

Applicability Assessment of Two Meteorological Datasets in Data-Scarce Regions: A Case Study of the Hutubi River Basin (Postprint)

Authors: Sun Mingyue

Date: 2022-01-26T18:01:15+00:00

Abstract

In data-scarce regions, the lack of observational data for precipitation, temperature, and other variables affects the accuracy of flood forecasting, thereby impacting flood control, drought resistance, water resources planning, and related efforts. Therefore, conducting applicability analysis of existing precipitation and temperature data in such regions is essential. Using the HBV model incorporating snowmelt runoff simulation in the area above the Shimen hydrological station in the Hutubi River basin, comparative analysis and applicability analysis for snowmelt-type runoff simulation were performed on the China Land Surface Precipitation and Temperature Daily Value $0.5^{\circ}\times 0.5^{\circ}$ gridded dataset and the meteorological station observation dataset, based on comparisons between simulated snowmelt flood runoff and observed runoff. The results demonstrate that the snowmelt runoff simulation performance of the hydrological model driven by gridded data is generally superior to that driven by station observation data. The Nash coefficient for snowmelt runoff simulation by the gridded data-driven hydrological model is 0.792 during the validation period, whereas that for the station data-driven model is 0.433. Additionally, characteristics of snowmelt floods in the Hutubi River basin were analyzed, and possible causes of errors in simulated snowmelt runoff were examined in relation to the different driving datasets, providing support for improving the accuracy of snowmelt-type flood forecasting in data-scarce regions with snowmelt characteristics.

Full Text

Applicability Assessment of Two Meteorological Datasets in Data-Scarce Regions: A Case Study of the Hutubi River Basin

SUN Mingyue¹², LYU Haishen¹², ZHU Yonghua¹², LIN Yu¹², ZHANG Meijie¹²

¹State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, Jiangsu, China

²College of Hydrology and Water Resources, Hohai University, Nanjing 210098, Jiangsu, China

Abstract

In ungauged regions, the lack of observational data for precipitation and temperature affects the accuracy of flood forecasting, which in turn impacts flood control, drought relief, and water resources planning. Therefore, analyzing the applicability of existing precipitation and temperature data in ungauged areas is essential. This study applied the HBV (Hydrologiska Byråns Vattenbalansavdelning) model, which includes snowmelt runoff simulation, to the area upstream of the Shimen hydrological station in the Hutubi River Basin. Based on a comparison between simulated snowmelt flood runoff and measured runoff, we conducted a comparative analysis of two datasets: the China Meteorological Administration's gridded daily precipitation and temperature dataset and observational data from meteorological stations. The applicability of these datasets for snowmelt-type runoff simulation was evaluated.

The results demonstrate that the hydrological model driven by gridded data generally outperformed the model driven by station observation data in simulating snowmelt runoff. During the validation period, the Nash-Sutcliffe efficiency coefficient for snowmelt runoff simulation using gridded data was 0.792, compared to 0.433 when using station data. The study also analyzed characteristics of snowmelt floods in the Hutubi River Basin and examined potential causes of simulation errors in relation to the driving datasets. These findings provide support for improving the accuracy of snowmelt flood forecasting in ungauged regions with snowmelt characteristics.

Keywords: precipitation data; ungauged region; snowmelt flood; runoff simulation; applicability assessment; Hutubi River Basin

1. Materials and Methods

1.1 Study Area The study area comprises the region upstream of the Shimen hydrological station in the Hutubi River Basin (Fig. 1). Located on the northern slope of the central Tianshan Mountains and the southern edge of the Junggar

Basin, the basin lies between 43°07'–45°20' N and 86°05'–87°08' E. The catchment area above the Shimen station is 1840 km², with a river channel slope of 27.5‰ and an annual runoff volume of 3.39×10^8 m³, accounting for 43.9% of the total basin runoff. The Hutubi River exhibits highly uneven intra-annual runoff distribution, with summer concentrations. Surface runoff is primarily supplied by alpine glacier and snow melt, seasonal snowmelt, precipitation, and groundwater.

1.2 HBV Model The HBV model is a conceptual hydrological model developed by the Swedish Meteorological and Hydrological Institute in the 1970s. Its simple structure and clear mechanisms provide strong adaptability for watersheds with distinct hydrological processes. This study employed the HBV-96 daily model version. Notably, due to sparse meteorological stations in the study area and insufficient evapotranspiration data, the original method for calculating potential evapotranspiration (based on monthly mean temperature, monthly mean potential evapotranspiration, and daily mean temperature) was replaced with a modified Blaney-Criddle equation. This approach calculates modified potential evapotranspiration based on wind speed, humidity, solar radiation, and primarily daily mean temperature.

The model consists of four main modules: (1) snow accumulation and melt, (2) soil moisture and effective precipitation, (3) evapotranspiration, and (4) runoff response. The overall water balance is described by:

$$P - E - R = \frac{d(SP + SM + UZ + LZ + L)}{dt}$$

where P is precipitation, E is evapotranspiration, R is runoff, SP is snowpack water equivalent, SM is soil moisture, UZ is upper groundwater zone storage, LZ is lower groundwater zone storage, and L is lake storage.

Snow Accumulation and Melt Module: The model assumes snow accumulation and melt are controlled by the relationship between observed temperature (T) and a threshold temperature (TT). When $T < TT$, all precipitation accumulates as snow without generating runoff. When $T > TT$, snowmelt occurs, calculated using the degree-day method:

$$Sm = DD \cdot (T - TT)$$

where Sm is snowmelt amount (mm · d⁻¹), DD is the degree-day factor (mm · °C⁻¹ · d⁻¹), TT is the threshold temperature (°C), and T is daily mean temperature (°C).

Soil Moisture and Effective Precipitation Module: Based on soil moisture conditions during precipitation events, the model partitions precipitation into infiltration and surface runoff components. Effective precipitation is calculated as:

$$P_{eff} = P \cdot \left(\frac{SM}{FC} \right)^\beta$$

where P_{eff} is effective precipitation, FC is maximum soil water holding capacity, β is a shape factor, SM is actual soil moisture, and P is daily precipitation.

Evapotranspiration Module: Actual evapotranspiration (Ea) is calculated based on the relationship between actual soil moisture and the permanent wilting point (PWP):

$$Ea = ETP \cdot \min \left(1, \frac{SM}{PWP} \right)$$

where ETP is potential evapotranspiration.

Runoff Response Module: The model employs linear reservoir concepts to simulate runoff at the watershed outlet. The system comprises three virtual reservoirs: the first two simulate near-surface flow and interflow, while the third simulates baseflow. These reservoirs are interconnected by constant percolation rates, with total simulated runoff obtained by summing the outflows from the first and third reservoirs.

The HBV model contains 13 parameters, four of which are sensitive: degree-day factor (DD), field capacity (FC), interflow recession coefficient (k_1), and baseflow recession coefficient (k_2). Parameter calibration was conducted through manual tuning to understand parameter impacts, followed by genetic algorithm optimization for sensitive parameters.

1.3 Data Sources Two meteorological datasets were employed:

Dataset 1 (Station Observations): Data from three national meteorological stations (Tianshan Xidigou, Hutubi, and Xiaoquzi) and the Shimen hydrological station. None of the 699 national meteorological stations are located within the study area; the selected stations are the nearest available (39 km, 17 km, and 45 km from the basin boundary, respectively). Daily temperature and precipitation data were averaged arithmetically across these stations.

Dataset 2 (Gridded Data): The $0.5^\circ \times 0.5^\circ$ gridded daily precipitation dataset (SURF_{CLI}{CHN}{PRE}{DAY}{GRID}0.5) and temperature dataset (SURF_{CLI}{CHN}{TEM}{DAY}{GRID}0.5) from the China Meteorological Administration, based on high-density national stations and interpolated using the thin-plate spline method. Nine grid cells cover the study area (Fig. 1), with precipitation and temperature data area-weighted according to the proportion of the watershed within each cell.

Hydrological data (runoff, temperature, and precipitation) for the Shimen station were obtained from the Hutubi County Hydrological Bureau, covering 1990–2015.

1.4 Statistical Metrics Model performance was evaluated using the Nash-Sutcliffe Efficiency (NSE) coefficient and Mean Relative Error (MRE):

$$NSE = 1 - \frac{\sum(Q_{obs} - Q_{sim})^2}{\sum(Q_{obs} - \bar{Q}_{obs})^2}$$

$$MRE = \frac{\sum(Q_{obs} - Q_{sim})}{\bar{Q}_{obs}} \times 100\%$$

where Q_{obs} is observed discharge, Q_{sim} is simulated discharge, and \bar{Q}_{obs} is the mean observed discharge. NSE ranges from $-\infty$ to 1, with values closer to 1 indicating better fit. MRE assesses overall water balance, with values closer to 0 indicating better performance.

2. Results and Analysis

2.1 Comparison of Station and Gridded Data Temperature comparisons reveal that station data are consistently higher than gridded data across all timescales. At the daily scale, the maximum station temperature was 24.3°C (July 13, 2004) versus 17.5°C for gridded data (July 29, 2008), while minimum temperatures were -26.6°C and -24.8°C, respectively. Seasonally, summer shows the largest discrepancy, with station data averaging 16.1°C compared to 9.7°C for gridded data—a difference of 6.4°C. Spring and autumn differences are 4.4°C and 4.6°C, respectively, while winter shows the smallest difference (3.8°C). The annual mean temperature from station data (5.2°C) exceeds gridded data (1.7°C) by 3.5°C.

Precipitation patterns show gridded data generally exceeding station data. At the annual scale, station data average 494.2 mm compared to 526.2 mm for gridded data. Daily maximum precipitation also differs significantly: in 1999, station data recorded 36.5 mm while gridded data reached 65.0 mm. Seasonally, spring and summer show the largest differences, with gridded data averaging 1.2 mm and 2.4 mm daily compared to station data at 0.7 mm and 1.5 mm, respectively. Spatially, station precipitation increases with elevation within the study area, while gridded precipitation shows less clear elevation-dependent patterns.

2.2 Snowmelt Flood Runoff Simulation A fused dataset was created by arithmetically averaging station and gridded temperature and precipitation data for model parameter calibration. Using calibrated parameters, separate simulations were conducted with station and gridded datasets.

Calibration Period (1990–2005): The fused dataset yielded $NSE = 0.814$ and $MRE = 1.63\%$. Simulated summer flood peaks were generally underestimated, while spring snowmelt runoff was overestimated in some years (e.g., 1993, 1999, 2005). Winter simulations matched observations well in some years but were generally underestimated.

Validation Period (2006-2015): Gridded data significantly outperformed station data (Table 2). The gridded dataset achieved $NSE = 0.792$ and $MRE = 2.44\%$, while station data yielded $NSE = 0.433$ and $MRE = 14.64\%$. Gridded data better captured flood timing and magnitude, particularly for peak flows (Fig. 3). Station data consistently underestimated summer peaks, likely due to lower precipitation measurements at the selected stations. For example, during peak events on July 31, 2007 and August 1, 2010, station precipitation was 11.4 mm and 10.0 mm, respectively, compared to gridded precipitation of 21.0 mm and 11.4 mm. Conversely, station data overestimated spring flows due to higher temperature readings, causing excessive snowmelt calculations when temperatures first exceeded the melt threshold.

Winter simulations using both datasets underestimated low flows ($1.8\text{--}3.8 \text{ m}^3 \cdot \text{s}^{-1}$), with station data performing slightly worse. However, when observed winter flows were around $1.5\text{--}2.5 \text{ m}^3 \cdot \text{s}^{-1}$, gridded data simulations showed better agreement.

Table 1. HBV Model Parameters: Ranges and Calibrated Values

Parameter	Description	Range	Calibrated Value
TT ($^{\circ}\text{C}$)	Threshold temperature	-1.0-2.7	0.5
DD ($\text{mm} \cdot ^{\circ}\text{C}^{-1} \cdot \text{d}^{-1}$)	Degree-day factor	0.5-4.0	2.5
FC (mm)	Maximum soil water holding capacity	100-450	350
PWP (mm)	Soil moisture at potential evapotranspiration	90-180	150
k_0 (d^{-1})	Surface runoff recession coefficient	0.05-0.2	0.15
k_1 (d^{-1})	Interflow recession coefficient	0.01-0.2	0.08
k_2 (d^{-1})	Baseflow recession coefficient	0.001-0.1	0.05
$PERC$ ($\text{mm} \cdot \text{d}^{-1}$)	Percolation rate	0.01-0.05	0.03

Table 2. Flow Simulation Results Using Different Driving Datasets

Dataset	Calibration NSE	Calibration MRE (%)	Validation NSE	Validation MRE (%)
Fused Data	0.814	1.63	—	—
Station Data	—	—	0.433	14.64
Gridded Data	—	—	0.792	2.44

2.3 Snowmelt Flood Characteristics Analysis The largest floods occur in summer with rapid rising limbs and steep peaks. Daily mean discharge often increases abruptly—for example, on July 31, 2007, flow rose from $60.7 \text{ m}^3 \cdot \text{s}^{-1}$ to $103 \text{ m}^3 \cdot \text{s}^{-1}$ within one day, and on August 1, 2010, from $111 \text{ m}^3 \cdot \text{s}^{-1}$ to $171 \text{ m}^3 \cdot \text{s}^{-1}$. The 25-year average summer peak discharge at Shimen station is $116 \text{ m}^3 \cdot \text{s}^{-1}$. Both datasets underestimated these peaks, with station data showing greater bias.

Spring discharge gradually increases from winter baseflow ($1.8\text{--}3.8 \text{ m}^3 \cdot \text{s}^{-1}$) due to rising temperatures and snowmelt. Small spring peaks ($20\text{--}35 \text{ m}^3 \cdot \text{s}^{-1}$) occur in some years (e.g., 1993, 1999, 2005), with a long-term spring average of $5.66 \text{ m}^3 \cdot \text{s}^{-1}$. Station data consistently overestimated spring flows due to temperature biases.

Autumn discharge reflects recession from summer floods, gradually declining with some fluctuations. The long-term autumn average is $10.5 \text{ m}^3 \cdot \text{s}^{-1}$. Winter flows remain low ($1.8\text{--}3.8 \text{ m}^3 \cdot \text{s}^{-1}$) with minimal variation, averaging $2.80 \text{ m}^3 \cdot \text{s}^{-1}$.

Error sources primarily relate to data representativeness. The three meteorological stations and one hydrological station used for station data are located outside the study area. In mountainous regions, precipitation exhibits strong spatial variability due to orographic effects, and temperature shows significant vertical gradients. Consequently, these external stations poorly represent conditions within the watershed, particularly underestimating summer precipitation and overestimating spring temperatures. While gridded data, interpolated from high-density national stations, performed better, both datasets require local corrections based on elevation, slope, and aspect to further reduce simulation errors.

3. Conclusions

- 1) The hydrological model driven by gridded meteorological data demonstrated superior performance in snowmelt runoff simulation compared to station observation data. During the validation period, the gridded data achieved a Nash-Sutcliffe efficiency coefficient of 0.792, substantially higher than the 0.433 obtained with station data.

- 2) Analysis of flood characteristics revealed that the largest floods occur in summer with rapid rising limbs, while spring snowmelt produces smaller, more gradual peaks. Simulation errors arise primarily from data representativeness issues in this mountainous, data-scarce region. Station data underestimate summer precipitation and overestimate spring temperatures, while gridded data provide better overall representation but still require refinement.
- 3) For data-scarce mountainous watersheds like the Hutubi River Basin, gridded precipitation and temperature datasets demonstrate stronger applicability for snowmelt flood simulation than sparse station observations. Future work should focus on developing correction methods that account for local topographic factors (elevation, slope, aspect) to further improve simulation accuracy.

References

- [1] Abudu S, Cui C L, Saydi M, et al. Application of snowmelt runoff model (SRM) in mountainous watersheds: A review[J]. *Water Science and Engineering*, 2012, 5(2): 123-136.
- [2] Xu Kaili, Lyu Haishen, Liu Mingwen, et al. Numerical simulation of the ice jam stage in the Sanhuhekou bend reach of the Yellow River[J]. *Arid Zone Research*, 2021, 38(6): 1556-1562.
- [3] Qiu L, You J, Qiao F, et al. Simulation of snowmelt runoff in ungauged basins based on MODIS: A case study in the Lhasa River basin[J]. *Stochastic Environmental Research and Risk Assessment*, 2014, 28(6): 1577-1585.
- [4] Tang Rong, Wang Yuntao, Li Min, et al. Suitability evaluation of gridded precipitation dataset for the upstream of Nierji Reservoir[J]. *Journal of Water Resources and Water Engineering*, 2019, 30(5): 26-31, 9.
- [5] Han Chuntan, Wang Lei, Chen Rensheng, et al. Precipitation observation network and its data application in the alpine region of Qilian Mountains[J]. *Resources Science*, 2020, 42(10): 1987-1997.
- [6] Lu Aigang, Kang Shichang, Pang Deqian, et al. Different landform effects on seasonal temperature patterns in China[J]. *Ecology and Environment*, 2008, 17(4): 1450-1452.
- [7] Liu Junfeng, Chen Rensheng, Qin Wenwu, et al. Study on the vertical distribution of precipitation in mountainous regions using TRMM data[J]. *Advances in Water Science*, 2011, 22(4): 447-454.
- [8] Chen Rensheng, Kang Ersi, Ding Yongjian. Some knowledge on and parameters of China alpine hydrology[J]. *Advances in Water Science*, 2014, 25(3): 307-317.
- [9] Tuo Y, Duan Z, Disse M, et al. Evaluation of precipitation input for SWAT

- modeling in Alpine catchment: A case study in the Adige River Basin (Italy)[J]. *Science of the Total Environment*, 2016, 573: 66-82.
- [10] Zhang Junlan, Luo Ji, Wang Rongmei. Combined analysis of the spatiotemporal variations in snowmelt (ice) flood frequency in Xinjiang over 20 years and atmospheric circulation patterns[J]. *Arid Zone Research*, 2021, 38(2): 339-350.
- [11] Shen Y, Xiong A. Validation and comparison of a new gauge-based precipitation analysis over mainland China[J]. *International Journal of Climatology*, 2016, 36(1): 252-265.
- [12] Gao Hongkai, He Xiaobo, Ye Baisheng, et al. The simulation of HBV hydrology model in the Dongkemadi River Basin, headwater of the Yangtze River[J]. *Journal of Glaciology and Geocryology*, 2011, 33(1): 171-181.
- [13] Geng Junling, Gao Ling, Chen Jianjiang, et al. Analysis on the hydrological characteristics in the Hutubi River Basin, Xinjiang[J]. *Arid Zone Research*, 2005, 22(3): 371-376.
- [14] Li Yan. Change of river flood and disaster in Xinjiang during past 40 years[J]. *Journal of Glaciology and Geocryology*, 2003, 25(3): 342-346.
- [15] Kumar M, Hodnebrog Ø, Daloz A S, et al. Measuring precipitation in Eastern Himalaya: Ground validation of eleven satellite, model and gauge interpolated gridded products[J]. *Journal of Hydrology*, 2021, 599: 126252.
- [16] Gao Z, Long D, Tang G, et al. Assessing the potential of satellite-based precipitation estimates for flood frequency analysis in ungauged or poorly gauged tributaries of China's Yangtze River basin[J]. *Journal of Hydrology*, 2017, 550: 478-496.
- [17] Dembélé M, Schaeffli B, Van De Giesen N, et al. Suitability of 17 gridded rainfall and temperature datasets for large-scale hydrological modelling in West Africa[J]. *Hydrology and Earth System Sciences*, 2020, 24(11): 5379-5406.
- [18] Mazzoleni M, Brandimarte L, Amaranto A. Evaluating precipitation datasets for large-scale distributed hydrological modelling[J]. *Journal of Hydrology*, 2019, 578: 124076.
- [19] Zhang J, Li Y, Huang G, et al. Evaluation of uncertainties in input data and parameters of a hydrological model using a Bayesian framework: A case study of a snowmelt precipitation driven watershed[J]. *Journal of Hydrometeorology*, 2016, 17(8): 2333-2350.
- [20] Aghakouchak A, Habib E. Application of a conceptual hydrologic model in teaching hydrologic processes[J]. *International Journal of Engineering Education*, 2010, 26: 963-973.
- [21] Lyu H, Crow W T, Zhu Y, et al. The impact of assumed error variances on surface soil moisture and snow depth hydrologic data assimilation[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(11): 5116-5129.

[22] Chen Tingxing, Lyu Haishen, Zhu Yonghua. Analysis of flood characteristics in Xiying River Basin based on GEV distribution[J]. Arid Zone Research, 2021, 38(6): 1563-1569.

[23] Bergström S. Experience from applications of the HBV hydrological model from the perspective of prediction in ungauged basins[J]. IAHS, 2006, 307(1): 97-109.

[24] Jin Haoyu, Ju Qin, Xie Jiayao. The application of HBV model in the Nyang River Basin[J]. China Rural Water and Hydropower, 2019(6): 23-28, 34.

Figures

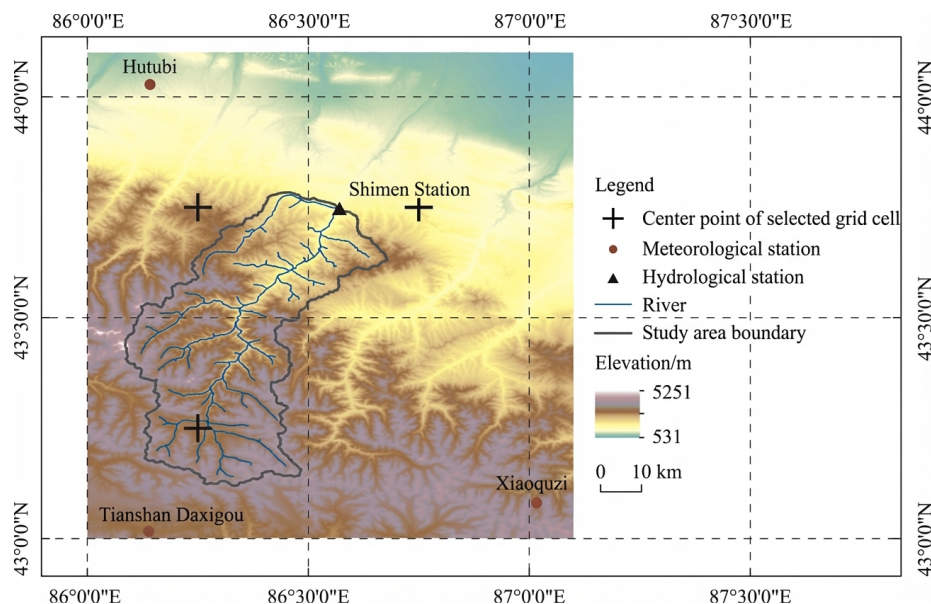


Figure 1: Figure 1

Source: ChinaXiv –Machine translation. Verify with original.