University of Birjand

Water Harvesting Research Vol. 8, No. 1, Spring & Summer 2025, Original paper p. 93-104

10.22077/JWHR.2025.9094.1170

The Impact of Snow Cover on River Discharge Simulation: Insights from the Barandozchay River Basin

Halleh Hashemi[a], Hossein Rezaie[b&*], Keivan Khalili[c], Amin Amini Rakan[d]

[a]MSc Student, Department of Water Engineering, Urmia University, Urmia, Iran.
[b]Professor, Department of Water Engineering, Urmia University, Urmia, Iran.
[c]Associate Professor, Urmia Lake Research Institute, Urmia University, Urmia, Iran.
[d]Ph.D. Department of Water Engineering, Urmia University, Urmia, Iran.

[*]Corresponding Author, E-mail address: h.rezaie@urmia.ac.ir
Received: 11 March 2025/ **Revised**: 18 May 2025/ **Accepted**: 30 May 2025

Abstract

This study presents a comprehensive analysis aimed at predicting the discharge of the Barandozchay River using machine learning algorithms and meteorological data from both satellite and ground sources over the period from 2002 to 2022. The research highlights the significance of incorporating snow cover data in enhancing predictive accuracy, particularly during the spring and summer seasons. Utilizing Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forest (RF), the study evaluates various parameters affecting river discharge, including temperature, precipitation, and solar radiation. The results indicate that the Random Forest model outperforms the others in accuracy and generalization, while SVM demonstrates improved predictive capabilities with the inclusion of snow cover data. Specifically, the integration of snow cover data significantly enhanced the simulation accuracy of river discharge. The SVM model showed notable improvements in evaluation metrics, with R^2 increasing from 0.64 to 0.72, MAE decreasing from 0.4 to 0.61, and RMSE reducing from 0.81 to 0.29 in the test data. Conversely, the RF model experienced an increase in error for the test data, but the correlation coefficient R^2 improved from 0.85 to 0.88. The findings underscore the necessity of employing advanced machine learning techniques for water resource management, especially in regions facing water crises due to climate change.
Keywords: Artificial Neural Networks, Barandozchay River, Random Forest, Support Vector Machines.

1. Introduction

Predicting river discharge is a critical aspect of water resource management, playing a vital role in supplying drinking water, agricultural needs, and flood control. This is especially relevant in countries like Iran, which face water crises and climatic changes, underscoring the need for precise predictions of water flows. Given that most water resources are influenced by factors such as rainfall, temperature, humidity, and snow cover, understanding and analyzing these factors is essential for forecasting river flows (Wang et al., 2020; Kim et al., 2021; Panahi et al., 2021; Nguyen et al., 2022). Additionally, the recent decline in lake and river water levels due to drought and mismanagement emphasizes the necessity of effective predictive models. The intensification of drought and unsustainable management of water resources have led to a significant decrease in the water level of Lake Urmia in northwestern Iran, bringing this ecosystem to a critical point of no return.

In the context of river discharge forecasting, the utilization of data collected from meteorological stations and satellite sensors has increasingly gained attention. Meteorological data, used as inputs for predictive models, can encompass various atmospheric parameters such as rainfall, temperature, and solar radiation. Alongside these data, information regarding snow cover also holds special significance as a natural source influencing water discharge level. Examining changes and management of water resources in lakes and rivers using diverse models and assessing the impacts of human activities and climatic changes has been the subject of numerous studies. Below is a summary of the key findings and studies that have been mentioned:

Research by Hassanzadeh et al. (2012) revealed that 65% of the decrease in lake level is attributed to excessive surface water consumption, 25% due to dam construction, and 10% owing to reduced precipitation. Recommendations for preserving the current level of the lake include improved water resource management and agricultural practices.

Shadkam et al. (2016) indicated that the reduction in water input to Lake Urmia from 1960 to 2010 was influenced by climatic changes and the development of water resources. Climatic changes accounted for 60% of the reduction in water input, while water resource development accounted for 40%. Xie et al. (2016) utilized the ARMA-GARCH model to study flow variations in the Yangtze River, demonstrating similarities in variability across nearby stations. Additionally, Banihabib et al. (2017) employed the DARIMA-NARX model, which significantly enhanced water flow prediction accuracy. Yazdani and Zolfaghari (2017) determined that using the JEN neural network model achieved greater accuracy compared to other models. Furthermore, the study by Alizade Govarchin Ghale et al. (2018) emphasized the substantial impact of human activities on the contraction of Lake Urmia.

The results of Chaudhari et al. (2018) showed that human water management significantly decreased water flow to the lake. Schulz et al. (2020) also highlighted the negative impacts of human activities and climatic changes on the Lake Urmia crisis. Radman et al. (2022) investigated the use of various models, including RNN and LSTM, to forecast changes in Lake Urmia, pointing out the superior accuracy of the LSTM model relative to the others. In the study by Barbuleshcu and Zhen (2024), the ELM model was identified as the most effective for predicting discharge in the Buzau River. Moreover, Wei et al. (2024) demonstrated that the use of wavelet transforms and CNNs could enhance water flow predictions.

These studies collectively illustrate the increasing emphasis on leveraging meteorological data and machine learning algorithms for predicting and managing water resources. In recent decades, machine learning algorithms have emerged as efficient tools for hydrological predictions and water flow simulation across various global regions. These algorithms are rapidly gaining traction due to their robust capacity for processing nonlinear data and their good generalization capabilities, especially when historical data is limited (Chau et al., 2005; Nayak et al., 2005; Awchi, 2014). This study aims to forecast the discharge of the Barandozchay River utilizing machine learning algorithms, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forest (RF).

2. Materials and Methods

2.1. Study area

The Barandozchay watershed in the West Azerbaijan Province is considered a strategic region concerning water resources. Covering an area of 1203 square kilometers, this watershed originates from the mountainous border between Iran and Turkey and flows into the Urmia Plain. The Barandozchay River is regarded as an essential water source in this region due to its supply from mountain peaks and favorable precipitation patterns.

2.2. Data

2.2.1. Meteorological data

This study utilized three types of meteorological data:

1. Ground-Based Data: Meteorological stations "Babaroud" provided parameters including average temperature, rainfall, humidity, and solar radiation.
2. Satellite Data (LARC) program: We used four sets of data in this research. Two of data are observations from Bibakran and Gasemlou hydro meteorology stations. Locations of these stations are shown in Figure 1. We also used two sets of data from NASA (LaRC) power project which produces meteorological data using MERRA-2 archive. This data is available in 0.05 decimal degree grids around the world. The datasets and the

parameters used from each set are shown in Table 2. For modeling snow covered area in the basin, 40 parameters of monthly observations are used from February 2000 to October 2019.

3. Modis Data: Snow cover extent in the Barandozchay watershed was extracted using eight-day MODIS (MOD10A2) data in HDF format, obtained from the National Snow and Ice Data Center (NSIDC). The images, corresponding to tile h21v05.006, were imported and georeferenced in ArcMap. The watershed boundary was delineated using digital elevation models (DEMs). Snow cover images were clipped to the watershed extent, and snow cover areas were calculated for a 20-year period (2002–2022). The snow cover data were reloaded in ArcGIS, and raster layers were used to generate time series for trend analysis and future modeling. Monthly averages of snow cover extent were prepared for further analysis.

2.2.2. Discharge data from the Barandozchay River

The discharge data from the Barandozchay River Basin were obtained from the Babarud Hydrometric Station, covering the period from 2002 to 2022, and were used as the output variable for the models. The location of the Babarud Hydrometric Station is illustrated in Figure 1.

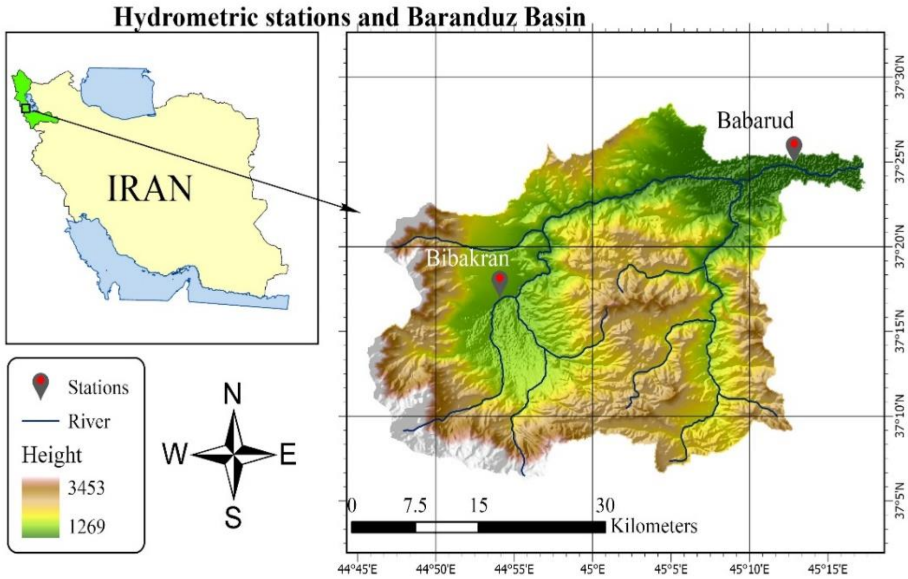

Fig. 1. Hydrometric stations and Barandoz Basin

Table 1. The geographic coordinates of the Babarud Hydrometric station

Latitude	Longitude	
37° 24′ 00″	45° 14′ 00″	Babarud
37° 17′ 00″	44° 54′ 00″	BiBakran

Tables 2 and 3 present the satellite-based meteorological parameters and ground-based meteorological parameters that were analyzed for their correlation with the discharge data from the Babarud Hydrometric Station in the Barandozchay Watershed. Based on the analysis of these correlations, the selected parameters include temperature (maximum, mean, and minimum), precipitation, solar radiation, and snow cover data. These variables were chosen as they exhibited significant relationships with the discharge measurements, making them critical inputs for the hydrological modeling process.

Table 2. Satellite meteorological parameters

Row	Parameters	Measurement Units
1	Mean Temperature	°C
2	Maximum Mean Temperature	°C
3	Minimum Mean Temperature	°C
4	Solar Radiation	MJ/m2·day
5	Infrared Radiation	MJ/m2·day
6	Relative Humidity	%
7	Dew Point Temperature	°C
8	Wind Speed at 2 Meters	m/s
9	Precipitation	Mm
10	Surface Pressure	kPa

Table 3. Ground-Based meteorological parameters

Row	Parameters	Measurement Units
1	Mean Temperature	°C
2	Maximum Mean Temperature	°C
3	Minimum Mean Temperature	°C
4	Evaporation	Mm
5	Precipitation	Mm
6	Maximum Temperature	°C
7	Minimum Temperature	°C
8	Relative Humidity at 12:30 PM	%
9	Relative Humidity at 6:30 AM	%
10	Relative Humidity at 6:30 PM	%

In this study, discharge data, along with satellite-based and ground-based meteorological data, were collected on a monthly scale for the period from 2002 to 2022. These datasets were utilized to investigate the impact of snow cover data on river discharge modeling using the R software.

2.3. Study process

The models utilized for predicting river discharge included Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forest (RF). Prior to employing these algorithms, a data preprocessing step was performed, which included:

a. Data Cleaning: Removal of incomplete data and imputation of missing values.

b. Data Standardization: The various measured parameter values were converted to a common scale to prevent any parameter from disrupting the model.

c. Correlation Analysis: This analysis determined which parameters had the most significant impact on discharge and exhibited higher correlations.

The relevant parameters studied are referenced in Tables 2 and 3. However, the main parameters used among these include precipitation, temperature (monthly average, minimum, and maximum), solar radiation, and snow surface data, which were extracted using MODIS. These parameters were evaluated based on the correlation analysis with watershed discharge using cross ACF for each parameter up to a 20-lag period.

2.4. Model Evaluation Method

The model evaluation criteria included:

1 - R^2: R^2 is typically used to assess how well the model fits the data, and its definition is as follows:

$$R^2 = 1 - \frac{\sum(Y_i - X_i)^2}{\sum(Y_i - \bar{Y})^2} \quad (1)$$

2 - RMSE: Root Mean Square Error, which measures the deviation of predictions from actual values.

$$RMSE = \sqrt{\frac{\sum_i^N (Y_i - X_i)^2}{N}} \quad (2)$$

3 - MAE: Mean Absolute Error, defined as the average of the absolute differences between predicted and actual values.

$$MAE = \frac{\sum_i^N |Y_i - X_i|}{N} \quad (3)$$

In the above relationships, Y_i and X_i represent the corresponding values of the observed and modeled series, respectively. Additionally, $\bar{Y}$ and $\bar{X}$ denote their means. N represents the number of values in the time series.

2.5. Modeling process

Modeling the Discharge of the Barandozchay River: Assessing the Influence of Snow Cover

Two distinct modeling frameworks were developed to investigate the impact of snow cover on the discharge of the Barandozchay River. Additionally, three different parameter selection strategies were implemented to optimize the modeling process. In the first approach, all available parameters were included in the model. The influence of snow cover was evaluated by running the model twice: once with snow cover data incorporated and once without. This allowed for a direct comparison of the model's performance with and without snow-related inputs.

In the second approach, the modeling was conducted using a set of parameters selected based on prior research and empirical studies conducted by the researchers. This approach relied on expert knowledge to identify the most relevant parameters for the model.

Finally, in the third approach, the Random Forest method was employed to select the most significant parameters. The model was then run twice: once with snow cover data included and once without, to assess the effect of snow cover on the model's output within this parameter selection framework. This multi-scenario approach provides a comprehensive evaluation of the role of snow cover in the hydrological modeling of the Barandozchay River, leveraging both empirical and data-driven methods for parameter selection.

2.6. Parameter selection using the random forest method

Random Forest is widely used for both prediction and variable selection due to its capacity to evaluate the relative importance of input variables during model construction. In this approach, variable importance is assessed based on the degree to which each variable reduces the impurity in the data when used to split nodes across the ensemble of decision trees. The most common measure of importance, known as the Mean Decrease in Impurity, reflects the cumulative reduction in variance (for regression) or classification error (for classification tasks) attributed to each variable. Following model training, variables are ranked according to their importance scores, and a selection is made by retaining those that exceed a defined threshold or by choosing the top-ranking variables.

This process helps to identify and retain the most relevant predictors, thereby improving model interpretability, reducing dimensionality, and potentially enhancing generalization performance. In this study the random forest methods give us 7 parameters to use in models. These parameters are, Snow, Precipitation, four temperature parameters and the daily solar radiation parameter (Genuer et al., 2010; Hapfelmeier and Ulm, 2013; Behnamian et al., 2017).

3. Results and Discussion

The results obtained from modeling and predicting the discharge of the Barandozchay River, utilizing artificial intelligence and meteorological data, are elaborated upon in the following section.

3.1. Data analysis

Initially, the collected data, including temperature, rainfall, solar radiation, and snow cover parameters, were analyzed to determine their correlation and influence on river discharge. Correlation analysis revealed that temperature (with a three-month lag), rainfall (with a two-month lag), and solar radiation (with a three-month lag) exhibited the highest correlation with river discharge and were included as inputs for the models. Additionally, snow cover data, with a lag of two to three months, significantly impacted river discharge, particularly during the spring and summer seasons.

3.2. Detailed analysis of tables and numerical evaluation

3.2.1. Case 1: Using all parameters

In this case, we trained and tested models using all available parameters, both with and without including the snow cover parameter. The key objective was to assess how snow cover influences the models' predictive performance.

Table 4. Modeling results using all parameters *without* the snow cover parameter for ANN, SVM, and RF models

	ANN		
	R^2	RMSE	MAE
Train	0.97	0.38	0.51
Test	0.83	0.61	0.4
	SVM		
Train	0.86	0.69	0.26
Test	0.82	0.55	0.38
	Random Forrest		
Train	0.94	0.24	0.13
Test	0.82	0.55	0.33

Table 5. Modeling results using all parameters *with* the snow cover parameter for ANN, SVM, and RF models

	ANN		
	R^2	RMSE	MAE
Train	0.99	0.15	0.7
Test	0.85	0.64	0.36
	SVM		
Train	0.91	0.54	0.26
Test	0.85	0.43	0.38
	Random Forrest		
Train	0.95	0.25	0.13
Test	0.74	0.52	0.3

According to Tables 4 and 5, integrating snow cover into the ANN model significantly improved its fit to the training data, with the R^2 value rising from 0.97 to 0.99. However, in the testing data, the increase in R^2 was only modest, from 0.83 to 0.85, indicating limited improvement in generalization. Additionally, while the RMSE dropped from 0.38 to 0.15 in the training data, it slightly increased from 0.61 to 0.64 in the testing data, suggesting a risk of overfitting. The MAE showed a minor enhancement in the testing data, decreasing from 0.40 to 0.36.

For the SVM model, the R^2 value in the training data increased from 0.86 to 0.91, yet it remained stable at 0.85 in the testing data. The RMSE improved from 0.69 to 0.54 in the training data and from 0.55 to 0.43 in the testing data, indicating a boost in predictive accuracy. The MAE stayed constant at 0.26 in the training data and 0.38 in the testing data, implying that snow cover did not significantly impact the model's error margins.

The Random Forest model exhibited a slight increase in R^2 from 0.94 to 0.95 in the training data, but the testing data showed a decline from 0.82 to 0.74, pointing to potential overfitting. The RMSE remained nearly unchanged in the training data (0.24 to 0.25) but decreased from 0.55 to 0.52 in the testing data, indicating a slight improvement in generalization. The MAE remained at 0.13 in the training data and decreased slightly from 0.33 to 0.30 in the testing data.

Overall, the inclusion of the snow cover parameter appears to enhance model performance, particularly for the ANN and SVM models. Although the Random Forest model showed some improvement in testing data, it also exhibited signs of overfitting. In conclusion, incorporating snow cover data proves beneficial for increasing the accuracy of discharge predictions.

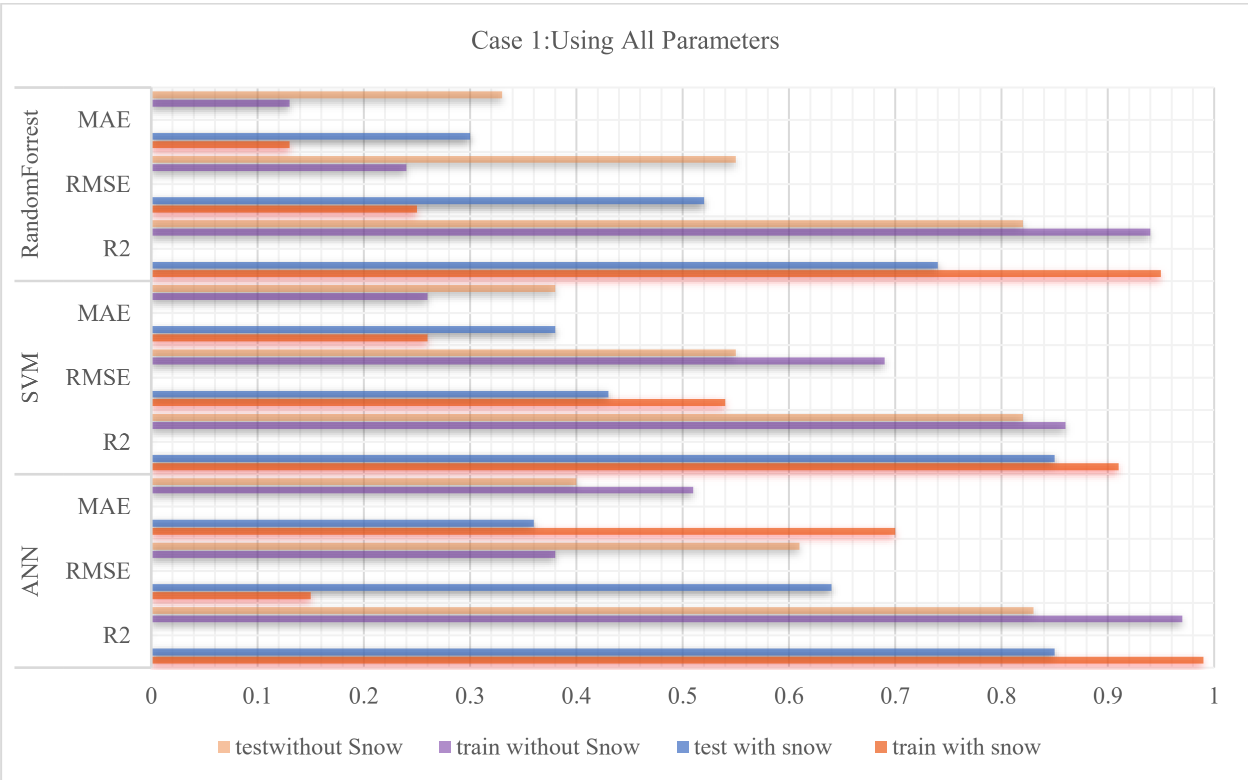

Fig 2. Modeling results using all parameters *with and without* the snow cover parameter for ANN, SVM, and RF models

The bar chart in Figure 2 illustrates the modeling results and the impact of the snow cover parameter on the predictive performance of the ANN, SVM, and RF models. The results indicate that the ANN model demonstrated a significant improvement in predictive capability when including snow cover in the training data. Similarly, the SVM model experienced notable enhancements in predictive accuracy, particularly in terms of RMSE values. In contrast, although the Random Forest model showed a slight improvement in accuracy with the inclusion of snow cover, the decrease in R^2 during testing suggests potential overfitting.

Overall, the bar chart clearly depicts the benefits of incorporating the snow cover parameter compared to not using it, highlighting its significance in improving streamflow predictions.

3.2.2. Case 2: Using selected parameters

In this case, an effort has been made to develop models using a selected set of parameters that includes various atmospheric variables associated with snow cover.

Table 6. Modeling results using the selected parameters for ANN, SVM, and RF models (With Snow)

	ANN		
	R^2	RMSE	MAE
Train	0.96	0.17	0.11
Test	0.8	0.7	0.4
	SVM		
Train	0.73	0.79	0.42
Test	0.69	0.59	0.27
	Random Forrest		
Train	0.92	0.29	0.15
Test	0.79	0.57	0.32

In this study, the researchers trained and tested models using a selected set of parameters that included various meteorological and snow cover-related variables.

As shown in Table 6, the ANN model performed well on the training data, achieving an R^2 of 0.96. However, its performance significantly declined when tested, with the R^2 dropping to 0.80, indicating that the model may be overfitting. This is further evidenced by an increase in RMSE from 0.17 in the training data to 0.70 in the testing data, and a rise in MAE from 0.11 to 0.40.

In contrast, the SVM model demonstrated relatively stable performance, with R^2 values of 0.73 for the training data and 0.69 for the testing data. Nevertheless, both the RMSE (0.79 in training and 0.59 in testing) and MAE (0.42 in training and 0.27 in testing) were higher than those of the other models, indicating lower overall accuracy.

The Random Forest model, on the other hand, achieved a good balance between training and testing performance, with R^2 values of 0.92 in the training data and 0.79 in the testing data. Its RMSE (0.29 in training and 0.57 in testing) and MAE (0.15 in training and 0.32 in testing) also showed a more balanced performance compared to the ANN and SVM models.

The analysis based on the results presented in Fig 3 illustrates the performance of the Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF) models when utilizing a selected set of atmospheric parameters related to snow cover. The bar chart visually emphasizes the disparity in model performance. The ANN model achieved a high training R^2 of 0.96 but demonstrated a significant decline to 0.80 on the testing data, indicating overfitting. This is further illustrated by the RMSE and MAE, both of which increased notably in the testing phase, underlining a loss of generalization. In contrast, the SVM model maintained relative stability across both datasets, but its overall accuracy lagged behind the other models.

The RF model showcased a favorable balance, with performance metrics that indicate a consistent capability to generalize from training to testing data. Hence, the bar chart effectively portrays these variances, particularly highlighting the Random Forest model's robustness. Consequently, it suggests that while the ANN may benefit from further tuning, the Random Forest model remains the most reliable choice among the three for predicting outcomes in this context.

In summary, while the ANN model exhibited signs of overfitting, the Random Forest model offered a more consistent performance across both training and testing datasets. The SVM model had the lowest accuracy among the three models. Therefore, we recommend the Random Forest model for this scenario, although further tuning of the ANN model might enhance its performance.

3.2.3. Case 3: Using parameters selected by the random forest method

In this case, the parameters selected using the Random Forest method were utilized for modeling, both with and without the snow cover variable. The inclusion of snow cover improved the training performance of the ANN model, as indicated by an increase in R^2 from 0.97 to 0.98. However, this came at the cost of testing performance, with R^2 dropping from 0.80 to 0.77, suggesting the model is likely overfitting. Additionally, the RMSE improved in the training data, decreasing from 0.22 to 0.13, but increased in the testing data from 0.61 to 0.75. Further supporting this overfitting scenario, the MAE also increased slightly in both training and testing datasets.

Table 7. Modeling results using parameters selected by the RF method *without* the snow cover parameter for ANN, SVM, and RF models

	ANN		
	R^2	RMSE	MAE
Train	0.97	0.22	0.12
Test	0.8	0.61	0.42
	SVM		
Train	0.74	0.62	0.3
Test	0.64	0.81	0.4
	Random Forrest		
Train	0.92	0.32	0.16
Test	0.85	0.28	0.25

Table 8. Modeling results using parameters selected by the RF method *with* the snow cover parameter for ANN, SVM, and RF models

	ANN		
	R^2	RMSE	MAE
Train	0.98	0.13	0.15
Test	0.77	0.75	0.51
	SVM		
Train	0.78	0.78	0.42
Test	0.72	0.72	0.29
	Random Forrest		
Train	0.93	0.26	0.14
Test	0.88	0.37	0.29

For the SVM model, the inclusion of snow cover resulted in moderate improvements in performance for both training and testing. The R^2 increased from 0.74 to 0.78 in the training data and from 0.64 to 0.72 in the testing data.

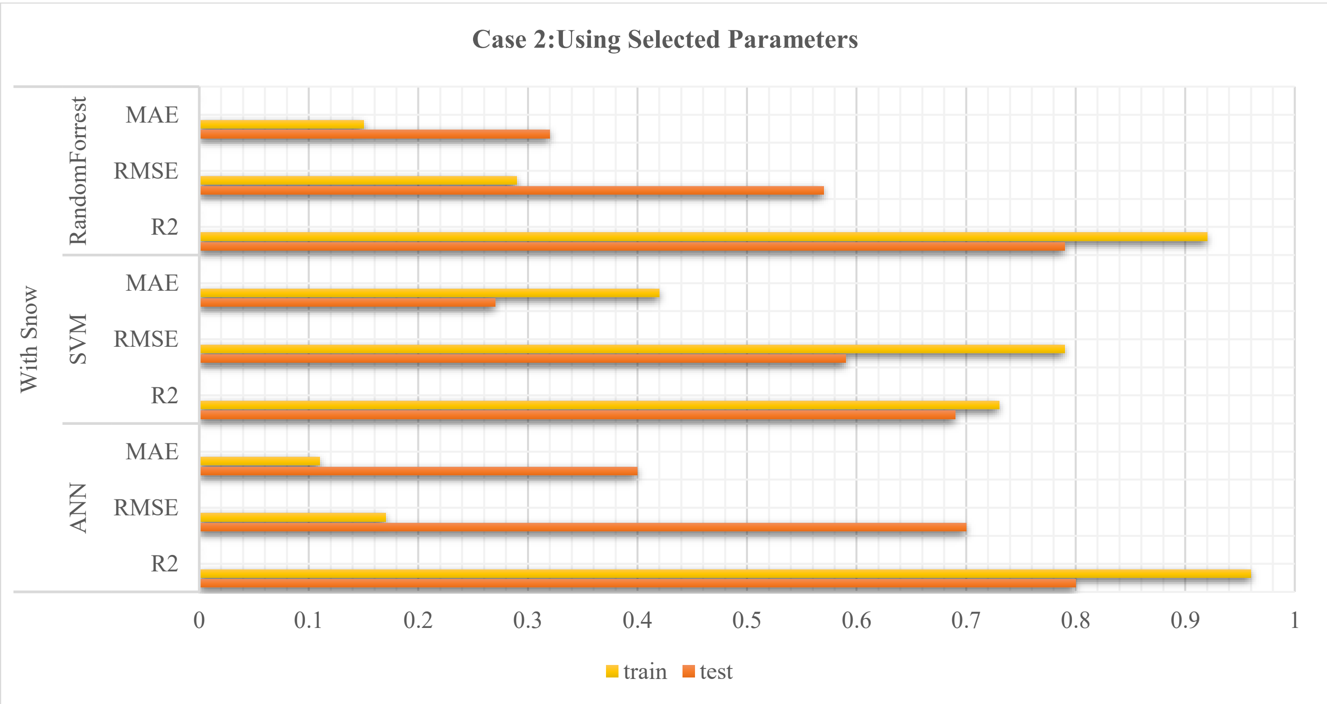

Fig 3. Modeling results using the selected parameters for ANN, SVM, and RF models

The RMSE decreased in the testing data from 0.81 to 0.61, while the MAE reduced from 0.40 to 0.29, indicating better generalization. The Random Forest model showed a slight enhancement in both training and testing performance with the inclusion of snow cover. The R^2 improved from 0.92 to 0.93 in the training data and from 0.85 to 0.88 in the testing data. Nevertheless, there was a slight increase in the RMSE in the testing data (from 0.28 to 0.37), and the MAE also increased slightly (from 0.25 to 0.29), indicating minor overfitting.

When analyzing the results of the river flow modeling, it was evident that while the inclusion of snow cover data in the Artificial Neural Network (ANN) improved the model's fit to the training data, it also led to a decrease in testing performance. This highlights a tendency toward overfitting, suggesting that the snow cover parameter may not be advantageous for the ANN model in this context. In contrast, the application of snow cover data in the Support Vector Machine (SVM) model significantly enhanced both training and testing performance, particularly by reducing RMSE and MAE, thereby indicating that the snow cover parameter helps the SVM model generalize better to unseen data.

The Random Forest model showed improvements in training fit with the inclusion of snow cover; however, there was a slight increase in testing error. This suggests that while the model does not display severe overfitting, the enhancement in testing performance is marginal, indicating that the positive impact of the snow cover parameter may be limited. In contrast, the SVM model benefits the most from including snow cover, demonstrating significant improvement in its ability to generalize to unseen data. Comparatively, the Random Forest model exhibits some gains as well, but these effects are less pronounced. The ANN model, however, tends to overfit in similar scenarios, highlighting the need for caution when using this parameter.

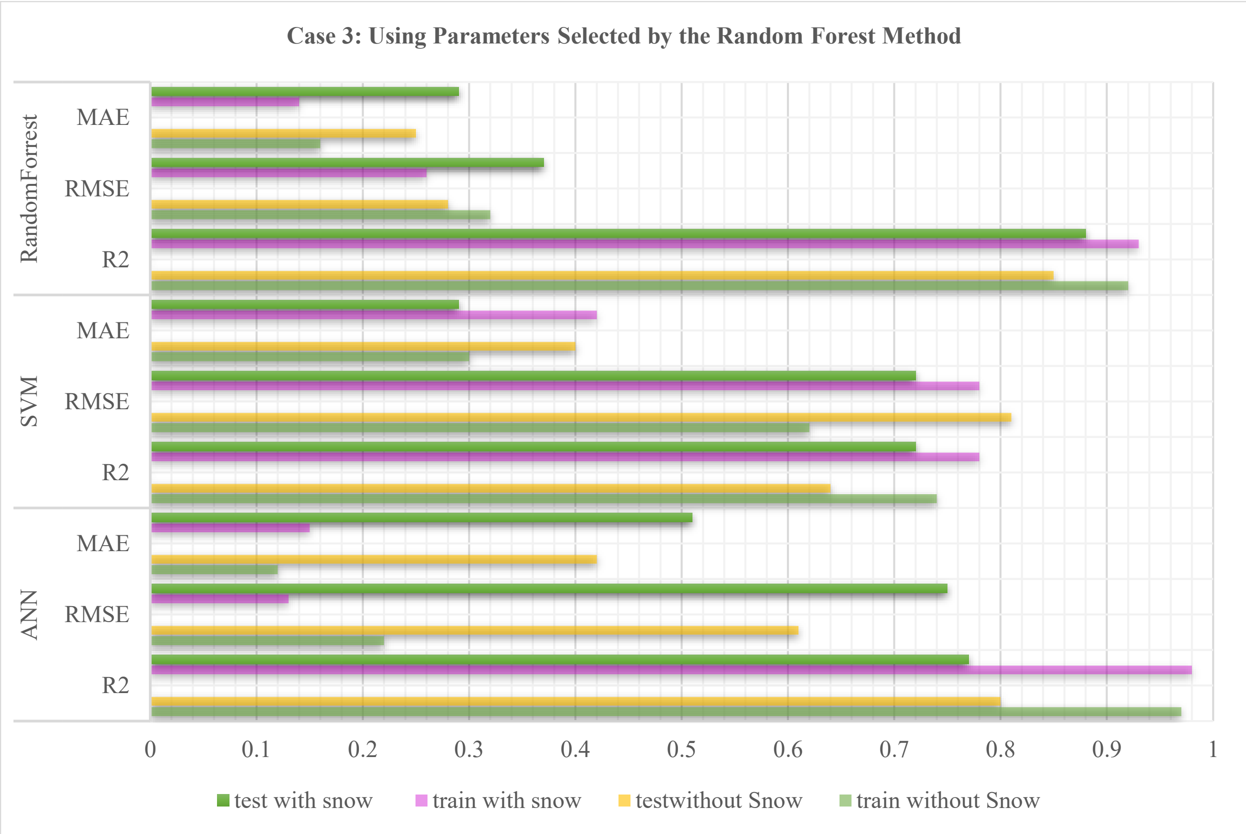

Fig 4. Comparison of the performance of ANN, SVM, and Random Forest models using parameters selected by the Random Forest method in Case 3

The bar chart illustrates the modeling results for the ANN, SVM, and Random Forest models utilizing parameters selected by the Random Forest method, with and without the inclusion of the snow cover variable. The data highlights notable trends in model performance. The ANN model demonstrated superior training R^2 when including the snow cover (0.98) compared to without it (0.97), yet its testing performance decreased from 0.80 to 0.77, indicating overfitting. In contrast, the SVM model benefited from snow cover, as evidenced by an increase in both training (from 0.74 to 0.78) and testing (from 0.64 to 0.72) performance metrics, suggesting better generalization. The Random Forest model also showed slight improvements in R^2 with snow cover, although the testing error increased slightly (0.37), reflecting less pronounced benefits. Overall, the chart effectively encapsulates the varying influences of the snow cover parameter on model accuracy, with the SVM exhibiting the most significant advantages in both training and testing settings (Fig 4).

These results align with the findings of Immerzeel et al. (2009), who stressed the importance of snow cover and melt in affecting river discharge in the Himalayan basin. This study supports the positive influence of incorporating snow cover data when modeling the discharge of the Barandozchay River, confirming its value as a crucial parameter in river flow simulations.

Moreover, Nijssen et al. (1997) employed the VIC-2L model to simulate river flow in the Columbia and Delaware basins, uncovering significant errors in the arid Columbia basin. In comparison, the Random Forest model in the present study outperforms the VIC-2L under similar arid conditions, primarily due to its capacity to model complex nonlinear relationships more effectively. Additionally, in the uniformly distributed precipitation and temperature conditions of the Delaware basin, the VIC-2L model achieved an acceptable relative error. However, the Random Forest model further demonstrated superior performance in watersheds with consistent hydrological conditions.

The findings of this study underscore the critical role of snow cover in enhancing the accuracy of river discharge predictions in the Barandozchay River basin. Incorporating snow cover data improved the performance of both the ANN and SVM models, particularly during the training phase, where the ANN model reached an R^2 of 0.99. However, the ANN model showed signs of overfitting, as the testing performance only slightly improved, with R^2 rising from 0.83 to 0.85. The SVM model exhibited more consistent improvements, with increases in both training and testing R^2 values, along with a significant reduction in RMSE and MAE, indicating better generalization to unseen data. The Random Forest model also experienced slight improvements in training performance, but there was a minor reduction in testing accuracy, suggesting limited benefits from including snow cover.

When using selected parameters, the ANN model again performed strongly during training but exhibited significant overfitting, with testing R^2 decreasing from 0.96 to 0.80. The SVM model maintained stable yet lower accuracy, while the Random Forest model delivered balanced performance, making it the most reliable option in this scenario. In cases where parameters were selected using the Random Forest method, incorporating snow cover enhanced the SVM model's generalization, with R^2 increasing from 0.64 to 0.72 in testing. However, the ANN model's testing performance declined, further highlighting overfitting. The Random Forest model demonstrated slight improvements but also minor increases in testing error, suggesting a limited positive impact of snow cover.

In summary, the SVM model gained the most from including snow cover, showing substantial improvements in its generalization capabilities. The Random Forest model displayed some enhancements, albeit less pronounced, while the ANN model exhibited a tendency to overfit, particularly when the snow cover parameter was included. These findings corroborate previous studies, such as that by Immerzeel et al. (2009), which emphasized the significance of snow cover in river discharge modeling, especially in snowmelt-dominated basins.

4. Conclusion

In conclusion, this study highlights the essential role of integrating satellite-derived snow cover data with machine learning techniques to improve the accuracy of river discharge predictions in the Barandozchay River basin. Including snow cover parameters significantly enhanced the performance of the models, especially during the spring and summer months, underscoring the importance of monitoring variations in snow cover for effective hydrological modeling.

Among the machine learning algorithms utilized, the Random Forest (RF) model proved to be the most effective, showing high accuracy with low error rates. This emphasizes its capability to manage non-linear and complex datasets. Correlation analysis further revealed that significant parameters, including precipitation, temperature, and snow cover, are the most influential factors affecting river discharge. This information provides valuable insights that can aid in water resource management and forecasting.

The findings of this research underscore the need for innovative and interdisciplinary approaches in hydrological forecasting, especially in light of climate change. By utilizing machine learning algorithms and integrating meteorological and remote sensing data, this study presents a solid framework for improving water flow predictions in areas dominated by snowmelt. This is particularly crucial for regions like the Barandozchay River basin, which face significant challenges regarding water resources. Our results also indicate the potential of machine learning models, particularly RF and SVM, to deliver reliable and accurate predictions, thus facilitating better decision-making in water resource management.

Future research should aim to optimize model parameters and investigate hybrid approaches that combine process-based and empirical models for even greater predictive accuracy. Moreover, the integration of advanced remote sensing data along with machine learning techniques could unveil promising opportunities for enhancing water resource management in similar basins. Overall, this study adds to the expanding knowledge on the application of machine learning in hydrology, laying a foundation for more effective planning and management of water resources in regions reliant on snowmelt.

In summary, this study not only improves predictions of the Barandozchay River discharge but also demonstrates that machine learning algorithms combined with various datasets can be powerful tools for hydrological forecasting amidst the challenges posed by climate change and water resource management. Ultimately, these findings can support informed and sustainable decision-making, optimize water resources and mitigate the adverse impacts of climate change.

5. Disclosure statement

No potential conflict of interest was reported by the authors.

6. References

Alizade Govarchin Ghale, Y., Altunkaynak, A., & Unal, A. (2018). Investigation anthropogenic impacts and climate factors on drying up of Urmia Lake using water budget and drought analysis. *Water Resources Management, 32,* 325-337.

Awchi, T. A. (2014). River discharges forecasting in northern Iraq using different ANN techniques. *Water resources management, 28,* 801-814.

Banihabib, M. E., Ahmadian, A., & Jamali, F. S. (2017). Hybrid DARIMA-NARX model for forecasting long-term daily inflow to Dez reservoir using the North Atlantic Oscillation (NAO) and rainfall data. *GeoResJ, 13*, 9-16.

Bărbulescu, A., & Zhen, L. (2024). Forecasting the River Water Discharge by Artificial Intelligence Methods. *Water, 16*(9), 1248.

Behnamian, A., Millard, K., Banks, S. N., White, L., Richardson, M., & Pasher, J. (2017). A systematic approach for variable selection with random forests: achieving stable variable importance values. *IEEE Geoscience and Remote Sensing Letters*, *14*(11), 1988-1992.

Chau, K. W., Wu, C. L., & Li, Y. S. (2005). Comparison of several flood forecasting models in Yangtze River. *Journal of Hydrologic Engineering, 10*(6), 485-491.

Chaudhari, S., Felfelani, F., Shin, S., & Pokhrel, Y. (2018). Climate and anthropogenic contributions to the desiccation of the second largest saline lake in the twentieth century. *Journal of Hydrology, 560*, 342-353.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern recognition letters, 31*(14), 2225-2236.

Hapfelmeier, A., & Ulm, K. (2013). A new variable selection approach using random forests. *Computational Statistics & Data Analysis, 60*, 50-69.

Hassanzadeh, E., Zarghami, M., & Hassanzadeh, Y. (2012). Determining the main factors in declining the Urmia Lake level by using system dynamics modeling. *Water Resources Management, 26*, 129-145.

Immerzeel, W. W., Droogers, P., De Jong, S. M., & Bierkens, M. F. P. (2009). Large-scale monitoring of snow cover and runoff simulation in Himalayan River basins using remote sensing. *Remote sensing of Environment, 113*(1), 40-49.

Kim, T., Yang, T., Gao, S., Zhang, L., Ding, Z., Wen, X., ... & Hong, Y. (2021). Can artificial intelligence and data-driven machine learning models match or even replace process-driven hydrologic models for streamflow simulation?: A case study of four watersheds with different hydro-climatic regions across the CONUS. *Journal of Hydrology, 598*, 126423.

Nayak, P. C., Sudheer, K. P., Rangan, D. M., & Ramasastri, K. S. (2005). Short-term flood forecasting with a neurofuzzy model. *Water Resources Research, 41*(4).

Nguyen, D. H., Le, X. H., Anh, D. T., Kim, S. H., & Bae, D. H. (2022). Hourly streamflow forecasting using a Bayesian additive regression tree model hybridized with a genetic algorithm. *Journal of Hydrology, 606*, 127445.

Nijssen, B., Lettenmaier, D. P., Liang, X., Wetzel, S. W., & Wood, E. F. (1997). Streamflow simulation for continental- scale river basins. *Water Resources Research*, *33*(4), 711-724

Panahi, F., Ehteram, M., Ahmed, A. N., Huang, Y. F., Mosavi, A., & El-Shafie, A. (2021). Streamflow prediction with large climate indices using several hybrids multilayer perceptrons and copula Bayesian model averaging. *Ecological Indicators, 133*, 108285.

Radman, A., Akhoondzadeh, M., & Hosseiny, B. (2022). Monitoring and predicting temporal changes of Urmia Lake and its basin using satellite multi-sensor data and deep-learning algorithms. PFG–Journal of Photogrammetry, *Remote Sensing and Geoinformation Science, 90*(3), 319-335.

Schulz, S., Darehshouri, S., Hassanzadeh, E., Tajrishy, M., & Schüth, C. (2020). Climate change or irrigated agriculture-what drives the water level decline of Lake Urmia. *Scientific reports, 10*(1), 236.

Shadkam, S., Ludwig, F., van Oel, P., Kirmit, Ç., & Kabat, P. (2016). Impacts of climate change and water resources development on the declining inflow into Iran's Urmia Lake. *Journal of Great Lakes Research, 42*(5), 942-952.

Wang, J., Wang, X., hui Lei, X., Wang, H., hua Zhang, X., jun You, J., ... & lian Liu, X. (2020). Teleconnection analysis of monthly streamflow using ensemble empirical mode decomposition. *Journal of Hydrology, 582*, 124411.

Wei, Yaxing, Huzaifa Bin Hashim, Sai Hin Lai, Kai Lun Chong, Yuk Feng Huang, Ali Najah Ahmed, Mohsen Sherif, and Ahmed El-Shafie. "Comparative analysis of artificial intelligence methods for streamflow forecasting." *IEEe Access* 12 (2024): 10865-10885.

Xie, H., Li, D., & Xiong, L. (2016). Exploring the regional variance using ARMA-GARCH models. *Water Resources Management, 30*, 3507-3518.

Yazdani, M. R., & Zolfaghari, A. A. (2017). Monthly river forecasting using instance-based learning methods and climatic parameters. *Journal of Hydrologic Engineering, 22*(6), 04017002.