Proceedings of the 5th International Conference on Materials Chemistry and Environmental Engineering
DOI: 10.54254/2755-2721/2025.23605

Prediction Model for Snowmelt Floods in the Hutubi River Based on Simulated Annealing Algorithm

Zerui Wang

Xinjiang University, Urumqi, China
1045686071@qq.com

Abstract: In response to the frequent occurrence of snowmelt floods in the Hutubi River and the lack of effective prediction models, this paper employs the simulated annealing algorithm to construct a new snowmelt flood prediction model by integrating multiple linear regression and BP (Backpropagation) neural network models. A comparison between the integrated prediction model and individual sub-models shows that the integrated model outperforms the single sub-models in both prediction accuracy and false alarm rate. The prediction accuracies of the two sub-models are 90.24% and 95.12%, respectively, while the integrated model achieves a 100% accuracy rate. The false alarm rates of the two sub-models are 39.73% and 35.62%, respectively, whereas the integrated model reduces the false alarm rate to 14.52%. Furthermore, after incorporating the time factor into the integrated prediction model, its predictive performance improves significantly, and the false alarm rate decreases further to 9.59%. The research results indicate that the integrated prediction model provides better prediction accuracy, and the time factor significantly impacts the model's performance. The constructed integrated model not only effectively combines the multiple linear regression and BP neural network models technically but also enhances the prediction accuracy of snowmelt floods in the Hutubi River while reducing the false alarm rate in practical applications.

Keywords: Multiple Linear Regression, BP Neural Network, Simulated Annealing False Alarm Rate.

1. Introduction

Snowmelt floods are a global challenge. Statistical studies indicate that from 1900 to 2020, snowmelt floods have caused economic losses amounting to $730 billion [1]. Given the severe economic impact of snowmelt floods on human activities and livelihoods, it is crucial to develop effective prediction models for snowmelt floods. The mechanisms behind snowmelt floods are complex, involving multiple factors such as meteorology, topography, soil, and vegetation. In particular, for small watersheds, snowmelt floods are more sensitive to factors like topography and soil. Therefore, constructing an effective warning model for snowmelt floods usually requires adapting to the specific conditions of the watershed and developing a customized prediction model.

Snowmelt flood prediction models are generally divided into two main categories: hydrological models and machine learning models. The HEC-HMS model, SWAT model, and NAM model are three hydrological models widely used for predicting snowmelt floods. The HEC-HMS model is based on the physical simulation of hydrological processes, predicting floods through the rainfall-runoff relationship and soil moisture balance [2]. The SWAT model involves land use, soil types, and

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.23605

meteorological conditions to improve the prediction accuracy [3]. The NAM model uses the instantaneous unit hydrograph theory to precisely simulate the rainfall-runoff process, enhancing prediction accuracy [4]. For machine learning, LSTM, Random Forest, and XGBoost models are widely used for snowmelt flood prediction. LSTM captures long-term dependencies in time-series data for dynamic flood prediction [5]. Random Forest integrates multiple decision trees and various features, improving accuracy and stability [6]. XGBoost optimizes with gradient boosting, efficiently handling high-dimensional data and nonlinear relationships, especially in complex watersheds, where accuracy improves [7].

These snowmelt flood prediction models have shown ideal predictive results in larger watersheds [8-9]. However, a study by John in 2020 [10] found that due to the topographical complexity, data scarcity, and parameter inapplicability in small watersheds, directly transferring these models significantly reduces prediction accuracy. For predicting snowmelt floods in small watersheds, scholar Meng Yu in 2021 [11] successfully used grey relational analysis and a multiple linear regression model to predict floods in the Yaoyang River Basin; in 2014, Zhang Jian [12] successfully applied a BP neural network snowmelt prediction model to predict snowmelt-induced flash floods. Additionally, when Kisi investigated snowmelt flood prediction in Turkey [13], he considered the specific timing of snowmelt flood events as a key factor and found that incorporating the time factor improved the model's accuracy.

The Hutubi River, as a typical small river, frequently experiences snowmelt floods during spring, But there is a lack of relevant research on snowmelt flood prediction for this river. Therefore, This paper builds a snowmelt flood prediction model for the Hutubi River by combining a multiple linear regression model and a BP neural network model, optimized with the simulated annealing algorithm.

The second part of this paper introduces the study area, the data used, and the constructed model; the third part presents the model's forecast validation results; and the fourth part provides the conclusion.

2. Data and method

2.1. Study area and data

The Hutubi River Basin is located in the central section of the northern slopes of the Eastern Tianshan Mountains, at the southern edge of the Junggar Basin, with a watershed area of 5,119 km² and a drainage area of 1,840 km² above the Shimen Hydrological Station [12]. Due to its climatic characteristics, including deep winter snow accumulation and rapid temperature rise in spring, the river is prone to snowmelt floods, making it a typical snowmelt flood river.

The data used in this study includes hydrological data from the Shimen Hydrological Station on the Hutubi River and meteorological data from Changji City, spanning nearly 23 years from 1978 to 2000. The data includes four parameters: daily runoff (i_1), daily precipitation (i_2), average temperature difference (i_3), and change in snow depth (i_4). The grey relational degree ξ_{ij} (see Equation 1) was calculated to determine the relationship between these four parameters (i_1, i_2, i_3, i_4) and the occurrence of snowmelt floods.

$$\xi_{ij} = \frac{min_i min_j \Delta_{ij} + \rho max_i max_j \Delta_{ij}}{\Delta_{ij} + \rho max_i max_j \Delta_{ij}} \quad (1)$$

Where Δ_{ij} represents the difference between the i-th comparative sequence (the four parameters, such as daily runoff) and the reference sequence (whether a snowmelt flood occurred); $min_i min_j \Delta_{ij}$ is the minimum difference between the four comparative sequences and the reference sequence; $max_i max_j \Delta_{ij}$ is the maximum difference between the four comparative sequences and the reference sequence; ρ is the distinguishing coefficient, which is set to 0.5 in this study.

DOI: 10.54254/2755-2721/2025.23605

The grey relational coefficients between the four parameters and the occurrence of snowmelt floods are calculated as shown in the table below:

Table 1: Grey relational coefficients between impact indicators and snowmelt flood occurrence in the Hutubi River

Impact Indicator	Grey Relational Coefficient
Daily Runoff	0.902465
Daily Precipitation	0.947493
Average Temperature Difference	0.836557
Change in Snow Depth	0.955369

Based on the data presented in the table above, it is evident that there is a significant correlation between these four parameters and the occurrence of snowmelt floods in the Hutubi River.

The data from 1978 to 1997 is used as the training and validation set, with 80% as the training set and 20% as the validation set. The data from 1998 to 2000 is used as an independent test set.

2.2. Model construction

This study builds a snowmelt flood prediction model by integrating a multiple linear regression model and a BP (Backpropagation) neural network model based on the simulated annealing algorithm. The simulated annealing algorithm is used to determine the optimal weight combination.

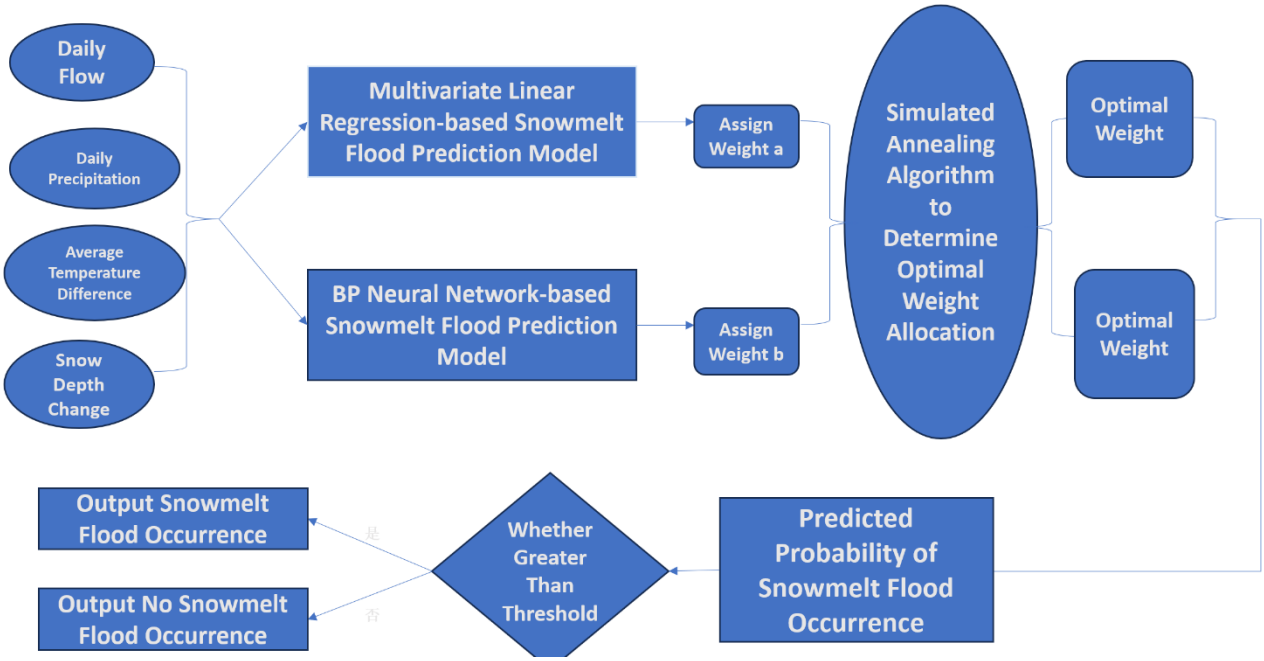

Figure 1: Overall research flowchart

Based on the four critical factors influencing snowmelt floods in the Hutubi River, a comprehensive multiple linear regression equation is meticulously established to model the intricate probability of snowmelt flood occurrence, taking into account these four pivotal impact factors. The probability of snowmelt flood occurrence is symbolically denoted as P_i, while the four independent variables—daily runoff, daily precipitation, average temperature difference, and the significant change in snow depth—are respectively represented as R, P, T, and S. Using $SPSS$ statistical software, the regression equation is calculated using the enter method, and the results are as follows:

DOI: 10.54254/2755-2721/2025.23605

$$P_i=0.147R+0.089P-0.021T+0.033S \quad (2)$$

The BP neural network model constructed in this study includes one hidden layer containing 150 neurons and introduces a $sigmoid$ activation function, which converts the four parameters into continuous values between 0 and 1. The cross-entropy loss function is used to measure the difference between the predicted snowmelt flood probability and the actual flood labels (see Equation 3).

$$Loss_{(y,t)} = -\left[tlog(y) + (1-t)\log(1-y)\right] \quad (3)$$

In Equation 3, y represents the predicted probability of flood occurrence, and t indicates whether a flood actually occurred on a given day. If a flood occurred on a particular day, then t= 1; if no flood occurred, then t= 0.

To prevent overfitting, $L2$ regularization is introduced by adding an $L2$ regularization term to the original loss function, resulting in a new loss function, $Loss_{L2}$. The formula for the new loss function is as follows:

$$Loss_{L2} = Loss + \alpha \sum_i w_i^2 \quad (4)$$

To achieve better predictive performance, the simulated annealing algorithm is used to combine the two models in a weighted manner. The following describes the optimization of the integrated snowmelt flood prediction model based on the simulated annealing algorithm:

Definitions: The snowmelt flood prediction model based on the multiple linear regression approach is denoted as f_1. The snowmelt flood prediction model based on the BP neural network method is denoted as f_2. The simulated annealing algorithm is applied as follows:

(1) Set Initial Temperature: The initial temperature is defined as T_0.

(2) Random Generation: A set of initial weights w_1 and w_2 is randomly generated, ensuring that they satisfy the following conditions:

$$w_1 + w_2 = 1 \quad (5)$$

The mathematical expression for the weighted integrated model is then presented as follows:

$$P_{weighting} = w_1 f_1 + w_2 f_2 \quad (6)$$

A threshold of 0.10 is set, meaning that if the predicted probability exceeds 0.10, it is determined that a flood event occurred on that day; otherwise, it is determined that no flood occurred. By comparing these predicted results with the actual flood records from 1998 to 2000, the accuracy of the initial weighted model, denoted as H_0, can be obtained.

The weights are continuously adjusted, and the model's prediction results are recalculated under the new weight combinations:

$$P'_{weighting} = w'_1 f_1 + w'_2 f_2 \quad (7)$$

(3) Difference Discussion: The model accuracy under $P'_{weighting}$ is denoted as H_0'. The difference between H_0' and H_0 is calculated and denoted as Δf. The next step involves discussing Δf and providing the probability expression for accepting a new solution:

When $\Delta f > 0$, it indicates that the model's accuracy is higher under the new weight combination. In this case, the new weight combination is accepted with 100% probability, i.e., $P = 1$.

When $\Delta f < 0$, it suggests that the model's accuracy is lower under the new weight combination. However, to avoid getting trapped in a local optimum, the new weight combination may still be

DOI: 10.54254/2755-2721/2025.23605

accepted with a certain probability. Based on the Boltzmann distribution and the principles of state transition probability, the acceptance probability expression is given as follows:

$$P = \exp\left(-\frac{\Delta f}{T}\right) \tag{8}$$

In the above equation, T represents the new temperature corresponding to the new weight combination.

(4) Cooling Iteration: Based on the above approach, a cooling coefficient is selected to continuously reduce the temperature, with each new temperature corresponding to a new set of weights. The new temperature is determined as follows:

$$T_{i+1} = \alpha T_i \tag{9}$$

Where α is the cooling coefficient. To improve the accuracy of the simulated annealing process, α is set to 0.97 in this study, allowing for an increased number of iterations and a gradual cooling process, thereby gradually approaching the optimal solution.

(5) Termination Criteria: The termination temperature is set at $e = 10^{-30}$. When the temperature $T < e$, the iteration stops, and the final solution is output. This solution represents the optimal weight combination and the corresponding model accuracy under this combination. Thus, the multi-model integrated optimized snowmelt flood prediction model based on the simulated annealing algorithm is successfully constructed. Through programming, the optimal weight combination determined in this study is:

$$w_1 = 0.466052212188104 \tag{10}$$

$$w_2 = 0.533947787811896 \tag{11}$$

3. Results and discussion

3.1. Prediction model results and analysis

To quantitatively evaluate the performance of the prediction model, the following validation table is presented. This table lists the overall prediction accuracy, sensitivity, and false alarm rate of the prediction model as well as the two sub-models that constitute it. This allows readers to comprehensively understand the results of the prediction model.

Table 2: Performance comparison of the prediction model and its sub-models

	Overall Prediction Accuracy	Sensitivity	False Alarm Rate
Multiple Linear Regression	0.6027	0.9024	0.3973
BP Neural Network	0.6438	0.9512	0.3562
Integrated Prediction Model	0.8548	1	0.1452

The overall prediction accuracy in the table refers to the model's general performance, including the number of days it correctly predicts no flood occurrence. Sensitivity refers to the proportion of predicted flood days relative to the actual flood days. The false alarm rate represents the instances where the model incorrectly predicts flood events.

The findings are as follows:

DOI: 10.54254/2755-2721/2025.23605

(1) The snowmelt flood prediction model based on the simulated annealing algorithm,which combines the multiple linear regression model and the BP neural network model, achieves better prediction accuracy and a lower false alarm rate compared to each of the individual sub-models. The false alarm rate of the integrated model is significantly lower, 25.21% less than that of the multiple linear regression model and 21.1% less than that of the BP neural network model.

(2) The prediction accuracy and false alarm rate of the two sub-models are very similar. The difference in sensitivity between the two sub-models is 4.88%, and the difference in their false alarm rates is 4.11%.

3.2. Improvement of the model by considering the time factor

Incorporating the time factor into the model draws on previous research by scholars. By capturing the temporal characteristics of snowmelt flood occurrences, the integrated prediction model based on the simulated annealing algorithm was modified accordingly.

Table 3: Comparison of the prediction model before and after incorporating the time factor

Prediction Model	Sensitivity	False Alarm Rate
Without Time Factor	1	0.1452
With Time Factor	1	0.0959

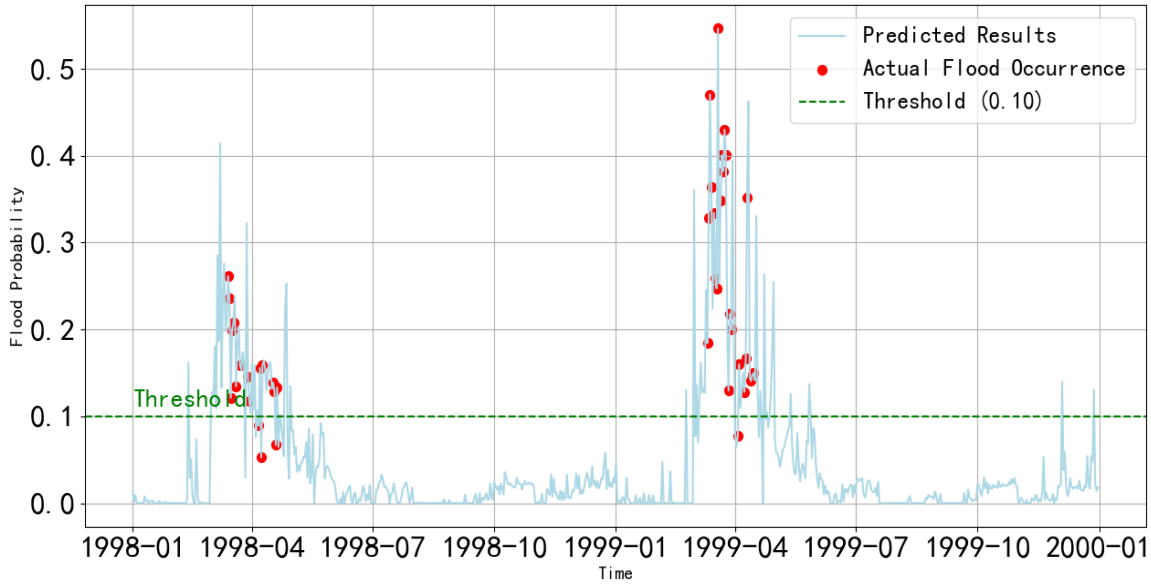

Figure 2: Comparison of predicted and actual results for snowmelt floods in the Hutubi River (1998-2000)

DOI: 10.54254/2755-2721/2025.23605

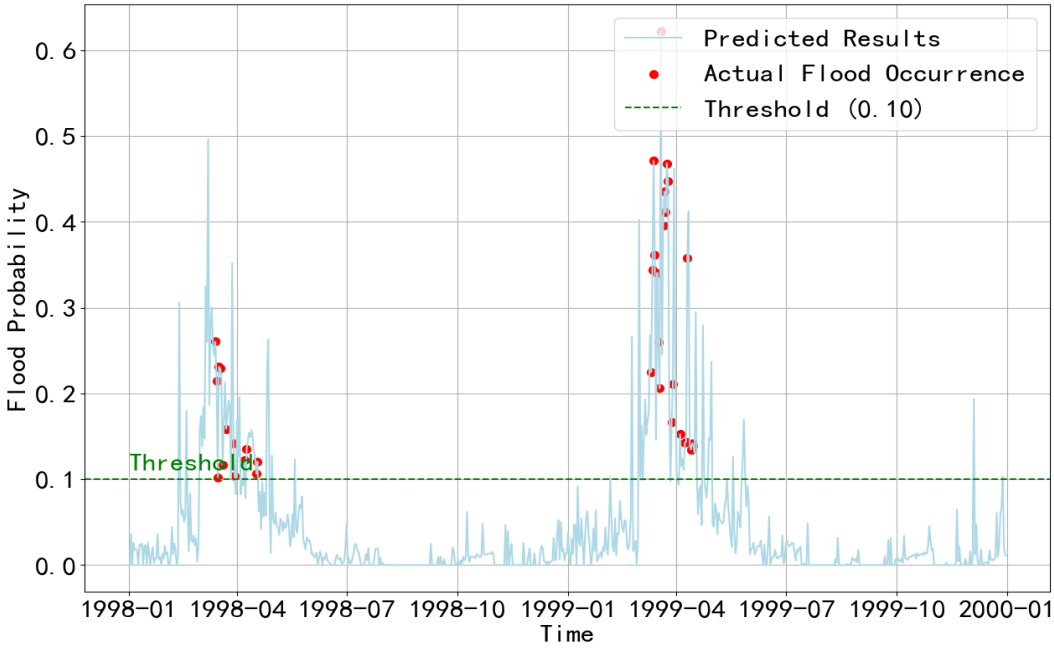

Figure 3: Comparison of predicted and actual results for snowmelt floods in the Hutubi River (1998-2000, considering the time factor)

The following observations can be made:

(1) Incorporating the time factor into the snowmelt flood prediction model construction further reduces the model's false alarm rate. The false alarm rate of the integrated prediction model decreases from 14.52% to 9.59%. This reduction in the false alarm rate enhances the overall predictive capability of the model.

(2) After considering the time factor, the model shows a significant improvement in the prediction probabilities output during forecasting. This improvement demonstrates that the model can more accurately capture the dynamic changes within the time series.

4. Conclusion

The objective of this study was to construct a snowmelt flood prediction model for the Hutubi River using data from the Shimen Hydrological Station and meteorological observations from Changji City. Based on the simulated annealing algorithm, an integrated prediction model combining a multiple linear regression model and a BP neural network model was developed.

An analysis of the constructed prediction model revealed that the two sub-models had similar predictive capabilities for the Hutubi River, a small watershed. The integrated prediction model, determined using the simulated annealing algorithm, showed only a slight improvement in flood detection sensitivity compared to the individual models, yet it achieved 100% sensitivity, meaning every actual flood event was correctly identified by the integrated model. Specifically, the prediction accuracies of the two sub-models were 90.24% and 95.12%, respectively, while the accuracy of the integrated prediction model reached 100%. Compared to the individual models, the integrated model showed a significant reduction in the false alarm rate, decreasing by over 20%. The false alarm rates of the two sub-models were 39.73% and 35.62%, respectively, while the integrated prediction model's false alarm rate was reduced to 14.52%.

Since the snowmelt floods in the Hutubi River occur primarily in March and April each year, this study incorporated the time factor to improve the prediction model. By capturing temporal characteristics, it was found that the model's predictive performance improved significantly.

DOI: 10.54254/2755-2721/2025.23605

Specifically, the false alarm rate of the integrated prediction model was further reduced from 14.52% to 9.59%.

References

[1] *Cui, M., Zhou, G., Zhang, D., & Zhang, S. (2022). Global snowmelt flood disasters and their impacts: 1900-2020.*

[2] *Gao, Y., Zhang, Z., Lai, L., & Liu, Y. (2019). Influence of parameter variation on basin flood simulation result of HEC-HMS model.*

[3] *Wang, X., & Shi, P. (2024). Research and evaluation of storm flood simulation based on SWAT and Xin'anjiang models.*

[4] *Sun, J., Liang, J., Zeng, Z., Fu, W., & Qin, J. (Year unavailable). Application of the NAM model coupled with hydrodynamic models in the Qingshitan Basin.*

[5] *Meng, H., Yao, C., Zheng, A., Yang, F., Li, J., Shi, Z., & Zhang, J. (Year unavailable). Flood simulation research using LSTM and its variants under different scenarios.*

[6] *Zuo, Y. (Year unavailable). Water disaster prediction model based on decision trees.*

[7] *Li, Y. (Year unavailable). Analysis and risk prediction of high-density urban waterlogging disaster factors based on explainable algorithms: A case study of Shenzhen.*

[8] *Selim, Ş., & İspirli, M. N. (2022). Predicting snowmelt runoff at the source of the mountainous Euphrates River Basin in Turkey for water supply and flood control issues using HEC-HMS modeling.*

[9] *Ma, M., Zhao, G., He, B., Li, Q., & Dong, H. (2021). XGBoost-based method for flash flood risk assessment.*

[10] *Doe, J., & Smith, J. (2020). Challenges in adapting large-scale flood prediction models to small watersheds: A case study in the Upper Colorado River Basin.*

[11] *Meng, Y. (2021). Grey relational analysis and multiple linear regression real-time flood prediction model: A case study of Dongbaichengzi Station in the Raoyang River Basin of the Liao River.*

[12] *Zhang, J. (2014). Design of a snowmelt-induced mountain flood early warning system.*

[13] *Kisi, O., & Genc, S. (2015). Predicting snowmelt-induced floods in Turkey using artificial neural networks.*