



Original scientific paper

UDC: 556.5:004.85(282.243.7)(437.611)
<https://doi.org/10.2298/IJGI251106012L>

Received: November 6, 2025

Reviewed: February 23, 2026

Accepted: June 8, 2026



COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR DISCHARGE PREDICTION: A CASE STUDY AT THE BRATISLAVA STATION ON THE DANUBE RIVER

Igor Leščešen¹ * , Pavla Pekárová¹ , Pavol Miklánek¹ , Zbyněk Bajtek¹

¹Institute of Hydrology, Slovak Academy of Sciences, Bratislava, Slovakia; e-mails: lescesen@uh.savba.sk; pekarova@uh.savba.sk; pavol.miklanek@savba.sk; bajtek@uh.savba.sk

Abstract: This study investigates the prediction of river flows for water resource management, flood and drought protection and hydropower planning, focusing on the limitations of traditional hydrological models due to complex data requirements. Machine learning (ML) models offer a solution by capturing non-linear relationships to improve prediction accuracy. We compare the performance of three data-driven models: Random Forest (RF), Gradient Boosting (GB), and Seasonal Autoregressive Integrated Moving Average (SARIMA), with a simple Linear Regression (LR) benchmark for monthly discharge forecasting at the Bratislava station on the Danube River. RF model performed best, with an MAE of 165.84 and an R^2 of 0.89. GB achieved moderate results ($MAE = 335.15$; $R^2 = 0.60$), while SARIMA performed below average ($MAE = 550.78$; $R^2 = 0.01$). Results are valid within a univariate autoregressive framework; multivariate extensions incorporating meteorological drivers are recommended for operational deployment. These results underline the potential of machine learning models, especially RF, for forecasting monthly river discharge up to 24 months in advance at the Bratislava gauging station on the Danube River. While these findings are based on a single-station, univariate analysis, they contribute to the growing body of evidence supporting the use of ensemble machine learning methods in operational hydrology for large regulated river systems. Transferability to other stations or basins would require further validation.

Keywords: river discharge forecasting; machine learning models; hydrological prediction; time series analysis; Danube River

1. Introduction

Predicting actual river discharge is crucial for effective risk assessment, developing flood and drought protection strategies and establishing early warning systems, hydropower generation, and irrigation planning (Pham et al., 2021; Rahimzad et al., 2021; Yaseen et al., 2016). Despite this importance, achieving high predictive accuracy remains challenging in many river basins worldwide, environmental data such as precipitation, temperature, and evapotranspiration are still scarce or of uncertain quality, whereas the Bratislava station on the Danube benefits from one of the longest and most complete discharge records in Central Europe (1901–2020), together with the access to long-term meteorological observations.

*Corresponding author, e-mail: lescesen@uh.savba.sk

Traditionally, river discharge prediction has relied on physically based hydrological models that simulate catchment processes. However, these models often require complex parameterization, extensive calibration, and high-quality data inputs. Their performance is limited by difficulties in capturing nonlinear relationships in river systems (Schneider et al., 2024; Yaseen, 2023).

To address these limitations, a data-driven approach that uses only river discharge as the input variable was adopted in this study. By selecting this approach, reliance was increased on empirical machine learning (ML) methods that are able to learn directly from observed patterns (Mehta et al., 2023). ML algorithms can be dynamically adapted to the new data and capable of revealing hidden patterns within them (Kumar et al., 2023).

Since the beginning of the 21st century, the application of different ML techniques has increased exponentially, motivated by the need for accurate, reliable and flexible tools that can enable hydrological forecasts under the changing climate. Early works utilized artificial neural networks which are known for their capability to seize nonlinear relationships between hydrological and meteorological data (Yaseen et al., 2016). In recent years, Deep learn models (e.g., long-short term memory—LSTM), known for effectively capturing temporal dependencies in time-series, have been introduced (Leščešen, Tanhapour et al., 2025). Deep learning models have shown excellent skills in forecasting extreme hydrological events (both floods and droughts; Leščešen, Pekárová et al., 2025; Rahimzad et al., 2021). In the last couple of years, ensemble learning methods such as random forest (RF) and gradient boosting (GB) have also shown reliable results by merging several weak learners to boost prediction accuracy and constancy (Kilinc et al., 2023). Hybrid approaches which combine ML and physical frameworks are increasingly being used to utilize the strengths of both modeling methods (Saleh et al., 2024).

Although these models are widely applied, ML approaches have some limitations. First, their interpretability is frequently lower than that of process-based models, and second, their performance is usually affected when available data are limited or uncertain (Kedam et al., 2024). Despite these challenges, they represent a promising path toward scalable, adaptable hydrological forecasting in the context of climate variability. In addition, multiple studies have confirmed that ML algorithms, including RF (Ali et al., 2020) and GB (Schoppa et al., 2020; Talukdar et al., 2023), outperform traditional hydrological models in capturing nonlinear relations between meteorological and hydrological variables. Time-series methods like the Seasonal Autoregressive Integrated Moving Average (SARIMA) model are considered a valuable tool for identifying seasonal and temporal trends in river discharge (Ahmadpour et al., 2023; Leščešen, Pekárová et al., 2025; Zhang et al., 2022) as well. However, all of these statistical approaches rely heavily on the assumption of data stationarity and may not adequately capture future variability.

The global ML-based literature on hydrological forecasting is vast, but geographically uneven. The majority of the studies are conducted on river basins in North America (Grenier et al., 2024; Konapala et al., 2020), East Asia (Wang et al., 2026; Yao et al., 2023), South Australia (Clark & Jaffrés, 2025; Dikshit et al., 2020), and Western Europe (Defontaine, 2024; Hauswirth et al., 2021; Khozani et al., 2025), while Central and South-Eastern Europe, particularly large river systems like Danube, are underrepresented. Given the Danube's strategic importance for regional water security, flood management, and navigation, this research aims to address that gap. The objective of this study is to evaluate the performance of the selected ML models in forecasting biennial streamflow of the Danube River at the Bratislava gauging station. By focusing on a data-rich yet underexplored region, this study contributes to a broader understanding of AI-driven hydrological modeling under climatic and anthropogenic pressures.

2. Study area, data and methods

2.1. Study area and data

The research was conducted on single hydrological station on the Danube River, Bratislava station (48.1397°N, 17.1082°E) with the catchment area of 131,338 km² (Halmová et al., 2021; Figure 1). Database comprises of mean monthly discharge values with no missing data, for the period 1901–2020, thus making it 1,440 continuous monthly observations. Descriptive statistics of the discharge data are summarized in Table 1.

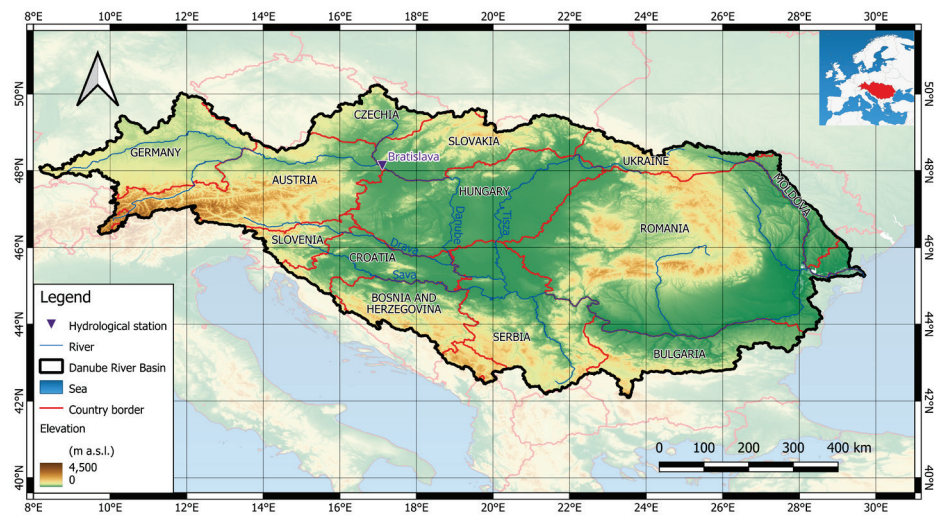


Figure 1. Bratislava gauging location within the Danube River Basin.

Table 1. Descriptive statistics of the mean monthly river discharge (m³/s) at the Bratislava gauging station on the Danube River for the period 1901–2020

Parameter	N	M (m ³ /s)	Mdn (m ³ /s)	Mode (m ³ /s)	SD (m ³ /s)	Variance	Skewness	Kurtosis	Range	Minimum (m ³ /s)	Maximum (m ³ /s)
Value	1,440	2,038	1,915	1,366	784.57	615,553.56	1.004	2.194	6,740	629	7,369

2.2 Data preprocessing and model setup

Before modeling, a thorough data quality and homogeneity evaluation was conducted to check that dataset met fundamental assumptions that are necessary for time series modeling. First, the time series was visually inspected for anomalies or abrupt variance changes that can be caused by measurements inconsistencies (Figure 2). Next, the statistical homogeneity was tested by applying two non-parametric tests, the Pettitt test and the Standard Normal Homogeneity Test (Firat et al., 2011). Both tests detect change points (if any) and evaluate if the mean of the series remains constant over time. The results of

these tests have shown that there were no statistically significant change points within the data series and thus the Bratislava discharge record can be considered homogeneous and suitable for use in long-term modeling applications. Outlier detection was performed using the interquartile range method (Wan et al., 2014). In the training set (1901–2004), 12 outliers were identified, representing 1.2% of the observations. All the identified extreme values were cross-checked against historical records of major flood and low-flow events. The most prominent outliers correspond to the well-documented floods in 1965 and 2013. All outliers were retained in the dataset to preserve natural hydrological variability and avoid artificial suppression of extreme events.

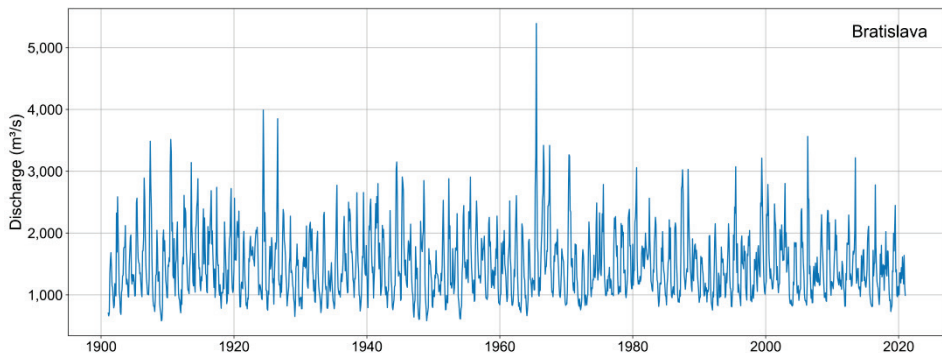


Figure 2. Time series of mean monthly discharge (m^3/s) at the Bratislava gauging station 1901–2020.

Note. The plot illustrates seasonal and interannual variability, as well as notable high- and low-flow events, highlighting the long-term hydrological dynamics of the Danube River at this location.

To prepare the dataset for machine learning, discharge values were standardized via z-score normalization to ensure a zero mean and unit variance. This transformation improves model convergence and prevents optimization algorithms from being influenced by the original data magnitudes. Because this study evaluates and forecasts river discharge, a univariate modeling framework was utilized, where past observations predict future discharge. Excluding meteorological or external variables (e.g., precipitation, temperature, evapotranspiration) allowed the models to focus exclusively on the autoregressive structure of the signals. Future extensions will incorporate exogenous predictors to explore multivariate influences under climate variability.

The full dataset was divided chronologically into training (80%; 1901–2004) and testing (20%; 2005–2020) subsets (Farfán-Durán & Cea, 2024; Zhu et al., 2025). To ensure temporal stability and avoid distributional shifts between periods, key statistical properties of the training and testing periods were compared. The training period had a mean discharge of $2,042.80 \text{ m}^3/\text{s}$ ($SD = 801.43 \text{ m}^3/\text{s}$, skewness = 1.025), while the testing period had a mean of $2,020.39 \text{ m}^3/\text{s}$ ($SD = 714.18 \text{ m}^3/\text{s}$, skewness = 1.118). The coefficients of variation were also similar (39.23% in training vs. 35.35% in testing). Highly comparable summary statistics confirm the chronological split preserved distributional stability, validating out-of-sample evaluation and increasing confidence in model performance. To optimize hyperparameters and prevent overfitting, a five-fold rolling-window cross-validation procedure was executed using scikit-learn's `TimeSeriesSplit`. This approach respects temporal ordering by ensuring each validation fold strictly succeeds its training fold, preventing future data leakage. Random

k-fold cross-validation was deliberately avoided, as it violates the sequential nature of hydrological time series. The objective was to forecast monthly discharge up to 24 months ahead. For data-driven models (RF, GB, Linear Regression [LR]), a recursive one-step-ahead prediction framework was formulated. A fixed-length sliding window of twelve months served as input, capturing the full annual cycle and flow persistence. The prediction for month $t+1$ is appended to the sequence, generating Q_{t+2} iteratively until the 24-month horizon. The baseline SARIMA model generated 24-step-ahead forecasts using fitted equations. However, this recursive multi-step approach inevitably causes error accumulation. By step 24, the input vector contains 23 previously predicted values, increasing uncertainty and limiting long-term interpretability. This known autoregressive limitation is acknowledged to ensure transparency for operational water-resource applications. Systematic sensitivity analysis of input window lengths was beyond this study's scope, and remains an important avenue for future research.

Three machine learning approaches were evaluated: RF, GB, and SARIMA, alongside LR as a baseline benchmark. RF and GB were selected for their proven robustness in hydrological applications, ability to model nonlinear relationships within small datasets, and modest computational demands. SARIMA serves as a traditional statistical benchmark representing the type of model still widely deployed in operational hydrology. Deep learning architectures, such as LSTM and Temporal Convolutional Networks (TCNs), were excluded because the available dataset contains only 1,440 monthly observations. This relatively limited size poses an overfitting risk for high-capacity models. Instead, our primary objective is benchmarking well-established, data-efficient machine learning techniques against a traditional time-series model in a univariate setting. Future work will extend this framework to deep learning using higher-resolution, multivariate data. The RF algorithm (Hasan et al., 2024) is an ensemble method aggregating predictions from multiple decision trees built on bootstrap samples, effectively capturing complex predictor interactions (Fang et al., 2024; Pham et al., 2021). Grid search optimized the RF hyperparameters: the number of trees ($n_{\text{estimators}} = 2000$), maximum tree depth ($\text{max_depth} = 12$), and the minimum samples per split ($\text{min_samples_split} = 2$). The final configuration was selected based on the lowest cross-validated *RMSE*. Random search and Bayesian optimization were not required given the small grid size (≤ 36 combinations per model); exhaustive evaluation was performed within acceptable computation time. The GB model constructs sequential trees, each correcting the residuals of its predecessors. This model is known for its flexibility, for minimizing a custom loss function, and for handling complex nonlinear patterns (Haghdoust et al., 2024; Kumar et al., 2023; Ni et al., 2020). This makes GB a valuable method for discharge forecast. GB hyperparameters were tuned via the same exhaustive grid search with five-fold cross-validation, testing learning rates $\in \{0.01, 0.05, 0.10\}$, boosting stages $\in \{100, 200, 500\}$, and maximum depths $\in \{3, 5, 7\}$ (27 combinations). The optimal configuration was learning rate = 0.05, $n_{\text{estimators}} = 500$, and $\text{max_depth} = 5$. Full grid ranges and selected values for both models are reported in Appendix Table 1. To handle the seasonality and autocorrelation found in river discharge, a SARIMA model (Khodakhah et al., 2022; Ahmadpour et al., 2022) was included. The general structure is denoted as SARIMA $(p, d, q) \times (P, D, Q, s)$, where p , d , and q are the non-seasonal components, and P , D , Q , and s are the seasonal components. The model parameters were determined autocorrelation (ACF) and partial autocorrelation (PACF) plots. Although a pure AR specification (without MA terms) is atypical for hydrological series, it was selected because the ACF/PACF of the training residuals showed exponential decay after seasonal differencing, and adding MA terms consistently increased Akaike information criterion (AIC). Then, fine-tuning of the model was

conducted by checking its AIC and Bayesian information criterion (BIC) results. As a verification step, an exhaustive AIC-minimisation grid search over $p, q \in \{0, 1, 2, 3\}$ and $P, Q \in \{0, 1, 2\}$ confirmed the $(2, 0, 0) \times (2, 0, 0, 12)$ specification as optimal for the training period. This method enabled us to select the most accurate model (Kabbilawsh et al., 2022; Valipour, 2015). Finally, we included a simple LR model as a baseline benchmark to quantify the added value of nonlinear ML and statistical approaches. LR assumes a linear relationship between past and future discharges, is computationally efficient, and remains widely used in hydrological practice as a first-order approximation. This makes LR easy to interpret but insufficient for capturing the complex, nonlinear behavior of river systems, and it therefore provides a useful lower bound against which to assess the gains achieved by RF, GB, and SARIMA.

To assess model performance, multiple evaluation metrics, that capture different aspects of predictive accuracy, were used. These included Mean Squared Error (*MSE*), Root Mean Squared Error (*RMSE*), Mean Absolute Error (*MAE*), and the Coefficient of Determination (R^2). In addition, two hydrologically oriented performance indicators were incorporated: the Nash–Sutcliffe Efficiency (*NSE*; Nash & Sutcliffe, 1970) and the Kling–Gupta Efficiency (*KGE*; Gupta et al., 2009). *NSE* evaluates the relative magnitude of residual variance compared to the variance of the observed data, indicating how well the simulated discharge matches the observed dynamics. *KGE* offers balanced evaluation by implementing correlation, bias, and variability components, which are crucial for hydrological forecasting. Both tests' values, *NSE* and *KGE*, are close to 1 which shows excellent model performance. Every analysis was in Python, by applying scikit-learn for ML models, pandas for data manipulation, and matplotlib for visualization.

3. Results

The evaluation of the river discharge forecasting models reveals significant differences in predictive performance across the various methodologies. Table 2 summarizes the key evaluation metrics, providing insights into the effectiveness of each model in accurately forecasting river discharge.

Table 2. Comparative evaluation metrics for four river discharge forecasting models, RF, GB, SARIMA, and LR, applied to the Bratislava discharge time series

Evaluation metric	RF	GB	SARIMA	LR
<i>MAE</i> (m ³ /s)	183.26	335.15	550.78	336.88
<i>MSE</i> (m ³ /s)	55,386.67	197,193.39	503,297.67	194,381.41
<i>RMSE</i> (m ³ /s)	235.34	444.06	709.43	440.89
R^2	0.89	0.60	0.01	0.61
AIC	–	–	22,581.74	–
BIC	–	–	22,613.37	–
<i>NSE</i>	0.88	0.59	–0.12	0.57
<i>KGE</i>	0.83	0.63	0.21	0.60

Note. AIC and BIC are reported only for the SARIMA model, for which a maximum-likelihood framework is defined. These metrics are not applicable to tree-based ensemble methods (RF, GB) or LR and have been omitted accordingly.

Test results presented in Table 2 clearly show that SARIMA model has underperformed, with high error metrics (*MAE*, *MSE* and *RMSE*) and low R^2 value, indicating that this model fails to capture fundamental discharge dynamics and thus is inadequate for modeling river discharge (Wang et al., 2017). GB model performs moderately, with acceptable accuracy

though still inferior to RF. The LR produces mixed results, with noticeably higher error metrics and weaker NSE, and KGE demonstrating its limited ability to handle nonlinear relationships in river discharge. The best performing model according to all of the test results is RF, with the lowest error metrics, highest R^2 , NSE and KGE values confirming its suitability for discharge forecasting.

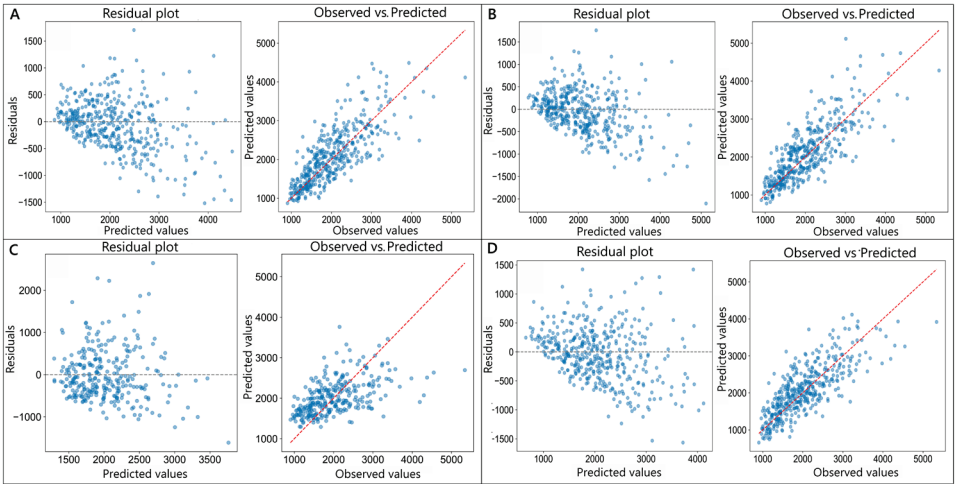


Figure 3. Residual analysis and scatter plots of observed versus predicted monthly discharge (m^3/s) for the four forecasting models: RF (A), GB (B), SARIMA (C), and LR(D).

Note. For all panels (A–D), x- and y-axes are scaled consistently across models to enable direct visual comparison. Residuals are plotted to illustrate prediction errors, model bias, and variance, while observed–predicted scatter plots assess model fidelity in reproducing the full range of discharge values.

Figure 3 visually compares observed vs. predicted values and residuals for all the four models. The RF (Panel A) clearly demonstrates the tight clustering around the line and symmetrical distribution of the residuals, further confirming its high accuracy. Panel C (SARIMA model) clearly illustrates models’ weakness with widely dispersed residuals and heteroscedastic, which is aligned with its weak forecasting capacity (Hansen & Nelson, 2003). GB (Panel B) performs moderately well, while LR (Panel D) displays systematic residual patterns, indicating lower robustness.

To further assess the hydrological realism of the simulations, a set of basic hydrological signatures computed from the observed and modeled discharge series was compared over the test period (2005–2020). These included the long-term mean monthly discharge, the coefficient of variation of monthly flows (CV), and the 10th and 90th percentile flows, which together characterize the central tendency, variability, and low- and high-flow behavior of the series. A summary of these signatures is provided in Table 3. RF model reproduces all signatures most closely, with only minor deviations from the observed mean and quantiles. GB and LR show slightly larger discrepancies, particularly in the low-flow quantile (10th percentile). SARIMA exhibits the largest biases, especially in variability (CV) and high-flow behavior (90th percentile), consistent with its poor performance in the error-based metrics. The two-sample Kolmogorov–Smirnov test showed no significant distributional differences between models and observations at $\alpha = 0.05$. However, this result requires caution: the test

has low power with the available sample size ($n \approx 192$) and a non-significant p -value does not prove equivalence. Given the large operational Q90 bias for SARIMA (–24%), percentage biases for Q10 and Q90 are reported in Table 3 for all models.

Table 3. Comparison of selected hydrological signatures computed from observed and modeled discharge series for the test period (2005–2020)

Model	$\bar{Q}$ (m ³ /s)	CV	Q10 (m ³ /s)	Q90 (m ³ /s)	KS p -value	Q10 bias (%)	Q90 bias (%)
Observed	2016.13	0.351	1226.12	2966.94	–	–	–
RF	2039.27	0.303	1340.09	2962.75	0.393 (n.s.)	+9.3	–0.1
GB	2012.34	0.325	1297.83	2805.28	0.495 (n.s.)	+5.8	–5.5
SARIMA	1708.61	0.261	1211.18	2266.54	0.172 (n.s.)	–1.2	–24.0
LR	2009.57	0.287	1283.08	2777.54	0.266 (n.s.)	+4.6	–6.4

Note. n.s. = not significant at $\alpha = 0.05$.

It can be concluded that RF provides the most accurate and reliable forecasts. Based on these findings, the RF model was further applied for 24-month discharge prediction at the Danube–Bratislava gauging station.

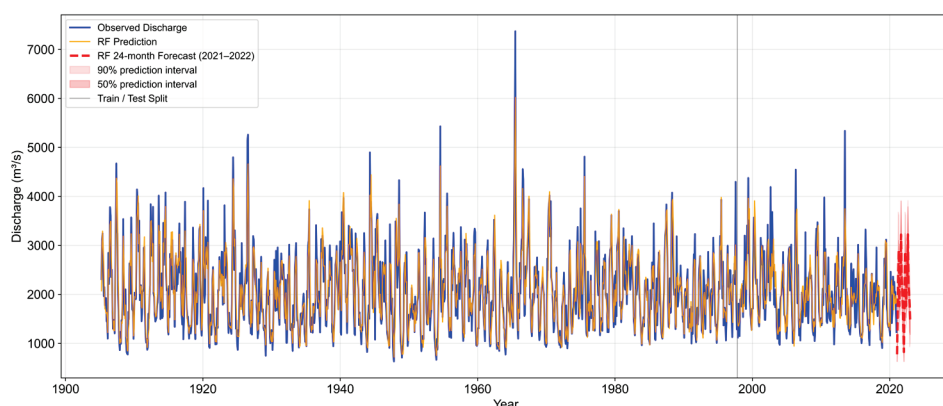


Figure 4. Comparison of observed mean monthly discharge (m³/s) and RF model predictions over the full study period (1901–2020), including a 24-month forecast horizon.

Note. The plot demonstrates the model's ability to reproduce seasonal patterns and extreme flow events, highlighting its predictive performance over short- and medium-term horizons. In addition to the point forecast, 50% and 90% prediction intervals are shown as shaded bands, derived from the distribution of predictions across 200 individual decision trees within the RF ensemble.

Operational hydrology requires probabilistic forecasts to support risk-based decisions on flood protection and water allocation. The 50% and 90% prediction intervals shown in Figure 4 represent the first approximation of forecast uncertainty quantified through the natural spread of individual decision trees in the RF ensemble. These intervals widen with lead time, consistent with the error-accumulation behavior demonstrated in Figure 6. Full probabilistic calibration (e.g., conformal prediction or Bayesian inference) would be required for operational deployment and is identified as a priority for future work.

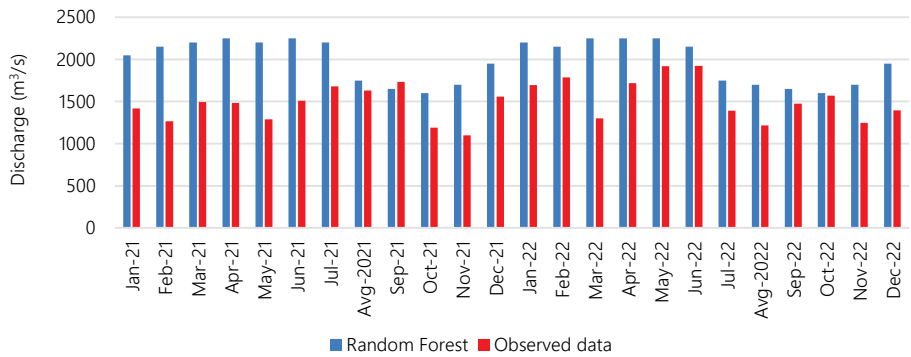


Figure 5. Comparison of forecasted and observed river discharge (m³/s) for 2021–2022 using the RF model.

Note. The figure shows monthly values alongside percentage differences, revealing seasonal biases, with winter overestimation and summer underestimation.

Figure 5 compares forecasted versus observed river discharge (m³/s) for 2021–2022 using the RF model, including monthly percentage differences. The mean absolute percentage error (*MAPE*) across the 24-month forecast period is 22%. Seasonal biases are clearly visible, with winter months (e.g., January and November) overestimating by up to +55% and summer months (e.g., July and August) underestimating by up to -34%. These results indicate notable seasonal discrepancies in model performance. While the overall mean error remains acceptable for broad hydrological applications, the magnitude of the seasonal biases (ranging from -34% to +55%) highlights the limitations of the current univariate approach and suggests that further refinement, particularly through the inclusion of meteorological predictors, should be addressed in future studies.

To quantify the impact of recursive multi-step forecasting, the degradation of RF predictive accuracy was examined as a function of lead time ($h = 1$ –24 months). Despite solid overall test-period performance ($MAE = 345.42$ m³/s, $RMSE = 455.17$ m³/s, $R^2 = 0.585$, $NSE = 0.585$, $KG\bar{E} = 0.794$), accuracy deteriorated markedly with increasing horizon. At $h = 1$, skill was already reduced ($MAE = 713.4$ m³/s, $RMSE = 926.0$ m³/s, $NSE = -0.656$), highlighting sensitivity to initial errors. Performance remained poor across all the horizons, with negative NSE values (e.g., $NSE = -0.309$ at $h = 6$). Degradation became especially severe at longer lead times, reaching $NSE = -0.968$ and $RMSE = 979.0$ m³/s at $h = 24$.

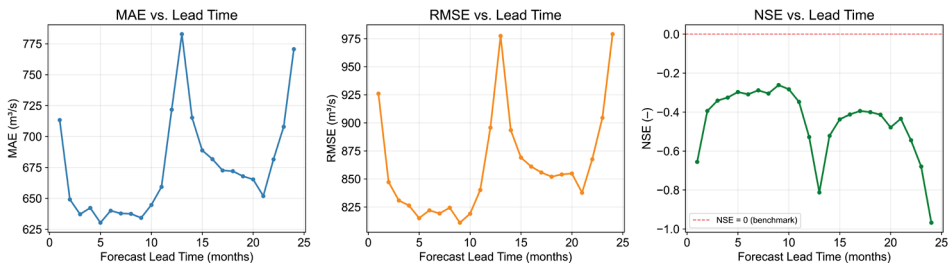


Figure 6. Degradation of forecast performance metrics (*MAE*, *RMSE*, and *NSE*) as a function of forecast lead time (1–24 months) for the RF model.

The rolling-origin lead-time analysis (Figure 6) reveals progressive degradation of RF recursive forecast accuracy across all the 24 horizons. *MAE* ranges from 632 to 781 m³/s, while *RMSE* fluctuates between 822 and 979 m³/s. *NSF* remains negative at all lead times (*NSE* = −0.66 at *h* = 1 to *NSE* = −0.97 at *h* = 24), indicating that the recursive multi-step forecast does not outperform a simple mean-discharge baseline. The non-monotonic error pattern reflects partial seasonal recovery when the predicted input window completes a full annual cycle.

4. Discussion

For the Danube River at the Bratislava station, the RF model produced the most favorable results in describing the nonlinear dynamics characteristic of hydrological time series, based on *R*² as well as error metrics (*MSE*, *RMSE*, *NSE*, *KGE*). These results align well with various regional and global studies that highlight the success of ensemble methods in modeling hydrological data series (Chen et al., 2024; Fernández-Delgado et al., 2019; Liu et al., 2024; Pham et al., 2021).

GB and LR exhibited only modest predictive skills based on the coefficient of determination and error metrics. These results highlight their limited ability to generalize across nonlinear and temporal river discharge patterns. Although GB benefits from sequential learning and error correction, it remains highly susceptible to overfitting, particularly when extreme events are present. This susceptibility likely explains its underperformance compared to RF (Friedman, 2001; Maulud & Abdulazeez, 2020). Conversely, several hydrological studies report contradictory findings, showing GB variants (XGBoost, LightGBM, CatBoost) outperforming RF in mountainous or snow-influenced catchments and extreme events (Kumar et al., 2023; Ni et al., 2020). This performance divergence stems from dataset characteristics: GB's sequential correction excels in highly non-stationary or topographically complex systems, whereas RF's bagging offers superior robustness against noise in medium-sized univariate datasets like the one used here (Kedam et al., 2024). A critical limitation of the current univariate approach is the total exclusion of meteorological predictors like precipitation, temperature, and evapotranspiration. This restricts operational applicability under future climate scenarios, as univariate models cannot account for shifting hydro-meteorological interactions (Aghakhani et al., 2025). Therefore, multivariate extensions incorporating these variables are essential and will be pursued in future work. The lead-time degradation analysis confirms that the recursive forecasting framework introduces substantial uncertainty, particularly beyond short lead times (Ji et al., 2025). Consistently negative *NSE* values across all horizons indicate that forecasts rapidly lose skill relative to climatological benchmarks. This behaviour is consistent with theoretical expectations for autoregressive recursive models, where error accumulation increases nonlinearly with forecast horizon (Cheng et al., 2020; Ji et al., 2025). These findings highlight a fundamental limitation of the applied approach and suggest that alternative strategies, such as direct multi-step forecasting or hybrid recursive-direct methods, may provide greater stability for long-term predictions (Ji et al., 2025; Kurian et al., 2020). Furthermore, incorporating exogenous predictors could help constrain error propagation by anchoring forecasts to physical drivers instead of relying solely on autoregressive memory (Kurian et al., 2020).

Figure 6 illustrates this lead-time-dependent degradation for the RF model. Both *MAE* and *RMSE* exhibit a clear increasing trend with forecast horizon, accompanied by pronounced fluctuations at intermediate and longer lead times that reflect episodic amplification of prediction errors. Meanwhile, *NSE* values decline rapidly and remain negative for most horizons, confirming that model performance quickly falls below the climatological benchmark beyond short lead times. This pattern aligns with broader hydrological forecasting literature,

where recursive multi-step approaches have been shown to accumulate errors nonlinearly as predictions are sequentially fed back into the model (Kumar et al., 2013; Mosavi et al., 2018). Specifically, Kumar et al. (2013) demonstrated that error propagation significantly limits the reliability of streamflow forecasts at seasonal horizons, while Mosavi et al. (2018) highlighted that machine learning models, despite strong short-term skill, often suffer substantial performance decay at longer lead times due to compounding uncertainties. In this study, forecast reliability is markedly reduced beyond approximately 6–12 months, reinforcing the need for alternative forecasting strategies when extended prediction horizons are required.

LR, being strictly a parametric approach, cannot effectively capture the complex interactions and temporal dependencies characteristic of river discharge. Similar findings have been reported for other European rivers (Ikeagwuani, 2021). The SARIMA model performed unexpectedly poorly ($R^2 = 0.01$, $NSE = -0.12$, $KGE = 0.21$). Stationarity of the monthly discharge series was formally tested using the Augmented Dickey-Fuller (ADF) test ($p < 0.01$ after first-order differencing) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test ($p > 0.05$ after differencing), confirming that the series became stationary. No Box-Cox transformation was required, as the variance was already stabilized after seasonal differencing. Model parameters were selected rigorously by examining ACF and PACF plots and minimizing AIC and BIC over the training period (1901–2004). Despite correct identification of the seasonal structure (order $(2, 0, 0) \times (2, 0, 0, 12)$), the drastic underperformance relative to ensemble machine-learning methods is not due to data errors, inadequate preprocessing, or suboptimal parameter selection. Rather, it reflects the fundamental limitations of linear stochastic models when applied to highly non-linear and regulated hydrological systems. The Danube at Bratislava is strongly influenced by upstream snowmelt dynamics, river regulation, and anthropogenic interventions, which introduce complex non-linear interactions that SARIMA cannot capture (Sirbu et al., 2026). In contrast, RF and GB can model these non-linear dependencies directly from the lagged discharge values, explaining their clear superiority in this univariate setting. In contrast, the rolling-origin evaluation (Figure 6, Table 4) confirms that RF recursive multi-step performance deteriorates to below the mean-discharge baseline ($NSE < 0$) at all horizons from $h = 1$ to $h = 24$, underscoring that strong one-step-ahead skill does not translate to reliable medium-range recursive forecasting without direct multi-step or multivariate extensions.

The observed seasonal discrepancies, summer underestimations and winter overestimations, indicate that model performance is moderately sensitive to climatic seasonality. This pattern closely aligns with documented European flow-regime shifts reported by Blöschl et al. (2019) and Trambly et al. (2021). Blöschl et al. (2019) showed that climate change has increased winter flood discharges in North-Western and Central Europe (due to higher winter precipitation and soil moisture) while decreasing summer low flows in Southern and Central Europe (due to rising evapotranspiration and reduced summer precipitation). Similarly, Trambly et al. (2021) documented a clear increase in the number and earlier occurrence of zero-flow days across Southern and Central Europe, driven primarily by intensified summer drying. The RF model's tendency to underestimate summer discharges and overestimate winter discharges therefore mirrors these observed continental-scale trends: the model underpredicts the growing summer low-flow stress and overpredicts winter flows, which are becoming more pronounced under changing climatic conditions. This consistency indicates that the seasonal biases are not merely model artifacts but reflect real shifts in the Danube's hydrological regime under climate variability (Khozani et al., 2024).

The exclusive use of discharge as a predictor in this study allowed for an assessment of univariate ML performance under data-limited conditions. The presented results demonstrate the efficiency of RF in reconstructing historical discharge dynamics. However, excluding meteorological variables such as precipitation, air temperature, and evapotranspiration is likely responsible for the residual errors, particularly during extreme events. Incorporating these meteorological variables, together with ERA5 reanalysis data (which can also serve as input or boundary conditions for regional climate model simulations), could enhance model interpretability and predictive ability under different climate change scenarios (Vilaseca et al., 2023). In addition to conventional error metrics, the hydrological-signature analysis (Table 3) further confirms that RF most accurately reproduces key runoff characteristics such as long-term mean flow, variability, and high- and low-flow quantiles, whereas SARIMA shows noticeable distortions of these signatures.

An additional extension involves using discharge observations from multiple gauging stations along the Danube and its tributaries. In a multivariate configuration, RF and GB can naturally incorporate simultaneous inputs from several upstream and downstream stations, allowing the models to exploit spatial correlations and travel-time effects along the river network (Jahanshahi et al., 2025). Previous studies have shown that such multi-station setups can improve forecasts at intermediate locations, particularly during flood waves and prolonged low-flow events (Hauswirth et al., 2021). In contrast, extending SARIMA to the multivariate case would require more complex formulations (e.g., vector autoregressive or SARIMAX models) with a much larger number of parameters, which can be difficult to estimate reliably for many stations (Galdelli et al., 2025). Evaluating these multivariate, spatially distributed configurations is beyond the scope of the present work, which intentionally focused on a single, well-gauged station, but it represents a promising direction for future research on the Danube River system.

From a hydrological perspective, the superior performance of RF is due to their ability to capture both temporal autocorrelation and nonlinear interactions that influence river discharge. The Bratislava section of the Danube River is governed by long-term precipitation patterns, snowmelt from upstream catchments, and anthropogenic regulation of the riverbed (Kron et al., 2019; Viero et al., 2019). The RF model successfully exploited historical discharges to account for these complex interactions, while purely statistical models did not adapt to changes in the hydrological regime.

The presented findings have significant implications for hydrological forecasting in Central and South-Eastern Europe. First, the results show that ensemble models can provide reliable outcomes and are therefore valuable operational tools for forecasting extreme events. Second, the highlighted limitations of SARIMA and other univariate statistical models indicate that caution is necessary when applying them.

The RF model is confirmed to be the most efficient for forecasting monthly discharge up to two years. This model demonstrated robustness, adaptability, and greater predictive skill compared with other ML or statistical approaches. This underscores the growing value of ML in modern hydrology by providing regionally adaptable and scalable solutions for managing hydrological extremes.

Despite these promising results, several limitations of this study should be acknowledged: (i) the analysis is restricted to a single gauging station and a univariate, monthly discharge time series, which constrains both the spatial and temporal generalizability of the findings and limits direct transferability to other parts of the Danube or other basins; (ii) only deterministic point forecasts were produced, without explicit quantification of predictive

uncertainty, which is essential for risk-informed decision-making in flood and drought management; A direct multi-step forecasting strategy, where separate models are trained for each prediction horizon, could reduce error accumulation and improve long-range stability. However, this approach was beyond the scope of the present study and should be systematically evaluated in future work; (iii) the models were calibrated and validated under historical conditions and may not fully capture future nonstationarities in the discharge regime arising from climate change, land-use alterations, or additional river regulation. Future research should extend this framework to multi-station and multivariate settings, integrate meteorological and climatic predictors as well as indicators of human influence, explore probabilistic and deep learning-based approaches, and explicitly test model robustness under nonstationary scenarios and for operational early warning applications.

5. Conclusion

This study demonstrates that RF significantly outperforms GB, SARIMA, and LR for monthly discharge forecasting at the Bratislava gauging station on the Danube River. RF achieved the highest accuracy ($R^2 = 0.89$, NSE = 0.88, KGE = 0.83) and best reproduced key hydrological signatures, confirming its ability to capture nonlinear relationships and temporal dependencies from univariate discharge data alone. In contrast, SARIMA performed very poorly, while GB and LR delivered only moderate results.

This study fills a significant geographical and methodological gap by providing the first comprehensive machine-learning assessment of monthly discharge forecasting for the strategically important Danube River at Bratislava, a large, regulated river in Central and South-Eastern Europe that has been notably underrepresented in the global hydrological modeling literature. By demonstrating the strong performance of RF over a 120-year observational record using only univariate data, the results offer new evidence on the practical value and robustness of ensemble learning methods for complex hydrological systems, thereby advancing data-driven approaches for operational water resource management.

The observed seasonal biases, underestimation in summer and overestimation in winter, mirror broader European flow-regime shifts under climate change. These findings highlight the superiority of ensemble machine-learning approaches for operational hydrological forecasting in regulated river systems.

Overall, RF provides a robust, efficient framework for monthly discharge prediction at Bratislava on the Danube River, benefiting water management, flood risk assessment, and hydropower planning across Central and South-Eastern Europe. For practitioners, twelve months of lagged discharge data offer an efficient, preliminary tool when meteorological data are unavailable. Forecast reliability remains high up to six months but decreases significantly beyond twelve months due to error accumulation. Operational deployment requires transition to probabilistic and multivariate frameworks. Future studies must incorporate meteorological drivers to enhance predictive skill under changing climate conditions.

Acknowledgements

This research was supported by the “Streamflow Drought Through Time” project funded by the EU NextGenerationEU through the Recovery and Resilience Plan of the Slovak Republic within the framework of project no. 09I03-03-V04-00186.

References

- Aghakhani, A., Robertson, D. E., & Pauwels, V. R. N. (2025). *Uncovering a Key Predictors for Enhancing Daily Streamflow Simulation Using Machine Learning*. EGUsphere. <https://doi.org/10.5194/egusphere-2025-553>
- Ahmadpour, A., Mirhashemi, S., Haghighat jou, P., & Foroughi, F. (2023). Correction: Comparison of the monthly streamflow forecasting in Maroon dam using HEC-HMS and SARIMA models. *Sustainable Water Resources Management*, 9(4), Article 99. <https://doi.org/10.1007/s40899-023-00862-x>
- Ali, M., Prasad, R., Xiang, Y., & Yaseen, Z. M. (2020). Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. *Journal of Hydrology*, 584, Article 124647. <https://doi.org/10.1016/j.jhydrol.2020.124647>
- Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., . . . Živković, N. (2019). Changing climate both increases and decreases European river floods. *Nature*, 573(7772), 108–111. <https://doi.org/10.1038/s41586-019-1495-6>
- Chen, S., Feng, Y., Li, H., Ma, D., Mao, Q., Zhao, Y., & Liu, J. (2024). Enhancing runoff predictions in data-sparse regions through hybrid deep learning and hydrologic modeling. *Scientific Reports*, 14(1), Article 26450. <https://doi.org/10.1038/s41598-024-77678-y>
- Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., & Pain, C. C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590, Article 125376. <https://doi.org/10.1016/j.jhydrol.2020.125376>
- Clark, S. R., & Jaffrés, J. B. D. (2025). Associations between deep learning runoff predictions and hydrogeological conditions in Australia. *Journal of Hydrology*, 651, Article 132569. <https://doi.org/10.1016/j.jhydrol.2024.132569>
- Defontaine, T. (2024). *Extending Flood Forecast Lead Time Using Machine Learning on Limited and Heterogeneous Datasets* [Doctoral dissertation, Institut National Polytechnique de Toulouse]. <https://cnrs.hal.science/tel-04749435>
- Dikshit, A., Pradhan, B., & Alamri, A. M. (2020). Temporal Hydrological Drought Index Forecasting for New South Wales, Australia Using Machine Learning Approaches. *Atmosphere*, 11(6), Article 585. <https://doi.org/10.3390/atmos11060585>
- Fang, W., Ren, K., Liu, T., Shang, J., Jia, S., Jiang, X., & Zhang, J. (2024). An evaluation of random forest based input variable selection methods for one month ahead streamflow forecasting. *Scientific Reports*, 14(1), Article 29766. <https://doi.org/10.1038/s41598-024-81502-y>
- Farfán-Durán, J. F., & Cea, L. (2024). Streamflow forecasting with deep learning models: A side-by-side comparison in Northwest Spain. *Earth Science Informatics*, 17(6), 5289–5315. <https://doi.org/10.1007/s12145-024-01454-9>
- Fernández-Delgado, M., Sirsat, M. S., Cernadas, E., Alawadi, S., Barro, S., & Febrero-Bande, M. (2019). An extensive experimental survey of regression methods. *Neural Networks*, 111, 11–34. <https://doi.org/10.1016/j.neunet.2018.12.010>
- Firat, M., Dikbas, F., Koc, A. C., & Gungor, M. (2012). Analysis of temperature series: Estimation of missing data and homogeneity test. *Meteorological Applications*, 19(4), 397–406. <https://doi.org/10.1002/met.271>
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/AOS/1013203451>
- Galdelli, A., Fronzi, D., Narang, G., Mancini, A., & Tazioli, A. (2025). Groundwater level forecasting using data-driven models and vadose zone: A comparative analysis of ARIMA, SARIMAX, Prophet, and NeuralProphet. *Applied Computing and Geosciences*, 28, Article 100304. <https://doi.org/10.1016/j.acags.2025.100304>
- Grenier, M., Boudreault, J., Raymond, S., & Boudreault, M. (2024). Projected seasonal flooding in Canada under climate change with statistical and machine learning. *Journal of Hydrology: Regional Studies*, 53, Article 101754. <https://doi.org/10.1016/j.ejrh.2024.101754>
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>

- Haghdoust, M., & Md Azamathulla, H. (2024). Predicting the drag coefficient of coastal trees using Support Vector Machines and boosting ensemble models. *Discover Water*, 4(1), Article 102. <https://doi.org/10.1007/s43832-024-00162-1>
- Halmová, D., Pekárová, P., Pekár, J., Miklánek, P., & Bačová Mitková, V. (2021). Identification of the historical drought occurrence on the Danube River and its tributaries. *Acta Hydrologica Slovaca*, 22(2), 237–247. <https://doi.org/10.31577/ahs-2021-0022.02.0027>
- Hansen, J. V., & Nelson, R. D. (2003). Time-series analysis with neural networks and ARIMA-neural network hybrids. *Journal of Experimental & Theoretical Artificial Intelligence*, 15(3), 315–330. <https://doi.org/10.1080/0952813031000116488>
- Hauswirth, S. M., Bierkens, M. F. P., Beijk, V., & Wanders, N. (2021). The potential of data driven approaches for quantifying hydrological extremes. *Advances in Water Resources*, 155, Article 104017. <https://doi.org/10.1016/j.advwatres.2021.104017>
- Ikeagwuani, C. C. (2021). Estimation of modified expansive soil CBR with multivariate adaptive regression splines, random forest and gradient boosting machine. *Innovative Infrastructure Solutions*, 6(4), 1–16. <https://doi.org/10.1007/s41062-021-00568-z>
- Jahanshahi, A., Asadi, H., & Gupta, H. (2025). A data fusion approach to enhancing runoff simulation in a semi-arid river basin. *Environmental Modelling & Software*, 190, Article 106468. <https://doi.org/10.1016/j.envsoft.2025.106468>
- Ji, Y., Lim, Y., Kim, D., Sung, J., & Kang, B. (2025). Effects of Forecasted Rainfall on Direct and Recursive LSTM-Based River Water Level Predictions. *Journal of Flood Risk Management*, 18(4), Article e70147. <https://doi.org/10.1111/jfr3.70147>
- Kabbilawsh, P., Kumar, D. S., & Chithra, N. R. (2022). Forecasting long-term monthly precipitation using SARIMA models. *Journal of Earth System Science*, 131(3), 1–11. <https://doi.org/10.1007/s12040-022-01927-9>
- Kedam, N., Tiwari, D. K., Kumar, V., Khedher, M. K., & Salem, M. A. (2024). River stream flow prediction through advanced machine learning models for enhanced accuracy. *Results in Engineering*, 22, Article 102215. <https://doi.org/10.1016/j.rineng.2024.102215>
- Khodakhah, H., Aghelpour, P., & Hamed, Z. (2022). Comparing linear and non-linear data-driven approaches in monthly river flow prediction, based on the models SARIMA, LSSVM, ANFIS, and GMDH. *Environmental Science and Pollution Research*, 29(15), 21935–21954. <https://doi.org/10.1007/s11356-021-17443-0>
- Khozani, Z. S., Precht, E., & Ionita, M. (2025). Weekly streamflow forecasting of Rhine river based on machine learning approaches. *Natural Hazards*, 121(4), 4135–4153. <https://doi.org/10.1007/s11069-024-06962-x>
- Kilinc, H. C., Ahmadianfar, I., Demir, V., Heddam, S., Al-Areeq, A. M., Abba, S. I., Tan, M. L., Halder, B., Marhoon, H. A., & Yaseen, Z. M. (2023). Daily scale river flow forecasting using hybrid gradient boosting model with genetic algorithm optimization. *Water Resources Management*, 37(9), 3699–3714. <https://doi.org/10.1007/s11269-023-03522-z>
- Konapala, G., Kao, S.-C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, 15(10), Article 104022. <https://doi.org/10.1088/1748-9326/aba927>
- Kron, W., Eichner, J., & Kundzewicz, Z. W. (2019). Reduction of flood risk in Europe – Reflections from a reinsurance perspective. *Journal of Hydrology*, 576, 197–209. <https://doi.org/10.1016/j.jhydrol.2019.06.050>
- Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., & Caloiero, T. (2023). Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow Prediction Models. *Water*, 15(14), Article 2572. <https://doi.org/10.3390/w15142572>
- Kurian, C., Sudheer, K. P., Vema, V. K., & Sahoo, D. (2020). Effective flood forecasting at higher lead times through hybrid modelling framework. *Journal of Hydrology*, 587, Article 124945. <https://doi.org/10.1016/j.jhydrol.2020.124945>
- Leščešen, I., Pekárová, P., & Bajtek, Z. (2025). Navigating hydrological extremes: SARIMA forecasting of minimum Danube River discharges. *Journal of the Bulgarian Geographical Society* 53, 29–47. <https://doi.org/10.3897/jbgs.e159722>
- Leščešen, I., Tanhapour, M., Pekárová, P., Miklánek, P., & Bajtek, Z. (2025). Long Short-Term Memory (LSTM) Networks for Accurate River Flow Forecasting: A Case Study on the Morava River Basin (Serbia). *Water*, 17(6), Article 907. <https://doi.org/10.3390/w17060907>

- Liu, Z., Zhou, J., Yang, X., Zhao, Z., & Lv, Y. (2024). Research on water resource modeling based on machine learning technologies. *Water*, 16(3), 1–26. <https://doi.org/10.3390/w16030472>
- Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(2), 140–147. <https://doi.org/10.38094/jastt1457>
- Mehta, D., Dhabuwala, J., Yadav, S. M., Kumar, V., & Azamathulla, H. M. (2023). Improving flood forecasting in Narmada river basin using hierarchical clustering and hydrological modelling. *Results in Engineering*, 20, Article 101571. <https://doi.org/10.1016/j.rineng.2023.101571>
- Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, 10(11), Article 1536. <https://doi.org/10.3390/w10111536>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., & Liu, J. (2020). Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *Journal of Hydrology*, 586, Article 124901. <https://doi.org/10.1016/j.jhydrol.2020.124901>
- Pham, L. T., Luo, L., & Finley, A. (2021). Evaluation of random forests for short-term daily streamflow forecasting in rainfall- and snowmelt-driven watersheds. *Hydrology and Earth System Sciences*, 25(6), 2997–3015. <https://doi.org/10.5194/hess-25-2997-2021>
- Rahimzad, M., Moghaddam Nia, A., Zolfonoon, H., Soltani, J., Danandeh Mehr, A., & Kwon, H. H. (2021). Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. *Water Resources Management*, 35(12), 4167–4187. <https://doi.org/10.1007/s11269-021-02937-w>
- Saleh, Md. A., Rasel, H. M., & Ray, B. (2024). A comprehensive review towards resilient rainfall forecasting models using artificial intelligence techniques. *Green Technologies and Sustainability*, 2(3), Article 100104. <https://doi.org/10.1016/j.grets.2024.100104>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/BJML/2024/007>
- Schneider, T., Ruby Leung, L., & Wills, R. C. J. (2024). Opinion: Optimizing climate models with process knowledge, resolution, and artificial intelligence. *Atmospheric Chemistry and Physics*, 24(12), 7041–7062. <https://doi.org/10.5194/acp-24-7041-2024>
- Schoppa, L., Disse, M., & Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. *Journal of Hydrology*, 590, Article 125531. <https://doi.org/10.1016/j.jhydrol.2020.125531>
- Sîrbu, N., & Rugină, A.-M. (2026). Short-Term Streamflow Forecasting for River Management, Using ARIMA Models and Recurrent Neural Networks. *Hydrology*, 13(3), Article 82. <https://doi.org/10.3390/hydrology13030082>
- Talukdar, S., Pal, S., Shahfahad, Naikoo, M. W., Parvez, A., & Rahman, A. (2022). Trend analysis and forecasting of streamflow using random forest in the Punarbhaba River basin. *Environmental Monitoring and Assessment*, 195(1), Article 153. <https://doi.org/10.1007/s10661-022-10696-3>
- Tramblay, Y., Rutkowska, A., Sauquet, E., Sefton, C., Laaha, G., Osuch, M., Albuquerque, T., Alves, M. H., Banasik, K., Beaufort, A., Brocca, L., Camici, S., Csabai, Z., Dakhlaoui, H., DeGirolamo, A. M., Dörflinger, G., Gallart, F., Gauster, T., Hanich, L., ... & Datry, T. (2021). Trends in flow intermittence for European rivers. *Hydrological Sciences Journal*, 66(1), 37–49. <https://doi.org/10.1080/02626667.2020.1849708>
- Valipour, M. (2015). Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*, 22(3), 592–598. <https://doi.org/10.1002/met.1491>
- Viero, D. P., Roder, G., Matticchio, B., Defina, A., & Tarolli, P. (2019). Floods, landscape modifications and population dynamics in anthropogenic coastal lowlands: The Polesine (northern Italy) case study. *Science of the Total Environment*, 651, 1435–1450. <https://doi.org/10.1016/j.scitotenv.2018.09.121>
- Vilaseca, F., Castro, A., Chreties, C., & Gorgoglione, A. (2023). Assessing influential rainfall–runoff variables to simulate daily streamflow using random forest. *Hydrological Sciences Journal*, 68(12), 1738–1753. <https://doi.org/10.1080/02626667.2023.2232356>

Wan, X., Wang, W., Liu, J., & Tong, T. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology*, 14, Article 135. <http://www.biomedcentral.com/1471-2288/14/135>

Wang, K., Song, Z., Yin, Z., Xu, J., & Cheng, S. (2026). Streamflow reconstruction at the source of the Yangtze River using ensemble machine learning. *Journal of Hydrology: Regional Studies*, 64, Article 103247. <https://doi.org/10.1016/j.ejrh.2026.103247>

Wang, X., Jiang, B., & Liu, J. S. (2017). Generalized R-squared for detecting dependence. *Biometrika*, 104(1), 129–139. <https://doi.org/10.1093/biomet/asw071>

Yao, S., Chen, C., Chen, Q., Zhang, J., & He, M. (2023). Combining process-based model and machine learning to predict hydrological regimes in floodplain wetlands under climate change. *Journal of Hydrology*, 626, Article 130193. <https://doi.org/10.1016/j.jhydrol.2023.130193>

Yaseen, Z. M. (2023). A new benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. *Knowledge-Based Engineering and Sciences*, 4(3), 65–103. <https://doi.org/10.51526/kbes.2023.4.3.65-103>

Yaseen, Z. M., Kisi, O., & Demir, V. (2016). Enhancing Long-Term Streamflow Forecasting and Predicting using Periodicity Data Component: Application of Artificial Intelligence. *Water Resources Management*, 30(12), 4125–4151. <https://doi.org/10.1007/s11269-016-1408-5>

Zhang, X., Wu, X., Zhu, G., Lu, X., & Wang, K. (2022). A seasonal ARIMA model based on the gravitational search algorithm (GSA) for runoff prediction. *Water Supply*, 22(8), 6959–6977. <https://doi.org/10.2166/ws.2022.263>

Zhu, Z., Zhou, Y., Chen, G., Zhao, X., Li, Q., & Zhang, W. (2025). Daily runoff prediction modelling using machine learning methods based on meteorological parameters. *Hydrology Research*, 56(12), 1241–1255. <https://doi.org/10.2166/nh.2025.069>

Appendix

Table 1. Hyperparameter grid search ranges and final selected values for the RF and GB models. Selection criterion: lowest cross-validated RMSE across five rolling-window folds

Model	Hyperparameter	Values tested	Final selected value
RF	n_estimators	500, 1000, 2000	2000
RF	max_depth	6, 9, 12, None	12
RF	min_samples_split	2, 5, 10	2
RF	Total combinations	36	–
GB	learning_rate	0.01, 0.05, 0.10	0.05
GB	n_estimators	100, 200, 500	500
GB	max_depth	3, 5, 7	5
GB	Total combinations	27	–

Note. RF and Bayesian optimization were not employed. Given the small number of combinations (≤ 36 per model) and the moderate dataset size (1,440 monthly observations), exhaustive grid search was computationally feasible and preferable for reproducibility. All hyperparameter tuning was performed exclusively on the training period (1901–2004) using five-fold rolling-window cross-validation; the test period (2005–2020) was not used at any stage of tuning.