDEI was near one. This is reasonable and encouraging, as this dataset is already bias-corrected (using observed precipitation). The precipitation correction factor for observed datasets was 1.34, indicating that the areal precipitation is underestimated. This underestimation is likely caused by a combination of undercatch of the precipitation gauges and the lack of representative observation stations. The undercatch depends on the precipitation gauge, wind, and precipitation phase (Mekonnen et al. 2015; Zhao et al. 2015), and the non-representativeness is caused by the limited number of observation stations at high altitudes. Better performance in discharge simulation by the application of a correction factor to precipitation was also shown by previous studies Larson & Peck (1974), Lawrence et al. (2009), and Lakew et al. (2017) and is a standard practice in the hydrologic analysis. The improvement of simulation from all forcing datasets versus prior results presented by Bhattarai et al. (2018) using the HBV model (Bergström 1995) in the same catchment is likely attributed to the distributed nature of Shyft.

Though we achieved high NSE, NSEsqrt, and KGE using WFDEI, ERA-I, and Observed + WFDEI forcing datasets with reasonable $D_v$ (Table 6), the simulations were still not able to capture the highest discharges as seen in the QQ-plots in Figure 9. As compared to high discharge, low discharge events were well captured by all forcing datasets.
Local simulated peak discharge during the monsoon season may be attributed to less representativeness of the input precipitation data. Particularly, the lack of ability to capture intense localized precipitation events could be the main reason for an inability to simulate high discharge concurrent with the observations. The CORDEX-SMHI simulation generated relatively higher discharge peaks as compared with the other datasets but has overall higher positive biases during the pre-monsoon season (Figure 3). The CORDEX-SMHI forcing datasets could provide an opportunity for the peak discharge analysis in the Himalayan catchment.

Discussion on the water balance analysis

The catchment water balance component analysis (Table 7) revealed notable differences among the forcing datasets. The discrepancies can, to some degree, be attributed to the uncertainty inherent in each forcing dataset. As mentioned by Nkiaka et al. (2017), different precipitation in each forcing dataset can strongly influence the optimized parameters that control the rates and threshold of hydrological processes in the catchment. Although measured evapotranspiration data are not available for comparison, the percentage of evapotranspiration estimates from WFDEI (24.79%) was found to be similar to the previous study by Sakai et al. (2004), Giertz et al. (2006), Sintondji et al. (2014), and Ragettli et al. (2015) in similar Himalayan catchments. The lowest evapotranspiration (399.5 mm/year) was found for the CORDEX-SMHI forcing dataset, which can be associated with high RH, and lower average shortwave radiation (Figure 3) than other forcing data. Lower evaporation (449.6 mm/year) and lower precipitation (1,771.1 mm/year) from the Observed + WFDEI than WFDEI may partly result from precipitation under-catch in the catchment.

Higher positive storage changes were found for the ERA-I and Observed + WFDEI forcing datasets as compared to CORDEX-SMHI and WFDEI. As compared to Observed + WFDEI forcing dataset, a lower storage change for ERA-I was observed and results from higher evapotranspiration in ERA-I. A change in storage was found smaller than the glacier melt for all forcing datasets. A smaller change in storage is due to the fact that it originates from snow that accumulates over the year and contributes to the glacier mass balance. Smaller changes in storage also indicate that the glacier mass balance was negative over the simulation period. Previous studies by Kulkarni (1992), Cogley et al. (2010), Bolch et al. (2012), Wagnon et al. (2013), and Gurung et al. (2016) also showed that the Himalayan glaciers are experiencing a higher retreat rate in recent decades. The highest glacier melt contribution was observed for ERA-I (19%), while the lowest glacier melt contribution is observed for WFDEI (4%). A study by Gupta et al. (2019) on the Marshyangdi River catchment (with 24% glacier cover area) shows that glacier melt contributes to 11.8% of the total discharge. A relatively lower percentage of snow and glacier melt contribution to the total discharge from the Modi River catchment (with 12% glacier cover area) is presented by Kayastha & Shrestha (2019). Similarly, a study by Nepal (2016) on the Dudhkoshi catchment (with 15% glacier cover area) shows that glacier melt contribution is 17% of the total discharge. A similar percentage of glacier melt contribution to the total discharge for Observed + WFDEI (16%) was observed for the Narayani river catchment, although the glacier cover area was only 8%. As suggested by Bonekamp et al. (2019), differences in glacier melt contribution to total discharge result from differences in meteorological forcing, as we also observed in our study.

Uncertainty in the model simulation and observation

Hydrological projections are subject to considerable uncertainty (Dobler et al. 2012) and are easily affected by various factors, including local and climatic conditions, optimized parameters (Shen et al. 2012), and the quality of forcing data (Teweldebrhan et al. 2018). Different factors also have a varying degree of impact on the discharge simulation. The challenge using a model with many model parameters is that we might get a good model fit but a less robust model for prediction and forecasting due to over-fitting. However, the uncertainty in the simulations will almost certainly be higher due to the increased uncertainty in the parameter values. In our study, the discharge was highly seasonal and particularly precipitation dependent. Inclusion of the precipitation correction factor as a calibrating parameter improves the predictability of the model at the expense of increases to the uncertainty in water balance components like evapotranspiration and snowmelt. However, only a single value of the precipitation correction
factor (p_corr_factor) was applied. More complex precipitation correction could include seasonality and impact of orographic effects, though we feel this would result in greater uncertainty. Nonetheless, uncertainties remain.

First, uncertainties in the observed data result from the uneven spatial distribution and few monitoring meteorological stations, which are mainly located in lower elevation regions (Figure 2). Furthermore, many observations are manual. This poses a unique source of uncertainty in measurements, not necessarily greater or less from automated. Furthermore, there are numerous sources of errors associated in the establishment of the river stage and parameterization of rating curves. During model calibration, we assumed that observed discharge is correct. Some of deviations between model simulations and observations might actually be explained by errors in the observations and not in the model or the forcings, to assess the uncertainty in streamflow observations with required detailed knowledge about local river profiles and data used to establish the rating curve and it is outside the scope of this paper. Second, uncertainties might originate from the estimation of AE and other water balance components, which again is influenced by the precipitation estimates and other forcing data used for driving the model. And thirdly, there are the uncertainties in the model formulation itself. The model in this study assumed that the discharge in the river depends on the amount of water stored in the catchment, and we did not consider the impact of industrial and households’ water consumption or any other regulation. Despite this, the simulation performance achieved by the model presented here is quite good, though the results are limited to the one catchment in Nepal. A further examination should evaluate whether the selected forcing datasets could be applied to simulated discharge at other mountain regions with longer time periods.

CONCLUSIONS

We aimed to identify the quality of discharge simulation for the Narayani catchment of Nepal based on different forcing datasets. The forcing datasets WFDEI, CORDEX-SMHI, ERA-I, and ground-based observation combined with WFDEI (i.e. Observed + WFDEI) were used to estimate the Narayani catchment discharge on a daily basis for the period of January 2000–December 2009. Because of the uniform coverage and data consistency, global forcing datasets from global and regional climate models were considered an important supplement to station data. In this study, the distributed Shyft hydrologic simulation platform was selected as a discharge simulation tool. The forcing data were interpolated using IDW for generating daily observed precipitation, while a gradient-based Kriging method was used for generating temperature fields matching the model resolution. In this study, Shyft was calibrated for the period of 2000–2004 and validation was done from 2005 to 2009.

Our analysis showed that large differences exist between different forcing datasets particularly in the amount of precipitation. Precipitation from ERA-I and CORDEX-SMHI was unrealistic leading to poor model performance. To improve the existing bias in precipitation particularly over the high mountain regions, further validation and algorithm improvements are required. With the application of a precipitation correction factor (p_corr_factor), a significant amount of bias in precipitation could be mitigated. Still, the performance of the bias-corrected forcing data, i.e. WFDEI, was better than the climate model datasets. Comparing model performance during model calibration and validation periods, relatively higher NSE, NSEsqrt, and KGE with lower $D_v$ were found for the Observed + WFDEI and WFDEI forcing datasets. The water balance analysis shows that higher evapotranspiration with large glacier melt is observed for the ERA-I forcing dataset, while the average evapotranspiration calculated from the WFDEI (24.79%) is similar to previous studies.

Therefore, based on the different results from the Shyft, we conclude that, in the data-poor Himalayan catchment, the WFDEI forcing dataset may be the best choice for water resource planning and hydropower inflow calculations. Discharge simulations resulting from the WFDEI forcing data were particularly promising for hydropower estimation and water resource assessment in data-scarce or ungauged regions. CORDEX-SMHI data captured higher peaks and may be suitable for the peak discharge analysis. However, to use CORDEX-SMHI and ERA-I in ungauged catchments for the water balance analysis, bias correction is required. Further analysis by implementing Shyft in the different Himalayan catchments for different forcing
datasets is a recommended further step to assess the regional extensibility of the current results.

**FUNDING**

This work is a contribution to the Strategic Research Initiative ‘Land Atmosphere Interaction in Cold Environments’ (LATICE) of the University of Oslo and partially supported through the Norwegian Research Council’s INDO program under the Hydrologic sensitivity to Cryosphere-Aerosol interaction in Mountain Processes (HyCAMP) project (NRF No. 222195)

**ACKNOWLEDGEMENT**

We are thankful to the Department of Hydrology and Meteorology, Government of Nepal for providing observed hydrological and meteorological datasets. We also gratefully acknowledge Special Issue Editor Dr. Kolbjørn Engeland and anonymous reviewer for their comments on the research and manuscript.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this paper is available online at https://dx.doi.org/10.2166/nh.2020.079.

**REFERENCES**


System for automated geoscientific analyses (SAGA) v. 2.1.4. Geoscientific Model Development Discussions 8 (2), 2271–2312.


Ghimire, S., Choudhary, A. & Dimri, A. P. 2018 Assessment of the performance of CORDEX-South Asia experiments for monsoon precipitation over the Himalayan region during present climate: part I. Climate Dynamics 50 (7), 2311–2334.


Immerzeel, W. W., van Beek, L. P. H. & Bierkens, M. F. P. 2010 Climate change will affect the Asian water towers. Science 328 (5984), 1382–1385.


Lombraña, J. U. 2017 Evaluation of Snow Simulations in SHyFT. Norwegian University of Science and Technology (NTNU), Trondheim.


