English-access research papers from ChinaXiv

A hybrid data-driven model integrating hydro-meteorological factors for snowmelt flood early warning in an arid mountain basin

WANG Zerui, LI Xiaoyang, LIU Yongqiang, WANG Weiping, LI Yaqian, ZHANG Yuanwei

2026-08-12

ChinaRxiv

View the original and related papers at https://chinarxiv.org/items/chinaxiv-202608.00023

Source: ChinaXiv. Machine translation. Verify with the original.

ChinaRxiv 中X

Abstract

High false alarm rates (FARs) in snowmelt flood forecasting persist, largely due to an insufficient understanding of the coupled effects of multi-source hydro-meteorological drivers and their inherent time lags. This study addressed this gap by developing a hybrid snowmelt-flood forecasting framework that combined multiple linear regression (MLR) and back propagation neural network (BPNN) models, with a simulated annealing (SA) algorithm employed to optimize the ensemble weighting. The developed hybrid model that incorporated daily hydro-meteorological inputs and temporal factor was validated using data from 1978 to 2011 for the Hutubi River Basin, an arid inland basin on the northern slope of the Tianshan Mountains, China. The findings demonstrated that the hybrid model achieved a specificity of 0.8549, significantly outperforming standalone MLR (0.6024) and BPNN (0.6436) models. Correspondingly, the FAR was reduced to 0.1451, which was substantially lower than that of MLR (0.3976) and BPNN (0.3564). Upon integrating temporal factor, the FAR was further reduced to 0.0959, markedly enhancing overall predictive robustness. Collectively, this study offers a robust methodological framework for optimizing snowmelt flood forecasting by effectively integrating multi-source data and temporal dependencies.

Full Text

J Arid Land (2026) 18(8): 1331–1353
doi: 10.1016/j.jaridl.2026.08.003; CSTR: 32276.14.JAL.20250514

A hybrid data-driven model integrating hydro-meteorological factors for snowmelt flood early warning in an arid mountain basin

WANG Zerui[1,2], LI Xiaoyang[1,2], LIU Yongqiang^1,2*^, WANG Weiping[1,2], LI Yaqian[1,2], ZHANG Yuanwei[1,2]

[1]College of Geography and Remote Sensing Sciences, Xinjiang University, Urumqi 830017, China;
[2]Xinjiang Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830017, China

Abstract: High false alarm rates (FARs) in snowmelt flood forecasting persist, largely due to an insufficient understanding of the coupled effects of multi-source hydro-meteorological drivers and their inherent time lags. This study addressed this gap by developing a hybrid snowmelt-flood forecasting framework that combined multiple linear regression (MLR) and backpropagation neural network (BPNN) models, with a simulated annealing (SA) algorithm employed to optimize the ensemble weighting. The developed hybrid model that incorporated daily hydro-meteorological inputs and temporal factor was validated using data from 1978 to 2011 for the Hutubi River Basin, an arid inland basin on the northern slope of the Tianshan Mountains, China. The findings demonstrated that the hybrid model achieved a specificity of 0.8549, significantly outperforming standalone MLR (0.6024) and BPNN (0.6436) models. Correspondingly, the FAR was reduced to 0.1451, which was substantially lower than that of MLR (0.3976) and BPNN (0.3564). Upon integrating temporal factor, the FAR was further reduced to 0.0959, markedly enhancing overall predictive robustness. Collectively, this study offers a robust methodological framework for optimizing snowmelt flood forecasting by effectively integrating multi-source data and temporal dependencies.

Keywords: snowmelt flood; multiple linear regression (MLR); backpropagation neural network (BPNN); simulated annealing (SA); Hutubi River Basin

Citation: WANG Zerui, LI Xiaoyang, LIU Yongqiang, WANG Weiping, LI Yaqian, ZHANG Yuanwei. 2026. A hybrid data-driven model integrating hydro-meteorological factors for snowmelt flood early warning in an arid mountain basin. Journal of Arid Land, 18(8): 1331–1353. https://doi.org/10.1016/j.jaridl.2026.08.003; https://cstr.cn/32276.14.JAL.20250514

1 Introduction

Snowmelt floods pose a significant global risk, causing an estimated 7.30×10^{11} USD in economic losses between 1900 and 2020 (Cui et al., 2022). Developing effective forecasting models is therefore critical for mitigating these socio-economic impacts. The mechanisms driving snowmelt floods are complex, involving interactions between meteorology, topography, soil, and vegetation (Stein et al., 2021). Temperature, precipitation, and snowpack depth are the most sensitive and direct drivers (Bean and Watts, 2024), while in small basins, terrain and soil characteristics amplify the flood response (Fileni et al., 2025). Consequently, reliable early-warning models for snowmelt

^*^Corresponding author: LIU Yongqiang (E-mail: liuyq@xju.edu.cn)
Received 2025-10-14; revised 2026-04-22; accepted 2026-04-29
© 2026 Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, and Science Press. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

http://jal.xjegi.com; https://www.keaipublishing.com/en/journals/journal-of-arid-land/

This version posted 2026-08-12.

floods must be tailored to the specific hydrological conditions of each basin.

Snowmelt-induced and rainfall-induced runoff processes exhibit distinct hydrological signatures in their generation mechanisms, response times, and seasonal timing. Snowmelt is controlled primarily by rising temperatures and solar radiation, which convert snowpack into liquid water discharged into the drainage network over a concentrated period, typically in spring or early summer. Rainfall-induced runoff, in contrast, originates from direct liquid precipitation, initiating a more immediate surface response. The two phenomena differ fundamentally in water source, formation mechanism, and temporal scale (Latif et al., 2020). When rainfall occurs on existing snowpack, a Rain-on-Snow (RoS) event can form. Under these scenarios, rainfall supplements streamflow and accelerates snowmelt by enhancing heat input and energy exchange within the snowpack, triggering larger and faster flood peaks (Hotovy et al., 2025).

The Hutubi River Basin, the focus of this study, is located in a typical arid region on the northern slope of the Tianshan Mountains, China. Its streamflow is primarily driven by mountain snowmelt, with liquid precipitation contributing minimally to the annual water balance. The flood regime is dominated by spring snowmelt events, which are characterized by concentrated runoff, steeply rising flood peaks, and high frequency due to large snow reserves, rapid spring temperature increases, and significant topographic relief. In certain years, the superposition of early spring rainfall on snowmelt amplifies flood risks (Zhao et al., 2020). Systematic research in such a basin is scientifically important for understanding flood formation mechanisms and enhancing regional flood management capabilities.

Snowmelt flood forecasting frameworks are generally categorized into process-based hydrological models and data-driven machine learning approaches, most of which have been validated in medium- to large-scale basins (Agnihotri and Coulibaly, 2020; Venegas-Cordero et al., 2023). Widely used hydrological approaches include the Variable Infiltration Capacity (VIC-CAS), Soil and Water Assessment Tool (SWAT), and Nedbør-Afledt Modstand (NAM) models (Xie et al., 2007; Teshome et al., 2020; Koltsida et al., 2023). The VIC-CAS model, a large-scale distributed framework employing Soil-Vegetation-Atmosphere Transfer Schemes (SVATS), is extensively applied in western China's cryosphere (Ding and Zhang, 2018). The SWAT model integrates land use, soil type, and meteorological data to improve hydrological cycle simulations and forecast accuracy (Janjić and Tadić, 2023). The NAM model further refines this by using instantaneous unit hydrograph theory to simulate rainfall-runoff processes with high precision (Tran Tuan, 2024).

Alongside process-based models, machine learning technologies have substantially advanced predictive capabilities for snowmelt runoff and flood events (Atashi et al., 2023a; Mirboluki et al., 2024). Approaches including Long Short-Term Memory (LSTM) networks, Support Vector Regression (SVR), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) have gained prominence for their capacity to map complex nonlinearities and temporal dependencies in snow-dominated systems (Adamowski and Prasher, 2012; Kratzert et al., 2018). LSTM-based models are particularly effective for modeling snowmelt dynamics, as they can explicitly represent the temporal memory and lag responses between temperature, snow accumulation, and runoff. For instance, Yu et al. (2023) demonstrated that integrating snowmelt temporal information into LSTM frameworks improved flood estimation, specifically by correcting peak discharge underestimation. Concurrently, tree-based ensemble methods like RF and XGBoost have proven highly applicable, offering robustness to multicollinearity and adept handling of complex variable interactions. Studies by Aryal et al. (2025) and Santos et al. (2025) confirmed their competitive accuracy, even with limited hydro-meteorological data in alpine arid regions. More recently, hybrid frameworks incorporating temporal awareness have gained attention. Duan et al. (2024) achieved substantial improvements in daily-scale predictions by integrating explicit temporal modules into deep learning architectures. Building on this, Thapa et al. (2024) found that Temporal Convolutional Network (TCN) models can more effectively capture seasonal variations and abrupt snowmelt events, outperforming traditional machine learning methods. These studies indicated that ML models with explicit temporal components represent a promising

direction for enhancing snowmelt flood prediction.

While data limitation and parameter applicability are universal challenges, these issues are often amplified in small watersheds (Blöschl and Sivapalan, 1995). The intricate topography and rapid hydrological responses of small basins make models highly sensitive to observational data density and parameter calibration. These factors significantly constrain the predictive performance of both hydrological and machine learning models at this scale (Argentin et al., 2025). Exploratory studies in small watersheds have been conducted. Lane et al. (2019) benchmarked lumped hydrological models across over 1000 small catchments in the UK, establishing regression-based methods, particularly multiple linear regression (MLR), as standard benchmarks for performance assessment under uncertainty (Lane et al., 2019; Yeganeh et al., 2024). Wang et al. (2017) integrated a backpropagation neural network (BPNN) with the semi-distributed Xin' anjiang model to improve flood forecasting in small watersheds, achieving higher accuracy for real-time peak flow and runoff prediction. Overall, a single model may perform well under specific conditions, but its accuracy and false alarm rate (FAR) remain limited in complex hydro-meteorological scenarios. Hybrid models that integrate multiple methods are therefore considered a promising approach to improve overall prediction performance. Therefore, this study sought to enhance snowmelt flood prediction in a small watershed by weighing and combining different models.

As a representative snowmelt flood-prone region in Northwest China, the Hutubi River Basin presents complex flood generation mechanisms driven by temperature, precipitation, and snowmelt, which have long posed severe challenges to regional flood control. The sudden onset and difficult prediction of spring snowmelt floods, in particular, negatively impact local socio-economic development and public safety. Existing machine learning models often underperform in these environments because the stochastic and nonlinear dynamics inherent in snowmelt-induced hydrology remain inadequately represented (Jiang et al., 2022). This study took the Hutubi River Basin as a representative testbed to develop a hybrid snowmelt-flood forecasting framework, integrating MLR and BPNN models via a simulated annealing (SA) algorithm. This research not only addresses a critical need for localized flood mitigation but also provides a replicable framework for snowmelt-induced runoff prediction in other data-scarce mountainous regions.

2 Materials

2.1 Study area

The Hutubi River Basin constitutes a pivotal hydrological system on the northern slope of the Tianshan Mountains, China. The basin spans from 86°05′E to 86°40′E longitude and 43°07′N to 43°45′N latitude, covering an area of approximately 10,255 km^2 (Fig. 1). Substantial high-elevation winter snowpack, coupled with rapid vernal thermal escalation, generates voluminous snowmelt over abbreviated temporal windows. Steep upstream topography facilitates rapid runoff concentration, whereas the gently sloping plains of the middle and lower reaches possess limited channel conveyance capacity. This spatial heterogeneity in terrain and hydrology renders the basin highly susceptible to snowmelt-induced flooding during the spring warming period.

2.2 Data sources

The meteorological data (daily precipitation (mm) and daily average temperature (°C)) for this study were sourced from the Changji Meteorological Bureau. This source was selected due to its geographical proximity and shared climatic conditions with the Hutubi River Basin, as both are situated on the middle northern slope of the Tianshan Mountains and influenced by similar atmospheric circulation patterns. Hydrological data (daily streamflow (m^3/s) and daily snow depth (cm)) were obtained from the Shimen Hydrological Station on the Hutubi River. The resulting datasets span the period 1978–2011 and comprise four key variables: daily streamflow, daily precipitation, average temperature difference (°C), and snow depth change (cm). Here, average temperature difference represents the change in average temperature between consecutive days, while snow depth change quantifies the day-to-day difference in snow depth.

Fig. 1 Overview of the Hutubi River Basin based on the elevation data and distribution of its drainage network

During 1978–2011, 31 severe flood events, totaling 122 flood days, were recorded in the Hutubi River Basin (Editorial Board of China Meteorological Disaster Series, 2006). Of these, 22 events (47 flood days) were classified as snowmelt-induced floods, while the remaining 9 events (75 flood days) were attributed to

Overview of the Hutubi River Basin based on elevation data and distribution of its drainage network. The map marks the Shimen Hydrological Station, drainage network, and elevation range from 1274 to 5253 m.

Figure 1: Overview of the Hutubi River Basin based on elevation data and distribution of its drainage network. The map marks the Shimen Hydrological Station, drainage network, and elevation range from 1274 to 5253 m.

rainfall. Snowmelt-induced floods occurred predominantly in April, corresponding with the seasonal snowpack ablation period confirmed by observed snow depth data. In contrast, rainfall-induced floods occurred primarily in July and August (Fig. 2). This historical record underscores the necessity of developing snowmelt-induced flood models specifically adapted to the hydrological regime of the Hutubi River Basin.

Time series plots of the primary hydro-meteorological elements from 1978 to 2011 illustrated their respective data ranges and interannual fluctuations (Fig. 3). Table 1 presents descriptive statistics (maximum, minimum, median, mean, and standard deviation) that summarize the dynamic characteristics of these hydro-meteorological elements.

3 Methods

3.1 Analytical framework

The overall workflow of this study was as follows: feature selection, feature preprocessing, hybrid model development, and model comparison and improvement. The feature selection stage involved the acquisition of raw data—daily streamflow, daily precipitation, average temperature difference, and snow depth change. For feature preprocessing, all raw inputs were standardized to eliminate dimensional disparities. Two- and three-dimensional distribution analyses were performed to visualize the relationships between these hydro-meteorological variables. For hybrid model development, we employed an SA algorithm in a weight-optimization framework to fuse the predictive outputs of MLR and BPNN models. For model comparison and improvement, we evaluated the effectiveness of the model through two comparative experiments: a comparison of model confusion matrices, and an assessment of model performance with and without the inclusion of a temporal factor.

Fig. 2 Monthly distribution and cumulative characteristics of snowmelt-induced and rainfall-induced floods. (a), monthly distribution of flood events; (b), monthly variation of flood days; (c), cumulative distribution of flood days.

3.2 Feature normalization and scaling

Prior to model development, we normalized all four hydro-meteorological variables using methods tailored to their specific distributional characteristics.

Fig. 2. Monthly distribution and cumulative characteristics of snowmelt-induced and rainfall-induced floods. (a), monthly distribution of flood events; (b), monthly variation of flood days; (c), cumulative distribution of flood days.

Figure 2: Fig. 2. Monthly distribution and cumulative characteristics of snowmelt-induced and rainfall-induced floods. (a), monthly distribution of flood events; (b), monthly variation of flood days; (c), cumulative distribution of flood days.

Fig. 3. Temporal variations in daily streamflow (a), precipitation (b), average temperature (c), and snow depth (d) in the Hutubi River Basin from 1978 to 2011

Figure 3: Fig. 3. Temporal variations in daily streamflow (a), precipitation (b), average temperature (c), and snow depth (d) in the Hutubi River Basin from 1978 to 2011

Daily streamflow data were transformed using the Box-Cox method to mitigate the positive skewness caused by extreme flood values. The transformation is defined as:

$$i_1^{(\lambda)} = \begin{cases} \dfrac{i_1^{(\lambda)} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln\left(i_1^{(\lambda)}\right) & \text{if } \lambda = 0 \end{cases}, \tag{1}$$

where $i_1^{(\lambda)}$ indicates the normalized daily streamflow; and λ is a transformation parameter optimized via maximum likelihood estimation to drive the post-transformation skewness toward

Fig. 3 Temporal variations in daily streamflow (a), precipitation (b), average temperature (c), and snow depth (d) in the Hutubi River Basin from 1978 to 2011

Table 1 Statistical characteristics of daily hydro-meteorological elements during 1978–2011

Element	Maximum	Minimum	Median	Mean	Standard deviation
Streamflow (m^3/s)	259.000	0.530	5.550	15.252	19.533
Precipitation (mm)	57.200	0.000	9.604	1.218	3.526
Average temperature (°C)	31.700	−27.900	7.700	6.232	12.177
Snow depth (cm)	45.665	0.000	0.679	4.581	8.024

zero. All input values of daily streamflow were required to be strictly positive.

Daily precipitation was normalized using min–max scaling, which projects the daily precipitation values onto the interval $[0, 1]$. The transformation is given by:

$$i_2^* = \frac{i_2}{\max(i_2)}, \qquad (2)$$

where i_2^* denotes the normalized daily precipitation; and $\max(i_2)$ is the historical maximum daily precipitation (mm). Days without precipitation were assigned a value of zero.

For the average temperature difference, we applied a symmetric min–max scaling to preserve sign information, ensuring a symmetric range of $[-1, 1]$. The normalization formula is:

$$i_3^* = 2\left(\frac{i_3 - \min(i_3)}{\max(i_3) - \min(i_3)}\right) - 1, \qquad (3)$$

where i_3^* denotes the normalized average temperature difference; $\max(i_3)$ indicates the maximum absolute average temperature difference (°C); and $\min(i_3)$ adopts the negative of the maximum absolute temperature difference (°C).

The snow depth change was normalized using a non-negative min-max scaling, formulated as:

$$i_4^* = \frac{i_4}{\max(i_4)}, \qquad (4)$$

where i_4^* denotes the normalized snow depth change; $\max(i_4)$ represents the historical maximum snow depth change.

3.3 Snowmelt flood forecasting and warning models

3.3.1 Assessment of predictive variables

This study conducted a Grey Relational Analysis (GRA) to validate the selection of the four hydro-meteorological variables. The GRA analysis quantified the degree of association between each variable and the binary flood-occurrence sequence, providing a quantitative basis for their inclusion as predictors.

The analysis began by defining a reference sequence P_m, where a snowmelt flood occurrence was coded as 1 and a non-occurrence as 0. The absolute deviation Δ_{mn} between each variable series x'_{mn} and this reference sequence was then computed:

$$\Delta_{mn} = \left| x'_{mn} - P_m \right|, \qquad (5)$$

where m denotes the number of observations (time series points); and n denotes the index of the hydro-meteorological variable.

Subsequently, the grey relational coefficients ξ_{mn} between each hydro-meteorological variable and the snowmelt-flood occurrence probability were computed as:

$$\xi_{mn} = \frac{\min_m \times \min_n \times \Delta_{mn} + \rho \times \max_m \times \max_n \times \Delta_{mn}}{\Delta_{mn} + \rho \times \max_m \times \max_n \times \Delta_{mn}}, \qquad (6)$$

where $\min_m$ and $\min_n$ indicate the global minimum of the absolute deviation Δ_{mn} across all time points m and all hydro-meteorological variables n, respectively, representing the minimum found within the entire matrix of deviations; $\max_m$ and $\max_n$ indicate the global maximum of the absolute deviation Δ_{mn} across all time points m and all hydro-meteorological variables n, respectively, representing the maximum found within the entire matrix of deviations; and ρ is the distinguishing coefficient, conventionally set to 0.5.

3.3.2 MLR model

This study developed a MLR model to establish the relationship between the four selected hydro-meteorological variables and the probability of snowmelt flood occurrence. Prior to model fitting, we assessed multicollinearity among the independent variables using the Variance Inflation Factor (VIF). A VIF value exceeding 10 is indicative of severe multicollinearity, which can lead to significant and non-negligible errors (O' brien, 2007). The VIF is calculated as:

$$\mathrm{VIF}_K = \frac{1}{1 - R^2_{1-3/K}}, \qquad (7)$$

where K indicates the K^{th} independent variable; and $R^2_{1-3/K}$ is the coefficient of determination obtained by regressing the K^{th} independent variable against the remaining three.

With the probability of snowmelt flood occurrence (P_f) as the dependent variable and the four hydro-meteorological variables (i_1-i_4) as independent variables, the MLR model is specified as:

$$P_f = \beta_0 + \beta_1 i_1 + \beta_2 i_2 + \beta_3 i_3 + \beta_4 i_4 + \mu, \qquad (8)$$

where β_0 is the intercept term; β_1-β_4 are the regression coefficients; and μ denotes the stochastic error term. The coefficients were estimated using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared differences between observed and predicted values.

3.3.3 BPNN model

The BPNN, a foundational machine learning model capable of capturing complex nonlinear relationships, was employed for flood prediction (Zhou et al., 2023). The BPNN' s multilayer architecture enables the abstraction and integration of input features, allowing it to effectively model the complex mechanisms of snowmelt flood processes. Accordingly, a BPNN-based snowmelt flood prediction model was constructed for the Hutubi River, with its architecture depicted in Figure 4.

Fig. 4 Architecture of the backpropagation neural network (BPNN)-based snowmelt flood forecasting model. i_1-i_4 indicate the input features (indicating normalized daily streamflow, daily precipitation, average temperature difference, and snow depth change, respectively); h_{i1}-h_{i4} indicate the hidden layer features; h'_{i1}-h'_{i4} indicate the cross-feature coupled hidden features; α_1-α_4 indicate the temporally extracted features; h_{fusion} indicates the fused feature vector; ReLU, Rectified Linear Unit; tanh, hyperbolic tangent; P_f, predicted probability of snowmelt flood occurrence.

The BPNN model constructed for this study comprised an input layer, three functional hidden layers, and an output layer. The input layer consisted of four nodes corresponding to daily streamflow, daily precipitation, average temperature difference, and snow depth change. Input data were standardized, and missing values were imputed using the median. The hidden layers included a feature-channel processing layer, a cross-feature coupling layer, and a temporal-feature extraction layer, all of which used the Rectified Linear Unit (ReLU) activation function to model nonlinear relationships and time-series features. The output layer predicted the flood probability, which was then classified against a predefined threshold. The model was trained using the Adam optimizer with a learning rate of 0.001 for a maximum of 1000 epochs. Overfitting was mitigated through Xavier weight initialization, early stopping (triggered if validation loss failed to improve for 50 consecutive epochs), and feature standardization.

To ensure both predictive accuracy and physical consistency, a composite loss function L was designed:

$$L = \frac{1}{N}\sum_{l=1}^{N} w_l (i_{1,l} - \hat{i}_{1,l})^2 + \alpha\|\Theta\|_2^2 + \beta\left\|\frac{\partial \hat{i}_1}{\partial Pre} - \varphi\right\|_2^2, \tag{9}$$

where N is the number of training days; w_l is the weight for the daily observation of the l^{th} training day; $i_{1,l}$ and $\hat{i}_{1,l}$ are the observed and predicted daily streamflow (m^3/s) of the l^{th} training day, respectively; α is the L2 regularization coefficient for model parameter Θ; β is the penalty multiplier for the physical constraint term; $\hat{i}_1$ is the predicted daily streamflow (m^3/s); Pre is the daily precipitation (mm); and φ is the predefined response threshold ($\text{m}^3/(\text{s}\cdot\text{mm})$).

3.3.4 Hybrid model fusion via SA optimization

To integrate the complementary strengths of the MLR model (f_1) and BPNN model (f_2), we implemented the SA algorithm to optimize their fusion weighting coefficients.

First, we initialized the temperature T_0. Next, we randomly generated initial weights w_1 and w_2 ($w_1 + w_2 = 1$). The fused probability ($P_{\text{weighting}}$) was defined as:

$$P_{\text{weighting}} = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}. \qquad (10)$$

We applied a classification threshold of 0.1, and probabilities exceeding this value were categorized as flood events. The initial accuracy H_0 was evaluated against historical observations (2009–2011). During the optimization phase, weights (w_1' and w_2') were iteratively adjusted to generate new predictions, the new fused probability ($P'_{\text{weighting}}$) was defined as:

$$P'_{\text{weighting}} = \frac{w_1' f_1 + w_2' f_2}{w_1' + w_2'}. \qquad (11)$$

Then, we defined the model accuracy under the new fused probability $P'_{\text{weighting}}$ as H_0'. The difference between H_0' and H_0 was calculated as Δf, and the acceptance probability P_a (which is the probability of accepting a candidate state in simulated annealing) was defined as:

$$P_a = \begin{cases} 1 & \text{if } \Delta f > 0 \\ \exp\left(-\dfrac{\Delta f}{T}\right) & \text{if } \Delta f < 0, \end{cases} \qquad (12)$$

where T is the temperature corresponding to the new weight combination.

After that, we iteratively reduced the temperature via a cooling schedule, with each new temperature corresponding to a new set of weights. The updated temperature satisfies:

$$T_{j+1} = cT_j, \qquad (13)$$

where T_{j+1} and T_j are the temperature values at the $(j+1)^{\text{th}}$ iteration and the j^{th} iteration, respectively; and c is the cooling coefficient. To enhance the precision of the SA search, this study set c to 0.96, thereby increasing the number of simulation iterations and facilitating a gradual convergence toward the optimal solution.

Finally, we set the termination temperature $e = 10^{-30}$. The iteration terminated when $T < e$, at which point the optimal weight combination and the corresponding maximum model accuracy were outputted.

3.3.5 Integration of temporal factor

To enhance the predictive accuracy of the SA-based fusion model, we incorporated the temporal factor to characterize the dynamics of snowmelt-induced floods. Such features are critical for capturing the onset, peak, and duration of melt-driven runoff events. Previous research has demonstrated that including temporal patterns can significantly improve hydrological model performance (Atashi et al., 2023b). For instance, Yan et al. (2023) highlighted that snowmelt timing is a key determinant in flood estimation for snow-dominated basins and that conventional rainfall-based timing models can substantially underestimate flood events. Their proposed method, which combined snowmelt timing, markedly improved estimation accuracy. Similarly, Duan et al. (2024) integrated a time module into a deep learning framework, which effectively boosted the forecast accuracy of daily-scale snow water equivalent. Furthermore, Thapa et al. (2020) confirmed that time-aware architectures like Temporal Convolutional Networks (TCNs) could effectively model snowmelt-runoff processes and outperformed conventional models.

Building upon these findings and to account for both regular cycles and irregular disturbances, the temporal index (t) was decomposed into a seasonal component (t_s) and an episodic component (t_e). Following Yan et al. (2023), a sine function was employed to capture the periodic changes throughout the year, enabling the model to represent the regular seasonality of snowmelt. For a given day of the year d and a period T_p (365 d), t_s can be calculated as:

$$t_s = \sin\left(\frac{2\pi d}{T_p}\right). \qquad (14)$$

Inspired by the time-module integration of Duan et al. (2024), we used an exponential decay function to quantify the impact of abrupt snowmelt events. It was defined as the cumulative exponentially-decayed count of consecutive days where temperature differences exceeded a predefined threshold:

$$t_e = \sum_{k=1}^{o} e^{-\lambda k}, \qquad (15)$$

where o is the total number of consecutive days where temperature differences exceeded a predefined threshold; k is the time step index representing the sequential day; and λ denotes the decay coefficient.

To integrate these dynamics into the SA fusion model, the two components were linearly combined and normalized to account for both seasonal patterns and lag characteristics:

$$t = at_s + (1-a)t_e = a \sin\left(\frac{2\pi d}{T_p}\right) + (1-a)\sum_{k=1}^{o} e^{-\lambda k}, \qquad (16)$$

where a is the weighting coefficient ($0 < a < 1$). Consequently, the augmented input vector incorporating the temporal factor was defined as:

$$\mathbf{I}' = [i_1, i_2, i_3, i_4, t]. \qquad (17)$$

This integrated temporal representation enriched the input space, allowing the model to capture complex hydrological responses.

3.4 Performance evaluation metrics

A rigorous evaluation protocol was established by partitioning the data into a training set (1978–2008) and an independent validation set (2009–2011) to assess out-of-sample performance. This study quantitatively evaluated the predictive performance of the models using several standard classification metrics derived from a confusion matrix. The four components of the matrix were: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Here, TP indicates the number of flood events correctly predicted as flood; TN indicates the number of non-flood events correctly predicted as non-flood; FP indicates the number of non-flood events incorrectly predicted as flood; and FN indicates the number of flood events incorrectly predicted as non-flood. Based on these components, we computed the following performance indicators: overall accuracy (OA), specificity, FAR, and area under the Receiver Operating Characteristic (ROC) curve (AUC). Here, OA refers to the overall correct prediction rate (including non-flood days); specificity is the proportion of actual non-flood days correctly identified; and FAR indicates the proportion of non-flood events incorrectly predicted as floods. AUC provides a comprehensive measure of a classifier' s ability to discriminate between flood and non-flood events based on the ROC curve (Wilks, 2011). The ROC curve plots the true positive rate (TPR) against the FAR across all possible classification thresholds. The relevant formulas are as follows:

$$\mathrm{OA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}, \qquad (18)$$

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}, \qquad (19)$$

$$\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}, \qquad (20)$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \qquad (21)$$

4 Results

4.1 Quantification of hydro-meteorological drivers and predictive contributions

The MLR analysis served as the basis for a probabilistic forecasting model, and the regression results clarified the relationships between different hydro-meteorological drivers and flood occurrence probability within the arid context of the Hutubi River Basin (Table 2). Daily streamflow (coefficient=0.147) and daily precipitation (coefficient=0.089) both exhibited highly significant positive effects, indicating that during the snowmelt season, direct precipitation and antecedent runoff are decisive factors in flood formation. The snow depth change also showed significant positive effect (coefficient=0.033), confirming that snow ablation directly augments downstream streamflow and thereby increases flood probability. In contrast, the average temperature difference yielded a negative coefficient (-0.021) with marginal significance, suggesting a weaker influence. This negative relationship may be attributable to the strong sunlight and large diurnal temperature differences in arid regions; while daytime warming promotes snowmelt, significant nocturnal cooling can inhibit the continuous release of meltwater, thus moderating the overall flood risk. Collectively, these findings align with the established hydro-meteorological characteristics of spring and summer snowmelt floods in the basin.

The GRA results revealed varying degrees of association between the variables and flood probability. All four variables exhibited high relational coefficients, validating their strong association with the probability of snowmelt flood occurrence and confirming their utility as reliable predictors. Notably, the grey relational coefficients for snow depth change (0.955) and daily precipitation (0.947) demonstrated the strongest correlations, identifying them as the most significant drivers of flood formation. Daily streamflow also demonstrated a strong association (0.902), underscoring the importance of streamflow recharge. In contrast, the coefficient of average temperature difference was 0.837; although it showed some correlation, the correlation was clearly lower than other indicators, suggesting its relatively weak role in explaining the occurrence of snowmelt floods. These results correspond to the low significance of the temperature difference coefficient observed in the MLR analysis.

Table 2 Standardized coefficients of hydro-meteorological predictors

Predictor	t-statistic	Significance	Coefficient
Daily streamflow	13.336	0.0001	0.147
Daily precipitation	8.124	0.0001	0.089
Average temperature difference	-1.947	0.0520	-0.021

Fig. 5

Figure 4: Fig. 5

Predictor	t-statistic	Significance	Coefficient
Snow depth change	3.073	0.0020	0.033

4.2 Visualization of normalized hydro-meteorological variables

Figure 5 illustrates the distribution characteristics and pairwise correlations of the four preprocessed hydro-meteorological variables. For instance, a strong negative correlation (with Spearman's rank correlation coefficient (ρ) of -0.77) was evident between the average temperature difference and snow depth change (Fig. 5c2 and d4), indicating that greater temperature fluctuations accelerate snow ablation. Conversely, daily streamflow exhibited a notable positive relationship with the average temperature difference (ρ=0.29; Fig. 5a2 and c4), reflecting the thermal control on meltwater generation. These visualized relationships provide a physically coherent basis for the subsequent regression analysis, confirming that these variables satisfy the distributional assumptions for statistical modeling and possess distinct correlation structures.

Fig. 5 Distribution characteristics of normalized hydro-meteorological predictors and pairwise interactive relationships of each predictor with other three predictors. (a1-a4), i_1; (b1-b4), i_2; (c1-c4), i_3; (d1-d4), i_4. Histograms with kernel density estimation (KDE) curves illustrate the data distribution of each predictor. Scatter plots display the actual normalized observations between predictors, with red Locally Weighted Scatterplot Smoothing (LOWESS) curves indicating the non-linear trend of these data points. ρ indicates the Spearman's rank correlation coefficient, which quantifies the strength and direction of the monotonic relationship between two predictors; values range from -1.00 to 1.00, where a value closer to -1.00 or 1.00 indicates a stronger negative or positive correlation, respectively.

4.3 Parameter optimization for the BPNN model

A composite-loss descent curve and a gradient-descent spiral were generated to visualize the training process and parameter evolution of the BPNN (Fig. 6). The loss value decreased rapidly during the initial training iterations (Fig. 6a), indicating that the model efficiently learned the primary data features and significantly reduced the prediction error. Subsequently, the loss curve flattened and stabilized, signifying that the model had entered a convergence phase near the optimal solution. This trajectory demonstrated the effectiveness of the training process and the robustness of the optimization strategy. Figure 6b depicts the optimization trajectory within the parameter space. The parameter updates consistently followed the negative gradient of the loss surface, confirming that

Fig. 6

Figure 5: Fig. 6

the optimization process effectively converged by descending the loss function to identify an optimal parameter configuration.

Fig. 6 Training performance and optimization trajectory. (a), loss function decay curve demonstrating model convergence efficiency; (b), contour plot of the parameter space optimization, showing the gradient descent path relative to the loss intensity distribution. The blue curve traces the iterative updates of the model parameters from their initial state, clearly showing the path toward the optimal solution. The background contours represent the loss function landscape for different parameter combinations.

4.4 Optimization of ensemble weights for hybrid models

The SA algorithm was employed to determine the optimal weight distribution between the MLR and BPNN models. The resulting optimal weights for the MLR and BPNN components were 0.466 and 0.534, respectively. The marginally higher weight assigned to the BPNN model suggested that nonlinear dynamics predominantly govern snowmelt floods in the basin. However, the comparable magnitude of the weights confirmed that a hybrid framework integrating both linear and nonlinear components is essential to fully characterize the system’ s hydrological complexity.

The determination of the optimal weight for the MLR (0.466) was based on the microscopic convergence mechanism of the SA algorithm and the actual curvature of the objective function landscape, and was validated through the joint extremum of all performance indicators. As shown in both global and localized views (Fig. 7), the objective function and Root Mean Square Error (RMSE) exhibited a global minimum precisely at the optimal weight (0.466). Meanwhile, for the indicator requiring maximization, the global peak and local apex of Nash-Sutcliffe Efficiency (NSE) also exactly lied at this point. Critically, the SA iteration trajectory (Fig. 7d1 and d2) demonstrated that after extensive, high-precision perturbation searches, the algorithm consistently converged on 0.466 as the optimal value. This process effectively distinguished the true global optimum from adjacent suboptimal points (e.g., 0.467), confirming 0.466 as a stable and computationally accurate solution.

4.5 Comparative analysis of the hybrid model and individual sub-models

The forecasting framework and its two constituent sub-models were quantitatively evaluated using overall accuracy, specificity, and FAR (Table 3). The hybrid model demonstrated substantially higher accuracy and a lower FAR

Confusion matrices for MLR, BPNN, and hybrid models for snowmelt flood classification during 2009–2011.

Figure 6: Confusion matrices for MLR, BPNN, and hybrid models for snowmelt flood classification during 2009–2011.

than individual sub-model. Notably, its FAR was 25.25% lower than that of the MLR model and 21.13% lower than that of the BPNN model. The two sub-models exhibited comparable performance, with a modest 4.12% difference in both specificity and FAR.

To further visualize these performance differences, we generated the confusion matrices for all three models for the validation period (Fig. 8). The hybrid model demonstrated the highest proficiency in positive sample identification. It achieved a TP count of 56 d, representing

Fig. 7 Optimization landscape and convergence analysis of the optimal weight for the multiple linear regression (MLR) model. (a1–d1), global distribution of the objective function landscape, RMSE, NSE, and trajectory optimization; (a2–d2), localized zoom-in views with fitted trend lines. Objective function landscape indicates the multi-objective cost function evaluated for model convergence (minimization objective); RMSE is the Root Mean Square Error, quantifying the deviation between observations and simulations (minimization objective); NSE is the Nash-Sutcliffe Efficiency, denoting hydrological predictive performance (maximization objective, optimal at 1.0); trajectory optimization denotes the iterative search path of model weights during optimization. The red star denotes the global optimum at the optimal weight of 0.466, consistently identified by all performance metrics. The red solid line denotes the polynomial regression trend-fit, representing the stochastic iteration trajectory. The large value of proximity to the global optimum indicates higher convergence density.

Table 3 Comparative performance evaluation of the predictive model and its sub-models

Model	Overall accuracy	Specificity	FAR
MLR	0.6260	0.6024	0.3976
BPNN	0.6671	0.6436	0.3564
Hybrid model	0.8658	0.8549	0.1451

Note: MLR, multiple linear regression; BPNN, backpropagation neural network; hybrid model, the integrated forecasting framework optimized via the SA algorithm.

Fig. 8 Confusion matrices of MLR, BPNN, and hybrid models for snowmelt flood classification during 2009–2011. TP, true positive; FP, false positive; FN, false negative; TN, true negative.

100.00% of total flood events, with zero FN, indicating that no flood events were missed. Regarding negative sample identification, the hybrid model yielded a TN count of 576 d, accounting for 85.50% of non-flood days, while maintaining FP counts of 98 d, a substantial improvement over the sub-models. Conversely, the MLR and BPNN models showed lower sensitivity in positive sample identification, with TP values of 51 d (91.07% of total flood events) and 53 d (94.64% of total flood events), and FN of 5 and 3 d, respectively. Furthermore, both sub-models generated significantly higher false alarms, with FP counts of 268 and 240 d, respectively, thereby compromising the reliability of non-flood predictions. These results suggested that model fusion can effectively synchronize complete flood detection with a marked reduction in false alarms. In the Hutubi River Basin, characterized by infrequent but prolonged snowmelt floods, the hybrid model could provide a more robust predictive tool for disaster mitigation. This analysis quantified the performance gains and established a scientific foundation for optimizing operational flood early-warning systems.

We further generated ROC curves to assess performance on unseen data independent of a specific threshold. Using the 1978–2008 training set and the 2009–2011 test set, ROC curves and their corresponding AUC values were calculated for the MLR, BPNN, and hybrid models (Fig. 9). The AUC values clearly established a performance hierarchy: hybrid model>BPNN model>MLR model.

4.6 Performance enhancement from temporal factor

Figures 10 and 11 provide a comparative visualization of the predictive accuracy of hybrid models (without and with temporal factor) for identical flood events. Furthermore, a probability threshold of 0.1 was selected to account for the low frequency of flood occurrences and to optimize model sensitivity, thereby improving the diagnostic utility of the visualizations. The comparative analysis revealed two primary improvements: (1) the integration of temporal factor substantially reduced the model’ s FAR from 0.1451 to 0.0959; and (2) this reduction in FAR corresponded to a significant increase in specificity, from 0.8549 to 0.9041. These results suggested that the model with temporal factor can more effectively capture the temporal dependencies inherent in the hydrological data, leading to enhanced overall predictive capability.

4.7 Sensitivity analysis and algorithmic robustness

4.7.1 Sensitivity analysis of hydro-meteorological variables

We further conducted a systematic sensitivity analysis to quantify the relative importance of the four key input variables (Fig. 12). The analysis involved systematically removing single variables or combinations of variables and calculating the resulting change in the hybrid model’ s predictive performance, thereby assessing both individual contributions and synergistic effects.

Fig. 9 Comparative analysis of area under the Receiver Operating Characteris-

ROC curve comparing MLR (AUC=0.79), BPNN (AUC=0.83), and hybrid model with temporal factor (AUC=0.92); axes: FAR and TPR.

Figure 7: ROC curve comparing MLR (AUC=0.79), BPNN (AUC=0.83), and hybrid model with temporal factor (AUC=0.92); axes: FAR and TPR.

Time-series comparison between predicted flood probability without temporal factor and observed flood events in the Hutubi River Basin during 2009–2011; probability threshold is 0.1.

Figure 8: Time-series comparison between predicted flood probability without temporal factor and observed flood events in the Hutubi River Basin during 2009–2011; probability threshold is 0.1.

tic (ROC) curve (AUC) values among MLR, BPNN, and hybrid models. TPR, true positive rate.

Fig. 10 Comparison between predicted flood probability (without temporal factor) and observed flood events in the Hutubi River Basin during 2009–2011. A red marker positioned above the probability threshold line signifies a successful prediction (a true positive), whereas a marker below it indicates a missed event (a false negative).

Fig. 11 Comparison between predicted daily flood probability (with temporal factor) and observed flood events in the Hutubi River Basin (2009–2011). A red marker positioned above the probability threshold line signifies a successful prediction (a true positive), whereas a marker below it indicates a missed event (a false negative).

Fig. 12 Sensitivity analysis of hydro-meteorological variables in the hybrid model by removing single variables or combinations of variables.

The sensitivity analysis results indicated that daily streamflow was the most critical variable. Removing daily streamflow alone caused the AUC to decrease by 0.0986, while removing it in combination with snow depth change resulted in the largest performance degradation (decrease in AUC (ΔAUC) of 0.2590). This confirmed that daily streamflow was the dominant variable for flood prediction in this framework. Furthermore, significant synergistic effects were observed; removing combinations such as {daily streamflow, average temperature differ-

Time-series comparison between predicted daily flood probability with temporal factor and observed flood events in the Hutubi River Basin during 2009–2011; probability threshold is 0.1.

Figure 9: Time-series comparison between predicted daily flood probability with temporal factor and observed flood events in the Hutubi River Basin during 2009–2011; probability threshold is 0.1.

Fig. 12. Sensitivity analysis of hydro-meteorological variables in the hybrid model by removing single variables or combinations of variables.

Figure 10: Fig. 12. Sensitivity analysis of hydro-meteorological variables in the hybrid model by removing single variables or combinations of variables.

ence, and snow depth change} or {daily streamflow, daily precipitation, and snow depth change} also led to substantial performance drops (with ΔAUC of 0.2328 and 0.2249, respectively). In contrast, removing daily precipitation, average temperature difference, or snow depth change individually had a relatively minor impact. Interestingly, some combinations lacking daily streamflow yielded slightly negative ΔAUC values, suggesting that while these variables contribute little on their own, they can enhance model robustness when combined with the primary variable.

4.7.2 Parameter sensitivity and optimization of the SA algorithm To systematically evaluate the stability and convergence of the SA algorithm, we analyzed its performance across a range of hyperparameter settings. A two-dimensional parameter space was defined by the initial temperature and cooling coefficient. The analysis generated a three-dimensional response surface illustrating the effect of these hyperparameters on the model's classification accuracy (Fig. 13). Specifically, the initial temperature varied from 50 to 200 in increments of 25, while the cooling coefficient was tested at 10 equally spaced points between 0.90 and 0.99. To mitigate the effects of stochasticity, the algorithm was executed five times for each hyperparameter combination, and the mean accuracy was recorded. The response surface displayed a smooth and continuous trend, indicating that the SA algorithm exhibits considerable robustness to hyperparameter variations for this application. The mean accuracy approached a local optimum as the initial temperature approached 150 and the cooling coefficient approached 0.96. This finding suggested that a moderate initial temperature combined with a gradual cooling strategy can optimize both search efficiency and final convergence. Consequently, this optimal

parameter set was adopted for all subsequent SA applications in this study.

Heatmaps (Fig. 14) provide further insight into the relationship between the hyperparameters and model performance. The higher mean accuracy was concentrated around the initial temperature of 150 and the cooling coefficient of 0.96 (Fig. 14a). The standard deviation of accuracy was represented by a red color gradient, where lighter shades indicate lower variance and higher model stability (Fig. 14b). The selected hyperparameter combination resided in a region of both high accuracy and high stability, demonstrating that the model is efficient and produces reliable, consistent results.

Fig. 13 Mean accuracy response surface under varying simulated annealing (SA) hyperparameters

(a) Mean accuracy

Cooling coefficient	50	75	100	125	150	175	200
0.99	0.9740	0.9740	0.9740	0.9740	0.9740	0.9740	0.9740
0.98	0.9723	0.9723	0.9723	0.9723	0.9723	0.9723	0.9723
0.97	0.9713	0.9713	0.9713	0.9713	0.9713	0.9714	0.9714
0.96	0.9711	0.9710	0.9710	0.9711	0.9711	0.9711	0.9711
0.95	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707
0.94	0.9705	0.9705	0.9705	0.9705	0.9706	0.9706	0.9705
0.93	0.9704	0.9704	0.9704	0.9705	0.9705	0.9705	0.9704
0.92	0.9702	0.9701	0.9702	0.9702	0.9701	0.9701	0.9701
0.91	0.9702	0.9701	0.9701	0.9701	0.9701	0.9701	0.9701
0.90	0.9701	0.9701	0.9701	0.9701	0.9701	0.9701	0.9701

(b) Stability

Cooling coefficient	50	75	100	125	150	175	200
0.99	2.87	2.87	2.87	2.87	2.87	2.87	2.87
0.98	26.70	26.70	26.70	26.70	26.70	26.70	26.70
0.97	35.12	34.98	34.82	34.80	34.71	34.58	34.98
0.96	35.39	34.82	35.12	35.32	35.28	35.35	35.34
0.95	38.24	38.63	38.42	38.31	38.51	38.23	38.23
0.94	38.64	38.91	38.77	38.74	38.72	38.55	38.21
0.93	39.65	39.77	39.65	39.84	39.67	39.88	40.14
0.92	40.77	40.76	40.68	40.95	40.78	40.98	40.65
0.91	40.67	41.06	40.83	41.17	41.13	41.04	40.89
0.90	41.23	40.96	41.13	40.74	41.57	40.72	40.35

Fig. 14 Sensitivity analysis of SA algorithm hyperparameters. (a), mean accuracy; (b), stability (indicated by the standard deviation of accuracy).

4.8 Comparative analysis of hybrid and baseline models

To evaluate the performance of the proposed hybrid model, we conducted a comprehensive comparative analysis among established baseline models (including classic regression methods (MLR, BPNN, RF, and XGBoost) and a sequence model (LSTM)) and a series of hybrid configurations optimized via SA algorithm to explore the efficacy of different model fusion strategies. Performance was quantified using RMSE and the coefficient of determination (R^2). The proposed hybrid model (SA-BPNN-MLR) achieved the highest performance, recording the lowest RMSE and the highest R^2 among all tested models (Fig. 15). It significantly outperformed the baseline models and other hybrid models,

Fig. 15

Figure 11: Fig. 15

Fig. 16

Figure 12: Fig. 16

indicating that the SA-optimized fusion of BPNN and MLR models is a highly effective approach for predicting snowmelt floods in this basin.

Fig. 15 Predictive performance comparison of various models during the period 2009–2011. (a), coefficient of determination (R^2); (b), RMSE. XGBoost, eXtreme Gradient Boosting; RF, Random Forest; LSTM, Long Short-Term Memory.

4.9 Quantitative strategy for balancing warning reliability and timeliness

This study analyzed the trade-off between the FAR, the warning probability threshold, and the resulting warning lead time to inform the optimization of flood warning systems (Fig. 16). The results indicated that as the threshold was lowered from 0.1 to 0.02, the FAR increased from 0.0959 to 0.3227, while the warning lead time extended from 1 to 10 d. This analysis demonstrated that by selecting an appropriate threshold, an effective warning lead time of 3–7d can be achieved while maintaining a low FAR (( )0.1400). These findings provide quantitative evidence for decision-makers to balance warning effectiveness with the costs of false alarms.

For the Hutubi River Basin, optimizing the FAR and early warning threshold can therefore yield significant economic and risk management benefits. Striking a strategic balance between these factors can reduce the economic impact of false alarms while minimizing the risk of missed flood events. This helps improve the resilience of flood defense systems, enhance regional economic stability, and strengthen overall flood protection strategies.

Fig. 16 Relationships between the FAR, warning probability threshold, and warning lead time. The blue solid line maps the decreasing predefined warning probability thresholds against the corresponding increase in FAR. Concurrently, the red dashed line correlates the warning lead time with the FAR. The optimal intersection where forecast accuracy and adequate preparation time are balanced can be identified by superimposing these two distinct operational parameter curves onto a single continuous FAR scale.

5 Discussion

A key contribution of this study is the development of a hybrid model that integrates MLR and a BPNN to predict snowmelt floods. In contrast to studies that rely on singular linear or nonlinear models (Jhong et al., 2018; Mosavi et al., 2018; Davenport et al., 2020; Nearing et al., 2021; Islam et al., 2022; Samantaray et al., 2023; Yan et al., 2023; Luppichini et al., 2024; Nearing et al., 2024; Brooks et al., 2025), this approach can leverage the respective strengths of both methodologies. MLR provides a stable, interpretable framework for capturing linear trends among hydro-meteorological variables, whereas the BPNN excels at modeling the complex, nonlinear dynamics inherent in snowmelt-runoff processes. This synthesis enhances predictive accuracy and promotes model robustness, which is particularly valuable in data-scarce small basins. Future work could extend this framework by incorporating additional machine learning algorithms, such as XGBoost or RF, to further refine predictive performance and generalization capabilities.

This study also advanced conventional modeling by explicitly incorporating the temporal factor, which are decomposed into seasonal and event-driven components. This dual representation enables the model to capture both the regular periodicity of snowmelt cycles and the abrupt hydrological responses to sudden temperature shifts. Consequently, the model' s accuracy and stability were significantly improved relative to baseline models that depend solely on instantaneous hydro-meteorological inputs. This addressed a noted limitation in prior research where linear models inadequately capture temporal dynamics (Davenport et al., 2020; Brooks et al., 2025) and where standard machine learning models face generalization challenges in small basins (Mosavi et al., 2018; Yan et al., 2023). This approach therefore could not only improve forecast accuracy but also enhance model interpretability.

The application of the SA algorithm for model fusion represents a key innovation. Most of the current model fusion frameworks rely on empirical or equal-weighting schemes, which can be subjective and sub-optimal (Tencate et al., 2016). In this study, the SA algorithm performed a systematic, global search to identify the optimal weights for the MLR and BPNN outputs. This data-driven process removed operator bias and enhanced scientific rigor. The resulting optimal weights—0.466 for MLR and 0.534 for BPNN—quantitatively affirmed the balanced and complementary contributions of both the linear and nonlinear model components.

While the proposed SA-BPNN-MLR hybrid model demonstrated strong predictive performance on historical data, its operational feasibility for real-time flood warnings warrants careful consideration. First, data availability presents a practical challenge. The model was trained on a continuous 1978–2011 dataset, but acquiring uninterrupted, real-time streams of daily streamflow, precipitation, average temperature, and snow depth presents challenges in the Hutubi River Basin due to reporting lags or sparse monitoring networks (Kuhaneswaran et al., 2025). Second, the model has distinct computational demands. The SA-BPNN component is computationally intensive during the training phase, although prediction inference is rapid. In contrast, the MLR component is lightweight and suitable for rapid execution (Kumar et al., 2025; McEachran et al., 2025). To facilitate near-real-time deployment, strategies such as periodic parameter updates or incremental learning could be implemented. Finally, system latency—encompassing data acquisition, model computation, and warning dissemination—remains a critical bottleneck. The operational viability of the model will ultimately depend on balancing these constraints with local data infrastructure and computational resources. Exploring model compression or high-performance computing solutions will be essential for enhancing its practical value.

Future research should prioritize validating the model' s generalizability across other small watersheds with diverse climatic and hydrological regimes. Furthermore, integrating real-time data streams alongside more advanced machine learning architectures could substantially enhance the model' s predictive skill and responsiveness. These findings yield broader implications for flood risk management, demonstrating that the synergistic fusion of complementary models and the inclusion of temporal dynamics are effective strategies for improving the reliability and accuracy of hydrological predictions.

6 Conclusions

This study developed a snowmelt flood prediction model for the Hutubi River Basin utilizing hydrological data from the Shimen Hydrological Station and meteorological observations from Changji Meteorological Bureau. An integrated forecasting framework was constructed by coupling MLR and BPNN models, with ensemble weights optimized via the SA algorithm. The analysis revealed that while both constituent sub-models exhibited robust predictive capabilities, the integrated hybrid model demonstrated superior sensitivity in flood detection. These findings confirmed that ensemble-based architectures substantially outperform traditional individual models when applied to complex hydrological conditions. By synergizing the complementary strengths of multiple algorithms and incorporating explicit temporal features, the proposed model could not only improve predictive accuracy but also effectively mitigate the high FAR typically associated with individual models. Consequently, this research provides a critical theoretical foundation and technical framework for developing high-precision early warning systems in future work. The integration of temporal dynamics and SA-optimized model fusion could significantly enhance snowmelt flood forecasting, offering a robust and scalable modeling framework for small watersheds characterized by distinct seasonal and event-driven hydrological dynamics.

Conflict of interest

LIU Yongqiang is an Editorial Board member of Journal of Arid Land and was not involved in the editorial review or the decision to publish this article.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the National Key Research and Development Program of China (2019YFC1510505) and the National Undergraduate Training Program for Innovation and Entrepreneurship of Xinjiang University (202510755008).

Author contributions

Conceptualization: WANG Zerui, LI Yaqian; Investigation: WANG Zerui, LI Xiaoyang, ZHANG Yuanwei; Data curation: WANG Zerui, ZHANG Yuanwei; Software: WANG Zerui, LI Xiaoyang; Writing - original draft preparation: WANG Zerui; Writing - review and editing: WANG Zerui, LIU Yongqiang, WANG Weiping; Funding acquisition: LIU Yongqiang, WANG Weiping; Resources: LIU Yongqiang, WANG Weiping; Supervision: LIU Yongqiang, WANG Weiping; Validation: LI Yaqian, ZHANG Yuanwei; Project administration: LIU Yongqiang; Visualization: WANG Zerui, LI Xiaoyang. All authors approved the manuscript.

References

Adamowski J, Prasher S O. 2012. Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. Journal of Water and Land Development, (17): 89–97.

Agnihotri J, Coulibaly P. 2020. Evaluation of snowmelt estimation techniques for enhanced spring peak flow prediction. Water, 12(5): 1290, doi: 10.3390/w12051290.

Argentin A L, Horton P, Schaefli B, et al. 2025. Scale dependency in modeling nivo-glacial hydrological systems: the case of the Arolla basin, Switzerland. Hydrology and Earth System Sciences, 29(6): 1725–1748.

Aryal Y. 2025. Assessing climate and watershed controls on rain-on-snow runoff using XGBoost-SHAP explainable AI (XAI). Geosciences, 15(12): 467, doi: 10.3390/geosciences15120467.

Atashi V, Barati R, Lim Y H. 2023a. Distributed Muskingum model with a Whale Optimization Algorithm for river flood routing. Journal of Hydroinformatics, 25(6): 2210–2222.

Atashi V, Barati R, Lim Y H. 2023b. Development of a distributed nonlinear Muskingum model by considering snowmelt effects for flood routing in the Red River. Scientific Reports, 13(1): 21356, doi: 10.1038/s41598-023-48895-8.

Bean B L, Watts E. 2024. Quantifying the impact of rain-on-snow induced flooding in the western United States. Water, 16(19): 2826, doi: 10.3390/w16192826.

Blöschl G, Sivapalan M. 1995. Scale issues in hydrological modelling: a review. Hydrological Processes, 9(3-4): 251-290.

Brooks P D, Solomon D K, Kampf S, et al. 2025. Groundwater dominates snowmelt runoff and controls streamflow efficiency in the western United States. Communications Earth & Environment, 6(1): 341, doi: 10.1038/s43247-025-02303-3.

Cui M Y, Zhou G, Zhang D H, et al. 2022. Global snowmelt flood disasters and their impact from 1900 to 2020. Journal of Glaciology and Geocryology, 44(6): 1898-1911. (in Chinese)

Davenport F V, Herrera-Estrada J E, Burke M, et al. 2020. Flood size increases nonlinearly across the western United States in response to lower snow-precipitation ratios. Water Resources Research, 56(1): e2019WR025571, doi: 10.1029/2019WR025571.

Ding Y J, Zhang S Q. 2018. Study on water internal recycle process and mechanism in typical mountain areas of inland basins, Northwest China: Progress and challenge. Advances in Earth Science, 33(7): 719-728.

Duan S H, Ullrich P, Risser M, et al. 2024. Using temporal deep learning models to estimate daily snow water equivalent over the Rocky Mountains. Water Resources Research, 60(4): e2023WR035009, doi: 10.1029/2023WR035009.

Editorial Board of China Meteorological Disaster Series. 2006. Records of Meteorological Disasters in China-Xinjiang Volume. Beijing: China Meteorological Press, 1-450. (in Chinese)

Fileni F, Fowler H J, Lewis E, et al. 2025. Flood risk assessments for small catchments under climate change—How can Scotland improve its policy for enhanced flood resilience and preparedness. Journal of Flood Risk Management, 18(1): e70035, doi: 10.1111/jfr3.70035.

Hotovy O, Nedelcev O, Seibert J, et al. 2025. Rain-on-snow events in mountainous catchments under climate change. Hydrology and Earth System Sciences, 29(17): 4199-4217.

Islam K I, Elias E, Brown C, et al. 2022. A statistical approach to using remote sensing data to discern streamflow variable influence in the snow melt dominated upper Rio Grande Basin. Remote Sensing, 14(23): 6076, doi: 10.3390/rs14236076.

Janjić J, Tadić L. 2023. Fields of application of SWAT hydrological model—A review. Earth, 4(2): 331-344.

Jhong Y D, Chen C S, Lin H P, et al. 2018. Physical hybrid neural network model to forecast typhoon floods. Water, 10(5): 632, doi: 10.3390/w10050632.

Jiang C, Parteli E J R, Xia Q, et al. 2022. A regional hydrological model for arid and semi-arid river basins with consideration of irrigation. Environmental Modelling & Software, 157: 105531, doi: 10.1016/j.envsoft.2022.105531.

Koltsida E, Mamassis N, Kallioras A. 2023. Hydrological modeling using the Soil and Water Assessment Tool in urban and peri-urban environments: the case of Kifisos experimental subbasin (Athens, Greece). Hydrology and Earth System Sciences, 27(4): 917-931.

Kratzert F, Klotz D, Brenner C, et al. 2018. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11): 6005-6022.

Kuhaneswaran B, Sorwar G, Alaei A R, et al. 2025. Evolution of data-driven flood forecasting: trends, technologies, and gaps—A systematic mapping study. Water, 17(15): 2281, doi: 10.3390/w17152281.

Kumar V, Sharma K V, Mangukiya N K, et al. 2025. Machine learning applications in flood forecasting and predictions, challenges, and way-out in the perspective of changing environment. AIMS Environmental Science, 12(1): 72-105.

Lane R A, Coxon G, Freer J E, et al. 2019. Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. Hydrology and Earth System Sciences, 23(10): 4011-4032.

Latif Y, Ma Y M, Ma W Q, et al. 2020. Differentiating snow and glacier melt contribution to runoff in the Gilgit River Basin via degree-day modelling approach. Atmosphere, 11(10): 1023, doi: 10.3390/atmos11101023.

Luppichini M, Vailati G, Fontana L, et al. 2024. Machine learning models for river flow forecasting in small catchments. Scientific Reports, 14: 26740, doi: 10.1038/s41598-024-78012-2.

McEachran Z, Ghosh R, Renganathan A, et al. 2025. Knowledge-guided machine learning for operational flood forecasting. Water Resources Research, 61(11): e2024WR039064, doi: 10.1029/2024WR039064.

Mirboluki A, Mehraraein M, Kisi O, et al. 2024. Groundwater level estimation using improved deep learning and soft computing methods. Earth Science Informatics, 17(3): 2587-2608.

Mosavi A, Ozturk P, Chau K. 2018. Flood prediction using machine learning models: literature review. Water, 10(11): 1536, doi: 10.3390/w10111536.

Nearing G, Cohen D, Dube V, et al. 2024. Global prediction of extreme floods in ungauged watersheds. Nature, 627: 559-563.

Nearing G S, Kratzert F, Sampson A K, et al. 2021. What role does hydrological science play in the age of machine learning? Water Resources Research, 57(3): e2020WR028091, doi: 10.1029/2020WR028091.

O' brien R M. 2007. A caution regarding rules of thumb for variance inflation factors. Quality & Quantity, 41(5): 673–690.

Samantaray S, Sahoo A, Agnihotri A. 2023. Prediction of flood discharge using hybrid PSO-SVM algorithm in Barak River

Basin. MethodsX, 10: 102060, doi: 10.1016/j.mex.2023.102060.

Santos L B L, Escobar-Silva E V, Satolo L F, et al. 2025. Machine learning-based hydrological models for flash floods: a systematic literature review. Smart Construction and Sustainable Cities, 3(1): 21, doi: 10.1007/s44268-025-00071-9.

Stein L, Clark M P, Knoben W J M, et al. 2021. How do climate and catchment attributes influence flood generating processes? A large-sample study for 671 catchments across the contiguous USA. Water Resources Research, 57(4): e2020WR028300, doi: 10.1029/2020WR028300.

Tencate A J, Kalivas J H, White A J. 2016. Fusion strategies for selecting multiple tuning parameters for multivariate calibration and other penalty based processes: a model updating application for pharmaceutical analysis. Analytica Chimica Acta, 921: 28–37.

Teshome F T, Bayabil H K, Thakural L N, et al. 2020. Verification of the MIKE11-NAM model for simulating streamflow. Journal of Environmental Protection, 11(2): 152–167.

Thapa S, Zhao Z B, Li B, et al. 2020. Snowmelt-driven streamflow prediction using machine learning techniques (LSTM, NARX, GPR, and SVR). Water, 12(6): 1734, doi: 10.3390/w12061734.

Thapa U, Pati B M, Thapa S, et al. 2024. Comparative analysis of snowmelt-driven streamflow forecasting using machine learning techniques. Water, 16(15): 2095, doi: 10.3390/w16152095.

Tran Tuan T. 2024. Multiple conceptual hydrological models for simulating streamflow in data-sparse river basins: an application of the Vietnamese Cau river basin. Water Practice & Technology, 19(7): 2944–2958.

Venegas-Cordero N, Cherrat C, Kundzewicz Z W, et al. 2023. Model-based assessment of flood generation mechanisms over Poland: the roles of precipitation, snowmelt, and soil moisture excess. Science of The Total Environment, 891: 164626, doi: 10.1016/j.scitotenv.2023.164626.

Wang J J, Shi P, Jiang P, et al. 2017. Application of BP neural network algorithm in traditional hydrological model for flood forecasting. Water, 9(1): 48, doi: 10.3390/w9010048.

Wilks D S. 2011. Statistical Methods in the Atmospheric Sciences (3rd ed.). Boston: Academic Press, 1–676.

Xie Z H, Yuan F, Duan Q Y, et al. 2007. Regional parameter estimation of the VIC land surface model: methodology and application to river basins in China.

Journal of Hydrometeorology, 8(3): 447–468.

Yan H X, Sun N, Wigmosta M S, et al. 2023. The role of snowmelt temporal pattern in flood estimation for a small snow-dominated basin in the Sierra Nevada. Water Resources Research, 59(10): e2023WR034496, doi: 10.1029/2023WR034496.

Yan K, Gao S, Wen J H, et al. 2023. A multi-factor combination model for medium to long-term runoff prediction based on improved BP neural network. Water, 15(20): 3559, doi: 10.3390/w15203559.

Yeganeh A, Ahmadi F, Wong Y J, et al. 2024. Shallow vs. deep learning models for groundwater level prediction: a multi-piezometer data integration approach. Water, Air, & Soil Pollution, 235(7): 441, doi: 10.1007/s11270-024-07247-6.

Yu X T, Xu Y P, Gu H T, et al. 2023. Multi-objective robust optimization of reservoir operation for real-time flood control under forecasting uncertainty. Journal of Hydrology, 620: 129421, doi: 10.1016/j.jhydrol.2023.129421.

Zhao Q D, Zhao C C, Qin Y, et al. 2020. Response of the hydrological processes to climate change in the Muzati River basin with high glacierization, southern slope of the Tianshan Mountains. Journal of Glaciology and Geocryology, 42(4): 1285–1298. (in Chinese)

Zhou M Q, Lu W J, Ma Q, et al. 2023. Study on the snowmelt flood model by machine learning method in Xinjiang. Water, 15(20): 3620, doi: 10.3390/w15203620.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv. Machine translation. Verify with the original.