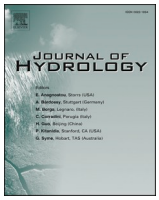
Journal of Hydrology 620 (2023) 129480

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

# Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran

Mohammad Akbarian [a], Bahram Saghafian [a,*], Saeed Golian [b]

[a] Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran 1477893855, Iran
[b] Irish Climate Analysis and Research UnitS (ICARUS), Department of Geography, Maynooth University, Maynooth, Co. Kildare, Ireland

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

*Keywords:*
Streamflow forecast
C3S data store
ECMWF
Ensemble
Recursive Feature Elimination (RFE)
Bayesian Networks (BN)
Machine learning (ML)

ABSTRACT

Seasonal hydrological forecasts play a critical role in water resources management. The Copernicus Climate Change Service (C3S) data store provides open access to monthly hydrological forecasts for up to six-months. This study aims to evaluate, for the first time, 1- to 3-month runoff forecasts using the European Centre for Medium-Range Weather Forecasts (ECMWF) ensembles of precipitation, runoff, and temperature in 1981–2015 period over a total of 30 s-level basins in Iran. We adopted the 5th, 50th and 95th ECMWF ensemble quantiles for each variable that represent low, medium and high probability of occurrence, respectively. Pearson correlation analysis (Pca), Recursive Feature Elimination (RFE) via random forest (RF) model, and Bayesian Networks (BN) feature selection algorithms were used in order to reduce input variable dimension and select potential predictors to be fed to the machine learning models. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Regression (SVR), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) machine learning models were used with Repeated K-Fold cross validation (rK-Fold CV) while model efficiency was evaluated using modified Kling-Gupta efficiency coefficient (KGE'), Nash-Sutcliffe Efficiency coefficient (NSE), and Normalized Root Mean Square Error (NRMSE). Results of this study revealed that C3S runoff ensembles have the highest impact on forecast accuracy of streamflow, followed by precipitation and temperature. Overall, model performance yield a best-to-worst ranking of ANN, XGBoost, RF, MLR, and SVR with KGE' values of 0.70, 0.68, 0.66, 0.57, and 0.41, respectively. The predictive performance of all models decreased with lead times beyond 1-month, where ANN and XGBoost outperformed other models with KGE' of 0.65 for 2-month lead time and 0.60 for 3-month lead time. The three superior models of XGBoost, ANN, and RF, were employed with RFE and BN FSAs most frequently across Iran's 30 s level basins in all lead times. Almost all models in the arid central region of Iran showed the lowest performance while highest skills were achieved in the western regions of Iran. Finally, for all models and over all regions, the model performance reduced by increase in lead-time.

## 1. Introduction

Given global water scarcity, particularly in the past decade, it is crucial to adopt the best water resource management practices to handle or avert consequent water crises (Greve et al., 2018). An important factor in water resource management is the accurate estimation of streamflow in order to plan for available water resources (Sharma & Machiwal, 2021). Long-term forecasts include weekly, monthly, seasonal, and even annual predictions and are crucial for operation of reservoirs, irrigation management systems, and hydropower generation (Liang et al., 2018). Improving the accuracy of long-term forecasts is significantly dependent on improvements in availability and sustainability of climate datasets and modeling tools (Wegayehu and Muluneh, 2022). Recent advances in meteorological forecasts open up new opportunities for improving long-term hydrological forecasting capabilities (Kilinc & Haznedar, 2022). Long-term forecasts are more complex to simulate compared to short-term forecasts (Karimi et al., 2016). The practical gap in access to accurate long term runoff predictions pushes research toward reliable hydroclimatological forecasts, particularly given the benefits received from even a slight increase in the accuracy of these forecasts to water resource management (Cheng et al., 2020).

Ensemble hydroclimatological forecasts are used to numerically produce a range of forecasts based on possible future atmospheric

* Corresponding author.
*E-mail address:* b.saghafian@gmail.com (B. Saghafian).

https://doi.org/10.1016/j.jhydrol.2023.129480
Received 16 August 2022; Received in revised form 26 March 2023; Accepted 1 April 2023
Available online 12 April 2023

conditions. Unlike single predictions, ensemble forecasting provides a more robust prediction by generating the range of possible outcomes and evaluating the confidence in future system states (Hawcroft et al., 2021). A prominent provider of ensemble forecasts is the Copernicus Climate Change Service (C3S) multi-system seasonal forecast (https://cds.climate.copernicus.eu/). The service, which has recently released several European center forecasts, is run on behalf of the European Union by ECMWF and offers seasonal forecasting protocols publicly available through the Climate Data Store. The database combines observations of weather systems and provides holistic past, present and future information on atmospheric conditions (Ingleby, 2015; Lopez, 2013). Given the recent launch of the C3S database, research on its evaluation is very limited, in particular over Iran. The evaluation of performance of ensemble forecast can provide context to decision makers' strategic choices for overcoming climate related risks (Nobakht et al., 2021).

Crochemore et al. (2017) evaluated precipitation and river flow forecasting performance in 16 basins in France. The study evaluated and post-processed seasonal precipitation forecasts from the open library of ECMWF (System 4) with 90-day lead time. They used linear scaling (LS) and different distribution mapping methods for post processing of monthly and annual raw precipitation data. The results demonstrated that applying post-processing techniques increase precipitation forecasting accuracy. Manzanas et al. (2019) applied simple Bias Adjustment (BA) methods, such as quantile mapping, as well as complicated ensemble Recalibration (RC) methods, such as non-homogenous Gaussian regression, to increase the accuracy of C3S forecasted precipitation and temperature. In particular, they evaluated UK Met Office (UKMO)-GloSea5 (Maclachlan et al., 2015), Météo France-System5 (Descamps et al., 2015), and ECMWF-SEAS5 (Johnson et al., 2019) seasonal forecasts with one-month lead time. The results revealed that both BA and RC methods correct large raw model biases effectively, with high model skill in confined regions and seasons. In these instances, RC bias correction outperformed BA methods (Manzanas et al., 2019). In another study, Gebrechorkos et al. (2022) evaluated the performance of precipitation forecasts from five potential climate models namely ECMWF, UK Met Office, Météo France, Deutscher Wetterdienst (Kaspar et al., 2015), and Centro Euro-Mediterraneo sui Cambiamenti Climatici (Nicolì et al., 2023). Multi-Source Weighted-Ensemble Precipitation was used as the reference data for model performance evaluation. Model performance was evaluated in daily, weekly, monthly, and seasonal timeframes, and at specific months and lead times. All models showed reliable predictions for 1-month forecasts. However, they showed rapid decline in performance with increase in lead times, in particular in drier regions and seasons. It was found that ECMWF followed by UK-Met were the most accurate models among other C3S products (Gebrechorkos et al., 2022).

Due to machine learning (ML) algorithms' significant nonlinear modeling capabilities in complex problems, they have been widely used for the monthly prediction of streamflow (Ali & Shahbaz, 2020). The data preprocessing, feature preprocessing, ML algorithm selection, and hyperparameter optimization stages are typically included in the machine-learning-based prediction process for streamflow prediction. One of the more common forms of data-driven models, multiple linear regression (MLR), has been shown to perform effectively for long-term forecasts (Krstanovic & Singh, 1991). However, due to the nature of the process, it makes the assumption that the relationships between input and output data are linear and the data has no multicollinearity whereas streamflow prediction is a highly nonlinear process and depends on a variety of known and unknown factors (Sudheer et al., 2014). When it comes to model selection, The Artificial Neural Networks (ANNs) (Kilinc & Haznedar, 2022), Deep Neural Networks (DNNs) (Apaydin et al., 2021; He et al., 2022; Kao et al., 2020, 2021; Maddu et al., 2022), Support Vector Regression (SVR) (Ni et al., 2020), Random Forest (RF) (Tyralis et al., 2021), Adaptive Boosting (AdaBoost) (Liu et al., 2014), Light Gradient Boosting Method (LGBM) (Szczepanek, 2022), and eXtreme Gradient Boosting (XGBoost) (Ni et al., 2020) are frequently used for streamflow prediction. Both DT and Gradient Boosting (GB) are utilized by XGBoost. It has many benefits, including its predictive algorithms are straightforward but still effective, simple to understand, and don't need as much data preparation (Ni et al., 2020). Artificial Neural Network (ANN)'s ability to process and model complex nonlinear time series has recently enabled its widespread applications in hydroclimatological studies, in contrast to application of physically-based models (Kilinc & Haznedar, 2022). The ability to work with large amounts of noisy data from nonlinear and dynamic systems, especially when the fundamental physical relationships are unknown, is one of the advantages ANN has over traditional modeling (Anusree & Varghese, 2016). ANN has been successfully applied in various hydrological simulations, known as an efficient and accurate tool in forecasting streamflow, precipitation, and water quality (Kilinc & Haznedar, 2022). RF is a decision tree-based model that addresses the overfitting issues with single decision trees while maintaining their prediction accuracy. The RF approach, in contrast to ANN and SVM, offers excellent computing speed while being simple to use (Schoppa et al., 2020). Multiple studies have reported complexities associated with flow forecasting, driven by the natural complexity, nonlinearity, and randomness of river systems (Smith et al., 2007). In order to reduce the number of predictors, different linear or non-linear Feature Selection Algorithms (FSAs) such as Pearson's correlation analysis (Pca) (Djibo et al., 2015), Recursive Feature Elimination (RFE) (Ferreira et al., 2021), and Bayesian networks (BN) (Das et al., 2022) are commonly used. In terms of ML algorithm calibration, the optimization of hyperparameters has a significant impact on the performance of ML models (Szczepanek, 2022). As a result, to optimize hyperparameters for ML algorithms, researchers have used different methods such as Grid Search (GS) (LaValle et al., 2016) or metaheuristics (Malik et al., 2020).

In Iran, ML algorithms have also been used in many hydroclimatological studies. For example, forecasts from ECMWF, UKMO, and National Centers for Environmental Prediction (NCEP) were evaluated for 13 synoptic stations in eight precipitation zones (Aminyavari et al., 2018). The evaluation was based on precipitation data from 2008 to 2016 with a 1- to 3-day lead times. Their results showed decrease in model performance with increase in lead time while different models performed differently over various regions, i.e., ECMWF in most regions, UKMO in mountainous regions, and NCEP around the Persian Gulf. Kolachian and Saghafian (2019) used a number of deterministic and probabilistic criteria to evaluate the performance of precipitation forecast for ECMWF Sub-seasonal to Seasonal (S2S) model with 1 month lead time for several synoptic stations and precipitation regimes in Iran. They found acceptable performance in wet months in most regions; however, absence of reliable raw forecasts in dry months was evident. No significant relationship was found between the precipitation regime and prediction skill. Furthermore, their results showed forecasting and post-processing capabilities vary significantly in different seasons and locations.

Nobakht et al. (2021) evaluated the ensemble precipitation forecasts of ECMWF, UKMO, and Météo France C3S models over 1993–2017 period in Iran's eight classified precipitation clusters with 1- to 3-month lead times. Probabilistic and non-probabilistic criteria were used for the evaluation. The results indicated all models performed better in the western precipitation regions, while in the northern region with humid climate models had poor skill scores. All forecasts were better at predicting upper-tercile events in dry seasons and lower-tercile events in wet seasons. Moreover, with increasing lead time, the forecast skills worsened. In terms of forecasting in dry and wet years, the predictions were generally close to observations, albeit they underestimated several severe dry periods and overestimated a few wet periods.

In another study, Meydani et al. (2022) applied weather forecast downscaling and rainfall-runoff modeling for daily reservoir inflow forecasts in the Urmia Lake basin in Iran. They utilized large scale weather forecasts from ECMWF and NCEP to evaluate various

downscaling methods, including Artificial Intelligence (AI) and Bayesian Belief Network (BBN) techniques, to derive local reservoir inflow forecasts. The results showed a hybrid downscaling approach, that combined Group method of data handling (GMDH) and Support vector regression (SVR), performed better than their non-hybrid counterparts in downscaling precipitation. Furthermore, the authors found BBN to outperform hybrid AI in forecasting the dynamics of precipitation in observed datasets.

In this study for the first time, the performance of ECMWF runoff forecasts over Iran is comprehensively assessed by evaluating the monthly ECMWF's seasonal forecasting system 5 (SEAS5) and its potential in predicting streamflow over 30 s level basins in Iran subject to different climate and hydrological regimes. Based on a set of initial "potential predictors" and their association (high correlation and statistical significance) with the target (monthly streamflow), a set of "potential predictors" is established. The "potential predictors" are then put through three Feature Selection Algorithms (FSA) including Bayesian Networks (BN), Recursive Feature Elimination (RFE), and Pearson correlation analysis (Pca), to create a set of "optimal predictors" for each scenario. These "optimum predictors" served as inputs for different ML algorithms. The study is also the first to use ECMWF runoff forecasts to predict monthly streamflow in Iran, according to the authors' knowledge. Additionally, this study investigates the use of five ML algorithms, notably XGBoost a novel intelligent method based on the gradient boosting algorithm to forecast the streamflow across Iran's 6 major- and 30 s-level basins.

## 2. Materials and methods

This section describes the study area, datasets, pre-processing and modeling techniques to develop streamflow forecasting with lead-times from 1 to 3 months over Iran. The data used are from the latest


**Fig. 1.** Spatial illustration of Iran's second level basins and the selected hydrometric stations.

generation of ECMWF's seasonal forecasting system 5 (SEAS5) and consist of ensemble forecasts of precipitation, runoff, and temperature from 1991 to 2015, with 1- to 3-month lead time. In order to reduce the dimension of input variables, Pearson's correlation analysis (Pca), Recursive Feature Elimination (RFE), and Bayesian Networks (BN) are used. MLR, ANN, RF, SVR, and XGBoost models with rK-Fold cross validation are adopted to model runoff in 30 Iranian basins. Lastly, recommendations are proposed for runoff forecast for different basins.

### 2.1. Study area

Iran with a total area of 1,648,195 km$^2$ is located in the mid-latitude northern region between 25°N and 40°N latitude, and 44°E and 64°E longitude in southwest Asia. Other than western and northern coastal areas, Iran's general climate is dominated by a mostly arid and semi-arid character. Precipitation varies mainly with latitude and topographical altitude of the region (Mansouri Daneshvar et al., 2019). The whole country may be divided into six major basins that are broken into a total of 30 s level basins (Saatsaz, 2020). Fig. 1 depicts the basins, rivers, and location of the hydrometeorological stations used in this study. In total, 571 hydrometric stations in the country have at least 240 months of data throughout the 1981–2015 study period. In this study, however, we selected the stations with the highest elevation in each basin, resulting in a total of 30 selected stations shown in Fig. 1. Table 1 provides characteristics of the basins including basin name, basin code, area, and climate class according to Köppen-Geiger climate classification (Raziei, 2022).

### 2.2. Datasets and Pre-Processing

#### 2.2.1. Hydro-Meteorological and ensemble seasonal forecast data

The observed station data spans the 1946 to 2015 period. Monthly averages of precipitation and runoff are shown in Fig. 2 for all 30- basins (in percentage).

C3S data store, launched in 2017 by ECMWF, regularly releases seasonal forecast products. The data store relies on forecasts from multiple organizations across Europe and updates monthly forecasts with up to six months lead time. The data includes forecasts created in real-time (since 2017) and retrospective forecasts (hindcasts) initialized at equivalent intervals during the 1981–2016 period. This study investigated monthly C3S precipitation, runoff, and temperature ensembles with 1- to 3-month lead time for the time period of 1981–2015 over the 25°–40°N 44°–64°E geographical area with a grid size of 0.25°×0.25° (with approximately 25-km spatial resolution).

The correlation coefficient between SEAS5 raw runoff, average precipitation, and temperature outputs with observed values over each basin is presented in Fig. 3 for 1- to 3-month lead time. It shows runoff ensemble demonstrates the highest correlation with observed values, followed by precipitation.

#### 2.2.2. Pre-Processing of raw product

In order to work with ECMWF SEAS5 ensemble outputs, the quantile classification method was applied in a way that for each variable, the -corresponding 5th, 50th and 95th quantiles were extracted to represent low, medium and high probability of occurrence values, respectively. These quantile values form potential predictors to drive machine learning (ML) models. Golian et al. (2010, 2011) used a similar quantile classification to derive rainfall thresholds with different probability of occurrence through a Monte Carlo framework. Fig. 4 illustrates the end-to-end data preparation process implemented in this study, starting with data retrieval from the ECMWF data store to preparation for runoff simulation. Ensemble precipitation, runoff, and temperature hindcasts were retrieved for each cell over the entirety of Iran from CDS website (https://cds.climate.copernicus.eu/) for lead-times from 1 to 3 months. Next, data in Gridded Binary format (GRIB) were converted and saved in a comma-separated values (CSV) format. These converted files include

**Table 1**
Characteristics of the studied basins.

| Major Basin No. | Major Basin Name | Basin No. | Basin Name | Area (km$^2$) | Climate classification* | Area (km$^2$) |
|---|---|---|---|---|---|---|
| 1 | Caspian Sea | 11 | Aras | 39,534 | Bsk | 174,618 |
| | | 12 | Talesh-Anzali Lagoon | 6,827 | Csa | |
| | | 13 | Sefidrud | 59,217 | Bsk | |
| | | 14 | Sefidrud - Haraz | 10,905 | Csa | |
| | | 15 | Haraz | 18,644 | Csa | |
| | | 16 | Gharasu and Gorgan | 13,061 | Bsk | |
| | | 17 | Atrak | 26,430 | Bsk | |
| 2 | Persian Gulf | 21 | West border | 39,667 | Csa | 424,515 |
| | | 22 | Karkhe | 51,643 | Csa | |
| | | 23 | Karun | 67,257 | Csa | |
| | | 24 | Jarahi & Zohre | 40,788 | BSh | |
| | | 25 | Heleh | 21,274 | BSh | |
| | | 26 | Mand | 47,654 | BSh | |
| | | 27 | Kol & Mehran | 62,918 | Bwh | |
| | | 28 | BandarAbbas-Sadij | 44,763 | Bwh | |
| | | 29 | South Baluchestan | 48,551 | Bwh | |
| 3 | Lake Urmia | 30 | Lake Urmia | 51,801 | Bsk | 51,801 |
| 4 | Central | 41 | Namak Lake | 92,563 | Bsk | 824,356 |
| | | 42 | Gavkhouni | 41,550 | Bsk | |
| | | 43 | Tashk, Bakhtegan, and Maharloo Lakes | 31,492 | BSh | |
| | | 44 | Abarkooh & Sirjan playas | 57,196 | Bsk | |
| | | 45 | Hamun-e Jaz Murian | 69,390 | Bwh | |
| | | 46 | Lut Desert | 206,222 | Bwh | |
| | | 47 | Central Desert | 226,523 | Bwh | |
| | | 48 | SiahKuh, RigZarin, and DeghSorkh Deserts | 48,912 | Bwh | |
| | | 49 | DorAnjir and Saghand Deserts | 50,508 | Bwk | |
| 5 | Eastern | 51 | Khaf Namak Zar | 32,980 | Bwk | 103,169 |
| | | 52 | Hamun Hirmand | 33,731 | Bwh | |
| | | 53 | Hamun Moshkil | 36,458 | Bwh | |
| 6 | Garagum | 60 | Garagum | 44,156 | Bsk | 44,156 |

* Köppen-Geiger climate classification (Raziei, 2022) (climate/weather/temperature): Bsk: arid/summer dry/cold arid, Csa: warm temperature/summer dry/hot summer, BSh: arid/steppe/hot arid, Bwh: arid/winter dry/hot arid, Bwk: arid/winter dry/cold arid.

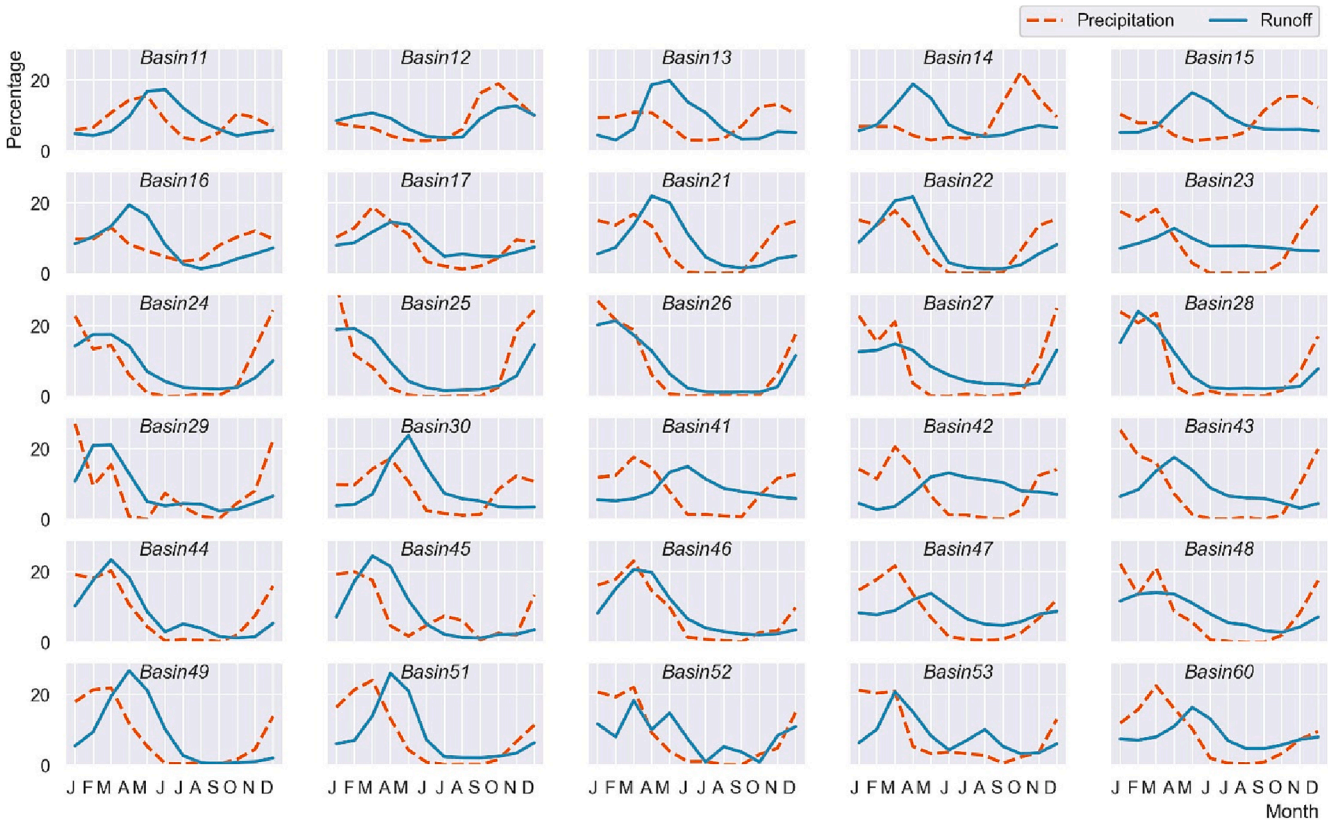



**Fig. 2.** Average observed monthly precipitation and runoff in percentage in 30 basins (The solid blue line and the dashed red line represent runoff and precipitation, respectively). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

information on precipitation, runoff, and temperature containing all 25 ensembles with 1- to 3-month lead time. Then, precipitation, runoff, and temperature were extracted for each cell and 5th, 50th and 95th quantiles were extracted along with the average of ensembles for each cell inside a particular basin (represented by ENS_5%, ENS_50%, ENS_95%, ENS_mean). Finally, for each variable and across all cells in a basin, the 5th, 50th and 95th quantiles and the mean values were calculated for each basin (represented by cell_5%, cell_50%, cell_95%, cell_mean).

### 2.3. Methods

#### 2.3.1. Selection of predictors

Given the large number of potential predictors (i.e., 36 for each lead time), dimension reduction of the input matrix is needed to simplify the model and reduce calculation time. The feature selection algorithms used to select the optimum predictors has some limitations, such as the risk of overfitting when the data size is small, the high computation time when the variables are many, and the trade-off between reducing variance and increasing bias by eliminating some relevant features (Munson & Caruana, 2009). Predictor selection is herein implemented through two methods of highly correlated variable exclusion and elimination by importance, where via a high-correlation filter, highly redundant predictors are identified and removed (Kuhn & Johnson, 2013).

In Pearson correlation analysis (Pca) method, the cross-correlation value is quantified between all potential predictor variable pairs, simultaneously, using Pca. Each predictor pairing with a Pearson's correlation of more than 95% is compared with other predictors, keeping one predictor among those with the highest correlation (Ferreira et al., 2021). Thus, for each predictor pair with a high Pearson correlation, those with the highest correlation with other predictors are dropped from further analyses. This eliminates redundant input variable predictors that are already captured by other variables.

The second method, elimination by importance, is implemented for predictor selection. By eliminating redundant and irrelevant inputs, the predictability and robustness of machine learning methods are increased while reducing the computational costs. A suitable method to identify the most important feature is the Recursive Feature Elimination (RFE) algorithm. In this approach, all possible combinations of predictive variables are used to apply the models, whereby the explanatory power of each predictor is identified through RFE. The algorithm then repeatedly eliminates variables below an importance criterion admitted by the models in each step of searching (Guyon et al., 2002; Kuhn & Johnson, 2013). Overall, RFE is performed on all variables considering 2 to 36 predictors as possible inputs to the simulation models. Selection of the optimal predictor set based on the Leave One Out Cross Validation (LOOCV) method is tested using RMSE as a performance criterion for each set. The ideal predictor set is selected as one with the least RMSE and fewest predictors. This study has used RFE algorithm based on Random Forest (RF) model.

Directed Acyclic Graph (DAG) is used to define the Bayesian Network (BN). It offers the joint probability distribution for a set of random variables and connected nodes that represent these random variables. BN demonstrates the causal relationship or nature between pairs of such variables (Dutta & Maity, 2020). Structure learning is necessary to develop and understand this structure, which comes in three major categories: score-based, constraint-based, and hybrid algorithms. The score-based Hill Climb search algorithm is one of the most common approaches taken to discover network structures and identify the best predictors (Scutari, 2017). This algorithm starts with a saturated graph and compares the score to the maximum score for each potential addition, deletion or reversal before creating a top scoring network for the BN. In this study, the library "CausalNex" (Beaumont et al., 2017) was utilized for creation of BN models and plotting DAG graphs with Bayesian Information Criterion scores being employed to choose the best

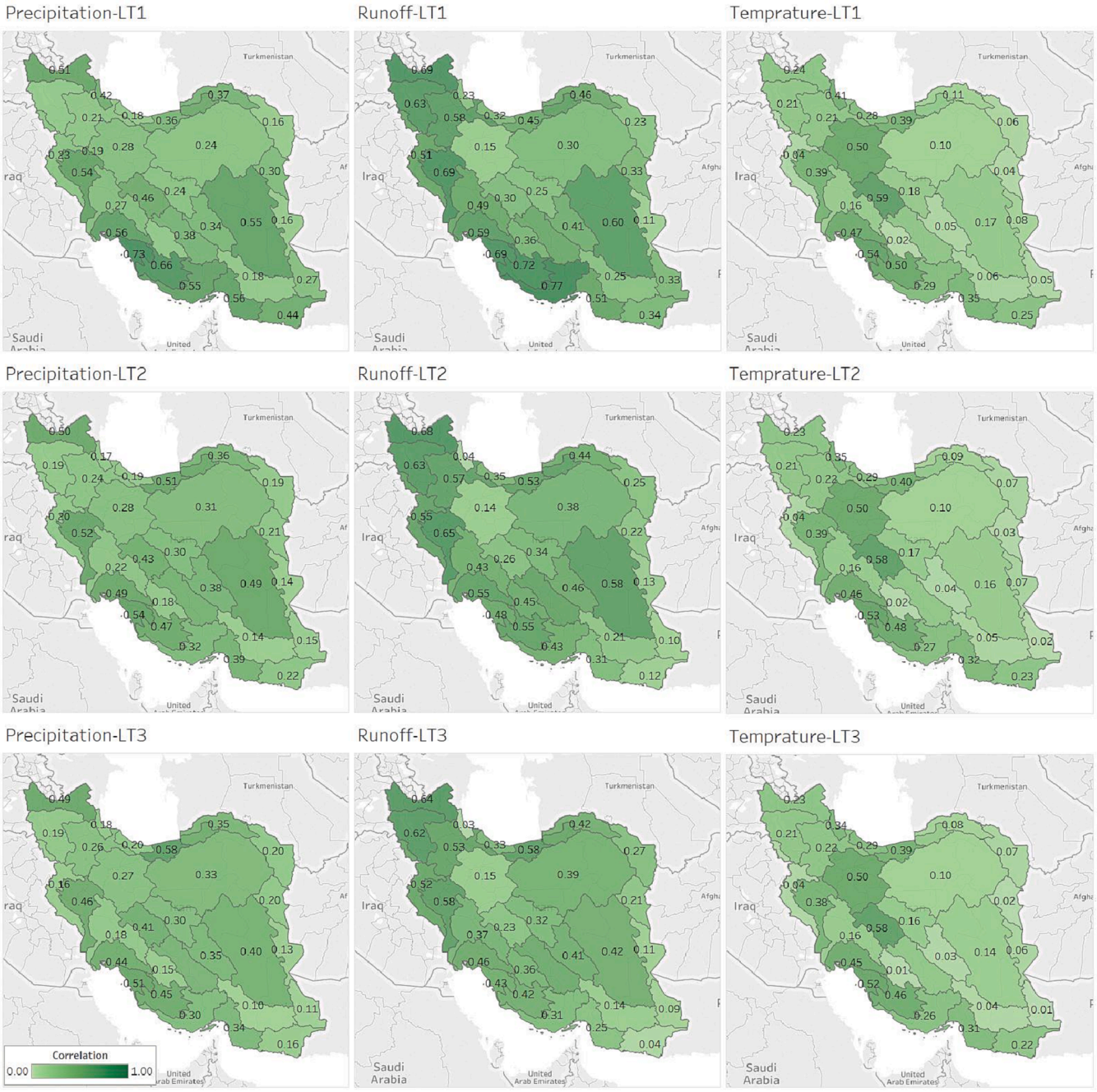



**Fig. 3.** Spatial correlation between observed runoff and ECMWF (C3S) ensemble mean of precipitation, runoff, and temperature for 1-to-3-month lead time.

prediction model (Leu & Bui, 2016). Once established, it can be analyzed in order to determine if variables are dependent or conditionally independent from one another based on the relationship to target variable. This aids in the selection process of inputs for Machine Learning models that comprise potential predictors directly related to targets. In essence, through use of DAG's and BNs, one is able to analyze data sets accordingly; this provides insight into cause and effect between variables while identifying key information that can be relayed back into ML models as best predictors possible under given conditions.

### 2.3.2. Simulation models

Five state-of-the-art ML models were used to develop models for the prediction of monthly streamflow in 30 s-level basins of Iran. Below is a brief description of each model. The selected five ML algorithms (MLR, ANN, SVR, RF, and XGBoost), are widely used for streamflow prediction.

The first technique employed is the Multiple Linear Regression (MLR) model, possibly the most recognizable data-driven forecasting technique. MLR creates a linear relationship between a continuous dependent variable $y$ and one or several independent variables $x_i$. Regression is the most widely used to identify variables $x_i$ with a relationship to the output $y$ (Araghinejad, 2014). For the MLR model implementation, 70% of the available data is used for training and 30% used for testing of the model.

Another simulation approach applied in this research is the Artificial Neural Network (ANN) with a unidirectional multilayer structure consisting of an input layer, multiple hidden layers, and an output layer. Each layer consists of several neurons connected to adjacent layers. Signals entering through the input layer are unidirectionally directed toward the output layer in ANN. The following transformation is applied for each neuron during signal flow in network:

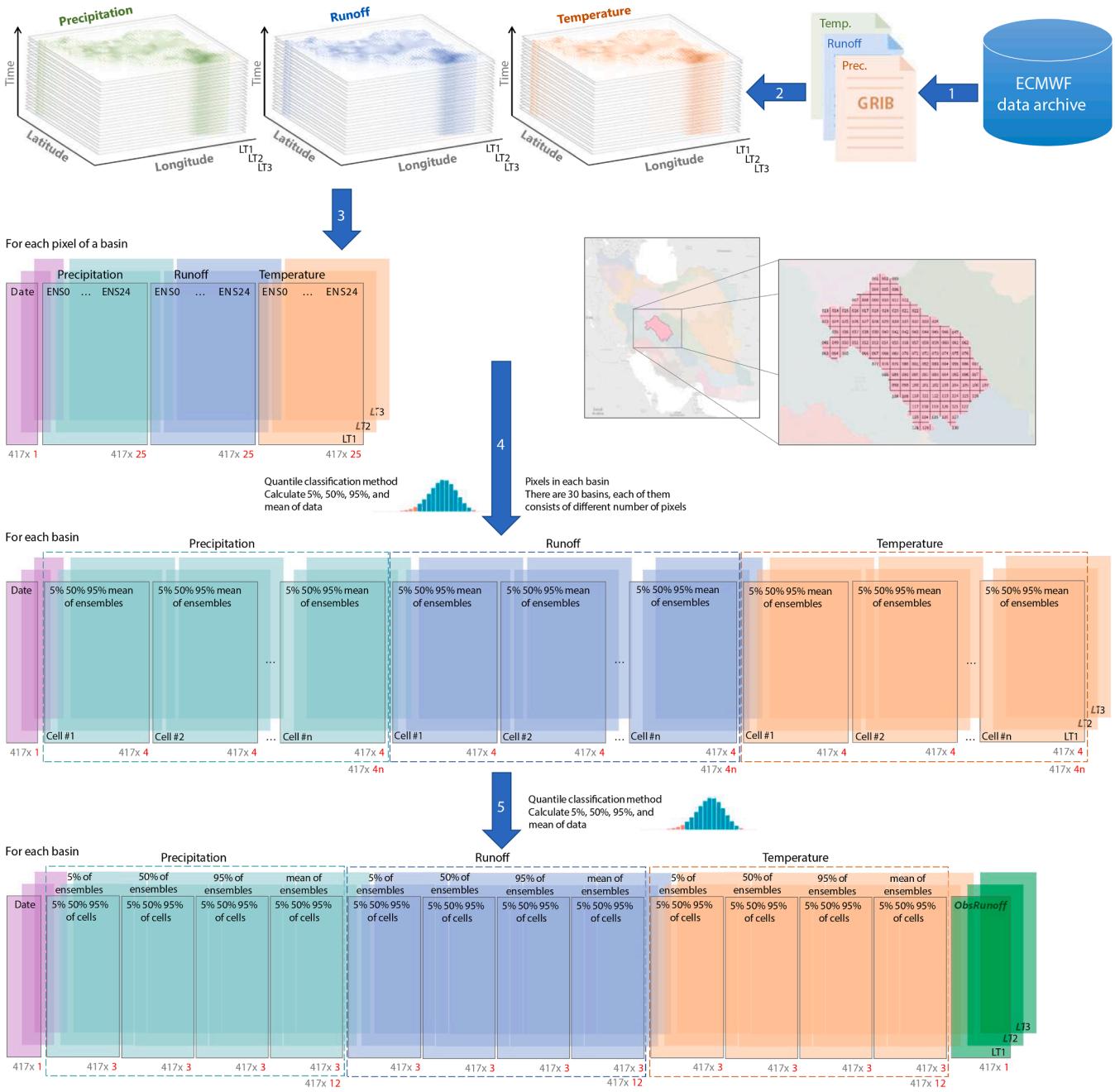



**Fig. 4.** End-to-end raw data pre-processing from data retrieval to runoff simulation.

$$y = f\left(\sum_{i=1}^{n}(x_i w_i + b_i)\right) \tag{1}$$

such that $y$ is the output, $x_i$ is the input, $n$ is the number of inputs, $w_i$ is the weight, $b_i$ is the bias, and f is the activation function of each input used to approximate any mapping between model inputs and outputs. Furthermore, this function normalizes neuron outputs to prevent extreme output values after several layers (Torre et al., 2020). The number of input layer neurons is based on the number of predictors considered for each basin. The number of hidden layer neurons is evaluated by forming networks of 5 to 15 neurons and selecting their optimal number by trial-and-error. The ratio of data used for training, validation and testing are 56%, 14% and 30%, respectively. The minimum network error for simulation cutoff was 0.000001, with maximum number of training failures of 30, and with Levenberg-Marquardt training functions.

Random forest (RF) fundamentally incorporates over-fitting and local convergence issues into multiple classifier forests through a single-classifier decision tree (DT) (Belgiu & Drăgu, 2016). The method used for resampling takes multiple samples from the original dataset, trains a DT for each bootstrap sample, combines these DTs, and averages the predicted values for all the combined DTs. The prediction value of the i-th DT, $y_i$, is evaluated using equation (2) with n representing the number of DTs, $x$ is the inputs and y as the prediction result of the RF model:

$$y = \frac{1}{n}\sum_{i=1}^{n} y_i(x) \tag{2}$$

Another model employed in this study is the Support Vector Regression (SVR). This model was created by generalization of the support vector machine (SVM) capabilities from classification to

regression problems (Noble, 2006). SVR inherits the core concepts of data fitting from SVM and has been widely applied in multiple areas recently. While SVM maximizes the distance to the sample point closest to the hyperplane, the SVR finds a hyperplane that minimizes the distance to the sample point furthest from the hyperplane. By turning the process of finding a hyperplane into a convex quadratic programming problem and solving it, SVR was able to realize nonlinear data modeling and obtain the hyperplane.

The final model employed in this study is a new intelligent algorithm for predicting streamflow based on the gradient boosting model. Recently eXtreme Gradient Boosting (XGBoost) has received praise in academia for its efficient performance, efficacy, and fast speed (Chen & Guestrin, 2016). In this method, a DT is selected by XGBoost as a weak learner. While training a weak learner, the algorithm increases the weight of previously misclassified data marginally, identifies the next weak learner, and adds another weak learner to correct the residuals of all previous weak learners. The result is finally obtained through the weighted sum of learners. By training the XGBoost-based model using the collected data, a strong learner can be used to predict the streamflow values obtained. Rk-Fold cross validation method is used to validate the developed models.

### 2.3.3. Repeated K-Fold cross validation

Once produced based on training data, the model accuracy is assessed using the test dataset. A classic validation method for machine learning methods is cross-validation, here employed by separating the dataset into its training and testing sets. High model accuracy in the training phase does not by default guarantee a similar performance with new data, emphasizing the importance of balancing between generalization and overfitting of the model outputs. Model underfitting implies the model lacks sufficient performance in both the training and testing phases. Most likely, this is because the model is not well-tuned or trained enough on the training set, resulting in high bias and low skills. Model overfitting means the model is too far tuned in the training set. As a result, the model performs well on the training set, but poorly on the test phase, resulting in low bias and high variance (Berrar, 2019; Cawley & Talbot, 2010).

The most important issue with separating data into training and test sets is that test datasets might not follow the distribution of the whole dataset. The K-fold cross validation process, as illustrated in Fig. 5, solves this by sampling all data in K rounds. K is defined as the number of folds and is typically between 3 and 10, but can also be any positive integer. The data is then divided into k equal parts. The algorithm, in k-1 step selects different groupings of folds for testing and separates the remaining folds for the training dataset. With this method, it is possible to train the model k-1 times independently, and measure performance scores k-1 times based on selected criteria. Finally, the average of all scores is evaluated.

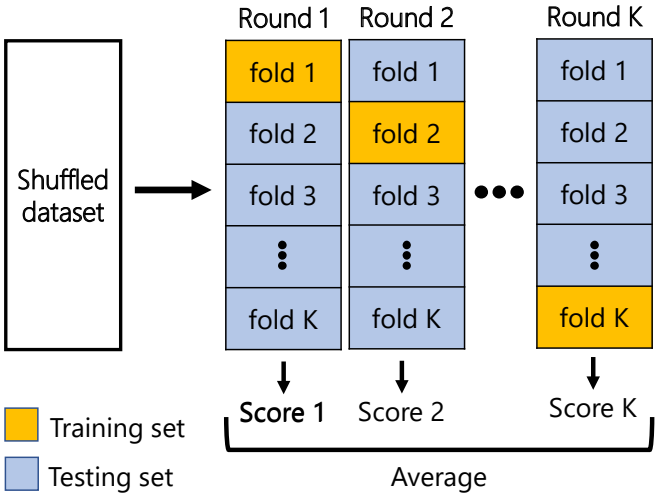



**Fig 5.** Schematic illustration of the K-fold cross-validation process.

Estimating model performance through k-fold CV may result in noisy estimates. This is due to the fact that every time the procedure is performed, a new division of the data enters the k-fold, leading to a different average estimate of the model performance. One way to reduce model performance noise is to increase the number of folds *k*. This reduces bias in performance of model estimates, while also increasing the variance of the outputs. An alternative approach is to repeat the k-fold CV process several times and report the average performance for all rounds. This approach is generally called repeated k-fold CV (Kim, 2009; Molinaro et al., 2005). The important point is that each repetition of k-fold cross-validation has to be performed on the same dataset but with different *k* folds. Repeated k-fold CV has the advantage of improving the average model performance through fitting and evaluation of other models. The process of such a method, similar to what is presented in Fig. 5, repeats multiple times as needed. The common number of repetitions are 3, 5, and 10. For example, if n repetitions with K number of folds are used to estimate model performance, n*K different models have to be fitted and evaluated. This approach is suitable for small to medium sized dataset and models which are not computationally extensive.

### 2.3.4. Evaluation procedure

A combination of criteria including KGE', NSE, and NRMSE are used to evaluate model performance. Modified Kling-Gupta efficiency (KGE') (Gupta et al., 2009; Kling et al., 2012) is a unique evaluation criterion used to express similarities between observed and simulated runoff. The KGE' and its three decomposed components (correlation, bias ratio, and variability ratio) are all dimensionless and defined by:

$$KGE' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{3}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{4}$$

$$\gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{5}$$

where $r$ is the Pearson correlation coefficient between simulation and observations, $\beta$ is the bias ratio of simulated and observed flow, $\gamma$ is the variability ratio between simulation and observed standard deviation, and $\sigma$ is the standard deviation. The important point is that according to KGE', the maximum score of KGE' is the least score of its components. This structure guarantees that the highest KGE' scores show good similarity between simulation and observation discharge. A KGE' = 1, an optimum value, demonstrates perfect agreement of simulations and observations. KGE' score is typically −0.41 for the mean flow (Knoben et al., 2019). In order to fully evaluate model performance and its reliability, Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) and Normalized Root Mean Square Error (NRMSE) (Janssen & Heuberger, 1995) analyses are also performed. These criteria have the advantage of being dimensionless and are used to compare the variety of basins, climates, flow regimes, and flow magnitudes. Equation (6) and equation (7) present NSE and NRMSE formulations, respectively.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{6}$$

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(R_i^p - R_i^o\right)^2}}{\overline{O}} \tag{7}$$

such that $R_i^s$ is the prediction of month $i$, $\overline{R^s}$ is the average of predictions, $R_i^o$ is observation for month $i$, and $\overline{R^o}$ is the average of observations. The relative magnitude of residual variance in comparison to the measured

data variance is provided by NSE. NSE varies from -∞ to 1, with 1 describing a highly performing model. NRMSE is a deterministic metric varying between 0 and ∞, with a perfect score of 0.

## 3. Results and discussion

### 3.1. Selection of optimum predictors

The high collinearity or interdependence of climatic variables often leads to redundant information and possibly deceptive results, where the data it provides is relevant to the analysis but not necessary, given its information is highly similar to that of another predictor. Utilizing these climatic variables requires careful consideration, especially when examining the connections between runoff and the climatic variables (Li et al., 2017). There are frequently connections that are overlooked in systems like hydro-meteorological issues with nonlinear and non-normal co-dependencies. However, by removing collinear/correlated and keeping orthogonal/uncorrelated variables, a correlation-based orthogonalization of datasets will reduce the problem's dimensionality. To remove irrelevant and redundant features and in order to make the ML models more accurate, Pca, a linear FSA and BN and RFE, two non-linear FSAs were employed in this study.

Feature selection was implemented for the 36 potential predictors through highly correlated variable exclusion using concurrent application of Pca, RFE, and BN algorithms. For each basin 36 potential predictors were calculated. Fig. 6 shows the heatmap of the cross correlation between predictors and predictant (runoff) for basin number 24 with 1-month lead time, as an example. The results show that temperature predictors are highly correlated. Amongst all 36 predictors, runoff-based predictors, followed by those of precipitation, are highly correlated with observed runoff.

FSA, is compared with two non-linear algorithms, i.e. RFE and BN. Rows with check marks in Tables 2, 3, and 4 show the features selected by different FSAs and used as input in the modeling stage. Selected predictors are for all 30 s level basins in Iran with 1-month lead time (LT1). As shown in Table 2, Pca optimum predictors demonstrate little dependence on temperature while runoff and precipitation show higher feature importance to the predictant.

The optimum selected predictors from the RFE method are shown in Table 3. It can be seen that except in basins 12, 27, 43, and 53, the most

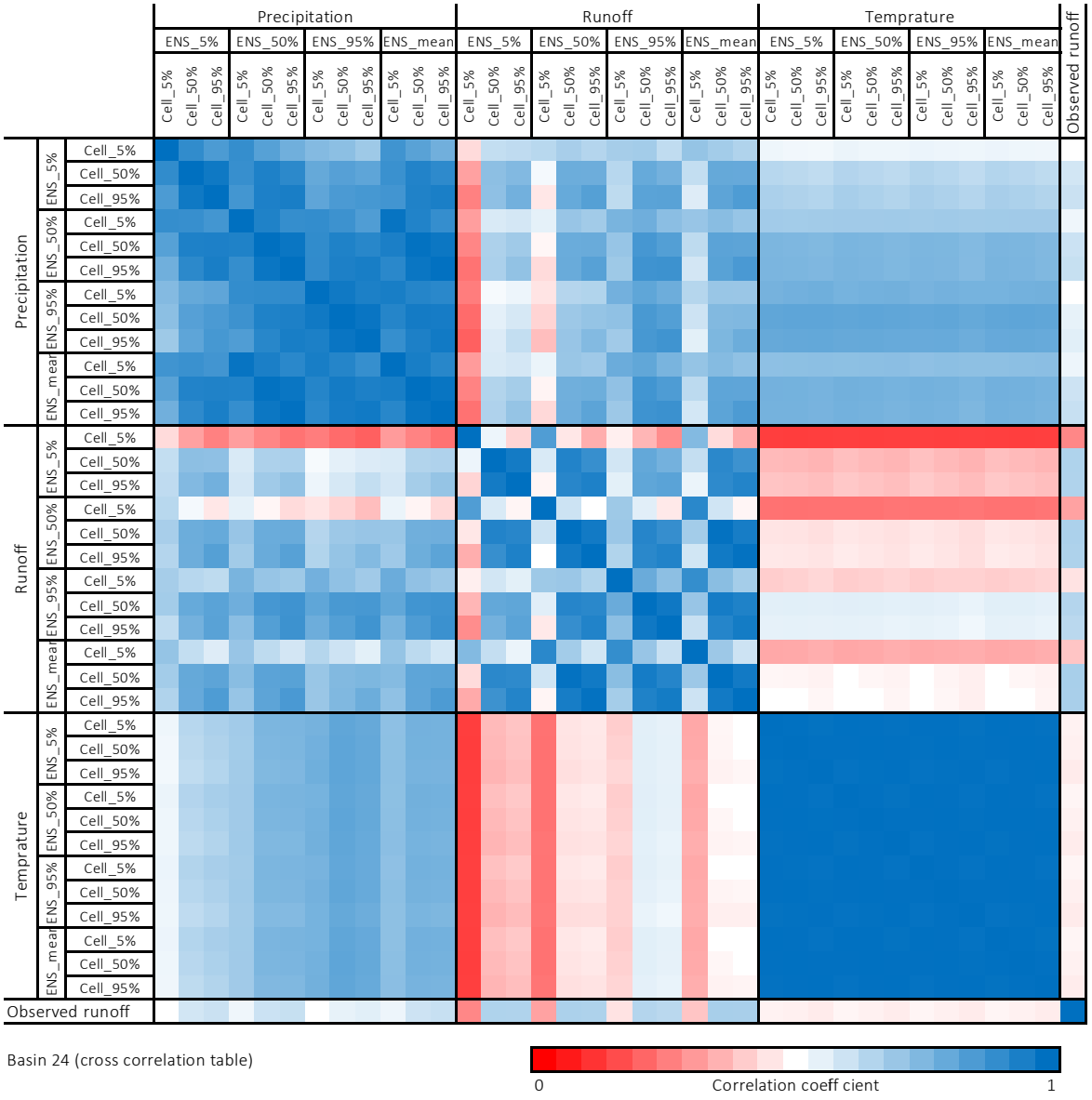



**Fig. 6.** Cross-correlation of potential predictors and the predictant (monthly observed runoff) for Basin #24 (ENS and Cell prefix represents ensemble and basin quantiles, respectively.) In order to create the ML prediction models for monthly streamflow, a total of 36 potential predictors are taken into account. Pca, a linear.

**Table 2**
Optimum predictors selected from Pca for 1-month lead time.

| Basin name | Selected predictors | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | R1 | R2 | R3 | R4 | R5 | R6 | R7 | R8 | R9 | R10 | R11 | R12 | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 | T11 | T12 |
| Basin11 | | | | | | | | | | | | | | ✓ | ✓ | | | | ✓ | | ✓ | | | | | | ✓ | | | | ✓ | | | | | |
| Basin12 | | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | ✓ | ✓ | | | | | ✓ | ✓ | | | | | | | | | | | | | | | | | |
| Basin13 | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | | | | ✓ | ✓ | | | | ✓ | | | | | | | | | | |
| Basin14 | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | ✓ | ✓ | | | | | | | | | | | |
| Basin15 | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | | | | ✓ | | | | | | | | | | | |
| Basin16 | ✓ | | | | | | ✓ | | | | | | ✓ | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | | | | | | | | | | | | | | |
| Basin17 | | ✓ | | | | | | ✓ | | | | | ✓ | ✓ | ✓ | | | | | | ✓ | ✓ | | | | | | | | | | | | | | |
| Basin21 | | ✓ | | | | | | ✓ | | | | | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | | | | | | | | | | | | | | | |
| Basin22 | | ✓ | | | | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | | | | | ✓ | ✓ | ✓ | | | ✓ | | | | | | | | | | | | |
| Basin23 | | | ✓ | | | | | | ✓ | | | | | ✓ | ✓ | | | | | ✓ | | | | | | | | | | | | | | | | |
| Basin24 | | | | | | | | | | | | | | | ✓ | | | | | | | | ✓ | | | | | | | | | | | | | |
| Basin25 | ✓ | ✓ | | | | | | ✓ | | | | | | | | ✓ | ✓ | | | ✓ | | | | | | | | | | | | | | | | |
| Basin26 | ✓ | ✓ | ✓ | | ✓ | | ✓ | | | | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | | | | | ✓ | | | | | | | | | | |
| Basin27 | ✓ | | ✓ | | ✓ | | | | | | | | ✓ | | | | | ✓ | | | ✓ | | | | | | | | | | | | | | | |
| Basin28 | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | | | | | | | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | | | | | | | | | | | | | |
| Basin29 | ✓ | | | | | | | | | ✓ | | | | | | | | | ✓ | | | | | | | | | | | | | | | | | |
| Basin30 | ✓ | ✓ | ✓ | ✓ | | | | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | ✓ | | | | | | | | | | | ✓ | | | |
| Basin41 | ✓ | ✓ | | ✓ | | | | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | | | | | | | | | ✓ | | | | | | |
| Basin42 | ✓ | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | | ✓ | ✓ | | | | | | | | | ✓ | | | | | | | |
| Basin43 | | | | | | | | | | | | | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | | ✓ | | | | | | | | | | | | | |
| Basin44 | | | | | | | | | | | | | | ✓ | | | | | | | | ✓ | ✓ | | | | | | | | | | | | | |
| Basin45 | | | ✓ | | | | | | | | | | | ✓ | | | ✓ | ✓ | | ✓ | ✓ | | | | | | | | | | | | | | | |
| Basin46 | | | | | | | | | | | | | | ✓ | | | | ✓ | | | ✓ | | | | | | | | | | | | | | | |
| Basin47 | ✓ | ✓ | ✓ | | | | | | ✓ | | | | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | | | | | | | | | | ✓ | | | |
| Basin48 | ✓ | | ✓ | | | | | | | | | | ✓ | ✓ | ✓ | | | | | | | ✓ | | | | | ✓ | | | | | | | | | |
| Basin49 | ✓ | | | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | | | | | | | | | | | ✓ | | | |
| Basin51 | | | | | | | | | | | | | ✓ | ✓ | | | | | ✓ | | | | | | | | | | | | | | | | | |
| Basin52 | | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | | | | ✓ | | | | | ✓ | ✓ | | | | | | | | | | | | | | | |
| Basin53 | | | | ✓ | | | | | | | | | ✓ | | | | | ✓ | | | | | | | | | | | | | | | | | | |
| Basin60 | ✓ | | ✓ | | | | | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | | | | | | | | | | ✓ | | | |

**Table 3**
Optimum predictors selected from RFE for 1-month lead time.

| Basin name | Selected predictors | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | R1 | R2 | R3 | R4 | R5 | R6 | R7 | R8 | R9 | R10 | R11 | R12 | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 | T11 | T12 |
| Basin11 | | | | | | | | | | | | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | | | | | | | | | | |
| Basin12 | | | ✓ | | | ✓ | | | | | | ✓ | | | | | | ✓ | | | | | | | | | ✓ | | | ✓ | | | ✓ | | ✓ | ✓ |
| Basin13 | | | | | | | | | | | | | ✓ | ✓ | ✓ | | ✓ | | | ✓ | | ✓ | ✓ | ✓ | | ✓ | | | | | | | | | | |
| Basin14 | | | | | | | | | | | | | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | | | | | | | | | |
| Basin15 | | | | | | | | | ✓ | ✓ | | | | | ✓ | | ✓ | | | ✓ | | | ✓ | | ✓ | ✓ | ✓ | | | | | | | | | |
| Basin16 | | | | | | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | | | | | | | | | | | | | |
| Basin17 | | | | | | | | | | | | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | | | | | | | | | | |
| Basin21 | | | | | | | | | | | | | ✓ | ✓ | ✓ | | | ✓ | | | | | | ✓ | ✓ | ✓ | ✓ | | | | | | | | | ✓ |
| Basin22 | | ✓ | | | ✓ | | | | | | | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | | | | | | | | | | | |
| Basin23 | | | | | | | | | | | | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | | | | | | | | | ✓ | | |
| Basin24 | | | | | | | | | | | | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | | | | | | | ✓ | ✓ | | | |
| Basin25 | ✓ | ✓ | | ✓ | ✓ | | | | | | ✓ | | | | | ✓ | ✓ | | | ✓ | | | ✓ | | | | | | | | | | | | | |
| Basin26 | ✓ | | | | | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | ✓ | ✓ | | | ✓ | | | | | | | | | |
| Basin27 | | | | | | | | | | | | | | | | | | ✓ | | | | | | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ |
| Basin28 | ✓ | ✓ | | | ✓ | | | | | | ✓ | | | ✓ | | | ✓ | ✓ | | ✓ | | | ✓ | | | | | | | | | | | | | |
| Basin29 | | ✓ | | ✓ | ✓ | | ✓ | | | ✓ | | | | | | | | ✓ | ✓ | | ✓ | | | ✓ | | | | | | | | | | | | |
| Basin30 | | | | | | | | | | | | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | | | | | | | | | | | |
| Basin41 | | | ✓ | | | | | | | | | | | ✓ | ✓ | | | ✓ | | | ✓ | | | | | | | ✓ | | ✓ | | | ✓ | | | ✓ |
| Basin42 | | | | | | | | | | | | | | ✓ | ✓ | | | | | | | | | | ✓ | ✓ | ✓ | | | ✓ | | ✓ | ✓ | | | ✓ |
| Basin43 | | | ✓ | | | | | | | | | | | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | | | | | | | | | | | | |
| Basin44 | | ✓ | | | | | | | | | | | | ✓ | | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | | | | | | | |
| Basin45 | | | ✓ | | | ✓ | | | ✓ | | | ✓ | | ✓ | ✓ | | | ✓ | | | ✓ | | | ✓ | | | | | | | | | | | | |
| Basin46 | | | | | | | | | | | | | | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | | | | | | | |
| Basin47 | | | | | | ✓ | | | | | | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | | | | | | | | | | | |
| Basin48 | | | ✓ | | | | | | ✓ | | | | | ✓ | ✓ | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | | | | | | | | | | | | |
| Basin49 | ✓ | ✓ | ✓ | ✓ | | | | | | | | | | ✓ | ✓ | | | ✓ | | | ✓ | | | ✓ | | | | | | | | | | | | |
| Basin51 | | | | | | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | | ✓ | | | ✓ | | ✓ | | | ✓ | | | | | | | | | | | | |
| Basin52 | | | | | | | ✓ | | | ✓ | | | | ✓ | ✓ | | | | ✓ | | | ✓ | ✓ | ✓ | | | | ✓ | | | | | | | | |
| Basin53 | | | ✓ | ✓ | | | | | | ✓ | | | | | | | | | | ✓ | | | | ✓ | | ✓ | | | ✓ | | ✓ | | | | ✓ | | |
| Basin60 | | | | | | | | ✓ | ✓ | | | | | | ✓ | | | ✓ | | | ✓ | | | ✓ | ✓ | | ✓ | | | ✓ | | | | | | |

**Table 4**
Optimum predictors selected from BN for 1-month lead time.

| Basin name | Selected predictors | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | R1 | R2 | R3 | R4 | R5 | R6 | R7 | R8 | R9 | R10 | R11 | R12 | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 | T11 | T12 |
| Basin11 | | | | | | | | | | | | | ✓ | | ✓ | | | | | | | ✓ | | ✓ | | | | | | | | | ✓ | | | |
| Basin12 | | | | | | | | | | | | ✓ | | | | ✓ | | ✓ | | ✓ | | | ✓ | | ✓ | ✓ | ✓ | | | | | | | | ✓ | |
| Basin13 | | | | | | | | | | | ✓ | | | | | ✓ | | ✓ | | ✓ | | | ✓ | | ✓ | ✓ | ✓ | | | | | ✓ | | ✓ | | |
| Basin14 | | | | | | | | | | | | | | | | ✓ | | | | | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ |
| Basin15 | | | | ✓ | | | ✓ | | | | | | | | | | | | | | ✓ | | | | ✓ | | | | | | | ✓ | | | | |
| Basin16 | | ✓ | | | | | ✓ | ✓ | | ✓ | | | | | | | ✓ | | | | | | ✓ | | | ✓ | | | | | | ✓ | ✓ | | ✓ | | |
| Basin17 | | ✓ | | | | | | | | | | | | | | | | | | | | ✓ | | | ✓ | | | | | | | | | | | |
| Basin21 | | | | | ✓ | | | | | | ✓ | | ✓ | ✓ | | | ✓ | | | | | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | |
| Basin22 | | ✓ | | | | | | ✓ | | | ✓ | | | ✓ | | | | ✓ | | | ✓ | | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Basin23 | | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Basin24 | | | | | | | | | | | | ✓ | ✓ | ✓ | | ✓ | | | ✓ | | | ✓ | | | | | | | | | | ✓ | | | | |
| Basin25 | | ✓ | | | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | | ✓ | | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | ✓ | | | | |
| Basin26 | | | | ✓ | | ✓ | | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | | | | | | ✓ | ✓ | ✓ | | | | | ✓ | | ✓ | | ✓ | | | |
| Basin27 | | | | | | | | | | | | | | | | | | | ✓ | | | | | | | | | ✓ | | | | | | | | |
| Basin28 | | | | | | | | | | | | | | | | | ✓ | | | | | | | | | ✓ | | | | ✓ | ✓ | | | | | |
| Basin29 | | | | | | | ✓ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Basin30 | ✓ | | | | | | | | | | ✓ | | | | | ✓ | ✓ | | | ✓ | | | | | ✓ | ✓ | ✓ | | | | | | ✓ | | | |
| Basin41 | | | | | | | | ✓ | | | ✓ | | | | ✓ | ✓ | | | | | | | | | ✓ | | ✓ | | ✓ | | | | ✓ | ✓ | ✓ | ✓ |
| Basin42 | | ✓ | | | ✓ | | | | | | ✓ | ✓ | | | ✓ | | ✓ | | | | | | | | | ✓ | | | | | | ✓ | ✓ | | | |
| Basin43 | | | ✓ | | | | | | | | | ✓ | | | ✓ | | ✓ | | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | ✓ |
| Basin44 | | | ✓ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Basin45 | | | | | | | | | | | | ✓ | | | | | | | | | | | ✓ | | | ✓ | ✓ | | ✓ | | | | | | | ✓ |
| Basin46 | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Basin47 | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | | ✓ | | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ |
| Basin48 | ✓ | ✓ | ✓ | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | | | ✓ | | | | | | ✓ | ✓ | | | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ |
| Basin49 | ✓ | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | | | ✓ | ✓ | ✓ |
| Basin51 | | | | | | ✓ | | | | | | ✓ | ✓ | ✓ | | ✓ | | | | | | ✓ | | | ✓ | | | | | | | | | | | |
| Basin52 | | ✓ | | | | | ✓ | | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | | | ✓ | | | | | | | | | | | |
| Basin53 | | | | | | | | | | | | | | ✓ | ✓ | ✓ | | | | | | ✓ | ✓ | ✓ | | | | | | | | | ✓ | | | |
| Basin60 | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |

frequent predictors selected by RFE algorithm are from the pool of runoff variables. Unlike Pca, neither temperature nor precipitation played a significant role in outputs of RFE algorithm in all 30 basins.

For each basin, a more specialized subset of the potential predictors is selected using BN as an FSA to identify optimum predictors for machine learning models. In order to get a clear picture of optimum predictors, a conditional independent structure for the target (streamflow) and potential predictors is established via BN. For each major first level basin in Iran, a sample basin's conditional independence structure is represented by DAG in Fig. 7 for 1-month lead time. The highest dependent potential predictors are those that exhibit direct edges (between parent and child) with the target and other predictors. This feature selection depends on climate, lead-time, data availability, and the causal relationship between predictors and the target variable (Noorbeh et al., 2020). For the streamflow of Basin17, for instance, direct edges with three potential predictors are illustrated, where out of 36 potential predictors, 3 optimum predictors are connected (P2, R10, and T1). Additionally, it shows that while each potential predictor can be thought of as a candidate predictor, some of the information they provide is redundant given the information provided by others, leading to a network model with only a few optimum predictors (node connections). Having fewer predictors can reduce overfitting that generally leads to better performance on new data (Das et al., 2022).

Table 4 shows the optimum predictors selected from BN algorithm for 1-month lead time. It is worth noting that while in some basins nearly all predictors are selected for modeling, there exist also basins where only a very few predictors are selected. Unlike in Pca and RFE algorithms, runoff predictors show no significance in their selection compared to precipitation and temperature using BN method.

The set of variables (potential predictors) with strong relationships to the target streamflow were taken into account for each of the different lead times (1–3 months) in order to create parsimonious models to predict streamflow. All the five ML models are developed using the selected predictors that are chosen by the three FSAs, i.e., Pca, RFE, and BN in each lead time for all the basins. The best FSA and model combinations are chosen according to the KGE' criteria in each basin to be used as the final set of ML models. Given the large number of ML models for all basins (5 ML models, 3 FSA, 3 lead times, and 30 basins), only ANN model results are shown because of their better performance. Fig. 8 shows the KGE' evaluation criteria for ANN models with 1 to 3-month lead time and for all basins created based on (a) BN, (b) RFE, and (c) Pca feature selection algorithm in training phase. For all lead times, (d) is the best FSA of each basin with the highest KGE' value. Table 5 shows the frequency of BN, RFE, and Pca algorithms selected for all five ML models totaling to 30 s level basins in Iran. Reflecting Fig. 8 for ANN, each of the BN, RFE, and Pca algorithms is respectively selected in 12, 16, and 2 basins for 1-month lead time, in 14, 15, and 1 basins for 2-month lead time, and in 11, 16, and 3 basins for 3-month lead time.

Table 5 shows that based on XGBoost, ANN, and RF models, RFE and BN feature selection methods were used most frequently, while Pca was used least frequently in all lead times. In contrast, for SVR model, BN was the least frequently selected predictor while for MLR, feature selection methods do not follow any apparent pattern across all lead times.

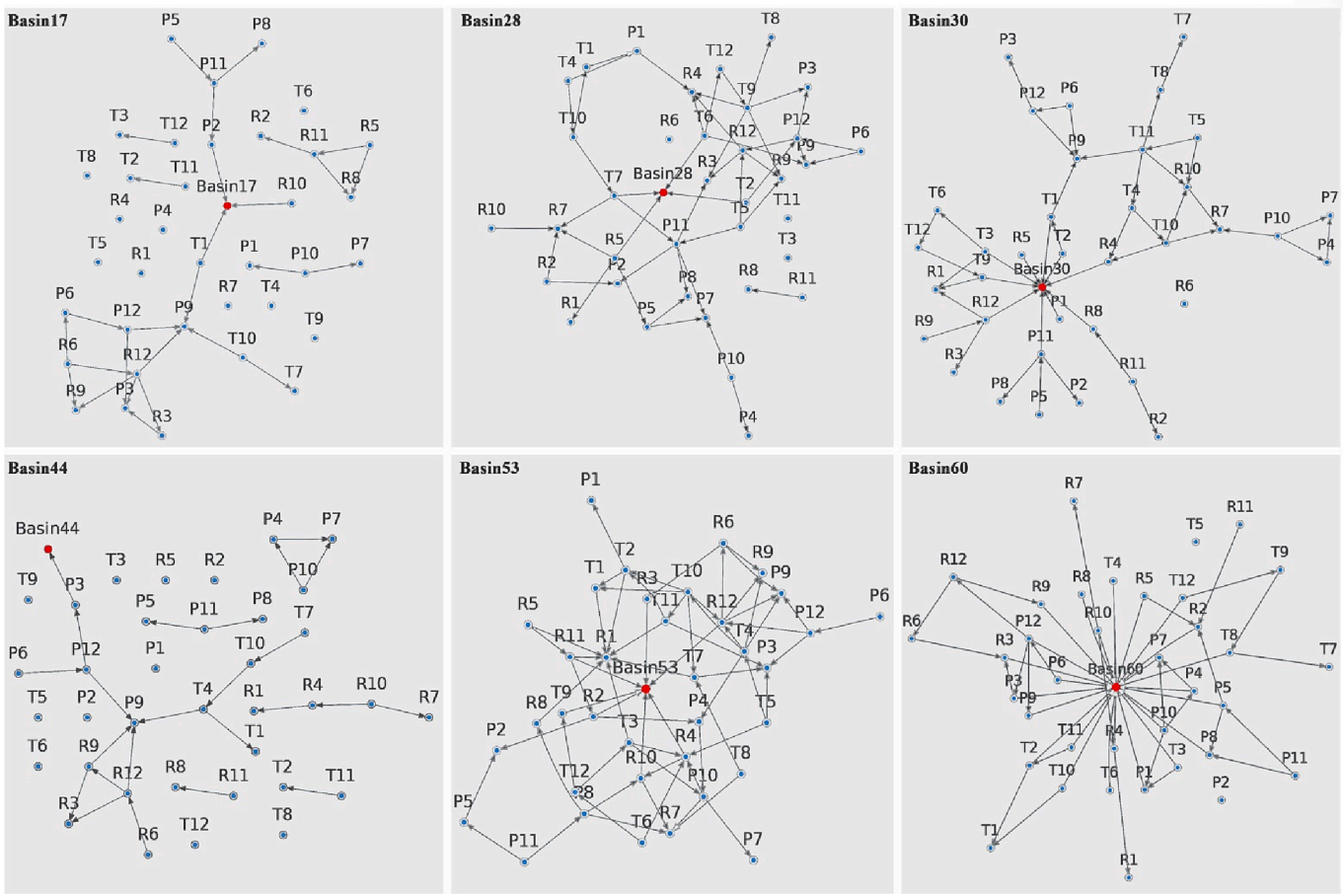



**Fig. 7.** BN represented by DAG developed for target and potential predictors for a lead time of 1 month. This structure, also referred to as the conditional dependence (independence) structure, visualizes the existing interplay between potential predictors. The potential predictors having direct edges with the streamflow variable are selected as the optimum predictors. for a representative basin from each major basin in Iran.

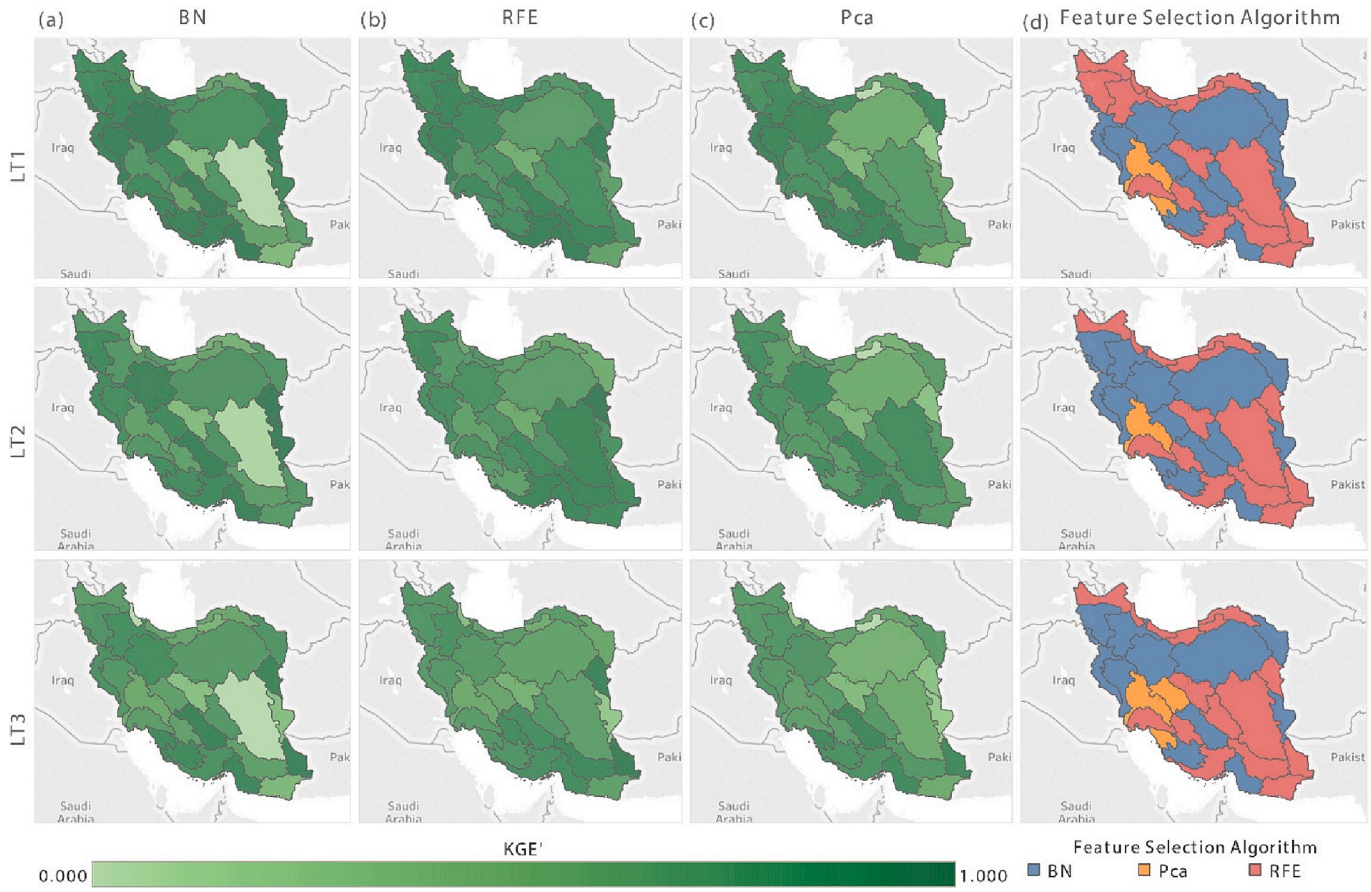



**Fig. 8.** Values of KGE' evaluation criteria for ANN models for 1 to 3-month lead time of all basins created based on (a) BN, (b) RFE, and (c) Pca feature selection algorithms in training phase. (d) is the best FSA of each basin with the highest KGE' values of the ANN models for all lead times.

**Table 5**
frequency of BN, RFE, and Pca algorithms selected for all five ML models across totaling to 30 s level basins in Iran.

| Lead time (month) | FSAs | Models | | | | |
|---|---|---|---|---|---|---|
| | | MLR | SVR | ANN | RF | XGBoost |
| | RFE | 3 | 11 | 16 | 15 | 11 |
| 1 | BN | 13 | 6 | 12 | 12 | 17 |
| | Pca | 14 | 13 | 2 | 3 | 2 |
| | RFE | 10 | 12 | 15 | 17 | 12 |
| 2 | BN | 14 | 8 | 14 | 8 | 14 |
| | Pca | 6 | 10 | 1 | 5 | 4 |
| | RFE | 9 | 13 | 16 | 15 | 10 |
| 3 | BN | 13 | 7 | 11 | 10 | 16 |
| | Pca | 8 | 10 | 3 | 5 | 4 |

### 3.2. Comparison of several ML prediction models

For the purpose of creating concise models for predicting streamflow, the set of variables (optimal predictors) with strong connections to the target streamflow were employed for each lead time (1–3 months). Monthly streamflow is simulated using MLR, SVR, ANN, RF, and XGBoost models in all 30 basins in Iran and compared with the average observed runoff over the period of 1981 to 2015. In order to improve prediction accuracy, tuning of the ML models' hyperparameters was performed. The Grid Search (GS) algorithm was used to optimize the hyperparameters of XGBoost, RF, and SVR models. In each model, a set of values for each hyperparameter were specified for the algorithm. 1000 models were created by the GS algorithm using different combinations of hyperparameters. Fig. 9 compares the simulated hydrograph with observed streamflow for one representative basin within each major first level basin with 1-month lead time. All models show acceptable results in simulating stream flow time series of the displayed basins.

Fig. 10 shows model evaluation results based on KGE', NSE and NRMSE criteria for the 30 studied basins with 1-month lead time. The color shading demonstrates model efficiency: the darkest color shows highest model efficiency and the lightest shows the least efficiency for all evaluation criteria. KGE' criterion shows ANN and XGBoosthave the highest prediction performance in all basins, followed closely by RF. It can be seen that almost all models performed poorly in subbasins of major first level basin 5, which is subject to arid climate located on the eastern border of Iran. NSE criterion shows all models have performed well in the training stage in almost all basins. It also shows that similar to the KGE' criterion, the best performance in the testing stage is delivered by ANN, XGBoost, and RF. Occurrence of negative NSE means that the average of observed values are more reliable than the simulated model predictions (Ferreira et al., 2021). NSE of less than 0.5 implies weak model performance (Moriasi et al., 2015). Accordingly, the low performance of MLR and SVR models, in particular in major first level basins 4, 5, and 6, is obvious. Given runoff variations across all basins, the dimensionless NRMSE criterion is used instead of RMSE for better evaluation of various models. Again, all models show acceptable performance in all basins except for major basin 5. Overall, the monthly results from all five models indicate better predictions in semi-humid and humid basins enjoying high flows as compared to the semi-arid and arid basins with low flow in central, eastern and southern regions; similar results were reported by Slater et al. (2017).

In order to better understand predictions of MLR, SVR, ANN, RF, and XGBoost models, monthly streamflow predictions are shown against their observed values during the study period in different basins

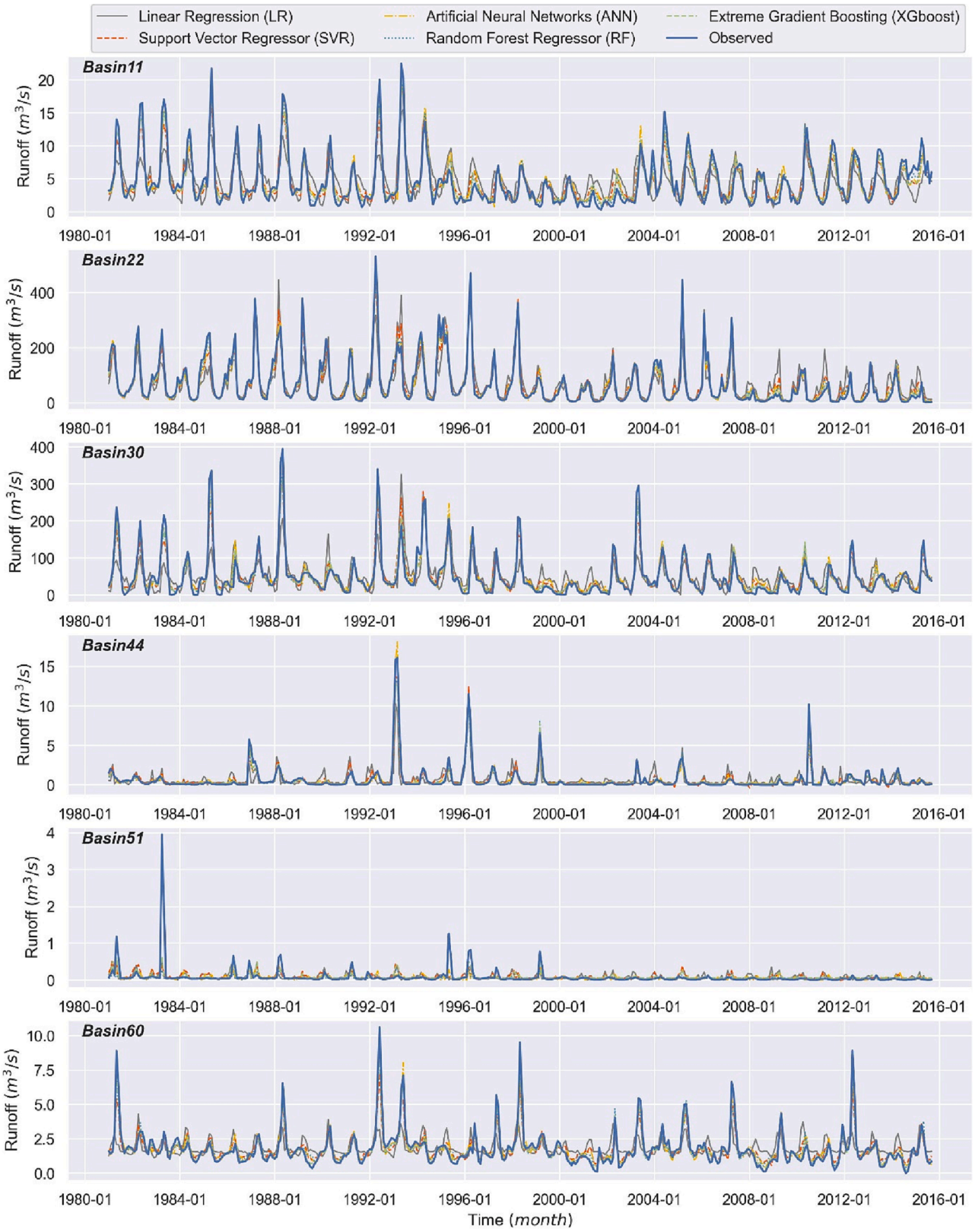



**Fig. 9.** Comparison of observed and simulated flows using XGBoost, RF, ANN, SVR, and MLR models for a sample basin within each major first level basin in Iran.

(Fig. 11). It can be seen that ANN, RF, and XGBoost models have higher prediction performance compared to SVR and MLR models, in particular in basins of high flow, similar to the results reported by Ferreira et al. (2021). Based on these scatter plots (Fig. 11), again it is confirmed that the highest bias occurs in major first level basin 5 which has one of the lowest streamflow among all six major basins. SVR and MLR models demonstrate a slightly larger spread of values especially in the area of low flows, the same conclusion reported by Szczepanek (2022). Lastly, the performance of ANN compared to other ML models in reducing the bias of data is evident.

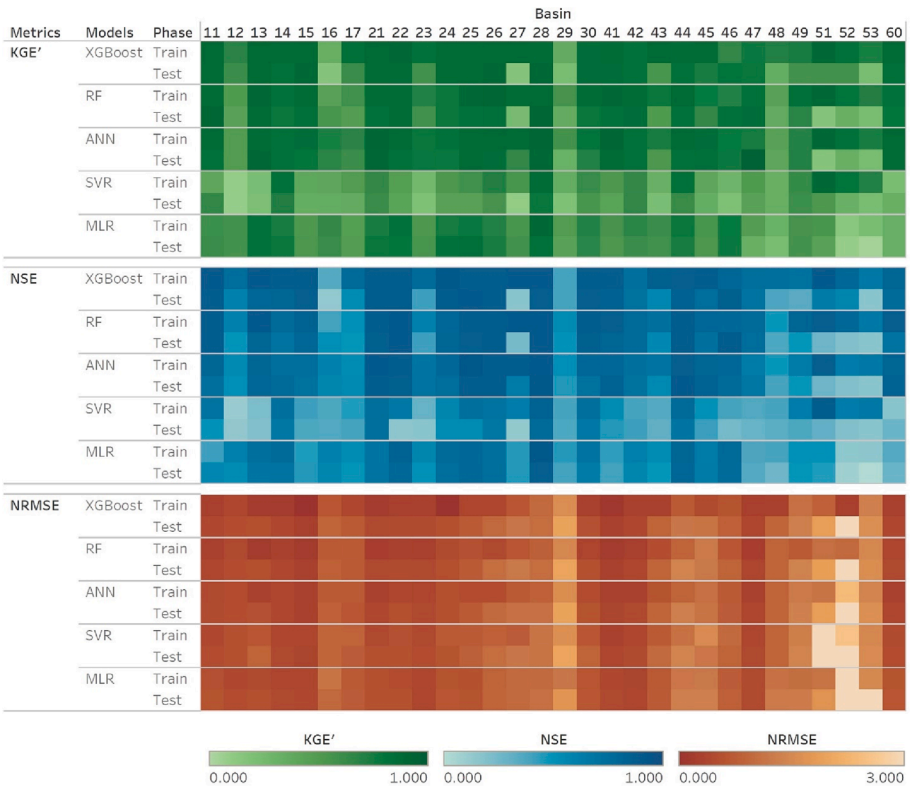



**Fig. 10.** Performance of all five evaluated ML models based on KGE', NSE, and NRMSE criteria during the training and testing phases for 30 basins with 1-month lead time.

Fig. 12 shows the spatial distribution of performance of all ML models based on KGE', NSE, and NRMSE criteria for training phase over all 30 basins in Iran for 1- to 3-month lead times. LT1, LT2, and LT3, represent 1-, 2-, and 3-month lead times, respectively. Similar to what was shown in Fig. 10, XGBoost, RF, and ANN offer better performance for 1-month lead time. All models show nearly the same performance with increasing lead time over all regions.

The spatial distribution of performance of all ML models are shown in Fig. 13 based on KGE', NSE, and NRMSE criteria for test phase over all 30 basins in Iran for 1- to 3-month lead times. It was found that XGBoost, RF, and ANN offer better performance for 1-month lead time predictions. All models show decrease in prediction performance with increasing lead time in all regions. This is similar to findings of Nobakht et al. (2021) and Wang et al. (2019). For longer lead times, ANN, XGBoost, and RFshow the best prediction performance on all evaluation criteria compared to other models. For 2- to 3-months lead times, ANN, XGBoost, and RFperform best in the western and northern basins for the KGE' criteria. None of the models delivered suitable prediction performance for central and southeastern basins with higher than 1-month lead time.

Next, the KGE', NSE, and NRMSE evaluation criteria were averaged over the six major first level basins of Iran, creating a single value for each major basin. Table 6 shows such averaged values with 1-, 2-, and 3-month lead times (LT1-LT3). For all MLmodels, the final 30% of the data was used for model testing, corresponding to the 2006–2015 period.

All models demonstrate good performance in runoff prediction in high streamflow basins, in particular in major basin 2 with the highest streamflow in Iran. However, in central arid regions with the largest land area and lowest precipitation (major basin 4), MLR and SVR models offer very poor prediction performance. While XGBoost performs better in the arid regions of eastern Iran, e.g. major basin 5, all models provide weaker streamflow predictions in these regions.

High rainfall variability is a characteristic of arid regions, and NWP models frequently overestimate rainfall amounts over these regions (Robertson et al., 2013). The ECMWF forecasting system will naturally be less vulnerable to more severe aridity conditions due to the stronger physical model structure that makes it less dependent on prior recurrence in moisture content and flows to foresee upcoming dynamics. On the other hand, with few observations, it might be challenging for NWP models to replicate the intricate meteorological processes that cause the high rainfall variability (Hapuarachchi et al., 2022). Empirical models built on observations and measurements will be limited in hydrometeorological forecasting in arid regions, mostly because the automated learning influencing the model behavior will not have mastered the range of dynamic behaviors underlying such climatic conditions. Therefore, it will continue to be challenging to increase forecasting capacity in arid basins, similarly reported by Nifa et al. (2023).

According to model evaluation criteria presented in Table 6 for 1-month lead time, ANN, XGBoost, and RF show better performance and achieved KGE' criteria values of 0.70–0.87, 0.68–0.86, and 0.66–0.80, respectively, while yielding NSE criteria values of 0.66–0.82, 0.66–0.86, and 0.65–0.86 in the prediction phase; except for major basin 5, an arid area with very.

low flows. As also observed in the scatterplots of Fig. 11, predictions over low flows regions such as the three subbasins in major basin 5 are underestimated by all analyzed models. The KGE' evaluation criteria values for ANN, XGBoost, and RF are 0.24, 0.24, and 0.46 and the NSE values are 0.21, 0.21, and 0.46, showing better performance of the XGBoost model in such a low flow simulation, while the greatest deviations from the observations can be noticed with the MLR model. Prediction models for low flow regions are generally found to exhibit low performance, similarly noted by Szczepanek (2022). Moreover, also in basins with high flow regime, high flow values are typically underestimated by all models, as confirmed by Szczepanek (2022). Overall, the models are compared and ranked based on the KGE' criteria from Table 6, with the order of ANN, XGBoost, RF, MLR, and SVR and mean KGE' values of 0.70, 0.68, 0.66, 0.57, and 0.41, respectively, in the prediction phase.

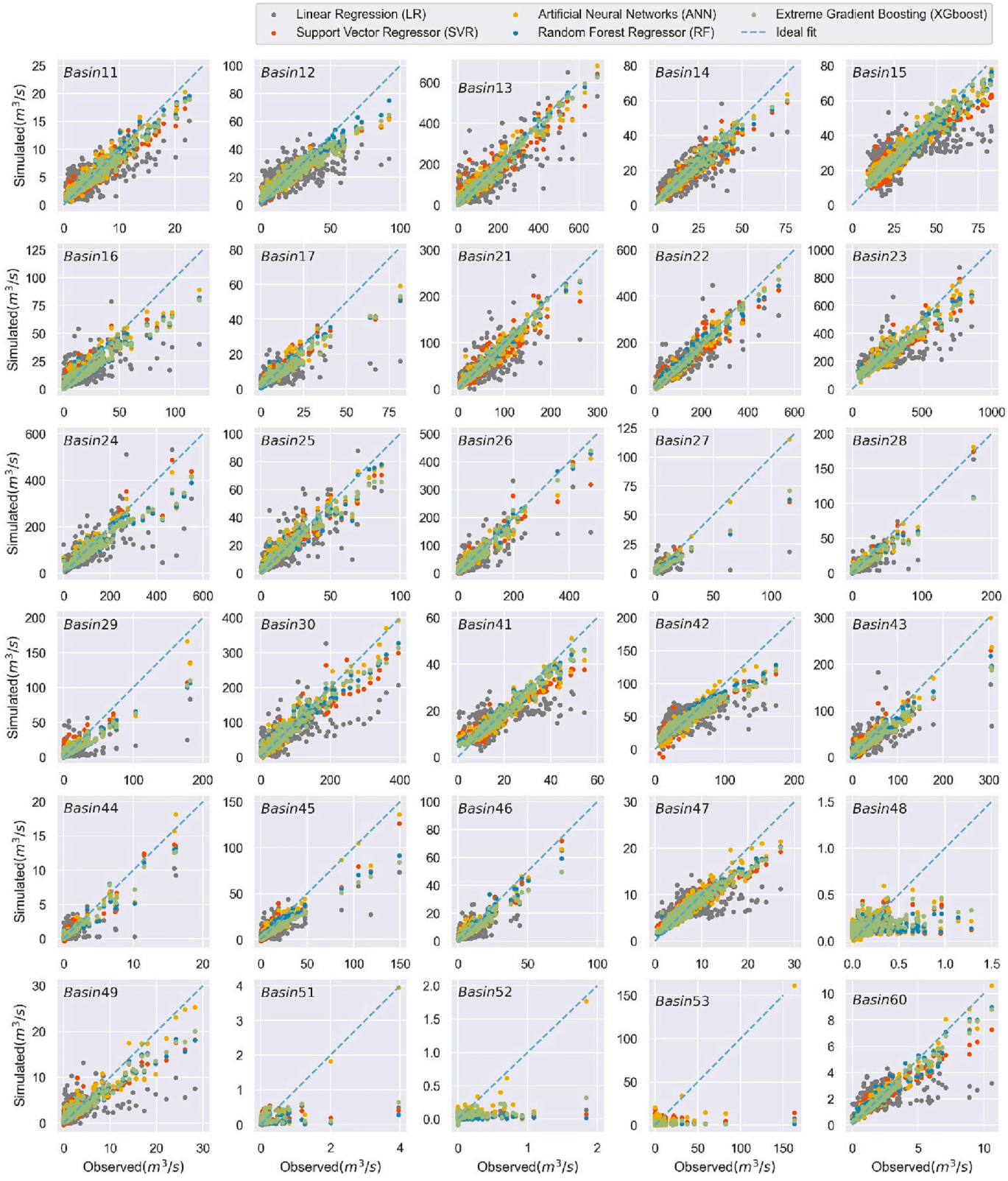



**Fig. 11.** Model simulated monthly streamflow against the corresponding observed values during the study period in all 30 basins.

The predictive ability of all the models decreases for 2 and 3-months lead time, compared to 1-month lead time forecasts. ANN and XGBoost outperformed other models in case of 2-months lead time with an average KGE' value of 0.65 for all basins, while resulting in an average KGE' value of 0.60 for 3-months lead time.

Overall, the best results are generated by ANN and XGBoost models and the worst by SVR and MLR in all major basins. In the prediction phase, ANN shows the best performance in major basins 1, 2, 4, and 6, while XGBoost shows the same in major basins 3 and 5. In ML algorithms, selection of the appropriate hyperparameters by GS method is a

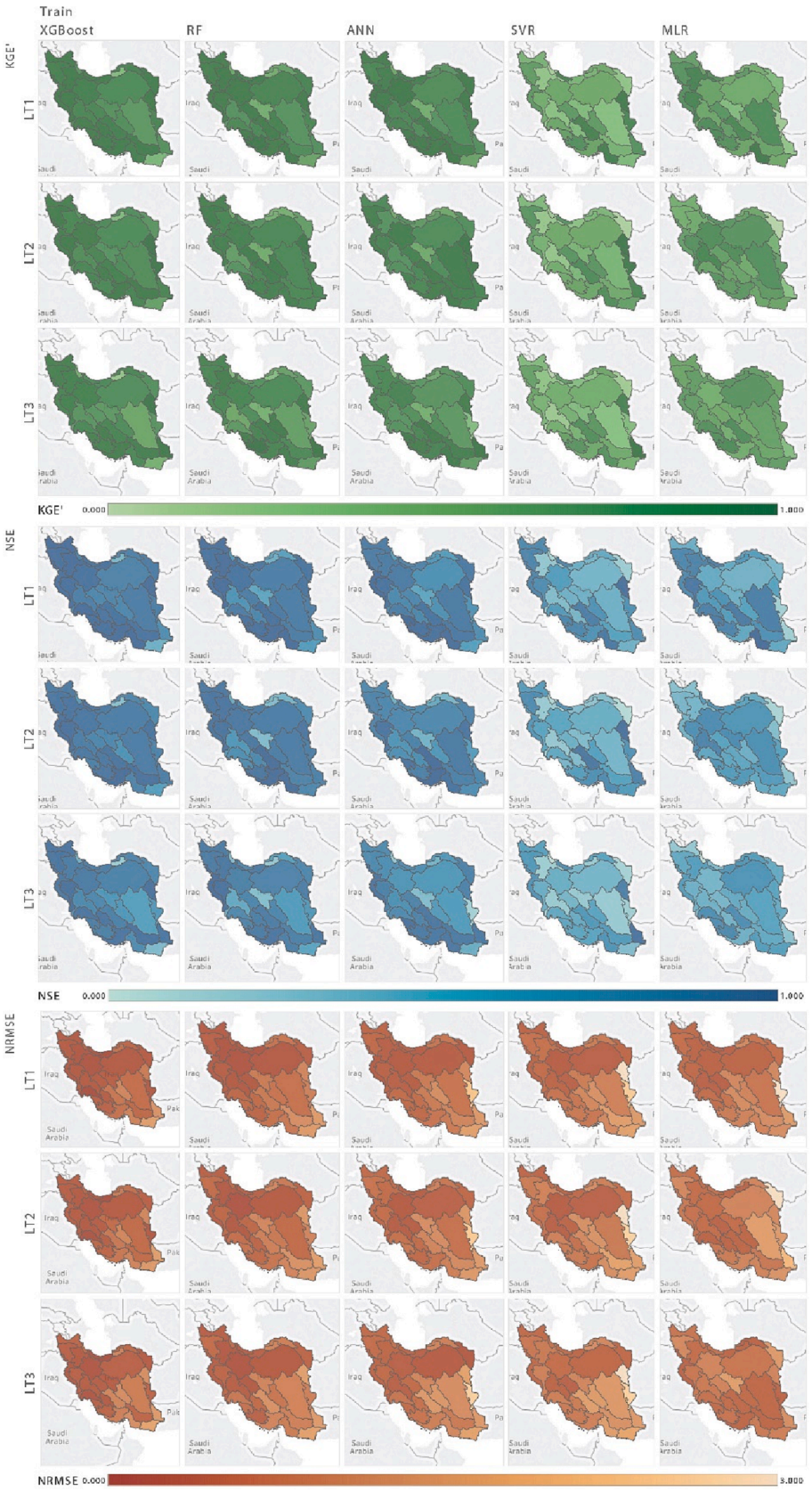



**Fig. 12.** Train phase performance for all models based on KGE', NSE, and NRMSE criteria in all 30 basins of Iran with 1- to 3-month lead times.

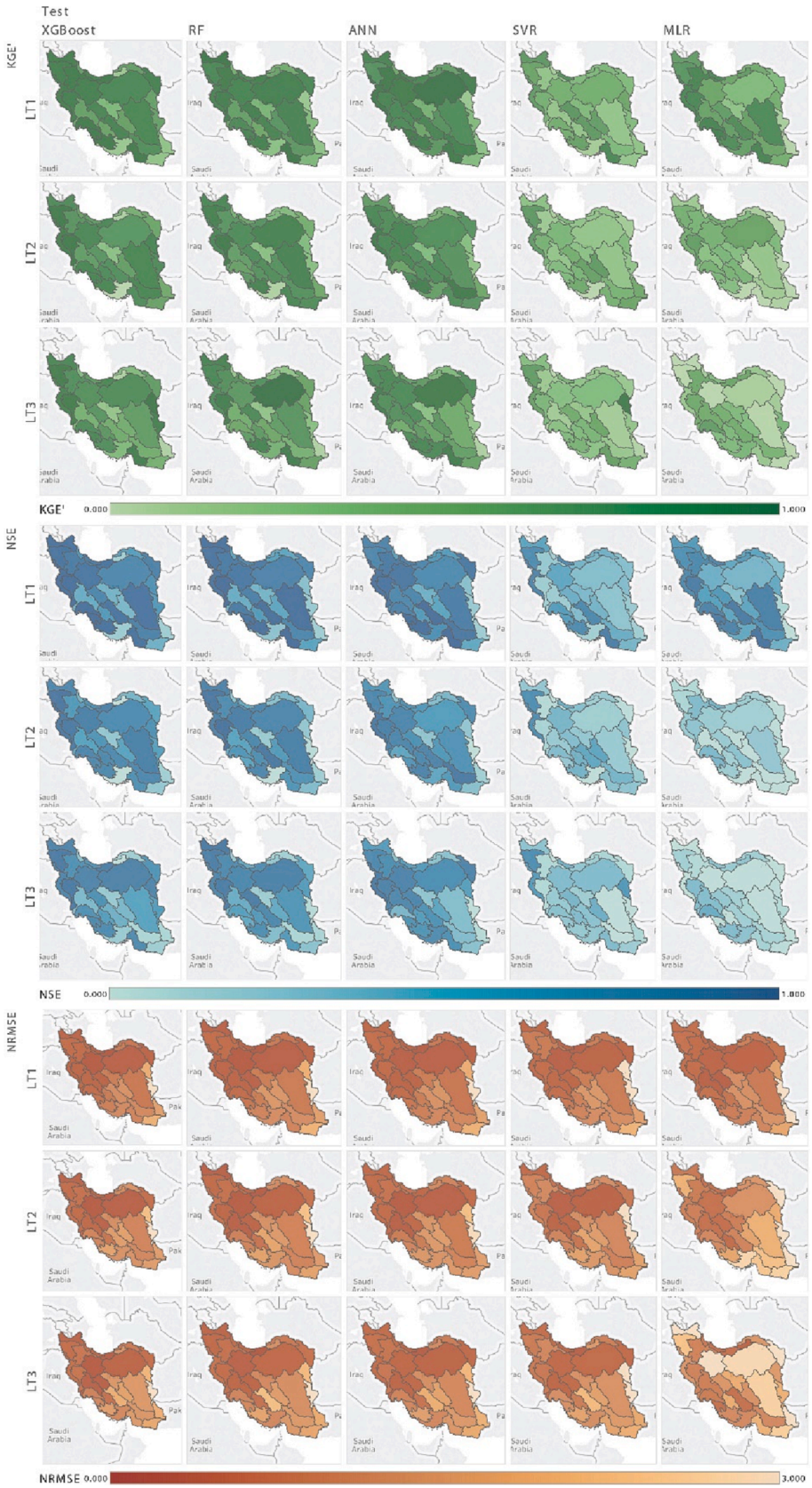



**Fig. 13.** Test phase performance for all models based on KGE', NSE, and NRMSE criteria in all 30 basins of Iran with 1- to 3-month lead times.

**Table 6**
Summary of average evaluation criteria for six major first level basins of Iran with 1- to 3-month lead times.

| Major Basin | Model | KGE' | | | | | | NSE | | | | | | NRMSE | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Train | | | Test | | | Train | | | Test | | | Train | | | Test | | |
| | | LT1 | LT2 | LT3 | LT1 | LT2 | LT3 | LT1 | LT2 | LT3 | LT1 | LT2 | LT3 | LT1 | LT2 | LT3 | LT1 | LT2 | LT3 |
| 1 | XGBoost | 0.77 | 0.72 | 0.67 | 0.70 | 0.56 | 0.52 | 0.77 | 0.70 | 0.66 | 0.66 | 0.48 | 0.45 | 0.27 | 0.37 | 0.40 | 0.52 | 0.68 | 0.66 |
| | RF | 0.72 | 0.64 | 0.60 | 0.70 | 0.54 | 0.50 | 0.70 | 0.62 | 0.60 | 0.67 | 0.49 | 0.45 | 0.38 | 0.51 | 0.51 | 0.51 | 0.68 | 0.63 |
| | ANN | 0.76 | 0.64 | 0.61 | 0.73 | 0.55 | 0.52 | 0.68 | 0.56 | 0.55 | 0.67 | 0.47 | 0.44 | 0.46 | 0.57 | 0.63 | 0.52 | 0.73 | 0.65 |
| | SVR | 0.40 | 0.33 | 0.31 | 0.39 | 0.21 | 0.19 | 0.42 | 0.32 | 0.33 | 0.37 | 0.18 | 0.16 | 0.57 | 0.71 | 0.76 | 0.60 | 0.79 | 0.77 |
| | MLR | 0.62 | 0.58 | 0.54 | 0.63 | 0.26 | 0.25 | 0.58 | 0.47 | 0.40 | 0.57 | 0.18 | 0.12 | 0.56 | 0.61 | 0.56 | 0.58 | 0.88 | 1.27 |
| 2 | XGBoost | 0.81 | 0.80 | 0.77 | 0.70 | 0.67 | 0.67 | 0.82 | 0.80 | 0.74 | 0.69 | 0.56 | 0.54 | 0.50 | 0.55 | 0.61 | 0.85 | 1.06 | 1.07 |
| | RF | 0.82 | 0.80 | 0.76 | 0.69 | 0.66 | 0.64 | 0.81 | 0.78 | 0.73 | 0.70 | 0.53 | 0.52 | 0.55 | 0.65 | 0.70 | 0.86 | 1.03 | 1.08 |
| | ANN | 0.84 | 0.82 | 0.78 | 0.77 | 0.74 | 0.72 | 0.82 | 0.77 | 0.74 | 0.77 | 0.60 | 0.60 | 0.62 | 0.69 | 0.76 | 0.85 | 1.00 | 1.07 |
| | SVR | 0.55 | 0.51 | 0.49 | 0.42 | 0.36 | 0.36 | 0.62 | 0.50 | 0.44 | 0.45 | 0.24 | 0.24 | 0.78 | 0.87 | 0.96 | 0.92 | 1.14 | 1.21 |
| | MLR | 0.70 | 0.63 | 0.63 | 0.71 | 0.22 | 0.29 | 0.69 | 0.54 | 0.53 | 0.67 | 0.11 | 0.15 | 0.73 | 0.83 | 0.67 | 0.82 | 1.59 | 1.29 |
| 3 | XGBoost | 0.86 | 0.89 | 0.86 | 0.86 | 0.90 | 0.87 | 0.89 | 0.84 | 0.86 | 0.86 | 0.84 | 0.81 | 0.25 | 0.28 | 0.33 | 0.55 | 0.53 | 0.58 |
| | RF | 0.86 | 0.88 | 0.81 | 0.79 | 0.84 | 0.80 | 0.87 | 0.84 | 0.81 | 0.86 | 0.84 | 0.78 | 0.33 | 0.56 | 0.52 | 0.54 | 0.51 | 0.48 |
| | ANN | 0.81 | 0.86 | 0.76 | 0.71 | 0.80 | 0.76 | 0.80 | 0.75 | 0.69 | 0.72 | 0.71 | 0.70 | 0.48 | 0.68 | 0.57 | 0.60 | 0.53 | 0.63 |
| | SVR | 0.62 | 0.68 | 0.58 | 0.54 | 0.65 | 0.61 | 0.74 | 0.67 | 0.58 | 0.66 | 0.65 | 0.65 | 0.64 | 0.84 | 0.75 | 0.86 | 0.98 | 0.84 |
| | MLR | 0.69 | 0.45 | 0.54 | 0.67 | 0.43 | 0.01 | 0.71 | 0.35 | 0.45 | 0.64 | 0.08 | 0.13 | 0.61 | 1.09 | 1.14 | 0.66 | 1.95 | 2.35 |
| 4 | XGBoost | 0.80 | 0.81 | 0.75 | 0.70 | 0.68 | 0.55 | 0.80 | 0.79 | 0.72 | 0.67 | 0.62 | 0.46 | 0.45 | 0.48 | 0.54 | 0.80 | 0.90 | 1.05 |
| | RF | 0.78 | 0.79 | 0.75 | 0.71 | 0.70 | 0.60 | 0.77 | 0.74 | 0.68 | 0.69 | 0.64 | 0.53 | 0.59 | 0.64 | 0.68 | 0.84 | 0.92 | 1.06 |
| | ANN | 0.77 | 0.79 | 0.74 | 0.75 | 0.72 | 0.61 | 0.72 | 0.70 | 0.64 | 0.68 | 0.62 | 0.50 | 0.68 | 0.75 | 0.85 | 0.82 | 0.91 | 1.04 |
| | SVR | 0.50 | 0.52 | 0.46 | 0.43 | 0.38 | 0.30 | 0.48 | 0.45 | 0.39 | 0.39 | 0.31 | 0.23 | 0.83 | 0.88 | 1.02 | 0.88 | 0.94 | 1.14 |
| | MLR | 0.59 | 0.63 | 0.55 | 0.54 | 0.30 | 0.26 | 0.57 | 0.55 | 0.46 | 0.53 | 0.09 | 0.03 | 0.78 | 0.69 | 0.75 | 0.87 | 1.22 | 1.86 |
| 5 | XGBoost | 0.82 | 0.90 | 0.84 | 0.46 | 0.69 | 0.60 | 0.78 | 0.81 | 0.77 | 0.46 | 0.43 | 0.45 | 0.79 | 0.93 | 1.03 | 2.26 | 2.31 | 2.67 |
| | RF | 0.81 | 0.85 | 0.79 | 0.24 | 0.36 | 0.29 | 0.77 | 0.80 | 0.74 | 0.21 | 0.14 | 0.23 | 1.15 | 1.32 | 1.32 | 2.22 | 2.28 | 2.63 |
| | ANN | 0.82 | 0.87 | 0.70 | 0.24 | 0.34 | 0.26 | 0.72 | 0.71 | 0.57 | 0.21 | 0.11 | 0.19 | 1.64 | 1.79 | 1.89 | 2.21 | 2.24 | 2.63 |
| | SVR | 0.80 | 0.87 | 0.69 | 0.35 | 0.34 | 0.38 | 0.73 | 0.75 | 0.60 | 0.22 | 0.13 | 0.22 | 2.58 | 3.07 | 3.10 | 4.71 | 5.44 | 5.50 |
| | MLR | 0.30 | 0.42 | 0.47 | 0.18 | −0.07 | 0.01 | 0.26 | 0.34 | 0.40 | 0.18 | −0.03 | −0.99 | 1.88 | 1.56 | 0.90 | 2.84 | 3.33 | 2.26 |
| 6 | XGBoost | 0.88 | 0.69 | 0.73 | 0.84 | 0.53 | 0.62 | 0.82 | 0.70 | 0.71 | 0.77 | 0.49 | 0.61 | 0.24 | 0.31 | 0.25 | 0.39 | 0.57 | 0.59 |
| | RF | 0.83 | 0.61 | 0.66 | 0.80 | 0.44 | 0.54 | 0.86 | 0.68 | 0.73 | 0.70 | 0.35 | 0.45 | 0.34 | 0.43 | 0.43 | 0.42 | 0.66 | 0.66 |
| | ANN | 0.89 | 0.73 | 0.75 | 0.87 | 0.70 | 0.75 | 0.82 | 0.74 | 0.76 | 0.82 | 0.66 | 0.74 | 0.27 | 0.46 | 0.23 | 0.38 | 0.55 | 0.48 |
| | SVR | 0.22 | 0.06 | 0.12 | 0.33 | 0.10 | 0.15 | 0.16 | 0.04 | 0.11 | 0.27 | 0.02 | 0.05 | 0.45 | 0.58 | 0.49 | 0.47 | 0.78 | 0.50 |
| | MLR | 0.34 | 0.02 | 0.52 | 0.34 | −0.24 | −0.25 | 0.35 | 0.10 | 0.44 | 0.27 | 0.01 | −0.16 | 0.59 | 2.99 | 1.48 | 0.62 | 3.81 | 3.24 |
| All | XGBoost | 0.80 | 0.79 | 0.75 | 0.68 | 0.65 | 0.60 | 0.80 | 0.77 | 0.72 | 0.66 | 0.55 | 0.50 | 0.44 | 0.51 | 0.56 | 0.87 | 1.01 | 1.09 |
| | RF | 0.78 | 0.76 | 0.72 | 0.66 | 0.61 | 0.56 | 0.77 | 0.73 | 0.69 | 0.65 | 0.52 | 0.48 | 0.57 | 0.67 | 0.70 | 0.88 | 1.01 | 1.09 |
| | ANN | 0.80 | 0.77 | 0.72 | 0.70 | 0.65 | 0.60 | 0.75 | 0.69 | 0.65 | 0.66 | 0.53 | 0.50 | 0.68 | 0.78 | 0.85 | 0.88 | 1.00 | 1.09 |
| | SVR | 0.52 | 0.50 | 0.45 | 0.41 | 0.33 | 0.30 | 0.53 | 0.46 | 0.41 | 0.39 | 0.25 | 0.22 | 0.91 | 1.05 | 1.12 | 1.20 | 1.41 | 1.48 |
| | MLR | 0.60 | 0.57 | 0.56 | 0.57 | 0.22 | 0.22 | 0.57 | 0.49 | 0.46 | 0.54 | 0.10 | −0.02 | 0.81 | 0.89 | 0.74 | 0.97 | 1.57 | 1.66 |

time-consuming process (Ni et al., 2020). Practice has shown that when there is a model with a large number of hyperparameters like XGBoost, the results may end up worse than those produced using a small number of parameters. This can potentially be explained by XGBoost model's optimization problems caused by the extensive list of parameters that need to be optimized (Szczepanek, 2022). The hyper-parameters in XGBoost have to be carefully tuned to achieve satisfactory forecasts and better generalization capability, despite the fact that it outperforms other tree-based models in terms of its ability to handle overfitting issues. This limitation may be the reason behind ANN's superior performance over XGBoost in some major basins. Existing studies have shown that each machine-learning algorithm has a certain scope of application, and there is currently no ML algorithm that performs best on any given dataset (Shi & Shen, 2022).

However, by comparing the result in all basins, especially by considering major basin 5, it can be seen that the fluctuation range of KGE' of XGBoost is significantly smaller than that of the other ML algorithms. This shows the XGBoost algorithm is more robust in capturing a wide set of characteristics across all basins compared to other ML algorithms. The superior performance of XGBoost over other models in particular SVR was likely related to its capacity to handle a larger space of features and non-linear relationships between features that can better capture the hydrological characteristics of river basins. Additionally, non-parametric models such as XGBoost have shown ability to identify complex relationships between different variables in river systems (Ni et al., 2020). The results also revealed that XGBoost outperformed RF in terms of accuracy and stability, likely due to its ability to account for nonlinear interactions between variables which often go undetected by other methods. This allows it to capture more complex relationships in the data which would otherwise be ignored. Tree-based machine learning models and boosting techniques displayed reasonably good results given they classify each variable based on their characteristics in making nodes and leaves. This allows these models to gradually improve performance starting with weak learners. Together these techniques further augment the potential of XGBoost as a viable alternative for streamflow forecasting applications.

The results of this study indicate that ECMWF precipitation, runoff, and temperature ensembles are suitable, although with varied degree of accuracy depending on the region, for flow forecasting in Iran. Furthermore, it was shown that runoff ensemble values contribute most significantly to basin streamflow forecasts. The two non-linear feature selections, i.e., RFE and BN, behaved similarly in selecting the best feature sets for all ML models over 30 s level basins. Finally, ANN and XGBoost broadly outperformed other ML models in all 30 basins and for all lead times.

## 4. Conclusion

This study evaluated runoff forecasting using ensemble products of ECMWF monthly precipitation, runoff, and temperature forecasts over Iran's basins with different climate zones for the period of 1981 to 2015. Using different linear and non-linear FSA i.e, Recursive Feature Elimination (RFE), Bayesian Networks (BN), and Pearson correlation analysis (Pca), best combination of inputs were selected to derive simulation models. The simulations of runoff were conducted through five ML models, namely eXtreme Gradient boosting (XGBoost), Random Forest (RF), Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Multiple Linear Regression (MLR) while results were compared to observed runoff in 30 basins in Iran. The findings of this study can help researchers beyond the geographical implementation in Iran. This analysis suggests that the modeling skill varied considerably according to climate, methodology, and lead-time.

This study found that ECMWF forecasts are efficient in prediction of runoff over a generally arid/semi-arid region, such as those found in Iran. In particular, the runoff ensemble followed by precipitation showed the highest importance in prediction of observed runoff in all basins. The ANN, followed by XGBoost and RF models had better fitting compared to SVR and MLR models in the training set in terms of most evaluation criteria in the majority of basins. In particular, the ANN, XGBoost, and RF models showed excellent performance in all basins in the wet months of the year. Overall, all models performed better over basins with higher runoff values, i.e. basins in the country's western region. In contrast, approximately all models had lower performance in basins with arid climate characteristics like basins in central Iran.

For the three superior models of XGBoost, ANN, and RF, RFE and BN FSAs were selected most frequently across Iran's 30 s level basins, while Pca was used least frequently in all lead times. Overall model performance based on the KGE' criteria yield a best-to-worst ranking of ANN, XGBoost, RF, MLR, and SVR (with KGE' values of 0.70, 0.68, 0.66, 0.57, and 0.41, respectively). The predictive performance of all models decreased with lead times beyond 1-month, where ANN and XGBoost outperformed other models (with KGE' of 0.65 for 2-month lead time and 0.60 for 3-month lead time).

Finally, an increase in lead time reduced the performance of all models although ANN and XGBoost performed better than other models for longer lead times based on all performance criteria. Furthermore, ANN, XGBoost, and RF demonstrated good performance in 2- to 3-month lead times over western and northern basins of Iran with high runoff values. In the arid central and southeastern Iran, however, except for XGBoost, all models showed poor performance, especially with more than 1-month lead times.

## CRediT authorship contribution statement

**Mohammad Akbarian:** Conceptualization, Methodology, Software, Writing – original draft, Visualization, Data curation. **Bahram Saghafian:** Conceptualization, Methodology, Software, Supervision, Writing – review & editing. **Saeed Golian:** Conceptualization, Methodology, Validation, Investigation, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## References

Ali, S., Shahbaz, M., 2020. Streamflow forecasting by modeling the rainfall–streamflow relationship using artificial neural networks. Modeling Earth Syst. Environ. 6 (3), 1645–1656. https://doi.org/10.1007/s40808-020-00780-3.

Aminyavari, S., Saghafian, B., Delavar, M., 2018. Evaluation of TIGGE Ensemble Forecasts of Precipitation in Distinct Climate Regions in Iran. Adv. Atmos. Sci. 35 (4), 457–468.

Anusree, K., Varghese, K.O., 2016. Streamflow Prediction of Karuvannur River Basin Using ANFIS, ANN and MNLR Models. Procedia Technol. 24, 101–108. https://doi.org/10.1016/J.PROTCY.2016.05.015.

Apaydin, H., Taghi Sattari, M., Falsafian, K., Prasad, R., 2021. Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions. J. Hydrol. 600, 126506 https://doi.org/10.1016/J.JHYDROL.2021.126506.

Shahab Araghinejad. (2014). Water Science and Technology Library ShahabbAraghinejad Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering. http://www.springer.com/series/6689.

Beaumont, P., Horsburgh, B., Pilgerstorfer, P., Droth, A., Oentaryo, R., Ler, S., Nguyen, H., Ferreira, G. A., Patel, Z., & Leong, W. (2017). *Causalnex*. https://github.com/quantumblacklabs/causalnex.

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011.

Berrar, D., 2019. Cross-Validation. Encyclopedia Bioinform. Comput. Biol.: ABC Bioinform. 1–3, 542–545. https://doi.org/10.1016/B978-0-12-809633-8.20349-X.

Cawley, G.C., Talbot, N.L.C., 2010. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J. Mach. Learn. Res. 11, 2079–2107.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. 10.1145/2939672.2939785.

Cheng, M., Fang, F., Kinouchi, T., Navon, I.M., Pain, C.C., 2020. Long lead-time daily and monthly streamflow forecasting using machine learning methods. J. Hydrol. 590, 125376 https://doi.org/10.1016/j.jhydrol.2020.125376.

Crochemore, L., Ramos, M.H., Pappenberger, F., Perrin, C., 2017. Seasonal streamflow forecasting by conditioning climatology with precipitation indices. Hydrol. Earth Syst. Sci. 21 (3), 1573–1591. https://doi.org/10.5194/hess-21-1573-2017.

Das, P., Sachindra, D.A., Chanda, K., 2022. Machine Learning-Based Rainfall Forecasting with Multiple Non-Linear Feature Selection Algorithms. Water Resour. Manag. 36 (15), 6043–6071.

Descamps, L., Labadie, C., Joly, A., Bazile, E., Arbogast, P., Cébron, P., 2015. PEARP, the Météo-France short-range ensemble prediction system. Q. J. R. Meteorolog. Soc. 141 (690), 1671–1685. https://doi.org/10.1002/qj.2469.

Djibo, A., Karambiri, H., Seidou, O., Sittichok, K., Philippon, N., Paturel, J., Saley, H., 2015. Linear and Non-Linear Approaches for Statistical Seasonal Rainfall Forecast in the Sirba Watershed Region (SAHEL). Climate 3 (3), 727–752. https://doi.org/10.3390/cli3030727.

Dutta, R., Maity, R., 2020. Temporal Networks-Based Approach for Nonstationary Hydroclimatic Modeling and its Demonstration With Streamflow Prediction. Water Resour. Res. 56 (8) https://doi.org/10.1029/2020WR027086.

Ferreira, R.G., da Silva, D.D., Elesbon, A.A.A., Fernandes-Filho, E.I., Veloso, G.V., de Fraga, M., S., & Ferreira, L. B., 2021. Machine learning models for streamflow regionalization in a tropical watershed. J. Environ. Manage. 280 (November) https://doi.org/10.1016/j.jenvman.2020.111713.

Gebrechorkos, S.H., Pan, M., Beck, H.E., Sheffield, J., 2022. Performance of State-of-the-Art C3S European Seasonal Climate Forecast Models for Mean and Extreme Precipitation Over Africa. Water Resour. Res. 58 (3) https://doi.org/10.1029/2021WR031480.

Golian, S., Saghafian, B., Maknoon, R., 2010. Derivation of Probabilistic Thresholds of Spatially Distributed Rainfall for Flood Forecasting. Water Resour. Manag. 24 (13), 3547–3559. https://doi.org/10.1007/s11269-010-9619-7.

Golian, S., Saghafian, B., Elmi, M., Maknoon, R., 2011. Probabilistic rainfall thresholds for flood forecasting: evaluating different methodologies for modelling rainfall spatial correlation (or dependence). Hydrol. Process. 25 (13), 2046–2055. https://doi.org/10.1002/hyp.7956.

Greve, P., Kahil, T., Mochizuki, J., Schinko, T., Satoh, Y., Burek, P., Fischer, G., Tramberend, S., Burtscher, R., Langan, S., Wada, Y., 2018. Global assessment of water challenges under uncertainty in water scarcity projections. Nat. Sustainability 1 (9), 486–494. https://doi.org/10.1038/s41893-018-0134-9.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91. https://doi.org/10.1016/J.JHYDROL.2009.08.003.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46 (1–3), 389–422. https://doi.org/10.1023/A:1012487302797.

Hapuarachchi, H.A.P., Bari, M.A., Kabir, A., Hasan, M.M., Woldemeskel, F.M., Gamage, N., Sunter, P.D., Zhang, X.S., Robertson, D.E., Bennett, J.C., Feikema, P.M., 2022. Development of a national 7-day ensemble streamflow forecasting service for Australia. Hydrol. Earth Syst. Sci. 26 (18), 4801–4821. https://doi.org/10.5194/hess-26-4801-2022.

Hawcroft, M., Lavender, S., Copsey, D., Milton, S., Rodríguez, J., Tennant, W., Webster, S., Cowan, T., 2021. The Benefits of Ensemble Prediction for Forecasting an Extreme Event: The Queensland Floods of February 2019. Mon. Weather Rev. 149 (7), 2391–2408. https://doi.org/10.1175/MWR-D-20-0330.1.

He, S., Guo, S., Zhang, J., Liu, Z., Cui, Z., Zhang, Y., Zheng, Y., 2022. Multi-objective operation of cascade reservoirs based on short-term ensemble streamflow prediction. J. Hydrol. 610 (April), 127936 https://doi.org/10.1016/j.jhydrol.2022.127936.

Ingleby, B., 2015. Global assimilation of air temperature, humidity, wind and pressure from surface stations. Q. J. R. Meteorolog. Soc. 141 (687), 504–517. https://doi.org/10.1002/QJ.2372.

Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecol. Model. 83 (1–2), 55–66. https://doi.org/10.1016/0304-3800(95)00084-9.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., Monge-sanz, B. M., & Park, S. (2019). *SEAS5 : the new ECMWF seasonal forecast system*. 1087–1117.

Kao, I.F., Zhou, Y., Chang, L.C., Chang, F.J., 2020. Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. J. Hydrol. 583, 124631 https://doi.org/10.1016/J.JHYDROL.2020.124631.

Kao, I.F., Liou, J.Y., Lee, M.H., Chang, F.J., 2021. Fusing stacked autoencoder and long short-term memory for regional multistep-ahead flood inundation forecasts. J. Hydrol. 598, 126371 https://doi.org/10.1016/J.JHYDROL.2021.126371.

Karimi, S., Shiri, J., Kisi, O., Shiri, A.A., 2016. Short-term and long-term streamflow prediction by using "wavelet–gene expression" programming approach. ISH J. Hydraulic Eng. 22 (2), 148–162. https://doi.org/10.1080/09715010.2015.1103201.

Kaspar, F., Zimmermann, K., Polte-Rudolf, C., 2015. An overview of the phenological observation network and the phenological database of Germany's national meteorological service (Deutscher Wetterdienst). Adv. Sci. Res. 11 (1), 93–99. https://doi.org/10.5194/ASR-11-93-2014.

Kilinc, H.C., Haznedar, B., 2022. A Hybrid Model for Streamflow Forecasting in the Basin of Euphrates. Water (Switzerland) 14 (1). https://doi.org/10.3390/w14010080.

Kim, J., 2009. Estimating classification error rate : Repeated cross-validation, repeated hold-out and bootstrap. Comput. Stat. Data Anal. 53 (11), 3735–3745. https://doi.org/10.1016/j.csda.2009.04.009.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrol. 424–425, 264–277. https://doi.org/10.1016/J.JHYDROL.2012.01.011.

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrol. Earth Syst. Sci. 23 (10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019.

Kolachian, R., Saghafian, B., 2019. Deterministic and probabilistic evaluation of raw and post processed sub-seasonal to seasonal precipitation forecasts in different precipitation regimes. Theor. Appl. Climatol. 137 (1–2), 1479–1493. https://doi.org/10.1007/s00704-018-2680-5.

Krstanovic, P.F., Singh, V.P., 1991. A univariate model for long-term streamflow forecasting – 1. Development. Stochastic Hydrology and Hydraulics 5 (3), 173–188. https://doi.org/10.1007/BF01544056/METRICS.

Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. In *Applied Predictive Modeling*. Springer New York. 10.1007/978-1-4614-6849-3.

LaValle, S. M., Branicky, M. S., & Lindemann, S. R. (2016). On the Relationship between Classical Grid Search and Probabilistic Roadmaps. Http://Dx.Doi.Org/10.1177/0278364904045481, 23(7–8), 673–692. 10.1177/0278364904045481.

Leu, S.S., Bui, Q.N., 2016. Leak Prediction Model for Water Distribution Networks Created Using a Bayesian Network Learning Approach. Water Resour. Manag. 30 (8), 2719–2733. https://doi.org/10.1007/S11269-016-1316-8/METRICS.

Li, Z., Xu, X., Xu, C., Liu, M., Wang, K., Yu, B., 2017. Annual runoff is highly linked to precipitation extremes in Karst catchments of Southwest China. J. Hydrometeorol. 18 (10), 2745–2759. https://doi.org/10.1175/JHM-D-17-0032.1.

Liang, Z., Li, Y., Hu, Y., Li, B., Wang, J., 2018. A data-driven SVR model for long-term runoff prediction and uncertainty analysis based on the Bayesian framework. Theor. Appl. Climatol. 133 (1–2), 137–149. https://doi.org/10.1007/s00704-017-2186-6.

Liu, S., Xu, J., Zhao, J., Xie, X., Zhang, W., 2014. Efficiency enhancement of a process-based rainfall–runoff model using a new modified AdaBoost.RT technique. Appl. Soft Comput. 23, 521–529. https://doi.org/10.1016/J.ASOC.2014.05.033.

Lopez, P., 2013. Experimental 4D-Var Assimilation of SYNOP Rain Gauge Data at ECMWF. Mon. Weather Rev. 141 (5), 1527–1544. https://doi.org/10.1175/MWR-D-12-00024.1.

Maclachlan, C., Arribas, A., Peterson, K.A., Maidens, A., Fereday, D., Scaife, A.A., Gordon, M., Vellinga, M., Williams, A., Comer, R.E., Camp, J., Xavier, P., Madec, G., 2015. Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. Q. J. R. Meteorolog. Soc. 141 (689), 1072–1084. https://doi.org/10.1002/qj.2396.

Maddu, R., Pradhan, I., Ahmadisharaf, E., Singh, S.K., Shaik, R., 2022. Short-range reservoir inflow forecasting using hydrological and large-scale atmospheric circulation information. J. Hydrol. 612, 128153 https://doi.org/10.1016/J.JHYDROL.2022.128153.

Malik, A., Tikhamarine, Y., Souag-Gamane, D., Kisi, O., Pham, Q.B., 2020. Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. Stoch. Env. Res. Risk A. 34 (11), 1755–1773. https://doi.org/10.1007/S00477-020-01874-1/METRICS.

Mansouri Daneshvar, M.R., Ebrahimi, M., Nejadsoleymani, H., 2019. An overview of climate change in Iran: facts and statistics. *Environmental*. Syst. Res. 8 (1) https://doi.org/10.1186/s40068-019-0135-3.

Manzanas, R., Gutiérrez, J.M., Bhend, J., Hemri, S., Doblas-Reyes, F.J., Torralba, V., Penabad, E., Brookshaw, A., 2019. Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. Clim. Dyn. 53 (3–4), 1287–1305. https://doi.org/10.1007/s00382-019-04640-4.

Meydani, A., Dehghanipour, A., Schoups, G., Tajrishy, M., 2022. Daily reservoir inflow forecasting using weather forecast downscaling and rainfall-runoff modeling: Application to Urmia Lake basin, Iran. Journal of Hydrology: Regional Studies 44. https://doi.org/10.1016/j.ejrh.2022.101228.

Molinaro, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics 21 (15), 3301–3307. https://doi.org/10.1093/bioinformatics/bti499.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. Trans. ASABE 58 (6), 1763–1785. https://doi.org/10.13031/trans.58.10715.

Munson, M.A., Caruana, R., 2009. On feature selection, bias-variance, and bagging. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 5782 (LNAI(PART 2)), 144–159. https://doi.org/10.1007/978-3-642-04174-7_10/COVER.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10 (3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J., 2020. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J. Hydrol. 586 (March), 124901 https://doi.org/10.1016/j.jhydrol.2020.124901.

Nicolì, D., Bellucci, A., Ruggieri, P., Athanasiadis, P.J., Materia, S., Peano, D., Fedele, G., Hénin, R., Gualdi, S., 2023. The Euro-Mediterranean Center on Climate Change (CMCC) decadal prediction system. Geosci. Model Dev. 16 (1), 179–197. https://doi.org/10.5194/gmd-16-179-2023.

Nifa, K., Boudhar, A., Ouatiki, H., Elyoussfi, H., Bargam, B., & Chehbouni, A. (2023). Deep Learning Approach with LSTM for Daily Streamflow Prediction in a Semi-Arid Area: A Case Study of Oum Er-Rbia River Basin, Morocco. *Water 2023, Vol. 15, Page 262*, *15*(2), 262. 10.3390/W15020262.

Nobakht, M., Saghafian, B., Aminyavari, S., 2021. Skill Assessment of Copernicus Climate Change Service Seasonal Ensemble Precipitation Forecasts over Iran. Adv. Atmos. Sci. 38 (3), 504–521. https://doi.org/10.1007/s00376-020-0025-7.

Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology 2006 24:12*, *24*(12), 1565–1567. 10.1038/nbt1206-1565.

Noorbeh, P., Roozbahani, A., Kardan Moghaddam, H., 2020. Annual and Monthly Dam Inflow Prediction Using Bayesian Networks. Water Resour. Manag. 34 (9), 2933–2951. https://doi.org/10.1007/S11269-020-02591-8/METRICS.

Raziei, T., 2022. Climate of Iran according to Köppen-Geiger, Feddema, and UNEP climate classifications. Theor. Appl. Climatol. 148 (3–4), 1395–1416. https://doi.org/10.1007/s00704-022-03992-y.

Robertson, D.E., Shrestha, D.L., Wang, Q.J., 2013. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. Hydrol. Earth Syst. Sci. 17 (9), 3587–3603. https://doi.org/10.5194/HESS-17-3587-2013.

Saatsaz, M., 2020. A historical investigation on water resources management in Iran. In: Environment, Development and Sustainability, Vol. 22, Issue 3. Springer, Netherlands. https://doi.org/10.1007/s10668-018-00307-y.

Schoppa, L., Disse, M., Bachmair, S., 2020. Evaluating the performance of random forest for large-scale flood discharge simulation. J. Hydrol. 590, 125531 https://doi.org/10.1016/J.JHYDROL.2020.125531.

Scutari, M., 2017. Bayesian Network Constraint-Based Structure Learning Algorithms: Parallel and Optimized Implementations in the bnlearn R Package. J. Stat. Softw. 77 (1), 1–20. https://doi.org/10.18637/JSS.V077.I02.

Sharma, P., Machiwal, D., 2021. Streamflow forecasting: overview of advances in data-driven techniques. Advances in Streamflow Forecasting: From Traditional to Modern Approaches 1–50. https://doi.org/10.1016/B978-0-12-820673-7.00013-5.

Shi, M., Shen, W., 2022. Automatic Modeling for Concrete Compressive Strength Prediction Using Auto-Sklearn. Buildings 12 (9). https://doi.org/10.3390/buildings12091406.

Slater, L.J., Villarini, G., Bradley, A.A., Vecchi, G.A., 2017. A dynamical statistical framework for seasonal streamflow forecasting in an agricultural watershed. Clim. Dyn. 1–17. https://doi.org/10.1007/s00382-017-3794-7.

Smith, D.M., Cusack, S., Colman, A.W., Folland, C.K., Harris, G.R., Murphy, J.M., 2007. Improved surface temperature prediction for the coming decade from a global climate model. Science 317 (5839), 796–799. https://doi.org/10.1126/SCIENCE.1139540/SUPPL_FILE/SMITH.SOM.PDF.

Sudheer, C., Maheswaran, R., Panigrahi, B.K., Mathur, S., 2014. A hybrid SVM-PSO model for forecasting monthly streamflow. Neural Comput. & Applic. 24 (6), 1381–1389. https://doi.org/10.1007/S00521-013-1341-Y.

Szczepanek, R., 2022. Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost. LightGBM and CatBoost. *Hydrology* 9 (12), 226. https://doi.org/10.3390/hydrology9120226.

Tyralis, H., Papacharalampous, G., Langousis, A., 2021. Super ensemble learning for daily streamflow forecasting: large-scale demonstration and comparison with multiple machine learning algorithms. Neural Comput. & Applic. 33 (8), 3053–3068. https://doi.org/10.1007/s00521-020-05172-3.

Wang, Q.J., Shao, Y., Song, Y., Schepen, A., Robertson, D.E., Ryu, D., Pappenberger, F., 2019. An evaluation of ECMWF SEAS5 seasonal climate forecasts for Australia using a new forecast calibration algorithm. Environ. Model. Softw. 122, 104550 https://doi.org/10.1016/J.ENVSOFT.2019.104550.

Wegayehu, E.B., Muluneh, F.B., 2022. Short-Term Daily Univariate Streamflow Forecasting Using Deep Learning Models. Adv. Meteorol. 2022, 1–21.