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

Khabat Khosravi*[1], Shaghayegh Miraki[2], Patricia M.Saco[3], Raziyeh Farmani[4]

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 (Q_t) 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 km^2 (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 m^3/s, respectively. The maximum and minimum discharges are 78 and 0.002 m^3/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

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

238 **3.4- Description of models**

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

242 ***3.4.1-Reduced error pruning tree (REPT)***

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

252 the most accurate sub-tree through the post-pruning method (Chen et al., 2009; Esposito et al.,
253 1999: Mohamed et al. 2013)

254 ***3.4.2-Boostrap aggregating (BA)***

255 The BA is known as one of the most effective ensemble methods in which repeated sampling
256 builds different data subsets, raising the extent and diversity of component learners by training
257 the data subsets (Opitz:1999). This model is capable of solving classification and regression
258 problems by reducing the defects of component learners and recognizing unstable classifiers. In
259 the algorithm, based on the core idea of BA, the training process is done through in the following
260 main steps: (i) generating bootstrap samples randomly and independently from the original
261 training dataset by replacing; (ii) repeating bootstrap samples several times to create a certain
262 amount of independent datasets; (iii) determining a weak learning algorithm to train various sub-
263 datasets and obtaining the sequence of predictive functions; and (iv) voting for outcomes to
264 select the outcome with the highest number of votes as the final result (Bauer and Kohavi,
265 1999). BA has extensively been combined with various weak classifiers to improve many base
266 learners, such as decision trees (Mert et al., 2014), SVMs (Pham et al., 2018), and naïve Bayes
267 trees (Pham and Prakash, 2017). In this study, the BA has been used to train the REPT base
268 learner for rainfall-runoff modeling.

269 ***3.4.3-Random committee (RC)***

270 Random committee is a meta-algorithm, which has proven very efficient for the enhancement of
271 the learning ability of most classifiers. This algorithm is able to construct a hybrid of base
272 classifiers. In the present study, the final estimation of a random tree was produced through
273 straight averaging probability prediction (Khosravi, 2018a). Although RCs use different numbers
274 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

297 also a more flexible and interpretable predictor than the general regression. In generic notation, a
298 brief description of the AR predictor is given by:

299 $$F(\boldsymbol{X}) = \sum_{m=0}^{M} \beta_m h(\boldsymbol{X}; a_m) \quad (1)$$

300 where $h(\boldsymbol{X}; \boldsymbol{a}_m)$ are unknown as the basic function and independent model output made by
301 inputs $\boldsymbol{X}$ and model parameters $\boldsymbol{a}_m$ at iteration m ($m = 1, \ldots, M$ and M is the number of
302 iterations), β_m are a set of basis coefficients at iteration m, and $F(\boldsymbol{X})$ is the AR algorithm output.
303 The best results are achieved when the standalone and AR algorithms' parameters (a_m and β_m)
304 are used in a stepwise method (i.e., each set of parameters is estimated at a particular iteration).
305 Fig 2 shows the conceptual model of developing the hybrid model by coupling AR with the
306 standalone algorithms (here as an example, REPT).
307
308 Fig 2. Conceptual model for the development of the AR hybrid algorithms (Mitchell, 1997)

309 ***3.4.6. Disjoint aggregating (Dagging)***

310 The Disjoint Aggregating (Dagging) is a resampling integration and group-sampling technique
311 that was proposed by Ting and Witten (1997). Dagging and Bagging work in a similar way but in
312 the Dagging method, the training dataset is used for classification of several disjoint subsets by
313 using separate samples rather than bootstrap samples (Chen et al., 2019b). In the Dagging model,
314 the majority voting combines several classifiers to build the final prediction and improve the
315 accuracies of the basic classifiers (Kotsianti and Kanellopoulos 2007). In order to build a robust
316 model, the weak learners are trained on various subsets of the training set (Onan et al., 2016).
317

318 **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, $\overline{Q_t^{pred}}$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

340 **4.1- Best Input Combination**

341 Selecting the optimal input variables is the first step in developing a predictive artificial
342 intelligence model. Also as the effectiveness of each input variable is not equal, and some of
343 them might have a null or negative effect on the results, it is necessary to determine the
344 effectiveness of each of the input variables. In the present study the *r* approach between inputs
345 and outputs has been applied to determine the effectiveness of each input variable (Fig 3). The
346 results reveal that rainfall values (R_t) are the most effective variables for Q_t prediction (r = 0.56),
347 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}
348 (r = 0.07). Accordingly, and as expected, it was found that the greater the lag-time, the lower the
349 r value and the predictive effectiveness of the variable.

350

351 Fig 3. r value between input variables and streamflow

352

353 Eight different input combinations scenarios were constructed and compared (Table 4), the
354 effectiveness of each input combination was evaluated in the training and validation phases using
355 the estimated r. The results of the r value during the validation phase showed that the best input
356 combination scenario is different for every model developed. Combination No. 7 (i.e.
357 combination of all input variables) had the highest r values and proved to be the most effective
358 for the standalone REPT model. For all the ensemble based models, with the exception of RC-
359 REPT, the input scenario No. 8 (i.e. combination of R_t, R_{t-1}, and Q_{t-1} variables) was identified as
360 the most effective scenario. For the RC-REPT model, combination No. 3, which is the
361 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 week 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 m3 s-1) was superior to ANN (R^2 = 0.78 and RMSE = 7.40 m^3/s) and MARS (R^2 = 0.79 and RMSE = 7.47 m^3/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

Abrahart, R.J., See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrol.Process. 14 (July 1999), 2157–2172.DOI:10.1002/ 1099-1085(20000815/30)14:11/12<2157::AID-HYP57>3.0.CO;2-S..

Ahmed JA, Sarma AK (2007) Artificial neural network model for synthetic streamflow generation.Water ResourManag 21:1015–1029. doi:10.1007/s11269-006-9070

Afan HA, El-Shafie A, Yaseen ZM, et al. (2014) ANN Based Sediment Prediction Model Utilizing DifferentInput Scenarios. Water Resour Manag 29:1231–1245. doi:10.1007/s11269-014-0870-1.

Allawi, M.F., El-Shafie, A., 2016. Utilizing RBF-NN and ANFIS methods for multi-leadahead

prediction model of evaporation from reservoir. Water Resour. Manage. http://dx.doi.org/10.1007/s11269-016-1452-1.

Anusree K., Varghese, K.O. 2016. Streamflow Prediction of Karuvannur River Basin Using ANFIS, ANN and MNLR Models. Procedia Technology, 24: 101-108.

Auria L, Moro R A (2009) Support Vector Machines (SVM) as a Technique for Solvency Analysis. DIW Berlin Discussion Paper No. 811. Available at SSRN: https://ssrn.com/abstract=1424949 or http://dx.doi.org/10.2139/ssrn.1424949.

Asim, Y., Shahid, A.R., Malik, A.K. and Raza, B., 2018. Significance of machine learning algorithms in professional blogger's classification. Computers & Electrical Engineering, 65, pp.461-473

Ayele G, Teshale E, Yu B, Rutherfurd I, Jeong J (2017) Streamflow and sediment yield prediction for watershed prioritization in the Upper Blue Nile River Basin, Ethiopia Water 9:782.

Box GEP, Jenkins GM (1970) Time Series Analysis, Forecasting and Control, 1st editio. Holden-Day, SanFrancisco, CA.

Breiman, L., and Friedman, J. 1985. Estimating Optimal Transformations for Multiple Regression and Correlation. Journal of the American Statistical Association, 80 (391):580-599.

Bui. DT, K Khosravi, M Karimi, G Busico, ZS Khozani, H Nguyen,2020a. Enhancing nitrate and strontium concentration prediction in groundwater by using new data mining algorithm Science of The Total Environment, 136836.

Bui. DT, K Khosravi, J Tiefenbacher, H Nguyen, N Kazakis., 2020b.Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Science of The Total Environment, 137612

Bui, D., Pradhan, B., Nampak, H., TQ Bui, QA Tran, QP Nguyen. 2016. Hybrid Artificial Intelligence Approach Based on Neural Fuzzy Inference Model and Metaheuristic Optimization for Flood Susceptibility Modelling in A High-Frequency Tropical. Journal of Hydrology

Burges C (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 2(2):121–167

Bray, M., Han, D., 2004. Identification of support vector machines for runoff modelling. J. Hydroinf., 265–280

Cigizoglu, H.K., 2005. Application of generalized regression neural networks to intermittent flow forecasting and estimation. J. Hydrol. Eng. 10 (4), 336–341. http://dx.doi.org/10.1061/(ASCE)1084-0699(2005) 10:4(336).

Carlson RF, MacCormick AJA, Watts DG. 1970. Application of linear random models to four annual streamflow series. Water Resources Research 6(4):1070–1078.

Chen, X., Chau, K., Wang, W., 2015. A novel hybrid neural network based on continuity equation and fuzzy pattern-recognition for downstream daily river discharge forecasting. J. Hydroinformatics 17, 733–744. doi:10.2166/hydro.2015.095.

Chen, J., Wang, X., Zhai, J., 2009. Pruning decision tree using genetic algorithms. In: Artificial Intelligence and Computational Intelligence, 2009. AICI'09. International Conference on. IEEE, pp. 244–248.

Chen, W., M. Panahi, K. Khosravi, H.R. Pourghasemi, F. Rezaie, and D. Parvinnezhad. 2019a. Spatial prediction of groundwater potentiality using ANFIS ensembled with teaching-learning-based and biogeography-based optimization. Journal of Hydrology 572:

435–448.
Chen, W., Pradhan, B., Li, S., Shahabi, H., Rizeei, H. M., Hou, E., & Wang, S. (2019b). Novel Hybrid Integration Approach of Bagging-Based Fisher's Linear Discriminant Function for Groundwater Potential Analysis. [journal article]. *Natural Resources Research, 28*(4), 1239-1258, doi:10.1007/s11053-019-09465-w.
Chiang ,Y.M, L.C. Chang, F.J. Chang. 2004, Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling, Journal of Hydrology 290 (3–4) (2004) 297–311.
Choubin, B., H. Darabi, O. Rahmati, F. Sajedi-Hosseini, and B. Klove. 2018. River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. Science of the Total Environment 615: 272–281.
Costabile P, Costanzo C, Macchione F, Mercogliano P (2012) Two-dimensional model for overland flow simulations: A case study. Eur Water 38:13–23

Danandeh Mehr, A., Kahya, E., S ahin, A., Nazemosadat, M.J., 2014. Successive-station monthly streamflow prediction using different artificial neural network algorithms. Int. J. Environ. Sci. Technol. http://dx.doi.org/10.1007/s13762-014- 0613-0
Deo, R.C., Sahin, M., 2016. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environ. Monit. Assess. 188 (2), 1–24. http://dx.doi.org/10.1007/s10661-016-5094-9.
De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2
Esposito, F., Malerba, D., Semeraro, G., Tamma, V., 1999. The effects of pruning methods on the predictive accuracy of induced decision trees. Appl. Stoch. Model. Bus. Ind. 15, 277–299.
Fahrmeir, L., Kneib, T., Lang, S., 2004. Penalized additive regression for spacetime data: a Bayesian perspective. Stat. Sin. 14, 731e761.
French, M.N., Krajewski ,W.F., Cuykendall, R.R.,1992., Rainfall forecasting in space and time using a neural network, Journal of Hydrology 137 (1–4) (1992) 1–31.
Ghorbani, M.A., Zadeh, H.A., Isazadeh, M., Terzi, O., 2016. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. Environ. Earth Sci. 75 (6). http://dx.doi.org/10.1007/ s12665-015-5096.
Granata, F., Saroli, M., Marinis, G., Gargan, R. 2018. Machine Learning Models for Spring Discharge Forecasting. Geofluid, https://doi.org/10.1155/2018/8328167

Ghorbani, M.A., Deo, R.C., Karimi, V., Yaseen, Z.M., Terzi, O., 2017. Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. Stoch. Environ. Res. Risk Assess. 1–15. https://doi.org/10.1007/s00477-017-1474-0
Gupta H V, Sorooshian S, Yapo P O (1999) status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. Journal of hydrologic engineering, (4):2 10.1061/(ASCE)1084-0699.
Guven, A., 2009. Linear genetic programming for time-series modelling of daily flow rate. J. Earth Syst. Sci. 118, 137–146. http://dx.doi.org/10.1007/s12040-009-0022-9.
He et al. 2014. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region.February 2014 .Journal of Hydrology 509:379–386.10.1016/j.jhydrol.2013.11.054

635 Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Trans.
636 Pattern Anal. Mach. Intell. 20, 832–844. https://doi.org/10.1109/34.709601
637 Hooshyaripor, F., Tahershamsi, A., Golian, S., 2014. Application of copula method and neural networks
638 for predicting peak outflow from breached embankments. J. Hydro-Environment Res. 8, 292–303.
639 https://doi.org/10.1016/j.jher.2013.11.004

640 Huang, S., Chang, J., Huang, Q., Chen, Y., 2014. Monthly streamflow prediction using modified
641 EMD-based support vector machine. J. Hydrol. 511, 764–775.
642 doi:10.1016/j.jhydrol.2014.01.062.
643 Hintze, J.L., and R.D. Nelson. 1998. Violin plots: A box plot-density traces synergism. Journal
644 of the American Statistical Association 52, no. 2: 181–184.
645 Hipel KW, Mcleod AI. 1994. Time Series Modeling of Water Resources and Environmental
646 Systems. Elsevier: Amsterdam; 463–465.
647 Hsu, K.-L., Gupta, H.V., Gao, X., Sorooshian, S., Imam, B., 2002. Self-organizing linear output
648 map (SOLO): an artificial neural network suitable for hydrologic modeling and analysis.
649 Water Resour. Res. 38 (12). http://dx.doi.org/10.1029/ 2001WR000795.
650 Hussain, D., Khan, A.A., 2020. Machine learning techniques for monthly river flow forecasting of Hunza
651 River, Pakistan. Earth Sci. Informatics. doi:10.1007/s12145-020-00450-z

652 Kasiviswanathan, K.S., He, J., Sudheer, K.P., Tay, J., 2016. Potential application of wavelet
653 neural network ensemble to forecast streamflow for flood management. J. Hydrol. 536,
654 161–173. doi:10.1016/j.jhydrol.2016.02.044.

655 Kakaei Lafdani و E., Moghaddam Nia و A., Ahmadi, A., Jajarmizadeh, M., Ghafari Gosheh, M.
656 2013. Stream Flow Simulation using SVM, ANFIS and NAM Models (A Case Study).
657 https://www.researchgate.net/publication/257001623_
658 Kisi, O., Ozkan, C., Akay, B., 2012. Modeling discharge–sediment relationship using neural
659 networks with artificial bee colony algorithm. J. Hydrol. 428–429, 94–103.
660 http://dx.doi.org/10.1016/j.jhydrol.2012.01.026.
661 Khosravi, K., Nohani, E., Maroufinia, E. and Pourghasemi, H.R., 2016. A GIS-based flood
662 susceptibility assessment and its mapping in Iran: a comparison between frequency
663 ratio and weights-of-evidence bivariate statistical models with multi-criteria decisionmaking
664 technique. Natural Hazards, 83(2), pp.947-987.
665 Khosravi, L Mao, O Kisi, ZM Yaseen, S Shahid.,2018a., Quantifying hourly suspended sediment
666 load using data mining models: case study of a glacierized Andean catchment in Chile.
667 Journal of Hydrology 567, 165-179
668 Khosravi, M Panahi, D Tien Bui.,2018b.Spatial prediction of groundwater spring potential
669 mapping based on an adaptive neuro-fuzzy inference system and metaheuristic
670 optimization.Hydrology & Earth System Sciences 22 (9)
671 Khosravi, K., BT Pham, K Chapi, A Shirzadi, H Shahabi, I Revhaug,2018c. A comparative
672 assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz
673 watershed, northern Iran. Science of the Total Environment 627, 744-755
674 Khosravi, P Daggupati, MT Alami, SM Awadh, MI Ghareb, M Panahi., 2019.Meteorological
675 data mining and hybrid data-intelligence models for reference evaporation simulation: A
676 case study in Iraq.Computers and Electronics in Agriculture 167, 105041.
677 Khosravi, K., JR Cooper, P Daggupati, BT Pham, DT Bui. 2020. Bedload transport rate prediction:
678 Application of novel hybrid data mining techniques. Journal of Hydrology, 124774

679 Khosravi, R Barzegar, S Miraki, J Adamowski, P Daggupati, .2020a., Stochastic Modeling of
680 Groundwater Fluoride Contamination: Introducing Lazy Learners.Groundwater.
681 Khosravi, JR Cooper, P Daggupati, BT Pham, DT Bui., 2020b.Bedload transport rate prediction:
682 Application of novel hybrid data mining techniques. Journal of Hydrology, 124774.
683 Khozani, Z.S., Khosravi, K., Pham, B.T., Kløve, B., Mohtar, W., Melini, W.H. and Yaseen,
684 Z.M., 2019. Determination of compound channel apparent shear stress: application of novel
685 data mining models. Journal of Hydroinformatics
686 Kotsianti SB, Kanellopoulos D (2007) Combining bagging, boosting and dagging for
687 classification problems. Presented at the international conference on knowledge-based and
688 intelligent information and engineering systems, Springer, pp 493–500
689 Liu, S.N., Shi, H.Y., 2019. A recursive approach to long-term prediction of monthly precipitation using
690 genetic programming. Water Resources Management, 33(3), 1103-1121.
691 Legates DR, McCabe GJ (1999) Evaluating the use of "goodness-of-fit" measures in
692 hydrologic and hydroclimatic model validation. Water Resources Research, 35: 233–241.
693 Long H, Zhang Z, Su Y (2014) Analysis of daily solar power prediction with data-driven
694 approaches. Applied Energy 126:29–37. https ://doi.org/10.1016/j.apene rgy.2014.03.084.
695 Meshgi, A., Schmitter, P., Chui, T.F.M., Babovic, V., 2015. Development of a modular
696 streamflow model to quantify runoff contributions from different land uses in tropical urban
697 environments using Genetic Programming. J. Hydrol. 525, 711–723.
698 doi:10.1016/j.jhydrol.2015.04.032
699 Melesse, A.M.S., M.E. Ahmad, X.W. McClain, and Y.H. Lim. 2011. Suspended sediment load
700 prediction of river systems: An artificial neural network approach. Agriculture Water
701 Management 98, no. 5: 855–866.
702 Mert A, Kılıç N, Akan A. 2014. Evaluation of bagging ensemble method with time-domain
703 670 feature extraction for diagnosing of arrhythmia beats. Neural .Comput. Appl; 24: 317-
704 326.
705 Mitchell, T. 1997. Machine learning, McGraw Hill. 414 pages. ISBN 0070428077

706 Mohamed, W.N.H.W., Salleh, M.N.M., Omar, A.H. 2012. A comparative study of reduced error
707 pruning method in decision tree algorithms. Control System, Computing and Engineering
708 (ICCSCE), 2012 IEEE International Conference on, IEEE, pp. 392-397.
709 Mohamed, W.N.H.W., Salleh, M.N.M., Omar, A.H., 2013. A comparative study of Reduced
710 Error Pruning method in decision tree algorithms, in: Proceedings - 2012 IEEE International
711 Conference on Control System, Computing and Engineering, ICCSCE 2012.
712 https://doi.org/10.1109/ICCSCE.2012.6487177.
713 Mielniczuk, J., Teisseyre, P., 2014. Using random subspace method for prediction and variable
714 importance assessment in linear regression. Comput. Stat. Data Anal. 71, 725–742.
715 https://doi.org/10.1016/j.csda.2012.09.018.
716 Moriasi D, Arnold J, Liew M W Van, Ron B, Harmel R D, Veith T L (2007) Model
717 evaluation guidelines for systematic quantification of accuracy in watershed
718 simulations. Transactions of the ASABE, 50:885–900
719 Nanni, L., Lumini, A., 2008. Random subspace for an improved BioHashing for face
720 authentication. Pattern Recognit. Lett. 29, 295–300.
721 https://doi.org/10.1016/j.patrec.2007.10.005
722 Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid
723 wavelet-Artificial Intelligence models in hydrology: a review. J. Hydrol.
724 514, 358–377. http://dx.doi.org/10.1016/j.jhydrol.2014.03.057.

Nourani, V., Alami, M.T., Aminfar, M.H., 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng. Appl. Artif. Intell. 22, 466–472. http://dx.doi.org/10.1016/j.engappai.2008.09.003.

Najafi B, Ardabili SF (2018) Application of ANFIS, ANN, and logistic methods in estimating biogas production from spent mushroom compost (SMC) Resources, Conservation and Recycling 133:169-178

Nohani, E., M Moharrami, S Sharafi, K Khosravi, B Pradhan, BT Pham. 2019. Landslide susceptibility mapping using different GIS-based bivariate models. Water 11 (7), 1402.

Onan, A., Korukoğlu, S., & Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. Expert Systems with Applications, 57, 232-247.http://dx.doi.org/10.1016/j.eswa.2016.03.045.

Opitz, D.; Maclin, R. Popular ensemble methods: An empirical study. J. Artif. Intell. Res. 1999, 11, 169–198.

Pham, B.T., Khosravi, K., Prakhsh, I. 2017. Application and comparison of decision tree-based machine learning methods in landside susceptibility assessment at Pauri Garhwal Area, Uttarakhand, India. Environ Processes, 4 (3):711–730.

Quej, Victor H. , Javier Almorox, Javier A. Arnaldo, Laurel Saito.2017.ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment

Rashidi, S., Vafakhah, M., Lafdani, E., Javadi, M. 2016. Evaluating the support vector machine for suspended sediment load forecasting based on gamma test. Arabian Journal of Geosciences, 583:140-152.

Rajurak M P, Kothyari U C, Chaube U C (2000) Modeling of the daily rainfall-runoff relationship with artificial neural network. Journal of hydrology, 285: 96-113 https://doi.org/10.1016/j.jhydrol.2003.08.011.

Rezaie-Balf, M., Zahmatkesh, Z. & Kim, S (2017) Soft Computing Techniques for Rainfall-Runoff Simulation: Local Non–Parametric Paradigm vs. Model Classification Methods. Water Resource Management, 31: 3843–3865. https://doi.org/10.1007/s11269-017-1711-9.

Singh VP, Cui H (2015) Entropy Theory for Streamflow Forecasting. Environ Process 2:449–460. doi:10.1007/s40710-015-0080-8

Salas, J.D., 1980. Applied Modeling of Hydrologic Time Series. Water Resources Publication.

Sharafati, A., Khosravi, K., Khosravinia, p., Ahmed, P., Salman SA, ZM Yaseen. 2019.The potential of novel data mining models for global solar radiation prediction. International Journal of Environmental Science and Technology, 1-18. https://doi.org/10.1007/s13762-019-02344-0.

Shiri, J. and Kişi, Ö., 2012. Estimation of daily suspended sediment load by using wavelet conjunction models. Journal of Hydrologic Engineering, 17(9), pp.986-1000.

Salih SQ, A sharafati, K Khosravi, H Faris, O Kisi, H Tao, M Ali, ZM Yaseen., 2019.River suspended sediment load prediction based on river discharge information: application of newly developed data mining models. Hydrological Sciences Journal.

Sihag, P., Kumar, M., Singh, B. 2020. Assessment of infiltration models developed using soft computing techniques. Geology, Ecology, and Landscapes, 1-11

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres, 106(D7), pp.7183-719210.1007/s40710-016-0147-1.

770 Tama, B.A., & Comuzzi, M. (2019). An empirical comparison of classification techniques for
771 next event prediction using business process event logs. *Expert Systems with Applications*,
772 *129*, 233-245.https://doi.org/10.1016/j.eswa.2019.04.016.
773 Termeh, SV. K Khosravi, M Sartaj, SD Keesstra, FTC Tsai, R Dijksma. 2019. Optimization of
774 an adaptive neuro-fuzzy inference system for groundwater potential mapping.
775 Hydrogeology Journal 27 (7), 2511-2534.
776 Taormina, R., Chau, K., Sethi, R., 2012. Artificial neural network simulation of hourly
777 groundwater levels in a coastal aquifer system of the Venice lagoon. Eng. Appl.Artif. Intell.
778 25, 1670–1676. http://dx.doi.org/10.1016/j.engappai.2012.02.009.
779 Tapoglou, E., Trichakis, I.C., Dokou, Z., Nikolos, I.K., Karatzas, G.P., 2014. Groundwater-level
780 forecasting under climate change scenarios using an artificial neural network trained with
781 particle swarm optimization. Hydrol. Sci. J. 59, 1225–1239.
782 http://dx.doi.org/10.1080/02626667.2013.838005.
783 Ting KM, Witten IH (1997) Stacking bagged and dagged models. (Working paper 97/09).
784 Hamilton, New
785 Zealand: University of Waikato, Department of Computer Science
786 Valipour, M., 2012. Critical areas of Iran for agriculture water management according to the
787 annual rainfall. Eur. J. Sci. Res. 84, 600–908
788 Valipour, M., Montazar, A., 2012a. Optimize of all effective infiltration parameters in furrow
789 irrigation using visual basic and genetic algorithm programming. Aust. J.
790 Basic Appl. Sci. 6, 132–137.
791 Valipour, M., Montazar, A., 2012b. Sensitive analysis of optimized infiltration parameters in
792 SWDC model. Adv. Environ. Biol. 6, 2574–2581
793 Valipour, M., 2015. Long-term runoff study using SARIMA and ARIMA models in the United
794 States. Meteorol. Appl. http://dx.doi.org/10.1002/met.1491.
795 Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA,
796 and the autoregressive artificial neural network models in forecasting
797 the monthly inflow of Dez dam reservoir. J. Hydrol. 476, 433–441. http://dx.doi.
798 org/10.1016/j.jhydrol.2012.11.017.
799 Waseem Ahmad, M., Reynolds, J., Rezgui , Y. 2018. Predictive modelling for solar thermal
800 energy systems: A comparison of support vector regression, random forest, extra trees and
801 regression trees. Journal of clean production, 203:810-821.
802 https://doi.org/10.1016/j.jclepro.2018.08.207
803 Wu, C.L., Chau, K.W., Li, Y.S., 2009. Predicting monthly streamflow using datadriven models
804 coupled with data-preprocessing techniques. Water Resour. Res.
805 45, 1–23. http://dx.doi.org/10.1029/2007WR006737.
806 Xia, J., Dalla Mura, M., Chanussot, J., Du, P., He, X., 2015. Random Subspace Ensembles for
807 Hyperspectral Image Classification with Extended Morphological Attribute Profiles. IEEE
808 Trans. Geosci. Remote Sens. 53, 4768–4786. https://doi.org/10.1109/TGRS.2015.2409195
809 Yaseen, Z.M., El-Shafie, A., Afan, H.A., Hameed, M., Mohtar, W.H.M.W., Hussain, A., 2015.
810 RBFNN versus FFNN for daily river flow forecasting at Johor River,
811 Malaysia. Neural Comput. Appl. http://dx.doi.org/10.1007/s00521-015-1952-6.
812 Yaseen, Z. M., Ebtehaj ,I., Bonakdari, H.,. Deo, R.C,
813 Danandeh Mehr, A., Melini ,W.H., Mohtar, W., Diop, L, Elshafie, A, Vijay P. Singh., 2017.
814 Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. Journal of
815 Hydrology. S0022-1694(17)30602-9. http://dx.doi.org/10.1016/j.jhydrol.2017.09.007

Zhang, X., Peng, Y., Zhang, C., Wang, B., 2015. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J. Hydrol. 530, 137–152. doi:10.1016/j.jhydrol.2015.09.047.

Zhang, X., Jia, Y., 2007. A linear discriminant analysis framework based on random subspace for face recognition. Pattern Recognit. 40, 2585–2591. https://doi.org/10.1016/j.patcog.2006.12.002

Figure caption

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

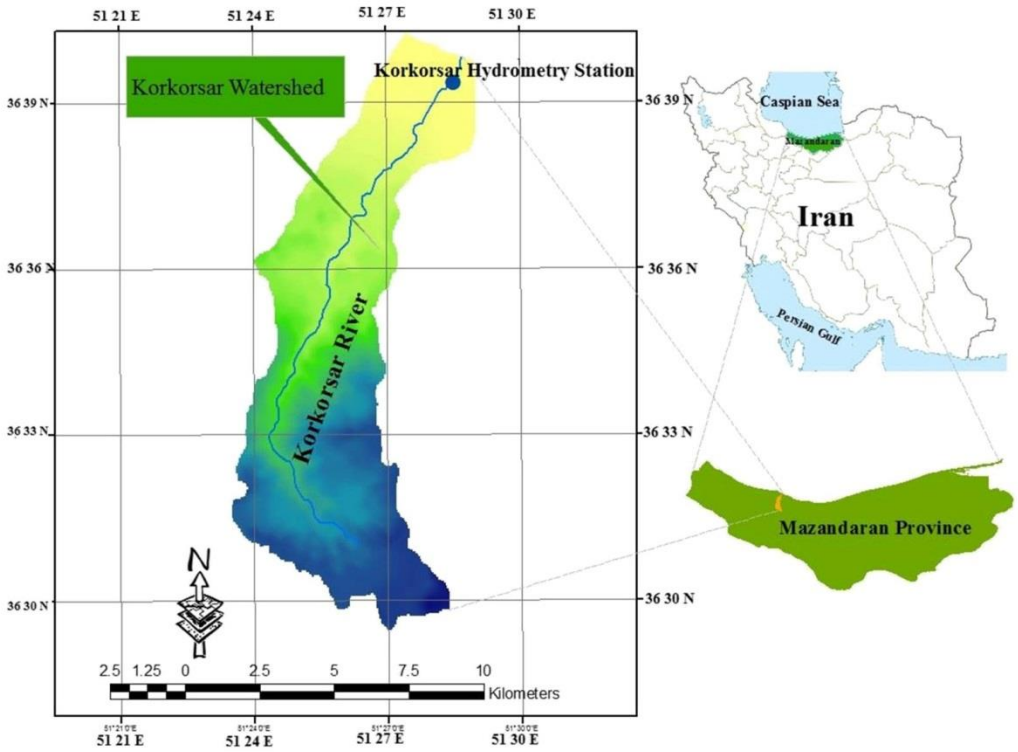
Fig 2. Conceptual model of working AR hybrid algorithms (Mitchell, 1997)

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

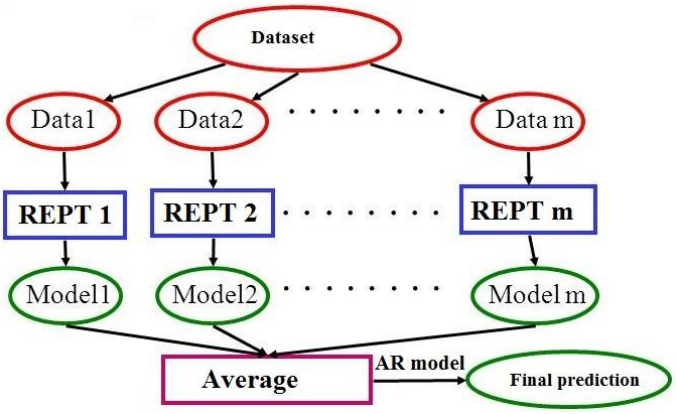


835

836 Fig 1. Location of the Kurkursar River and the Hydrometry station in the catchment (after

837 Rashidi et al. 2016)

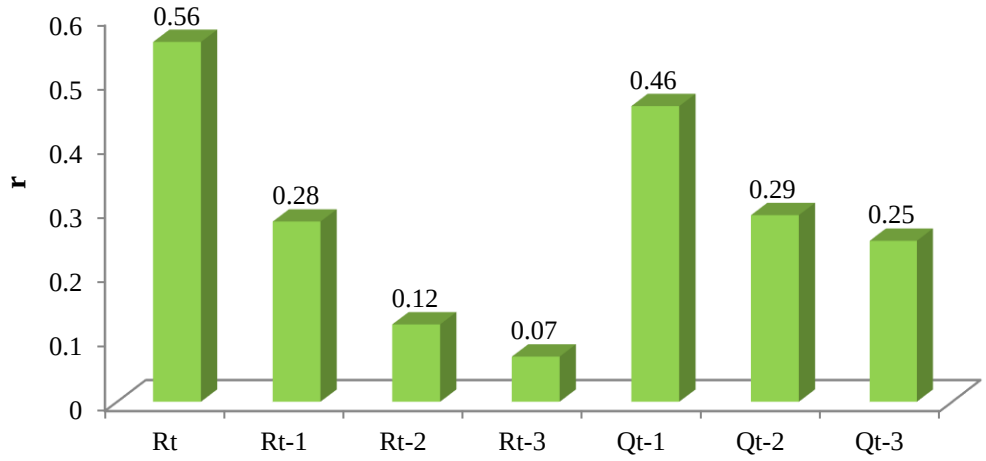
838



839

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

841

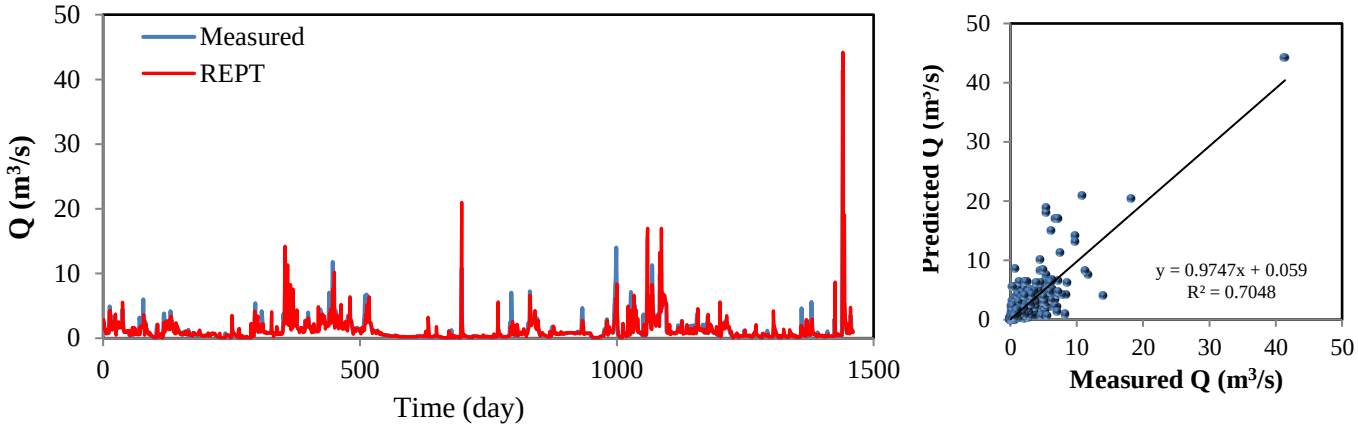


842

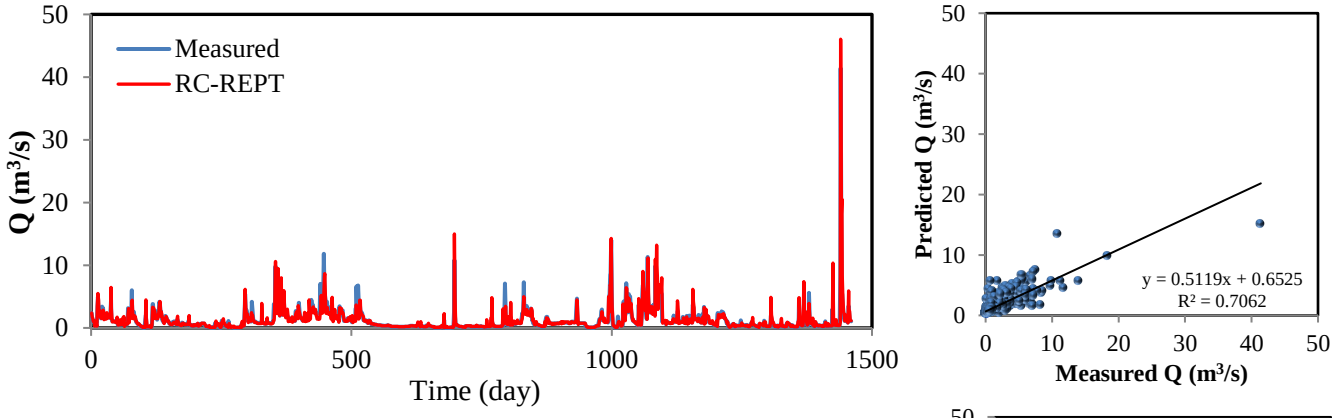
843 Fig 3. r value between input variables and streamflow

844

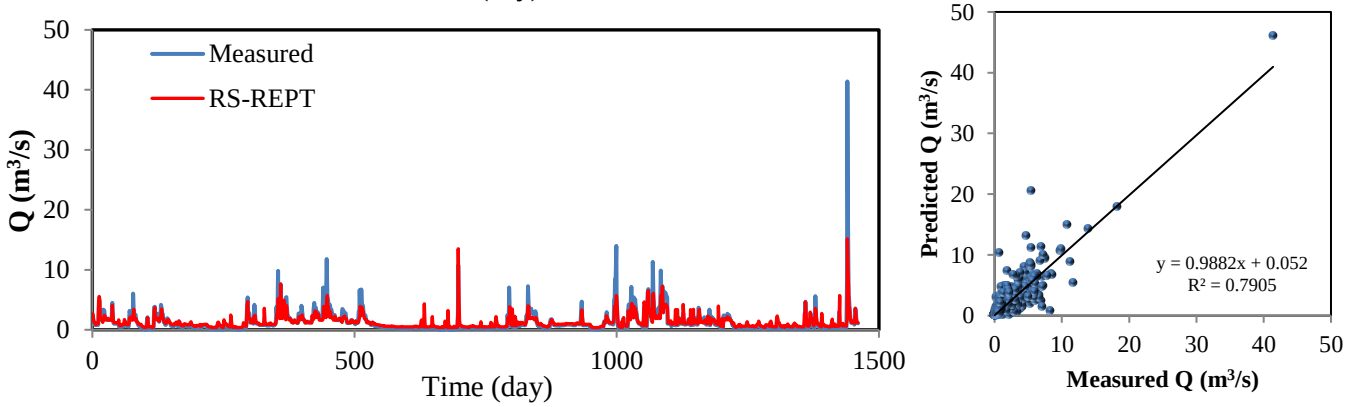
845 .



846

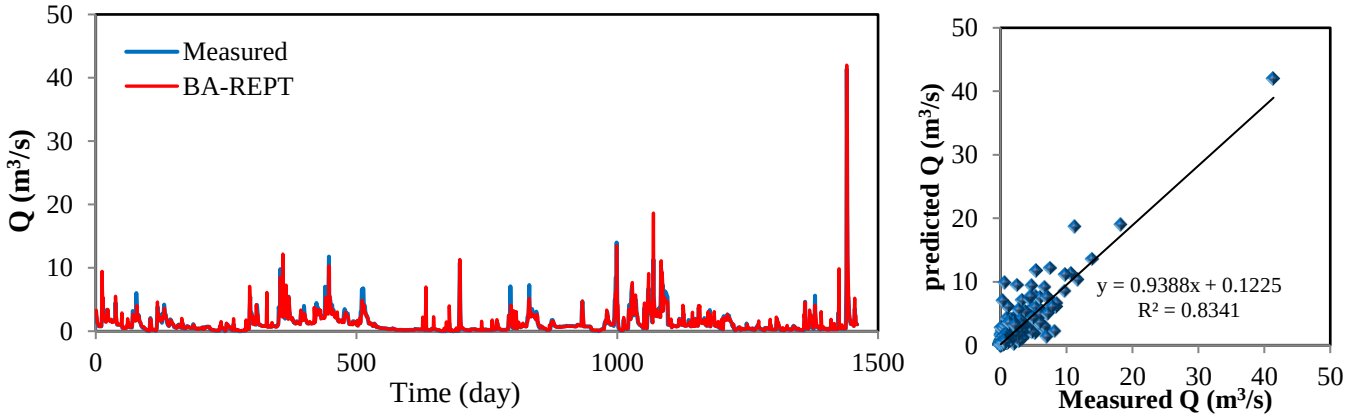


847

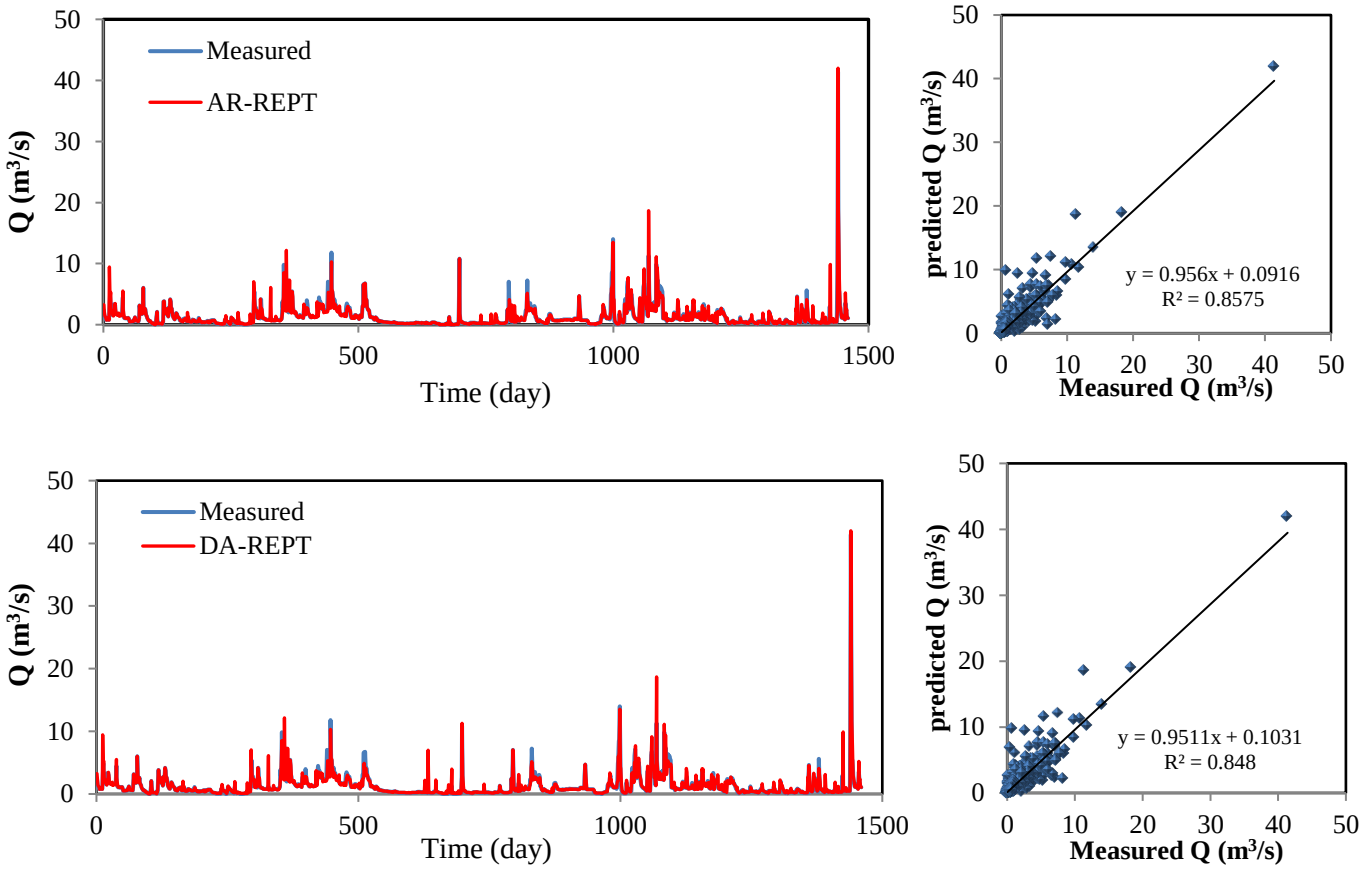


848

849



850



851

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

853 validation period

854

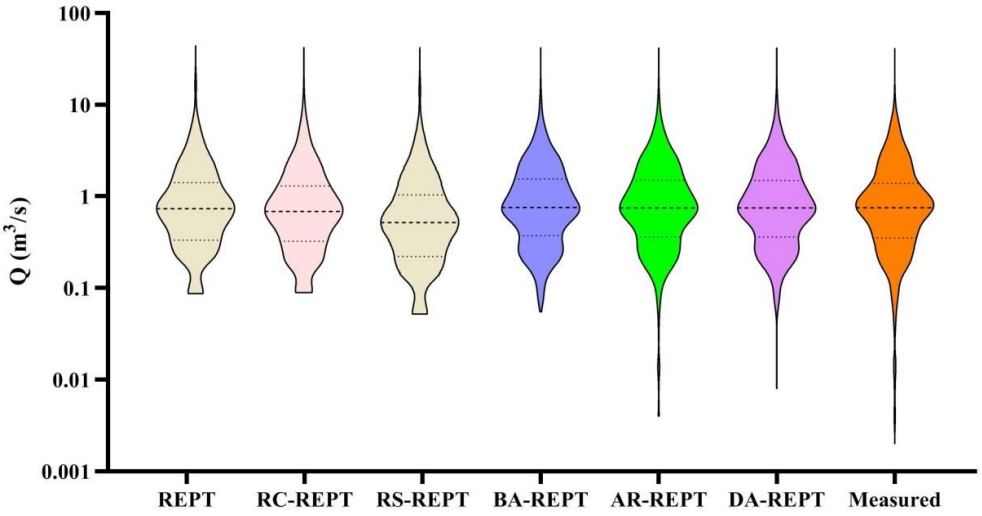
855

856

857

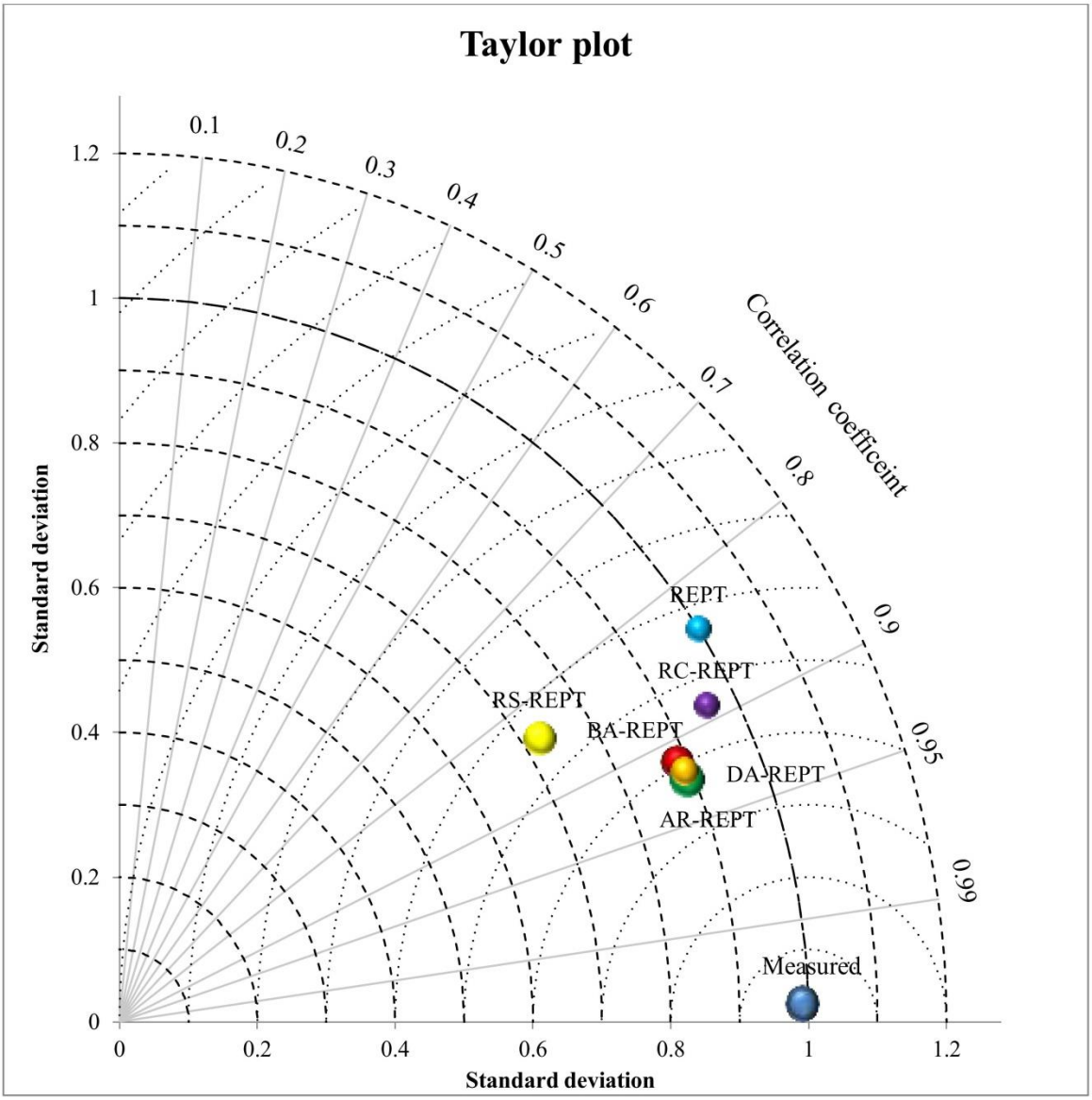
858

859



860

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



862

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

864

865

866

867

Table caption

Table 1. Basic statistics for the training and testing datasets

Table 2. Different input scenarios

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

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

Table 5. Model evaluation quantitatively during validation phase

Table 6. Model performance ranking

Table 1. Basic statistics for the training and testing datasets

	Training dataset				Validation dataset			
Variables	Min	Max	Mean	Std. deviation	Min	Max	Mean	Std. deviation
Rainfall	5.00	149.00	3.47	11.43	0.00	147.00	3.42	10.69
Streamflow	0.73	73.10	1.28	2.25	0.002	41.40	1.21	1.74

890 Table 2. Different input scenarios

No.	input combination	Output
1	R_t	Q_t
2	R_t, Q_{t-1}	Q_t
3	R_t, Q_{t-1}, Q_{t-2}	Q_t
4	R_t, Q_{t-1}, Q_{t-2}, R_{t-1}	Q_t
5	R_t, Q_{t-1}, Q_{t-2}, R_{t-1}, Q_{t-3}	Q_t
6	R_t, Q_{t-1}, Q_{t-2}, R_{t-1}, Q_{t-3}, R_{t-2}	Q_t
7	R_t, Q_{t-1}, Q_{t-2}, R_{t-1}, Q_{t-3}, R_{t-2}, R_{t-3}	Q_t
8	R_t, R_{t-1}, Q_{t-1}	Q_t

891

892

893

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

Equation	Ref.	Range	Performance
$R^2 = 1 - \left(\frac{\sum_{i=1}^{i=N}\left(Q_t^{obs} - Q_t^{pred}\right)^2}{\sum_{i=1}^{i=N}\left(Q_t^{obs}\right)^2}\right)$	Najafi and Ardabili (2018)	0.7≤ R2 ≤ 0.1 0.6 ≤ R2 ≤0.7 0.5 ≤ R2 ≤ 0.6 0 ≤ R2 ≤ 0.5	Very good Good Satisfactory Unsatisfactory
$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}\left(Q_t^{pred} - Q_t^{obs}\right)^2}$	(Najafi and Ardabili 2018)		The lower the better
$MSE = \frac{1}{N}\sum_{i=1}^{i=N}\left(Q_{t_i}^{pred} - Q_{t_i}^{obs}\right)^2$	(Najafi and Ardabili 2018)		The lower the better
$NSE = 1 - \frac{\sum_{i=1}^{i=N}\left(Q_t^{pred} - Q_t^{obs}\right)^2}{\sum_{i=1}^{N}\left(Q_t^{pred} - \overline{Q_t^{pred}}\right)^2}$	Moriasi et al, 2007	0.75 < *NSE* ≤ 1.00 0.65 < *NSE* ≤ 0.75 0.50 < *NSE* ≤ 0.65 0.4 < *NSE* ≤ 0.50 *NSE* ≤ 0.4	Very good Good Satisfactory Acceptable Unsatisfactory
$PBIAS = \frac{\sum_{i=1}^{i=N}\left(Q_t^{pred} - Q_t^{obs}\right)}{\sum_{i=1}^{i=N} Q_t^{pred}}$	Legates et al, 1999	*PBIAS* < ±10% ±10% ≤ *PBIAS* <±15% ±15%≤ *PBIAS* < ±25% *PBIAS* ≥ ±25%	Very good Good Satisfactory Unsatisfactory

$\text{RSR} = \frac{\sqrt{\sum_{i=1}^{i=N}\left(Q_t^{pred} - Q_t^{obs}\right)^2}}{\sqrt{\sum_{i=1}^{i=N}\left(Q_t^{pred} - \overline{Q_t^{pred}}\right)^2}}$	Gupta et al. 1999	$0 \le RSR \le 0.50$ $0.50 < RSR \le 0.60$ $0.60 < RSR \le 0.70$ $RSR > 0.70$	Very good Good Satisfactory Unsatisfactory

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

Input No.	REPT		BA-REPT		RC-REPT		RS-REPT		AR-REPT		DA-REPT	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	0.79	0.63	0.65	0.62	0.79	0.63	0.79	0.63	0.66	0.63	0.67	0.64
2	0.99	0.75	0.96	0.81	0.99	0.75	0.9	0.66	0.96	0.82	0.97	0.83
3	0.99	0.73	0.95	0.84	0.99	0.88	0.99	0.75	0.95	0.85	0.96	0.85
4	0.99	0.71	0.81	0.85	0.99	0.82	0.98	0.74	0.83	0.85	0.85	0.86
5	0.99	0.77	0.95	0.83	0.99	0.79	0.99	0.81	0.96	0.84	0.97	0.85
6	0.99	0.69	0.96	0.86	0.99	0.82	0.98	0.69	0.97	0.87	0.97	0.88
7	0.99	0.74	0.96	0.86	0.99	0.8	0.87	0.82	0.97	0.87	0.98	0.88
8	0.99	0.73	0.93	0.91	0.99	0.81	0.87	0.84	0.97	0.92	0.97	0.93

Table 5. Model evaluation quantitatively during validation phase

Models	R^2	RMSE (m^3/s)	MAE (m^3/s)	NSE	PBIAS (%)	RSR
REPT	0.704	1.25	0.420	0.600	-3.300	0.630
BA-REPT	0.834	0.73	0.300	0.820	-2.100	0.420
RC-REPT	0.706	1.000	0.450	0.650	-4.900	0.590
RS-REPT	0.790	0.88	0.350	0.740	-2.800	0.500
AR-REPT	0.857	0.682	0.258	0.845	-3.150	0.392
DA-REPT	0.848	0.706	0.274	0.834	-3.602	0.406

Table 6. Model performance ranking

Models	R^2	RMSE (m^3/s)	MAE (m^3/s)	NSE	PBIAS (%)	RSR	Sum	**Rank**
REPT	6	6	5	6	4	6	33	**6**
BA-REPT	3	3	3	3	1	3	16	**3**

RC-REPT	5	5	6	5	6	5	32	**5**
RS-REPT	4	4	4	4	2	4	22	**4**
AR-REPT	1	1	1	1	3	1	8	**1**
DA-REPT	2	2	2	2	5	2	15	**2**

903