Short-term River Streamflow Modeling Using Ensemble-based Additive Learner Approach

Khabat Khosravi*1, Shaghayegh Miraki2, Patricia M.Saco3, Raziyeh Farmani4

1- Department of Watershed Management Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.
2- Department of Watershed Management Engineering, Sari Agricultural and Natural Resources University, Sari, Iran.
3- Civil, Surveying and Environmental Engineering and Centre for Water Security and Environmental Sustainability, The University of Newcastle, Australia.
4- College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK.

*Corresponding author: K.Khosravi (Khabat.khosravi@gmail.com)

Abstract

Accurate streamflow (Qt) prediction can provide critical information for urban hydrological management strategies such as flood mitigation, long-term water resources management, land use planning and agricultural and irrigation operations. Since the mid-20th century, Artificial Intelligence (AI) models have been used in a wide range of engineering and scientific fields, and their application has increased in the last few years. In this study, the predictive capabilities of the reduced error pruning tree (REPT) model, used both as a standalone model and within five ensemble-approaches, were evaluated to predict streamflow in the Kurkursar basin in Iran. The ensemble-approaches combined the REPT model with the bootstrap aggregation (BA), random committee (RC), random subspace (RS), additive regression (AR) and disjoint aggregating (DA) (i.e. BA-REPT, RC-REPT, RS-REPT, AR-REPT and DA-REPT). The models were developed using 15 years of daily rainfall and streamflow data for the period 23 September 1997 to 22 September 2012. A set of eight different input scenarios was constructed using different combinations of the input variables to find the most effective scenario based on the linear correlation coefficient. A comprehensive suite of graphical (time-variation graph, scatter-plot, violin plot and Taylor diagram) and quantitative metrics (root mean square error (RMSE), mean
absolute error (MAE), Nash-Sutcliff efficiency (NSE), Percent of BIAS (PBIAS) and the ratio of RMSE to the standard deviation of observation (RSR)) was applied to evaluate the prediction accuracy of the six models developed. The outcomes indicated that all models performed well but the AR-REPT outperformed all the other models by rendering lower errors and higher precision across a number of statistical measures. The use of the BA, RC, RS, AR and DA models enhanced the performance of the standalone REPT model by about 26.82%, 18.91%, 7.69%, 28.99% and 28.05% respectively.

Keywords: Streamflow prediction, ensemble-based model, AR-REPT algorithm, Iran.

1. Introduction

Accurate prediction of streamflow and its patterns is vital for water resources planning and management projects. The major benefit of streamflow prediction is to reduce the impact of floods on infrastructure, property, and public health by issuing warnings for impending flood events. Moreover, streamflow prediction can provide hydrologists significant information to develop the sustainable design of water infrastructure, examine the behavior of rivers for operational purposes, water quality assessment, the estimation of water prices, and the adoption of sustainable agricultural practices (Guven, 2009; Yaseen et al., 2015). However, this is not an easy task due to its complexity, dynamic character, randomness, and non-linearity which results from the effect of the numerous physical mechanisms and characteristics involved in its generation, namely interception and stemflow, soil characteristics, catchment characteristics (topography and shape), land-use and land-cover, evapotranspiration and climate change.
Methods for streamflow prediction can be classified into short-term and long-term forecasting categories. Short-term (real-time) flood forecasting (e.g., hourly and daily) is important for the development of early warning systems (Chiang et al., 2004; Guven, 2009; Yaseen et al., 2015), while long-term forecasting (e.g., weekly, monthly, and annually) is of vital importance for the appropriate planning of reservoir management, erosion control, efficient hydro-power generation, irrigation management decisions, scheduling water releases, and many other applications.

Generally, there are four main methodologies used for streamflow forecasting, namely empirical formulas, statistical models, conceptual/physically-based modeling and data-driven techniques. Empirical formulas (i.e. Creager, Fuller, and Dicken) are simple methods developed based on specific datasets and for given catchment conditions, thus, they are not accurate for use in other catchments and therefore not very popular among hydrologists. The coefficients in these equations are determined for a specific catchment and hence, they are not reliable for other catchments, particularly those located in different climates. Since the publication of the pioneering study by Box and Jenkins in 1970, the classical statistical black-box time-series models, such as Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Average (ARIMA), Auto Regressive Integrated Moving Average with exogenous input (ARIMAX), (simple) Linear Regression (LR), and Multiple Linear Regression (MLR) have been frequently employed for streamflow forecasting (Box and Jenkins, 1970; Salas, 1980; Wu et al., 2009; Valipour et al., 2012a,b; Valipour, 2012, 2015; Valipour et al., 2013;). However, these models fail to capture the non-stationary and non-linear character of streamflow processes. Conceptual or physically-based hydrological models can be used for understanding the complex generation of runoff in a catchment. Physically-based distributed approaches, like MIKE SHE,
are developed based on pixel-scale catchment characteristics, and therefore require great effort, the inclusion of numerous hydrological variables, and therefore a large amount of input data (Costabile et al. 2012). These models are difficult to implement due to the great volume of data required for calibration, which is especially challenging to obtain in developing countries like Iran (Yaseen et al. 2017).

In recent decades, data-driven or soft computing (SC) approaches have received increasing attention among hydrologists and have been used on a wide range of hydrological applications, including streamflow forecasting, to formulate the non-linear relationships between the predictor and predicted variables. They require less input data and fewer parameters to develop the model (Carlson et al. 1970; Hipel and Mcleod 1994; Ahmed and Sarma 2007; Afan et al. 2014; Singh and Cui 2015). SC models are very successful and common tools used for prediction and forecasting due to its ability to extract patterns from historical data, which is used for the prediction of future events. They also have a non-linear and more flexible structure and the ability to predict complex phenomena with high accuracy. They generally lead to excellent results with good level of agreement between predicted and observed data when used for streamflow forecasting (Chen et al., 2015; Deo and Sahin, 2016; Huang et al., 2014; Kasiviswanathan et al., 2016; Meshgi et al., 2015; Zhang et al., 2015). Examples of such methods, that have been used in streamflow prediction, are Artificial Neural Network (ANN), Support Vector Machine (SVM), Adaptive Neuro Fuzzy Inference System (ANFIS), Genetic Algorithm (GA), Gene Expression Programming (GEP), Extreme Learning Machine (ELM), and Neuro-Wavelet Techniques (Nourani et al. 2014; Yaseen et al. 2015; Malik et al. 2020). Moreover, these are widely adopted to solve hydrological problems including rainfall and
precipitation forecasting (Nourani et al., 2009), sediment transport (Kisi et al., 2012), and groundwater modeling (Taormina et al., 2012; Tapoglou et al., 2014; French et al., 1992).

ANNs, which are the approach most widely used model in hydrological applications, can be classified into two types: (1) supervised (e.g., Feed Forward Back Propagation (FFBP), Radial Basis Function (RBF), Multi-Layer Perceptron (MLP), and Generalized Regression Neural Network (GRNN)), and (2) unsupervised (e.g., Self-Organizing Map (SOM)), which has been employed in numerous studies for streamflow modeling and forecasting (Abrahart and See, 2000; Allawi and El-Shafie, 2016; Bray and Han, 2004; Cigizoglu, 2005; Danandeh Mehr et al., 2014; Deo and Sahin, 2016; Ghorbani et al., 2016; Hsu et al., 2002; Singh et al. 2018; Liu and Shi. 2019). Recently, the popularity of ANNs has decreased due to their low convergence speed and low generalization ability, especially when the data record is short for training purposes, and/or the range of the testing data is out of that used for training (Hooshyaripor et al. 2014), which leads to local minimum and the random initial determination of weights in each simulation. Adaptive neuro-fuzzy inference systems (ANFIS) were developed to overcome most of the limitations of ANNs, through the use of an ensemble approach (i.e. coupling ANNs with fuzzy logic). Anusree and Varghese (2016) compared the modeling performance of ANN and ANFIS for daily streamflow prediction in the Karuvannur catchment and they reported an improvement in accuracy of ANFIS over ANN, but this approach also has weaknesses. Limitations in ANFIS arise from inefficiencies when defining the weights of membership functions, which significantly affect the model’s predictive potential (Bui et al. 2016). Support vector machine (SVM) is a type of neuron based approach which has been reported to be more powerful than ANFIS and ANNs. Quej et al. (2017) applied ANFIS, ANN and SVM models for the prediction of solar radiation in the Yucatán Peninsula, México, and found that the SVM
model outperforms the other two models. The superiority of SVM over ANFIS and ANN algorithms has also been reported for streamflow prediction (Kakaei Lafdani et al., 2013; He et al., 2014). In addition, the lack of transparency in the results is another unfavourable characteristic of some methods like the SVM, which can be tackled by adopting non-parametric techniques (Auria and Moro, 2009). SVM is also susceptible to hyper-parameter selection (Burgess, 1998; Waseem Ahmad et al., 2018).

Though significant progress has been accomplished in the development and application of SC models to hydrology, researchers are still exploring novel, robust, and more reliable approaches to overcome the shortcomings of the traditional approaches. New developments in data mining as well as the recent advances in modeling capabilities of AI techniques promise to offer more reliable approaches for solving regression problems in a diverse range of scientific realms (Sharafati et al, 2019; Khosravi et al. 2019). For example, in the field of hydrology, Random forest (RF) methods have been recently applied to predict apparent shear stress in channels (Khozani et al., 2019), random tree (RT) algorithms were used to predict nitrate and strontium in groundwater (Bui et al., 2020a) and solar radiation (Sharafati et al., 2019), Reduced Error Pruning Tree (REPTree), random committee (RC), M5P and Instance-Based K-nearest (IBK) neighbours were applied for the prediction of fluoride concentration in groundwater (Khosravi et al., 2020a), Cross-Validation Parameter Selection (CVPS) and Randomizable Filtered Classification (RFC) were used to predict water quality (Bui et al., 2020b), the Bagging Algorithm (BA) was used to predict bedload transport rates (Khosravi et al., 2020b), and the Kstar model applied to predict suspended sediment load prediction (Salih et al., 2019). In addition to the studies mentioned above, additional research have focussed in the comparison of these new the algorithms over traditional SC ones, demonstrating their superiority in several
studies across a wide range of hydrology applications. For example, Sihag et al. (2020) reported a superior performance of the RF model over the SVM and ANN for modeling infiltration processes. They showed that the RF model has a better prediction potential than the SVM and MLP models. Granata et al. (2018) found that the performance of the RF and M5P models was superior to that of the SVM model for the prediction of spring discharges in Italy.

A more recent development to achieve higher prediction performance is based on the use of ensemble-based approaches, which have attracted the attention of researchers all over the world due to their flexibility and benefit of combining multiple models with well-known advantages. For example, Khosravi et al. (2020b) demonstrated an enhanced performance of an ensemble-based approach by combining the RF, M5P, RT and REPT with the BA model. Bui et al. (2020b) used an ensemble-based approach by combining the BA and CVPS models with RF, M5P, RT and REPT, and found out that hybridization can improve the models’ prediction potential.

In this study we explore the use of a new set of ensemble-based algorithms to investigate possible improvements in their prediction potential with respect to standard stand-alone models frequently used for streamflow prediction in the past studies. This will allow us to overcome not only the limitations of the more traditional approaches (low convergence speed, low generalization potential for short input data time series), but also enhance the performance of more advanced stand-alone models and avoid the problem of overfitting and complex tree selection. The REPT algorithm has been successfully applied in many different areas of hydrology, and has therefore been selected in the current study as the base-model for streamflow prediction in the Kurkursar catchment (Iran). The REPT algorithm integrates Reduced Error Pruning (REP) with a Decision Tree algorithm. The REP has the advantage of reducing the complexity of the tree structure without reducing the model performance, and preventing over-
fitting problems. We pursue the main goal of improving prediction potential of the standalone REPT model by developing and investigating the performance of five ensemble-based algorithms, generated by combining the REPT model with the BA, random committee (RC), random subspace (RS), additive regression (AR) and disjoint aggregating (DA) algorithms. Although the BA-REPT, RC-REPT and RS-REPT have been already employed in hydrological applications, AR-REPT and DA-REPT models are new ensemble-based models which have not yet been previously used hydrology or any other geosciences, application.

2- Study area

The Kurkursar catchment is located in northern Iran between longitudes 51°29' and 51°42' E and latitudes 36°29' and 36°40' N, has an area of 75,495 km² (Fig 1). The catchment’s mean altitude is about 860 m, and the overall mean slope is around 21.80%. The catchment has a skewed shape and is bounded on the north by the Caspian Sea, on the south by the Mashlak River and a portion of the Chalus catchment, and on the west by the Chalus catchment. Various geomorphological forms such as flood plains, alluvial fans, and sedimentary dams have been identified in the catchment. The region has a Mediterranean climate with average annual rainfall and streamflow of 900 mm and 1.2 m³/s, respectively. The maximum and minimum discharges are 78 and 0.002 m³/s, respectively. Seventy-nine percent of the heavy precipitation events take place during November to April. The annual maximum daily rainfall occurs with a frequency of 33% in autumn, 42% in winter, and 25% in spring seasons (Rashidi et al. 2016). Temperatures during the year range from below 0°C during winter to 45°C during summer. The Kurkursar catchment is located upstream of Noshahr city, so accurate prediction of its discharge is imperative for the
establishment of meaningful flood mitigation plans for this city. Some recent disastrous flood events have caused irreparable damages to infrastructure in the study area, including bridges, rural buildings, main roads, water and gas pipelines, and the agricultural sector.

Fig 1. Location of the Kurkursar River and the Hydrometry station in the catchment (after Rashidi et al. 2016)

3-Methodology

3.1-Data collection and preparation

Data for a 15-year period (23 September 1997 to 22 September 2012), including precipitation and discharge for the Kurkursar river, were collected and compiled from the Mazandaran Meteorological Organization and regional water authority. The daily rainfall ($R_t$) time series data were obtained from four meteorological weather stations equipped with rainfall gauges, and streamflow ($Q_t$) data was sourced from the Kurkursar hydrometric station situated at outlet of the Kurkursar catchment. Spatially averaged values of rainfall data for the catchment were obtained using the Thiessen polygon approach (Melesse et al. 2011) and used as an input to the model.

The entire dataset was divided into two periods, the first portion comprising 70% of the data (from 23 September 1997 to 22 September 2008) was used as training dataset for model development, and the second portion including the remaining 30% of the data (from 23 September 2008 to 22 September 2012) was used as validation dataset (Ayele et al. 2017; Khsoarvi et al. 2018c; Termeh et al. 2019; Nohani et al. 2019) for model evaluation purposes. Some basic statistical information on the training and validation datasets is presented in Table 1.
Table 1. Basic statistics for the training and testing datasets

3.2. Different input variables combination scenarios

Auto-correlation and linear correlation coefficients (r) were applied to define effective input variables for streamflow prediction. Rainfall (R_t) and streamflow (Q_t) values, and their values at lag-times of 1, 2, 3 days (i.e. R_{t-1}, R_{t-2}, R_{t-3} and Q_{t-1}, Q_{t-2}, Q_{t-3}) were used to identify the best input variable combination. Based on the r-values, eight different input scenarios were constructed and examined (Table 2). Each scenario was constructed according to coefficient value between input and target variables. The first scenario was based on the assumption that the variable with the highest r-value is able to predict streamflow with high accuracy. Next, the variable with the second highest r-value was added to the first scenario to build the second scenario. This approach was continued until the variable with the lowest r-value was added to construct the last input scenario (Table 2). During this stage, each model was implemented using default operator values, just to determine the impact of each input scenario on the results. The r criterion was applied to determine the most effective scenario, with higher r leading to the most effective the input scenario.

Table 2. Different input scenarios

3.3-Determination of optimal operator values
Optimal values for the various operators were obtained through a trial and error method (Kisi et al. 2012, 2016; Khosravi et al, 2018a, Sharafati et al, 2019) using the Waikato Environment for Knowledge Analysis (WEKA 3.9) software developed by University of Waikato, New Zealand. At first each model was implemented using default values of the operators, and then a wide range of higher and lower values was examined until the optimum values for each operator was determined. The Root Mean Square Error (RMSE) metric was applied to obtain optimum operators, with lowest RMSE leading to optimum values.

3.4- Description of models

As mentioned in the introduction, a standalone and five ensemble-based data-mining algorithms, namely REPT, BA-REPT, RC-REPT, RS-REPT, AR-REPT and DA-REPT, were employed to model and predict streamflow at the Kurkursar River outlet using the WEKA software 3.9.

3.4.1- Reduced error pruning tree (REPT)

The REPT model is well known as the fastest tree learner, and it is developed as a combination of the Reduced Error Pruning (REP) and the Decision Tree (DT) learning algorithm methods. It is constructed as a decision or regression tree based on information gains or reductions in the variance (Mohamed et al., 2012). First, the DT algorithm is utilized to simplify the modeling process using the training dataset for the production of a large number of trees with various sizes. The best tree, based on the accuracy of data classification and avoids misclassification, is chosen. Then, the REP is employed to reduce the complexity of the structure of the chosen decision tree and to prevent overfitting problems (Mohamed et al., 2012). The REP algorithm is one of the simplest and most popular pruning techniques, and aims at eliminating some branches to obtain
the most accurate sub-tree through the post-pruning method (Chen et al., 2009; Esposito et al., 1999; Mohamed et al. 2013)

3.4.2-Bootstrap aggregating (BA)

The BA is known as one of the most effective ensemble methods in which repeated sampling builds different data subsets, raising the extent and diversity of component learners by training the data subsets (Opitz:1999). This model is capable of solving classification and regression problems by reducing the defects of component learners and recognizing unstable classifiers. In the algorithm, based on the core idea of BA, the training process is done through in the following main steps: (i) generating bootstrap samples randomly and independently from the original training dataset by replacing; (ii) repeating bootstrap samples several times to create a certain amount of independent datasets; (iii) determining a weak learning algorithm to train various sub-datasets and obtaining the sequence of predictive functions; and (iv) voting for outcomes to select the outcome with the highest number of votes as the final result (Bauer and Kohavi, 1999). BA has extensively been combined with various weak classifiers to improve many base learners, such as decision trees (Mert et al., 2014), SVMs (Pham et al., 2018), and naïve Bayes trees (Pham and Prakash, 2017). In this study, the BA has been used to train the REPT base learner for rainfall-runoff modeling.

3.4.3-Random committee (RC)

Random committee is a meta-algorithm, which has proven very efficient for the enhancement of the learning ability of most classifiers. This algorithm is able to construct a hybrid of base classifiers. In the present study, the final estimation of a random tree was produced through straight averaging probability prediction (Khosravi, 2018a). Although RCs use different numbers of seeds, the classifiers are created based on similar data. The algorithm in this study was applied
in 10 iterations with a number of seeds of 1 using the same parameters as those in the development of the RT model.

3.4.4 Random subspace (RS)

RS is a combination of a data mining and a parallel learning algorithm introduced by Ho (1998). This model is similar to the BA, which is known as a classic integrated algorithm, as it builds a decision tree using the classifier that has the highest level of precision and accuracy based on the training data (Mielniczuk and Teisseyre 2014). The only difference between RS and BA is that in the former, the training subset is created based on the original randomly selected training set (Mielniczuk and Teisseyre, 2014; Xia et al., 2015). The features of the series for each training sub-classifier in the final prediction results are obtained through a combination of voting methods (Zhang and Jia, 2007). The operation of the sub-classifiers relies on integrated learning diversity. The subcategories of RS are employed to specify the differences in the training performances of sub-classifiers and the adopted ensemble learning method is used to pool samples with various spatial specifications (Nanni and Lumini, 2008).

3.4.5 Additive regression (AR)

Structured additive regression (Fahrmeir et al., 2004) is a nonparametric regression method which was proposed first by Breiman and Friedman in 1985. The AR is considered as an indispensable section of the alternating conditional-expectations algorithm. This algorithm is able to provide a generalization of the generalized linear and additive models by building a restricted class of non-parametric regressions. During each iteration of the AR algorithm, the standalone model (i.e. REPT) is fit to the residuals from the former iteration. Finally during the last iteration, the final prediction is generated from adding all previous predictions together. It is
also a more flexible and interpretable predictor than the general regression. In generic notation, a brief description of the AR predictor is given by:

\[ F(X) = \sum_{m=0}^{M} \beta_m h(X; a_m) \]  

(1)

where \( h(X; a_m) \) are unknown as the basic function and independent model output made by inputs \( X \) and model parameters \( a_m \) at iteration \( m \) (\( m = 1, ..., M \) and \( M \) is the number of iterations), \( \beta_m \) are a set of basis coefficients at iteration \( m \), and \( F(X) \) is the AR algorithm output. The best results are achieved when the standalone and AR algorithms’ parameters (\( a_m \) and \( \beta_m \)) are used in a stepwise method (i.e., each set of parameters is estimated at a particular iteration). Fig 2 shows the conceptual model of developing the hybrid model by coupling AR with the standalone algorithms (here as an example, REPT).

Fig 2. Conceptual model for the development of the AR hybrid algorithms (Mitchell, 1997)

3.4.6. Disjoint aggregating (Dagging)

The Disjoint Aggregating (Dagging) is a resampling integration and group-sampling technique that was proposed by Ting and Witten (1997). Dagging and Bagging work in a similar way but in the Dagging method, the training dataset is used for classification of several disjoint subsets by using separate samples rather than bootstrap samples (Chen et al., 2019b). In the Dagging model, the majority voting combines several classifiers to build the final prediction and improve the accuracies of the basic classifiers (Kotsianti and Kanellopoulos 2007). In order to build a robust model, the weak learners are trained on various subsets of the training set (Onan et al., 2016).

3. 5. Model evaluation and comparison
In order to evaluate the performance of each of the models developed, and to compare their efficiency, six statistical metrics, namely the coefficient of determination ($R^2$), RMSE, Mean Absolute Error (MAE), Nash-Sutcliffe Efficiency (NSE), Percentage of bias (PBIAS), and the ratio of RMSE to the standard deviation of the observations (RSR) were utilized for validation period (Table 3).

Table 3. Different indicators used for streamflow ($Q_t$) prediction performance ($Q_t^{obs}$ and $Q_t^{pred}$ are the measured and predicted values of $Q_t$, respectively, $\bar{Q}_t^{pred}$ is the mean predicted value of $Q_t$, and $N$ is the sample size number of data)

In addition to the statistical metrics, three commonly used graphical approaches, namely scatter plots, Taylor diagrams (Taylor 2001) and Violin plots were used to visually compare the performances of the models. In the scatter plot, a lower scatter of cloud points around the 1:1 line indicates higher model accuracy. The Violin plot shows the mean, median, maximum, and minimum predicted values, and similar shapes of the violin plots for the predicted and observed values indicate high model performance. This approach allows for a better evaluation of the models in terms of their accuracy for predicting extreme values. The Taylor diagram incorporates the linear correlation coefficient ($r$), standard deviation, and RMSE simultaneously. The advantages of this comprehensive assessment have made it a popular criterion for visualizing overall model performance (Choubin et al. 2018).

4- Results and analysis
4.1- Best Input Combination

Selecting the optimal input variables is the first step in developing a predictive artificial intelligence model. Also as the effectiveness of each input variable is not equal, and some of them might have a null or negative effect on the results, it is necessary to determine the effectiveness of each of the input variables. In the present study the \( r \) approach between inputs and outputs has been applied to determine the effectiveness of each input variable (Fig 3). The results reveal that rainfall values (\( R_t \)) are the most effective variables for \( Q_t \) prediction (\( r = 0.56 \)), followed by \( Q_{t-1} (r = 0.46) \), \( Q_{t-2} (r = 0.29) \), \( R_{t-1} (r = 0.28) \), \( Q_{t-3} (r = 0.25) \), \( R_{t-2} (r = 0.12) \) and \( R_{t-3} (r = 0.07) \). Accordingly, and as expected, it was found that the greater the lag-time, the lower the \( r \) value and the predictive effectiveness of the variable.

Fig 3. \( r \) value between input variables and streamflow

Eight different input combinations scenarios were constructed and compared (Table 4), the effectiveness of each input combination was evaluated in the training and validation phases using the estimated \( r \). The results of the \( r \) value during the validation phase showed that the best input combination scenario is different for every model developed. Combination No. 7 (i.e. combination of all input variables) had the highest \( r \) values and proved to be the most effective for the standalone REPT model. For all the ensemble based models, with the exception of RC-REPT, the input scenario No. 8 (i.e. combination of \( R_t \), \( R_{t-1} \), and \( Q_{t-1} \) variables) was identified as the most effective scenario. For the RC-REPT model, combination No. 3, which is the combination of \( R_t \), \( Q_{t-1} \), and \( Q_{t-2} \) variables, is the optimum input scenario. These results are in
agreement with the findings of variable importance using r, which showed that variables with low r are not effective for the prediction of complex phenomena like streamflow, which has also been reported in previous studies (Yaseen et al. 2017). It can be concluded that specifying the optimal value for each operator and choosing the best input variable play a decisive role in the predictive potential of each model. The results shown in Table 4 indicate that best input scenario (i.e. gray shadow) has about 18.18%, 31.86%, 29.54%, 26.19%, 32.6% and 31.18% higher predictive potential (for REPT, BA-REPT, RC-REPT, RS-REPT, AR-REPT and DA-REPT, respectively ) than the worst input scenario, and this obviously highlights the importance of the selection of the best input scenario on the results.

Discrepancy between the models’ performance at this stage, results from the different computational structure of each algorithm. It shows that for the standalone REPT model to achieve high prediction potential, the input needs to involve all different variables, and that even in this case, it still has a lower performance that the ensemble-based models. From a modeling perspective, the best model should have two characteristics; (1) high performance and (2) require a lower number of input variables, as sometimes, measuring some input variables is difficult and time consuming. Thus, our result show that the standalone model (i.e. REPT) is not the best option as it involves a high number of input variables.

Table 4. Efficiency of different input scenarios for the training and validation phases based on the r metric

4.2- Evaluation of the developed models
After identifying the best input combination scenario and the optimal value for each operator, the performance of each of the models developed using the training dataset, was assessed using the validation dataset in the evaluation stage. The results of this evaluation were used to estimate and compare model performance (Khosravi et al 2016; 2018a,b; 2019; 2020a; Chen et al, 2019a). Time-variation graphs and scatter plots for the measured and predicted streamflow values are presented in Figure 4. These results indicate that the AR-REPT model is remarkably more accurate in capturing peak-flow values than the other models, as indicated by less scatter in the cloud points. Therefore, it has the best predictive potential for the estimation of streamflow, while the standalone REPT model, with the highest scatter in cloud points, has the worst prediction ability. Performance of the other models has an acceptable level of accuracy.

Fig 4. Time-variation graph and scatter plot for the measured vs. predicted values over the validation period

Figure 5 presents the results for the violin plots obtained for all models, and shows the maximum, minimum, median ($Q_{50}$), first quartile ($Q_{25}$), and third quartile ($Q_{75}$) of the predicted $Q_t$ and the measured $Q_t$ values. It can be seen that the BA-REPT, DA-REPT and AR-REPT ensemble based models are able to predict the maximum streamflow values accurately. Maximum streamflow is the source of flooding and its prediction with high accuracy is very important for the design of flood mitigation plans. The outcomes indicate that the AR-REPT model has the higher prediction accuracy for lower streamflow values. Overall, the violin plots show that the plot corresponding to the AR-REPT ensemble based model has a closer shape to that of the plot for the measured values.
Fig 5. Violin plots for the measured and predicted streamflow values

For further analysis of the efficiency of the developed models, a Taylor diagram is presented in Figure 6. This diagram confirms the superiority of the AR-REPT model compared with the other data-driven models, and that the REPT model has the lowest prediction potential. It shows that the r values obtained using measured data and the prediction results obtained using the AR-, DA- and BA- ensemble-based models vary between 0.90 and 0.95, while the r values for the remaining models vary between 0.80 and 0.90.

Fig 6. Taylor diagram to visualize the performance of the models

Six quantitatively statistical performance criteria to assess the performance of the developed models over the evaluation period are presented in Table 5. Based on the $R^2$ criteria, all developed algorithms have a very good performance (Ayele et al., 2017). Since $R^2$ is standardized for detecting the differences between the mean and variance of the measured and predicted values, and it is highly sensitive to outliers, it is not capable of evaluating the performance of the models by itself (Legates and McCabe, 1999; Shiri and Kisi, 2012), so other metrics (NSE and RSR) are used to further assess model performance.

Based on the NSE and RSR metrics, REPT, RC-REPT and RS-REPT have a good performance, while the remaining algorithms have a very good prediction capability. Over all, the AR-REPT
model ($R^2 = 0.857$, $RMSE = 0.682$ m$^3$/s, $MAE = 0.258$ m$^3$/s, $NSE = 0.845$, $PBIAS = -3.150$ and $RSR = 0.392$) outperforms all the other models, followed by DA-REPT ($R^2 = 0.848$, $RMSE = 0.706$ m$^3$/s, $MAE = 0.274$ m$^3$/s, $NSE = 0.834$, $PBIAS = -3.602$ and $RSR = 0.406$), BA-REPT ($R^2 = 0.834$, $RMSE = 0.730$ m$^3$/s, $MAE = 0.300$ m$^3$/s, $NSE = 0.820$, $PBIAS = -2.100$ and $RSR = 0.420$), RS-REPT ($R^2 = 0.790$, $RMSE = 0.880$ m$^3$/s, $MAE = 0.350$ m$^3$/s, $NSE = 0.740$, $PBIAS = -2.800$ and $RSR = 0.500$), RC-REPT ($R^2 = 0.706$, $RMSE = 1.000$ m$^3$/s, $MAE = 0.450$ m$^3$/s, $NSE = 0.650$, $PBIAS = -4.900$ and $RSR = 0.590$) and REPT ($R^2 = 0.704$, $RMSE = 1.200$ m$^3$/s, $MAE = 0.420$ m$^3$/s, $NSE = 0.600$, $PBIAS = -3.300$ and $RSR = 0.630$) (Table 5 and 6) (In Table 6, the best model gets a lower rank for each criteria).

Table 5. Quantitative model evaluation for the validation phase

Table 6. Model performance ranking

5. Discussion

Streamflow in the Kurkursar catchment is predicted in the present study through several standalone and ensemble-based machine learning algorithms. Our results show that the ensemble-based models developed have a higher performance than the standalone model. Based on the NSE metric, the BA-, RC-, RS-, AR- and DA- models enhance the performance of the standalone REPT model by about 26.82%, 18.91%, 7.69%, 28.99% and 28.05% respectively. In addition, our results show that all the models developed in this study have a reasonably good capability to predict the streamflow. The main reason is that Kurkursar catchment is located in
North of Iran, and has a Mediterranean climate and thus rainfall patterns are fairly regular, while the rainfall patterns in arid and semi-arid regions display larger temporal variable leading to more complex rainfall-runoff processes and lower predictability of streamflow patterns based on rainfall.

The fact that the AR model outperforms the other models, is likely to be due to the use of the 1D smoother approach to construct a restricted class of non-parametric regression. This results in the model to be less affected by the Curse of dimensionality. The DA algorithm utilizes the majority voting approach to build the final prediction (Tama and Comuzzi, 2019), and through this, a robust and reliable prediction can be achieved by the weak learners being trained on a different subset of the training set (Onan et al., 2016). Using multiple weak learners in combination, the BA has a better performance than the RC and RS ensemble-based models by reducing (1) variance and (2) over-fitting through the bootstrap procedure. In most cases, hybrid models are more flexible and can provide better prediction than individual models. Therefore, ensemble-based models are more reliable and accurate, especially for predictions of complex hydrological processes like the one in the present study (Ghorbani et al., 2017; Yaseen et al., 2017, De’ath and Fabricius, 2000).

Results from previous studies using SC approaches have also reported excellent model performance for streamflow prediction. For example, Kisi et al. (2012) used ANN, Gene Expression Programming (GEP), and ANFIS algorithms for streamflow modeling in Turkey and observed that GPE performed better than the other two models. The $R^2$ values for obtained for ANN, GEP, ANFIS, and MLR were 0.97, 0.93, 0.80, and 0.70, respectively. Rajurkar et al. (2000) modelled daily flows in India using ANN hydrological models. They obtained an $R^2$ of 0.92 and NSE of 0.702 for the best ANN model. Rezaie-Balf et al. (2017) simulated rainfall-
runoff processes in the Tajan catchment in northern Iran using a tree algorithm of model tree (MT), ANN and multivariate adaptive regression splines (MARS). They concentrated on the effects of data input size, involving the number of effective input variables for rainfall-runoff processes and the number of data values in the time series, on the quality of the runoff simulation. They found that the data mining model (i.e. MT) ($R^2 = 0.80$ and $RMSE = 6.70 \text{ m}^3/\text{s}$) was superior to ANN ($R^2 = 0.78$ and $RMSE = 7.40 \text{ m}^3/\text{s}$) and MARS ($R^2 = 0.79$ and $RMSE = 7.47 \text{ m}^3/\text{s}$). It is important to note that the direct comparison of different models and their prediction potential applied to datasets from different catchments is very difficult. Each catchment has different characteristics and the main processes driving streamflow generation can be very different across various catchments, so direct comparison is problematic. For example, predictions from the effective ensemble-based model used in this study resulted in an $R^2$ of 0.85, while Kisi et al. (2012) obtained an $R^2$ of 0.97 for streamflow prediction using an ANN model. However, as mentioned in the introduction, several studies using ANN models have reported much lower predictive performance. An important factor that affects the results in our study is that the model is used to predict short-term or real-time flows, while many of the previous studies have focused on long-term predictions. In addition, differences in algorithm structure lead to very different model behavior, which are also affected by parameter selection, nature of the amount data, data quality and length of the dataset (Asim et al. 2018).

There are several sources of uncertainty in the present study such as those stemming from the analysis of a single study area; the randomized splitting of data into training and testing sets; the uncertainty in the quality of input data and also from input variable selection. Also the lack of availability of additional input variables such as evaporation, temperature, soil moisture data, etc. are the main limitation of the current study, as additional input variables are highly desirable for
streamflow prediction. As a final note, each model has its own advantages and disadvantages and they are different in terms of structure and complexity (Khosravi et al., 2018c). Different artificial intelligence models, empirical equations, and physically based models have been utilized in various studies for the simulation of hydrological processes. It is important to employ a variety of physically based models to predict the streamflow (e.g., SWAT, Wetspa, HEC-HMS, SimHyd), validate their outputs using data mining algorithms, and, ultimately, based on the characteristics of the models including complexity, data requirements, and accuracy, identify the best model for future research. Further work is needed to investigate the generalization power of these algorithms for streamflow prediction, which will include the application of the algorithms developed in this study for other catchments with similar physical and climatic conditions. Also, results can be used to generate management strategies for flooding at this particular catchment.

6. Conclusions

Streamflow prediction is essential for flood impact assessment, and the implementation of useful flood management plans, however, due to the non-linear and chaotic nature of streamflow generation processes, it remains a challenging task. To date, no universal guidelines have been reported for streamflow prediction. In the present study, a standalone REPT model and five ensemble-based data mining models (BA-REPT, RC-REPT, RS-REPT, DA-REPT and AR-REPT) were employed to predict streamflow in the Kurkursar catchment in Iran. The main findings of the study can be summarized as follows:

- $R_t$ is the most effective variable for streamflow prediction followed by $Q_{t-1}$, $Q_{t-2}$, $R_{t-1}$, $Q_{t-3}$, $R_{t-2}$ and $R_{t-3}$.
• The greater the lag-time of each input variable, the lower its correlation coefficient and effectiveness for prediction purposes.

• Due to the different structures of the models, the best input combination was not the same for all the models applied.

• The AR-REPT ensemble-based model outperformed all the other models, followed by DA-REPT, BA-REPT, RC-REPT, RS-REPT and REPT.

• The BA, RC, RS, AR and DA models enhance the performance of the standalone REPT model by about 26.82%, 18.91%, 7.69%, 28.99% and 28.05% respectively.

• The PBIAS values showed that all the models overestimated streamflow.

• The violin-plots showed that the AR-REPT and DA-REPT ensemble-based models were the best for predicting extreme streamflow values.

Authorship contribution statement

KK: conceptualization, model implementation and software, formal analysis, manuscript writing and editing. SM: manuscript writing and interpretation of results. PMS: provided critical insights on the analysis and contributed to writing and editing the manuscript and RF: Review and editing.

References


for predicting peak outflow from breached embankments. J. Hydro-Environment Res. 8, 292–303.
https://doi.org/10.1016/j.jher.2013.11.004


Hipel KW, Mcleod Al. 1994. Time Series Modeling of Water Resources and Environmental

map (SOLO): an artificial neural network suitable for hydrologic modeling and analysis.


neural network ensemble to forecast streamflow for flood management. J. Hydrol. 536,

Kakaei Lafdani, E., Moghaddam Nia, A., Ahmadi, A., Jajarmizadeh, M., Ghafari Gosheh, M.
https://www.researchgate.net/publication/257001623_

http://dx.doi.org/10.1016/j.jhydrol.2012.01.026.

susceptibility assessment and its mapping in Iran: a comparison between frequency
ratio and weights-of-evidence bivariate statistical models with multi-criteria decisionmaking

Khosravi, L Mao, O Kisi, ZM Yaseen, S Shahid.,2018a., Quantifying hourly suspended sediment
load using data mining models: case study of a glacierized Andean catchment in Chile.
Journal of Hydrology 567, 165-179

Khosravi, M Panahi, D Tien Bui.,2018b.Spatial prediction of groundwater spring potential
mapping based on an adaptive neuro-fuzzy inference system and metaheuristic
optimization.Hydrology & Earth System Sciences 22 (9)

Khosravi, K., BT Pham, K Chapi, A Shirzadi, H Shahabi, I Revhaug,2018c. A comparative
assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz
watershed, northern Iran. Science of the Total Environment 627, 744-755

Khosravi, P Daggupati, MT Alami, SM Awadh, MI Ghareb, M Panahi., 2019.Meteorological
data mining and hybrid data-intelligence models for reference evaporation simulation: A

Khosravi, K., JR Cooper, P Daggupati, BT Pham, DT Bui. 2020. Bedload transport rate prediction:
Application of novel hybrid data mining techniques. Journal of Hydrology, 124774


**Figure caption**

**Fig 1.** Location of the Kurkursar River and the Hydrometry station in the catchment (after Rashidi et al. 2016)

**Fig 2.** Conceptual model of working AR hybrid algorithms (Mitchell, 1997)

**Fig 3.** $r$ value between input variables and streamflow

**Fig 4.** Time-variation graph and scatter plot for the measured vs. predicted values over the validation period

**Fig 5.** Violin plots for the measured and predicted streamflow values

**Fig 6.** Taylor diagram to visualize the performance of the models
Fig 1. Location of the Kurkursar River and the Hydrometry station in the catchment (after Rashidi et al. 2016)

Fig 2. Conceptual model of working AR hybrid algorithms (Mitchell, 1997)
Fig 3. r value between input variables and streamflow
Fig 4. Time-variation graph and scatter plot for the measured vs. predicted values over the validation period

\[
y = 0.956x + 0.0916 \\
R^2 = 0.8575
\]

\[
y = 0.9511x + 0.1031 \\
R^2 = 0.848
\]
Fig 5. Violin plots for the measured and predicted streamflow values
Fig 6. Taylor diagram to visualize the performance of the models
Table 1. Basic statistics for the training and testing datasets

Variable | Training dataset | Validation dataset |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>Min 5.00 Max 149.00 Mean 3.47 Std. deviation 11.43</td>
<td>Min 0.00 Max 147.00 Mean 3.42 Std. deviation 10.69</td>
</tr>
<tr>
<td>Streamflow</td>
<td>Min 0.73 Max 73.10 Mean 1.28 Std. deviation 2.25</td>
<td>Min 0.002 Max 41.40 Mean 1.21 Std. deviation 1.74</td>
</tr>
</tbody>
</table>
Table 2. Different input scenarios

<table>
<thead>
<tr>
<th>No.</th>
<th>input combination</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R&lt;sub&gt;t&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;t&lt;/sub&gt;</td>
</tr>
<tr>
<td>2</td>
<td>R&lt;sub&gt;t&lt;/sub&gt;, Q&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;t&lt;/sub&gt;</td>
</tr>
<tr>
<td>3</td>
<td>R&lt;sub&gt;t&lt;/sub&gt;, Q&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;t&lt;/sub&gt;</td>
</tr>
<tr>
<td>4</td>
<td>R&lt;sub&gt;t&lt;/sub&gt;, Q&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-2&lt;/sub&gt;, R&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;t&lt;/sub&gt;</td>
</tr>
<tr>
<td>5</td>
<td>R&lt;sub&gt;t&lt;/sub&gt;, Q&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-2&lt;/sub&gt;, R&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;t&lt;/sub&gt;</td>
</tr>
<tr>
<td>6</td>
<td>R&lt;sub&gt;t&lt;/sub&gt;, Q&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-2&lt;/sub&gt;, R&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-3&lt;/sub&gt;, R&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;t&lt;/sub&gt;</td>
</tr>
<tr>
<td>7</td>
<td>R&lt;sub&gt;t&lt;/sub&gt;, Q&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-2&lt;/sub&gt;, R&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-3&lt;/sub&gt;, R&lt;sub&gt;t-2&lt;/sub&gt;, R&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;t&lt;/sub&gt;</td>
</tr>
<tr>
<td>8</td>
<td>R&lt;sub&gt;t&lt;/sub&gt;, R&lt;sub&gt;t-1&lt;/sub&gt;, Q&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>Q&lt;sub&gt;t&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Table 3. Different indicators used for streamflow prediction (Q<sub>t</sub>) (Q<sub>t</sub><sup>obs</sup> and Q<sub>t</sub><sup>pred</sup> are the measured and predicted values of Q<sub>t</sub>, respectively, Q<sub>t</sub><sup>pred</sup> is the mean predicted value of Q<sub>t</sub>, and N is the sample size number of data)

<table>
<thead>
<tr>
<th>Equation</th>
<th>Ref.</th>
<th>Range</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&lt;sup&gt;2&lt;/sup&gt; = 1 − \left(\frac{\sum_{i=1}^{N}(Q_{t,i}^{obs} - Q_{t,i}^{pred})^2}{\sum_{i=1}^{N}(Q_{t,i}^{obs})^2}\right)</td>
<td>Najafi and Ardabili (2018)</td>
<td>0.7 ≤ R&lt;sup&gt;2&lt;/sup&gt; ≤ 0.1</td>
<td>Very good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.6 ≤ R&lt;sup&gt;2&lt;/sup&gt; ≤ 0.7</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5 ≤ R&lt;sup&gt;2&lt;/sup&gt; ≤ 0.6</td>
<td>Satisfactory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 ≤ R&lt;sup&gt;2&lt;/sup&gt; ≤ 0.5</td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_{t,i}^{pred} - Q_{t,i}^{obs})^2}</td>
<td>(Najafi and Ardabili 2018)</td>
<td></td>
<td>The lower the better</td>
</tr>
<tr>
<td>MSE = \frac{1}{N}\sum_{i=1}^{N}(Q_{t,i}^{pred} - Q_{t,i}^{obs})^2</td>
<td>(Najafi and Ardabili 2018)</td>
<td></td>
<td>The lower the better</td>
</tr>
<tr>
<td>NSE = 1 − \left(\frac{\sum_{i=1}^{N}(Q_{t,i}^{pred} - Q_{t,i}^{obs})^2}{\sum_{i=1}^{N}(Q_{t,i}^{pred})^2}\right)</td>
<td>Moriasi et al, 2007</td>
<td>0.75 &lt; NSE ≤ 1.00</td>
<td>Very good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.65 &lt; NSE ≤ 0.75</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50 &lt; NSE ≤ 0.65</td>
<td>Satisfactory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4 &lt; NSE ≤ 0.50</td>
<td>Acceptable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NSE ≤ 0.4</td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>PBIAS = \frac{\sum_{i=1}^{N}(Q_{t,i}^{pred} - Q_{t,i}^{obs})}{\sum_{i=1}^{N}Q_{t,i}^{pred}}</td>
<td>Legates et al, 1999</td>
<td>PBIAS &lt; ±10%</td>
<td>Very good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>±10% ≤ PBIAS &lt; ±15%</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>±15% ≤ PBIAS &lt; ±25%</td>
<td>Satisfactory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PBIAS ≥ ±25%</td>
<td>Unsatisfactory</td>
</tr>
</tbody>
</table>
RSR = \sqrt{\frac{\sum_{i=1}^{N} (Q_{t}^{\text{pred}} - Q_{t}^{\text{obs}})^2}{\sum_{i=1}^{N} (Q_{t}^{\text{pred}} - Q_{t}^{\text{pred}})^2}}

Table 4. Efficiency of different input scenarios for the training and validation phases based on the correlation coefficient metric

<table>
<thead>
<tr>
<th>Input No.</th>
<th>REPT</th>
<th>BA-REPT</th>
<th>RC-REPT</th>
<th>RS-REPT</th>
<th>AR-REPT</th>
<th>DA-REPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.79</td>
<td>0.63</td>
<td>0.65</td>
<td>0.62</td>
<td>0.79</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>0.99</td>
<td>0.75</td>
<td>0.96</td>
<td>0.81</td>
<td>0.99</td>
<td>0.75</td>
</tr>
<tr>
<td>3</td>
<td>0.99</td>
<td>0.73</td>
<td>0.95</td>
<td>0.84</td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>4</td>
<td>0.99</td>
<td>0.71</td>
<td>0.81</td>
<td>0.85</td>
<td>0.99</td>
<td>0.82</td>
</tr>
<tr>
<td>5</td>
<td>0.99</td>
<td>0.77</td>
<td>0.95</td>
<td>0.83</td>
<td>0.99</td>
<td>0.79</td>
</tr>
<tr>
<td>6</td>
<td>0.99</td>
<td>0.69</td>
<td>0.96</td>
<td>0.86</td>
<td>0.99</td>
<td>0.82</td>
</tr>
<tr>
<td>7</td>
<td>0.99</td>
<td>0.74</td>
<td>0.96</td>
<td>0.86</td>
<td>0.99</td>
<td>0.8</td>
</tr>
<tr>
<td>8</td>
<td>0.99</td>
<td>0.73</td>
<td>0.93</td>
<td>0.91</td>
<td>0.99</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Table 5. Model evaluation quantitatively during validation phase

<table>
<thead>
<tr>
<th>Models</th>
<th>R²</th>
<th>RMSE (m³/s)</th>
<th>MAE (m³/s)</th>
<th>NSE</th>
<th>PBIAS (%)</th>
<th>RSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>REPT</td>
<td>0.704</td>
<td>1.25</td>
<td>0.420</td>
<td>0.600</td>
<td>-3.300</td>
<td>0.630</td>
</tr>
<tr>
<td>BA-REPT</td>
<td>0.834</td>
<td>0.73</td>
<td>0.300</td>
<td>0.820</td>
<td>-2.100</td>
<td>0.420</td>
</tr>
<tr>
<td>RC-REPT</td>
<td>0.706</td>
<td>1.000</td>
<td>0.450</td>
<td>0.650</td>
<td>-4.900</td>
<td>0.590</td>
</tr>
<tr>
<td>RS-REPT</td>
<td>0.790</td>
<td>0.88</td>
<td>0.350</td>
<td>0.740</td>
<td>-2.800</td>
<td>0.500</td>
</tr>
<tr>
<td>AR-REPT</td>
<td>0.857</td>
<td>0.682</td>
<td>0.258</td>
<td>0.845</td>
<td>-3.150</td>
<td>0.392</td>
</tr>
<tr>
<td>DA-REPT</td>
<td>0.848</td>
<td>0.706</td>
<td>0.274</td>
<td>0.834</td>
<td>-3.602</td>
<td>0.406</td>
</tr>
</tbody>
</table>

Table 6. Model performance ranking

<table>
<thead>
<tr>
<th>Models</th>
<th>R²</th>
<th>RMSE (m³/s)</th>
<th>MAE (m³/s)</th>
<th>NSE</th>
<th>PBIAS (%)</th>
<th>RSR</th>
<th>Sum</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>REPT</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>BA-REPT</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>---</td>
</tr>
<tr>
<td>RC-REPT</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>RS-REPT</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>AR-REPT</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>15</td>
<td>2</td>
</tr>
</tbody>
</table>