



Assessment of machine learning model performance for seasonal precipitation simulation based on teleconnection indices in Iran

Jalil Helali¹ · Mehdi Mohammadi Ghaleni² · Seyed Asaad Hosseini³ · Ali Lotfi Siraeei⁴ · Vahideh Saeidi⁵ · Farshad Safarpour⁶ · Mojgan Mirzaei⁷ · Mohammad Lotfi⁸

Received: 4 January 2022 / Accepted: 19 July 2022 / Published online: 28 July 2022
© Saudi Society for Geosciences 2022

Abstract

Precipitation is one of the most important factors affecting the climate, hydrological processes, and living environment. Hence, the precipitation forecast is significant for water resource exploitation and preparation for extreme climatic events such as drought and floods. In this context, teleconnection indices are commonly used as predictors across the globe. However, most studies have focused on investigating the correlation between seasonal precipitation and teleconnection indices from the meteorological stations by using some limited models to simulate the precipitation. This study evaluated the use of 40 teleconnection indices by exploiting 4 machine learning models (ML), namely generalized regression neural network (GRNN), multi-layer perceptron (MLP), least squares support vector machine (LSSVM), and multilinear regression (MLR) to forecast and model seasonal precipitation in a larger scale than meteorological stations, specifically main basins, and sub-basins of Iran. For that purpose, the seasonal precipitations in 6 main basins, including 30 sub-basins, were selected based on 717 stations for the period 1987–2015. First, the correlations between 40 teleconnection indices and the seasonal precipitation of sub-basins were measured by Pearson correlation to determine their significance using a correlation matrix. Then, the most significant (predictor) variables with time lags of 1 to 6 months (for each season) were extracted by a stepwise procedure per sub-basin and considered as input for the four ML models. Finally, the performances of the models were assessed based on coefficient of determination (R^2), root mean square error (RMSE), mean bias error (MBE), and scatter index (SI) statistic tests. According to the results, the most significant correlation between teleconnection indices and autumn, winter, and spring precipitations occurred with time lags of 1–4, 4–6, and 1–4 months, respectively, in some teleconnection indices. Evaluation of the simulation, the LSSVM model, performed excellently followed by MLP for most of the sub-basins, while GRNN and MLR models showed a poor simulation performance between 1987 and 2015. The results showed that the LSSVM had more accuracy and less RMSE in the training period than other models, while the MLP and GRNN models' RMSE were less than the MLR and LSSVM. Therefore, the MLP and GRNN are recommended in modeling and forecasting seasonal precipitation in Iranian sub-basins. The overall results confirmed that comprehensive study of teleconnection indices and seasonal precipitation based on all basins and sub-basin was appropriate for the teleconnection indices pattern using intelligent models and it might increase the accuracy and reliability of modeling and forecasting.

Keywords Precipitation · Teleconnection · Machine learning · Iran

Introduction

Precipitation is one of the most significant water resources to sustain the hydrological cycle. Therefore, its precise and reliable prediction is necessary for water resources planning and

management in many sectors (e.g., agriculture and industry). Consequently, more accurate precipitation prediction plays a major role to minimize the damages and losses of extreme events such as drought and floods. Numerous models are used to predict climatic variables, especially precipitation and temperature, which are commonly grouped into two groups: dynamical and statistical approaches (Doblas-Reyes et al. 2013; Navid and Niloy 2018; Schepen et al. 2012). Dynamical models make predictions using multiple equations that represent the fluid behavior in the atmosphere–ocean–land, such as general circulation models

Responsible Editor: Zhihua Zhang

✉ Seyed Asaad Hosseini
Hosseini.asad8@gmail.com

Extended author information available on the last page of the article

(GCM). In statistical models, prediction is based on the relationships between the target predictor variable and atmosphere–ocean–land data including time series models (Islam and Imteaz 2020), multivariate regressions (Nalley et al. 2019; Qian and Xu 2020; Hu et al. 2020), and methods based on artificial intelligence like the artificial neural network (Doblas-Reyes et al. 2013; Woldemeskel et al. 2014; Canchala et al. 2020; Helali et al. 2021a). Statistical models are relatively flexible in model construction and can improve the prediction accuracy depending on the length of the statistical period and the type of available data, although they are unstable to dynamical models (Kim et al. 2020). The applications of machine learning models, especially artificial intelligence models in modeling natural phenomena, have been reported by several studies (Choubin et al. 2014; Modaresi et al. 2018a; Helali et al. 2021a). Studies have shown that using different statistical models such as regression and artificial intelligence methods as well as large-scale teleconnection indices can better predict climate variables such as precipitation (Hartman et al. 2016; Kim et al. 2020; Helali et al. 2021a), runoff (Lee et al. 2020), evapotranspiration (Asadi Oskouei and Helali 2021), and crop yield (Gonsamo et al. 2016; Heino et al. 2020).

Teleconnection indices (e.g., El Niño Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO)) are determined by the large-scale climate variables with repetitive and periodic patterns (annually or decadal returning period) to explain the effect of climate anomalies of a distant phenomenon on regional climate conditions (Gholami Rostam et al. 2020). Synoptic and dynamic analysis of precipitation anomalies also demonstrated the role of westward displacement (sea to land) and eastward (land to sea) components of Saudi Arabia's high-pressure system (Helali et al. 2021c). Besides, the correlation between the teleconnection indices and climatic and hydrological variables has been extensively studied throughout the world (Gong and Wang 1999; Mekanik et al. 2013; Yang et al. 2019; Helali et al. 2020; Ahmadi et al. 2019). These indices were exploited by the researchers to measure hydrological sensitivity to climate change (Modaresi et al. 2018a; Kingston et al. 2006; Loboda et al. 2006) and in water resources planning due to their effect on climate changes and their impact on improving accuracy and time lag predictions (Dariane et al. 2019; Helali et al. 2021a). Pourasghar et al. (2012) revealed that changes in the annual precipitation of the southern part of Iran were mainly influenced by ENSO and Indian Ocean Dipole (IOD) indices so that in the negative phase of IOD, the anomalous moisture flux is directed towards the north. As a result, it decreases the moisture injection from the south, while fluctuations in late winter and early spring precipitation are strongly affected by changes in the Mediterranean Sea. Choubin et al. (2014) represented that teleconnection indices would explain up to

81% of the variance in Bakhtegan-Maharloo basin droughts, and those indices could better predict the drought using the neuro-fuzzy model rather than the regression model. Kinouchi et al. (2018) used teleconnection indices in two basins in Thailand and predicted seasonal precipitation using the multilinear regression (MLR) model and highlighted that equatorial SOI (EQ-SOI) and indices related to sea surface temperature (SST) could be used to quantify seasonal precipitation in the positive phase of EQ-SOI. Moreover, it is found that the LSSVM hybrid model was able to more accurately predict monthly precipitation of the Yangtze River basin by using teleconnection indices as predictor variables (Tao et al. 2017). Dariane et al. (2019) concluded that Pacific Decadal Oscillation (PDO), TNI, and NINO3 were the most important indices in long-term precipitation prediction using a genetic algorithm in some meteorological stations in Iran. Dehghani et al. (2020) demonstrated that seasonal precipitation had the highest correlation with Southern Oscillation Index (SOI), NAO, and PDO in Iran's basins. The authors stated that those teleconnection indices could be used as predictor variables at the basin scale, as well. Kim et al. (2020) showed that teleconnection indices could more satisfactorily forecast the seasonal precipitation of most seasons in the Han basin of South Korea using a multivariate regression model. However, it failed to predict the summer precipitation accurately. Using teleconnection indices, Qian and Xu (2020) studied the autumn precipitation in the Yangtze River basin of China using Bayesian linear regression (BLR) and multilinear regression methods (MLR) and showed that the BLR method had more satisfactory results than the MLR method and can be used in planning water resource management of this basin. Based on a synoptic-dynamic analysis, Helali et al. (2021c) showed that the seasonal and annual rainfall anomalies in Iran basins at different intensities of the ENSO were more than the different phases of ENSO.

Literature review shows a case relationship between teleconnection indices and climatic variables, especially precipitation, and therefore, it cannot be used operationally in the comprehensive and regional management of precipitation and water resources (Ruigar and Golian 2015; Kinouchi et al. 2018; Kim et al. 2020). On the other hand, the correlation between teleconnection indices and precipitation has been studied on a station and point scale, with sometimes contradictory results in neighboring areas (Ghasemi and Khalili 2008; Ahmadi et al. 2019) due to the lack of disproportion of the spatial scales of teleconnection indices and studied stations. Moreover, most studies have investigated the correlation and prediction of precipitation with teleconnection indices in limited seasons (Dehghani et al. 2020; Qian and Xu 2020; Helali et al. 2021a). Using different forecasting models also showed that intelligent structure models are more efficient and accurate in

precipitation forecasting (Kinouchi et al. 2018; Xiao 2019; Kim et al. 2020; Zhu et al. 2020). Machine learning models (MLMs) are widely applied to solve hydrological problems. The significance of MLMs is the ability to plot the input–output patterns without prior knowledge of the factors affecting the forecast parameters (Najah et al. 2011; Hipni et al. 2013; Ridwan et al. 2021). The MLMs, the most popular intelligent methods, learn from the patterns of input datasets (training data) to model and predict non-linear variables and solve complex problems in prediction of natural hazard and phenomena (Kalantar et al. 2021; Seydi et al. 2022). The ML algorithms accurately and rapidly model complex features and a large number of inputs. The outperformance of ML-based algorithms was reported in many applications (Kalantar et al. 2021). MLMs are artificial intelligence (AI) used to induce regularities and patterns, providing easier implementation with low computation cost, as well as fast training, validation, testing, and evaluation, with high performance and simplicity compared to physical models (Mekanik et al. 2013). The continuous advancement of MLMs over the last two decades confirmed their suitability for precipitation with an acceptable rate (Mosavi et al. 2017, 2018). Anochi et al. (2021) evaluated the different machine learning models for precipitation prediction over South America. The results showed that the machine learning models could produce predictions for different climate seasons with minimum errors of 2 mm in most of the continent in comparison to satellite-observed precipitation patterns. Helali et al. (2021a) emphasized that the MLP model was more accurate than the MLR model in predicting spring rainfall in Iran's basins. Helali et al. (2021b) analyzed the last spring frost (LSF) and chilling (LSC) considering the ENSO in Iran. Their results highlighted that the probability and occurrence date of the LSF and LSC (with different base and critical temperatures in ENSO phases) were correlated by precedence and latency in the whole period.

Considering the climate and also the importance of seasonal precipitation in an arid and semi-arid region, this study tried to comprehensively analyze the significance and correlation between teleconnection indices and seasonal precipitation in basins and sub-basins of Iran to select the most appropriate predictor variable using 4 simulating models (i.e., generalized regression neural network (GRNN), multi-layer perceptron (MLP), least squares support vector machine (LSSVM), and multilinear regression (MLR)). Then, the efficiency of statistical and intelligent models in seasonal precipitation simulation using teleconnection indices (as predictor variables) was analyzed and evaluated by the coefficient of determination (R^2), root mean square error (RMSE), mean bias error (MBE), and scatter index (SI) statistic tests.

Data and methodology

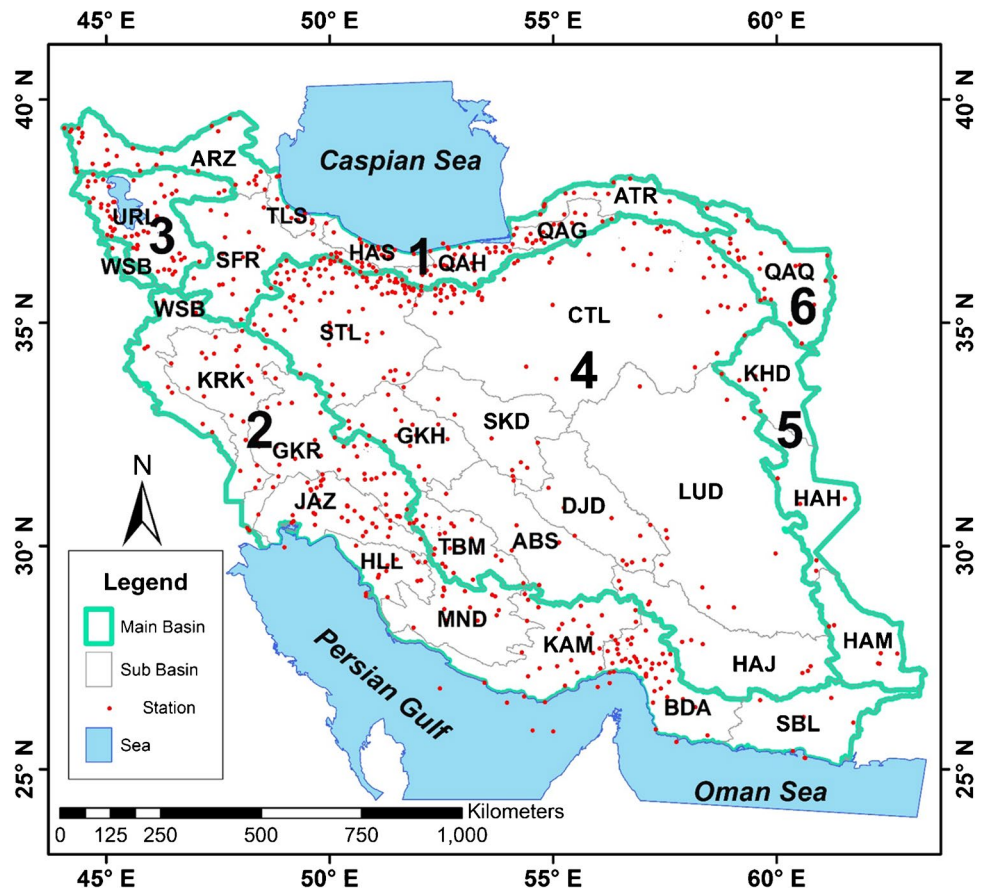
Study area

Iran is in arid and semi-arid part of the world, a country in Western Asia with mean annual precipitation of less than one-third of the global average (Salehi et al. 2019; Biabanaki et al. 2013). Iran is divided into 6 main basins and 30 sub-basins covering various climates (Fig. 1). The areas with a lower density of the stations are located in hot deserts, having almost no precipitation. From 1987 to 2015, the maximum (i.e., Talesh sub-basin) and minimum (Hamoon-Hirmand sub-basin) mean annual precipitations were 1253.6 mm and 82.3 mm, respectively. The monthly precipitation data was recorded by the Islamic Republic of Iran Meteorological Organization (IRIMO) and the Ministry of Energy. A total number of 717 stations (117 synoptic stations and 600 rain gage stations) were selected out of 2300 available stations in terms of sufficient statistical length (1987–2015) and continuous data with no gaps. The homogeneity of the dataset was examined against the run-test (Helali et al. 2020), and the amount of seasonal precipitation (autumn, winter, and spring) was determined per sub-basin using the Thiessen method (Table 1). The summer precipitation was excluded due to very limited or no precipitation in the arid and semi-arid climate of the region. According to studies by Masoudian (2005) and Masoudian and Ataei (2005) in Iran, precipitation mainly occurs in the autumn, winter, and spring seasons, while the summer precipitations are limited to some parts of Iran. Moreover, the spring starts in April and ends in June. The autumn begins in October and ends in December, and the winter arrives in January and ends in March. The characteristics (i.e., area, number of stations in each basin and sub-basin, mean autumn, winter, spring, annual precipitation, and the average altitude) of each sub-basin are shown in Table 1. Due to low precipitation in the arid zone, the summer precipitation was omitted from the modeling. The teleconnection indices for this study are listed in Table 2. The teleconnection data set were downloaded from the National Oceanic and Atmospheric Administration (NOAA) website (<http://www.cpc.ncep.noaa.gov/data/teledoc/teleindcalc.shtml>).

Selection of simulation indices

Dealing with large-scale datasets emphasizes the use of time-effective methods. Moreover, the optimization of the predictive variables will reduce the volume of data leading to higher and faster performance. The Pearson correlation method and correlation matrix were used to measure the relationship between 40 teleconnection indices and precipitation to determine and prioritize the most influential indices

Fig. 1 Geographical location of Iran's basins, sub-basins, and stations (the numbers of 1, 2, 3, 4, 5, and 6 are the main basins including the Caspian Sea, Persian Gulf-Oman Sea, Urmia Lake, Central Plateau, East Boundary, and Qara Qom, respectively.)



in the study area. Equation (1) explains the Pearson correlation coefficient (Salahi et al. 2017):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2)(\sum_{i=1}^n (y_i - \bar{y})^2)}} \quad (1)$$

where x_i denotes the independent (observation) and y_i is the dependent variables (simulation), and \bar{x} and \bar{y} are the mean of independent and dependent variables, respectively. The r value ranges from 1 to -1 representing perfect positive correlation to perfect negative correlation. Then, the percentage significant teleconnection indices (PST) were estimated based on p -value of ± 0.374 from the regression per basin and sub-basin. In this method, the datasets including the indices with a 1–6 months lag times and 40 indices with 240 variables were fed into the model by the correlation matrix and a stepwise procedure. This study aimed to investigate the effect of teleconnection indices on the seasonal precipitation of Iran basins and sub-basins. Considering the frequency and spatial distribution of the meteorological stations, an attempt was made to examine the correlations with time lags of 1 to 6 months (Helali et al. 2020, 2021c). Also, due to the importance of seasonal precipitation in the agricultural sector and forecasting crop yields, increasing the time lag by more than

6 months will associate with more uncertainty. That allowed us to define the significance of the relationship between predictive indices and the precipitation's occurrences. First, all 40 indices with 6-month time lags were used to determine seasonal large-scale teleconnection indices per sub-basin. Second, the 3 indices with the highest coefficient of determination (R^2) were fed into the models by a stepwise method (Table 4). Therefore, three variables with the highest coefficient of determination (R^2) were selected as the most influential indices for the seasonal precipitation (three seasons) of each sub-basin.

Simulation models

Machine learning (ML) methods are considered advanced and intelligent methods with a higher ability for non-linear data modeling and learning schemes for simulation (Ahmadi et al. 2020). The simulation model in ML techniques includes two phases of training and testing the model capability (Seyedzadeh et al. 2020). Therefore, four ML models including least squares support vector machine (LSSVM), two types of the artificial neural networks (i.e., the generalized regression neural network (GRNN), multi-layer perceptron (MLP)), and multilinear regression (MLR) were exploited to model seasonal precipitation per sub-basin.

Table 1 The Information and characteristics of main basins and sub-basins

Main basin name	Main basin abbreviation	Sub basin name	Sub basin abbreviation	Area (M ha)	Number of stations	Autumn Pr (mm)	Winter Pr (mm)	Spring Pr (mm)	Annual Pr (mm)	Elevation (m)
Basin 1-Caspian Sea	B1-CS	Aras	ARZ	4,099	32	72	59.5	125.6	300	1336.5
		Atrak	ATR	2,699	14	84.5	121.3	90.9	335.2	453.1
		Haraz-Sefidroud	HAS	1.1	14	452.8	246.3	164.7	1091.5	216.2
		Qarasu-Gorgan	QAG	1,325	19	151.9	165.7	123.4	521.2	282.4
		Qarasu-Haraz	QAH	1,884	30	234.9	178.8	120.9	659.3	601.4
		Sefidroud	SFR	5,961	34	144.4	138.6	137.8	457.7	1556.0
		Talesh Rivers	TLS	0,696	8	496.6	278.4	184.4	1253.6	75.0
		Bandarabas-Sedij	BDA	4,068	45	38.7	130.6	20.1	201.1	407.0
		Great Karoun	GKR	43.6	44	157.8	257.9	100	517.4	1154.1
		Helleh	HLL	1.98	13	109.9	189	35.8	336.6	446.8
Basin 2-Persian Gulf and Oman Sea	B2-PG	Jarahi-Zohreh	JAZ	3,842	36	138.7	219.3	59.8	419.4	656.2
		Kal-Mehran	KAM	5,757	30	37.2	112.2	26.6	191	733.8
		Karkheh	KRK	5,011	24	128	168.5	95.2	393.5	1040.4
		Mand	MND	4,387	15	95.2	214.4	51.4	366.7	1217.2
		South Balouchestan	SBL	4,379	7	32.8	73.1	19.8	141.4	334.01
		West Border	WSB	3,889	13	148.5	213.9	118.5	484	1183.1
		Urmia Lake	URL	5,272	58	96	99.7	116.3	325.6	1530.9
		Abarkouh-Sirjan	ABS	5,357	7	38	90.9	37.1	168.3	1897.4
		Central Desert	CTD	22,418	55	47.8	85.7	66.2	211.7	1452.5
		Daranjir Desert	DJD	4,767	10	28	86.4	35.2	152.2	2022.2
Basin 3-Urmia Lake	B3-UL	Gav Khouni	GKH	3,988	18	91.3	142.8	75.5	312.1	1878.2
		Hamoon-Jazmorian	HAJ	6,352	11	21.1	84.1	24.1	144.8	1034.5
		Lut Desert	LUD	19,499	11	18.5	71.1	27.9	118.6	1372.4
		Siahkough Desert	SKD	4,706	8	30.4	59.7	36.7	128.4	1745.7
		Salt Lake	STL	9,135	92	98.9	123.1	107	341.2	1662.5
		Tashk-Bakhtegan-Maharlou	TBM	2,944	23	97.4	216.9	67.9	385.1	1858.2
		Hamoon-Hirmand	HAH	3,189	5	11.6	53.6	16.6	82.3	1007.2
		Hamoon-Meshkil	HAM	3,343	6	14.3	54.9	17.7	95.2	1197.0
		Khuf Desert	KHD	3,208	6	32.2	103.1	41.6	177.4	1515.6
		Qaraqom	QRQ	4,421	29	45.6	105.1	88.4	243.7	1172.6

Pr stands for precipitation

Table 2 Lists of teleconnection indices for the proposed study area (Ahmadi et al. 2019; Helali et al. 2020)

Teleconnection index	Definition	Teleconnection index	Definition
1	AMO	21	QBO
2	AO	22	SCN
3	CSST	23	SFlux
4	EA	24	SOI
5	EAWR	25	SSPOT
6	IOBSST	26	SST1.2
7	MEI	27	SST3
8	MSST	28	SST3.4
9	NAO2	29	SST4
10	NCP	30	SSTas
11	Niño1.2	31	SSTs
12	Niño3	32	TNA
13	Niño3.4	33	TNA-TSA
14	Niño4	34	TNI
15	NPI	35	TNli
16	PDO	36	TPI
17	PNA	37	TSA
18	POI	38	WHWP
19	POL	39	WP
20	PSST	40	WPSST

Defining the most important predictive indices, 70% of the data were randomly divided as the training samples and the remains as testing samples, and the corresponding pixels of the predictive indices were fed into the four machine learning (ML) algorithms. For each model, three seasonal outputs resulting from 30 iterations were generated. MATLAB software and programming (version 9.0, R2016a) was used to implement the GRNN, MLP, LSSVM, and MLR models. The descriptions of the models are provided below.

Generalized regression neural network

In ML family algorithms, artificial neural networks (ANN) try to simulate the biological neural networks' structures or tasks similar to human brain information processing (Duan et al. 2013; Seyedzadeh et al. 2020). In ANN, the neurons as the processor elements include two factors, namely the weight and activation function. The input variables are weighted after being mapped to neurons, and the outcome is used as input to the activation function to create the final output. Equipped with the radial basis function (RBF) and probabilistic structure, GRNN is a type of neural network model that simulates the dependent variables in a regression function problem. Therefore, it does not experience the local minima problem, which is common in other neural networks (Cigizoglu 2005). GRNN, as a three-layer neural network,

has many neurons in input and output layers equivalent to input and output vector dimensions, respectively. In contrast with the artificial neural network (ANN) in the middle layer, the number of neurons is defined by observed data for model calibration and testing (Araghinejad 2014; Modaresi et al. 2018a). The applied function (in the middle layer of this neural network) assigns a normal (Gaussian) function as Eq. (2).

$$f(X_r, t) = e^{-[I(t)]^2}$$

$$I(t) = \|X_r - X_t\| \times \frac{0.8326}{h} \quad (2)$$

$$t = 1.2. \dots .n$$

where $\|X_r - X_t\|$ measures Euclidean distance between real-time vector (X_r) and the observed vector of predictors related to the t th neuron (X_t). The constant of 0.8326 applies an optimization for spread parameter and h denotes the spread parameter presenting the spread of a radial basis function and adjusting the function for the best fit. The spread value is commonly 1.0. The larger values result in the smoother function approximation and the smaller ones lead to closely fitness. However, the spread value should be defined by users in the range of $h > 0$.

The model's output (Y_r) (forecasted seasonal precipitation) for the vector of (X_r) is determined according to a

kernel function of the normal performance outputs [f (X_{r,t})] as Eq. (3) (Modaresi et al. 2018a):

$$Y_r = \frac{1}{\sum_{t=1}^n f(X_{r,t})} \sum_{t=1}^n [f(X_{r,t}) \times T_t] \tag{3}$$

Multilayer perceptron

MLP is a common type of ANN with feed-forward network class including at least three layers, namely input, hidden, and output layer (Gholami Rostam et al. 2020). Several neurons in the hidden layer are determined based on weight and bias using the minimum RMSE in order to optimize the model performance (Widiasari et al. 2018). The functions used in the middle and output layers of the neurons perform the linear and sigmoid functions, respectively, presented in Eqs. 4 and 5 (Araghinejad 2014; Modaresi et al. 2018a):

$$f(x) = x \tag{4}$$

$$f(x) = \frac{1}{1 + e^{-\alpha x}} \alpha > 0 \tag{5}$$

In each neuron, the weight (*w*) and bias (*b*) are calculated for inputs of the neurons as (*w_jx_j + b_j*) where *j* = 1, 2... *m*. The optimum values are determined by the model calibration. To train and calibrate the neural network, a Feed Forward Back Propagation (FFBP) algorithm (Araghinejad 2014) is applied until the best forecasts are achieved, where the error function in Eq. 6 is minimized for the iterations (epochs):

$$E = \frac{1}{nc} \sum_{i=1}^{nc} e_i^2 \tag{6}$$

where E denotes the error function; *e_i* calculates the error of the model simulation for *i*th training pair data and *nc* is the number of training pairs.

Least square support vector machine

As a supervised learning mechanism, the support vector machine (SVM) uses structural risk minimization notion to minimize the model error (Dibike et al. 2001; Ghosh et al. 2019), whereas other methods (e.g., ANN) apply the empirical risk minimization principles (Seyedzadeh et al. 2020; Modaresi and Araghinejad 2014). The least square support vector machine (LSSVM) approach (Suykens and Osipov 2008) exploits linear equations in simulation algorithm and leads to higher performance by using effective kernel function (Modaresi et al. 2018b; Seyedzadeh et al. 2020). In the LSSVM method, a nonlinear mapping of *φ* in the trait space for X_{*t*} ∈ R^{*m*} as the input data and Y(X_{*t*}) ∈ R as the output data is calculated as follows (Suykens et al. 2002):

$$Y(X_t) = w^T \cdot \phi(X_t) + b \tag{7}$$

where *w* and *b* denote the weights and biases values of the regression function, respectively, calculated by minimization of the following function:

$$\min_{w,b,e_i} j(w,e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{t=1}^n e_t^2$$

$$S.T : w^T \phi(X_t) + b + e^t \tag{8}$$

$$t = 1.2. \dots .n$$

where *e* is the error, and gamma (*γ*) are the regularization parameters in the model to control the flatness of approximation function, and the optimum values are determined by users.

Multilinear regression

The regression model reliability can decline if a few variables are chosen. In this work, in order to establish the forecast models for the seasonal observations, a regression equation using three independent variables was formulated (Kim et al. 2020). For this study, the datasets from the past 28 years were used to estimate the forecast models for each season. The regression equation with three independent variables was determined as follows:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \varepsilon \tag{9}$$

where *Y* is the response (monthly precipitation); *α₀*, *α₁*, *α₂*, and *α₃* determine the regression coefficients. In addition, *X₁*, *X₂*, and *X₃* are the most correlated teleconnection indices (predictors); and *ε* is the model’s residual.

Evaluation indices of models

To evaluate the accuracy of the simulation models and define the level of performances, the coefficient of determination (Behar et al. 2015; Houshyar et al. 2018), RMSE (Ma and Iqbal 1984; Willmott and Matsuura 2006; Heydari Tasheh Kabood et al. 2020), MBE (Ma and Iqbal 1984), and Scatter Index (SI) (Li et al. 2013; Seyedzadeh et al. 2020) were used by the following equations:

$$R^2 = \left[\frac{\sum_{i=1}^n (O_i - \bar{O})(M_i - \bar{M})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 (M_i - \bar{M})^2}} \right]^2 \tag{10}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - M_i)^2}{n}} \quad (11)$$

$$MBE = \frac{\sum_{i=1}^n (M_i - O_i)}{n} \quad (12)$$

$$SI = \frac{RMSE}{\bar{O}} \quad (13)$$

In the above equations, M is the modeled data, O is the observational, n denotes the time series number, and \bar{O} represents the mean. Based on the above statistics, the perfect models are defined by R^2 values of 1 and RMSE and MBE values of zero. The efficiency of SI statistics in modeling and forecasting is classified in below (Li et al. 2013):

$$\text{if } \left\{ \begin{array}{ll} SI < 0.1 & \text{Excellent} \\ 0.1 < SI < 0.2 & \text{Good} \\ 0.2 < SI < 0.3 & \text{Fair} \\ SI > 0.3 & \text{Poor} \end{array} \right\} \quad (14)$$

Result

Structural of machine learning models

The structural characteristics and schematic views of the GRNN, MLP, LSSVM, and MLR models are shown in Table 3 and Fig. 2. Due to different results of training and testing in implementation of intelligent models, the average output of 10 consecutive executions for each model was reported as the final result of the models. To obtain the best accuracy, training of the GRNN, MLP, LSSVM, and MLR models was stopped after 50,000, 50,000, 50,000, and 1000 iterations, respectively.

Optimal GRNN and MLP architectures were chosen through trial and error approach, which generally stopped after 1000 epochs. LSSVM gets the advantage of the applying equality constraints (in exchange for traditional inequality constraints of SVM) and implements the sum of squared regression errors in the training process. For the MLR model parameter selection, the iterative process continues until an acceptable level of error is obtained. The dataset was divided into the training set (70% of the dataset) used for developing the ANN model and estimating the model parameters, and the testing dataset (30% of the dataset) to evaluate the accuracy of the predictive variables.

Percentage of significant teleconnection indices

The analysis of the PST is presented in Fig. 3 and the significant levels of correlation (in terms of sub-basin and

season) are illustrated in Fig. 4. Considering the lag time of 1–6 months, the PST with autumn precipitation in basin 1 (CS) was between 17.5 and 35.0% (maximum amount in the time lag of 1–3 months). While in basin 2, PST were estimated between 12.5 and 45.0% (maximum values in 1–3 month time lag), basin 3 (UL) between 5.0 and 37.5% (maximum values in 1–3 month time lag), basin 4 (CP) from 10.0 to 42.5% (maximum values in 1–4-month time lag), basin 5 (EB) between 5.0 and 42.5% (maximum values in 1–4 months lag times), and finally in basin 6 (QQ), it was recorded between 5.0 and 32.5% (maximum values at 6-, 3-, 2-, and 1-month time lags). Therefore, in most of the main basins, the PST had the highest values with autumn precipitation in 1–4 month time lags (Fig. 3A). According to the results, MEI, NINOs, SSTs, POI, SOI, and TPI (teleconnection indices) exhibited the most significant correlation with autumn precipitation basins in the abovementioned time lags. In addition, MEI, NINOs, SST, and TPI indices had positive correlations, whereas POI, SOI, TNA, WPSST, and CSST showed negative correlations (Fig. 4A).

Likewise, the PST with winter precipitation (in time lags of 4–6 months) had the highest value in most of the basins (Fig. 3B). Therefore, the PST of winter precipitation was observed in some basins and was limited to NAO2, PDO, TNA-TSA, AO, CSST, NCP, and MSST indices (Fig. 3B). The correlation between winter precipitation and NAO2, NCP, PDO, and TNA-TSA was positive, while its correlation with CSST, MSST, TSA, WPSST, EA, AO, and POL was negative (Fig. 4B).

Moreover, the highest PST in spring precipitation occurred in 1–4-month time lags in most of the basins (Fig. 3C). Consequently, SST, NINOs, MEI, NPI, PNA, POI, SOI, TNI, TPI, and TNi indices showed significant correlation with some basins in the 1–3-month time lags. The correlation with MEI, NINOs, SSTs, PNA, TNA, TPI, and TNi was positive, while the negative correlation was observed with CSST, NPI, POI, SOI, TNI, and WPSST (Fig. 4C).

The Pearson correlation results highlighted that the teleconnection indices with higher significant changes in most of the basins became dependent on the season and time lags (Fig. 4). The autumn precipitation had the highest PST correlations in the 6- and 5-month lag time of TNA (57% of sub-basins), the 4- and 3-month time lags of SOI (50% and 73% of sub-basins), the 2-month time lag of NINO 3.4 (77% of sub-basins), and 1-month time lag of NINO 4 (70% of sub-basins). It was observed that the pattern and PST correlations of teleconnection indices with winter precipitation changed in order to obtain a significant correlation between PDO, NAO2, MSST, TSA, SOI, and POL indices and a higher number of basins (i.e., 40%, 30%, 60%, 47%, 20%, and 40% of the basins, respectively) in 6-, 5-, 4-, 3-, 2-, and 1-month time lags. An examination of spring precipitation and its relationship with

Table 3 Characteristic of machine learning models study in the training and testing phases

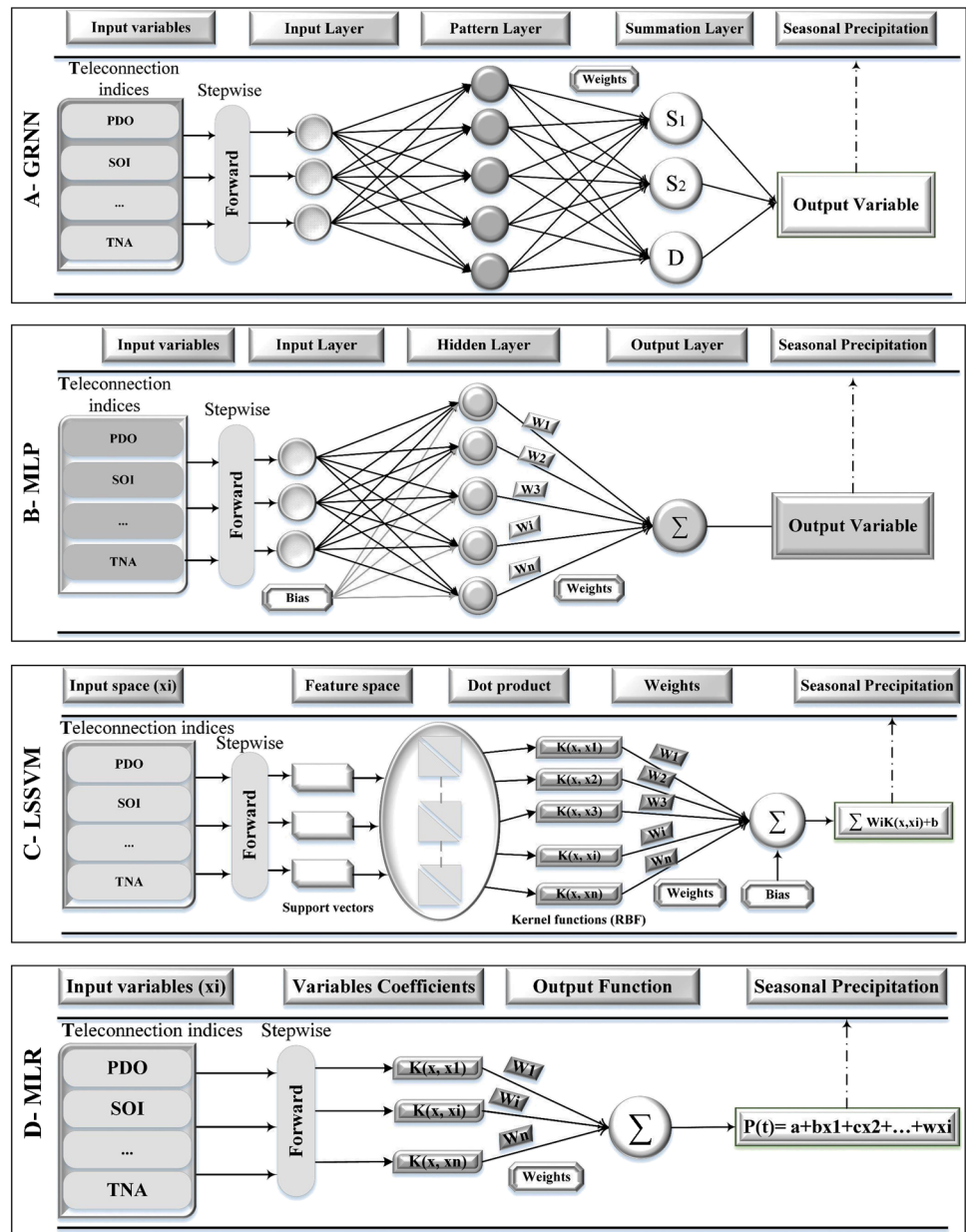
Generalized regression neural network (GRNN)	
Input variables	3 teleconnection indices
Function's nodes	Newgrnn
Neuron number in layer	5
Spread number of radial basis functions	2.0
Training output weights algorithm	Gradient decent
Data division function	Random data
Iteration number	50,000
Output variable	Seasonal precipitation
Multi-layer perceptron (MLP)	
Input variables	3 teleconnection indices
Hidden layer number	1
Neuron number in hidden layer	5
Activation function in hidden layer	Tangent sigmoid
Activation function in output layer	Pure line
Train function	Levenberg–Marquardt algorithm
Data division function	Dividerand
Iteration number	50,000
Output variable	Seasonal precipitation
Least squares support vector machine (LSSVM)	
Inputs variables	3 teleconnection indices
Function estimation	Gaussian
Kernel function	Radial basic function (RBF)
Tuning parameters (γ, σ^2)	$\gamma = 10$ and $\sigma^2 = 0.2$
Selection function	Randomize selection (Randper's Function)
Datasets ratio in train and test stages	70% and 30%
Iteration number	50,000
Target variable	Seasonal precipitation
Multilinear regression (MLR)	
Explanatory variable	3 teleconnection indices
Model selection method	Best model
Criterion	Mean square error (MSE)
Selection function	Randomize selection
Datasets ratio in training and testing stages	70% and 30%
Iteration number	1000
Dependent variable	Seasonal precipitation

teleconnection indices indicated that NINO 4 index had a significant correlation with higher percentage of basins in time lags of 6, 5, and 2 months (50% of sub-basins), SOI in 6, 4, and 3 months (50–53% of sub-basins), SST4 in 6, 5, and 2 months (50% of sub-basins), POI in 1 month (57% of sub-basins), and WPSST in 3 and 1 month (50% of sub-basins). Finally, the most significant independent predictors (in terms of time lag and teleconnection indices pattern) are calculated and listed in Table 4. Those indices were considered as the optimum inputs or main teleconnection variables for the different models reducing the volume of computation in seasonal precipitation simulation in Iran's sub-basins.

Accuracy assessment and evaluation of the models

The accuracy assessment results of the models in seasonal precipitation simulation (based on teleconnection indices) are summarized for autumn, winter, and spring precipitation in Tables 5, 6, and 7, respectively. In autumn precipitation, the changes of R2 values in CS, PG, UL, CP, EB, and QQ main basins were recorded between 0.47 and 0.94, 0.48 and 0.95, 0.74 and 0.90, 0.37 and 0.90, 0.45 and 0.95, and 0.35 and 0.84, respectively. Similarly, the RMSE ranged between 6.9 and 90.1 mm in CS, 11.1 and 58.1 mm in PG, 16.1 and 24.2 mm in UL, 5.3 and 62.5 mm in CP, 3.3 and 15.0 mm in EB, and 6.9 and 14.9 mm in QQ. Based on MBE statistics in CS, PG, UL, CP, EB, and QQ main basins, the MLP, LSSVM,

Fig. 2 A schematic of the flowchart used for the machine learning models



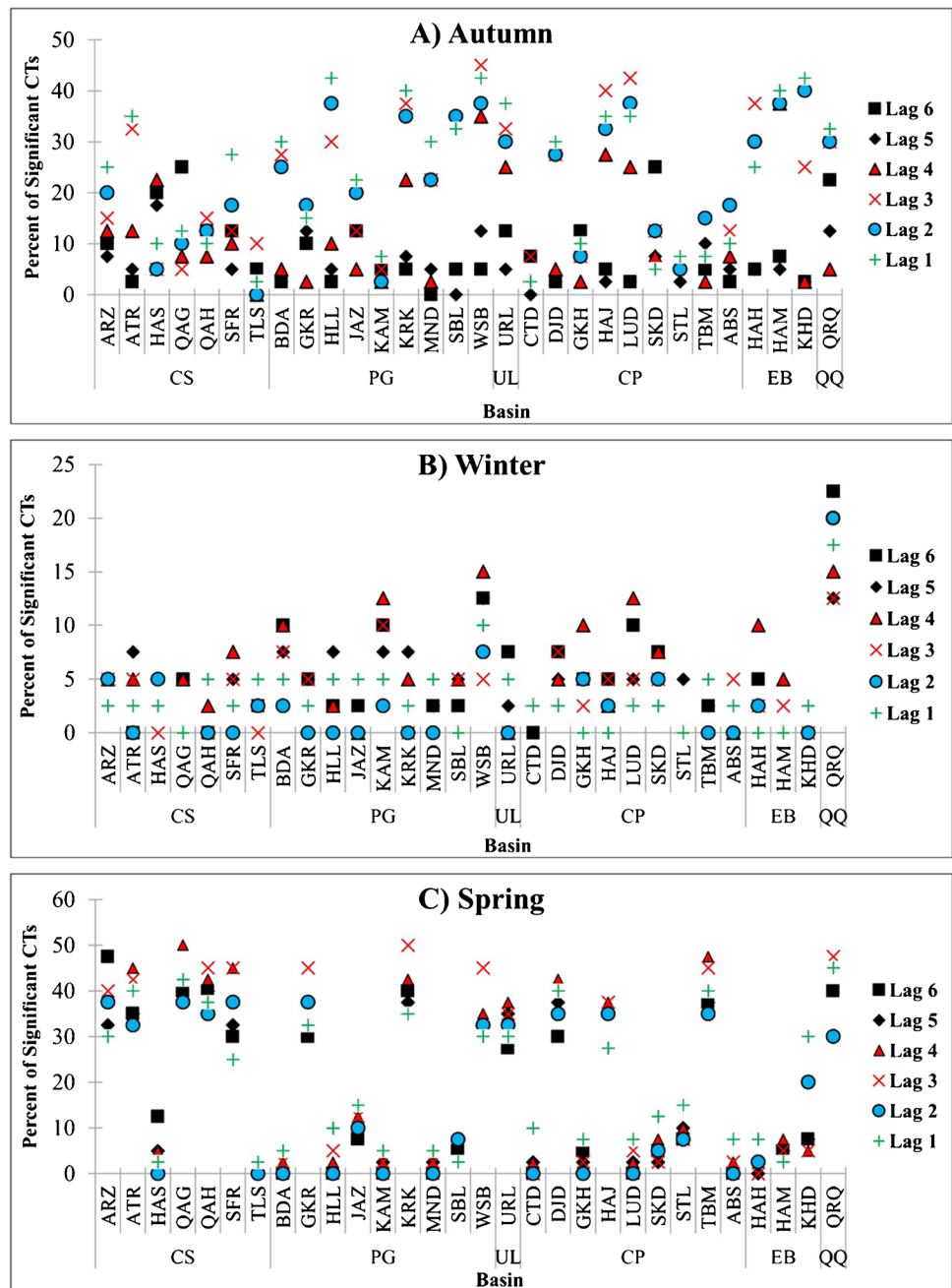
GRNN, GRNN, LSSVM, and MLR models had minimum values of MBE and there was no sign of overestimation or underestimation of the models in different sub-basins with specific teleconnection indices. A closer examination of the obtained results revealed that the best model based on R² and RMSE statistics in most of the sub-basins was the LSSVM model followed by the MLP. The rate of changes in the coefficient of determination in the LSSVM model was equal to 0.82–0.94, 0.86–0.95, 0.90, 0.82–0.90, 0.93–0.95, and 0.84 in CS, PG, UL, CP, EB, and QQ basins, respectively, and the corresponding RMSE values of 6.9–56.2, 11.1–28.2, 16.1, 5.3–29.6, 3.3–5.4, and 6.9 mm.

According to the MBE statistics for winter precipitation (Table 5), it was determined that the minimum errors in CS,

PG, UL, CP, EB, and QQ basins were obtained by GRNN, MLR, GRNN, MLR, LSSVM, and GRNN models, respectively. The analysis of the models based on R² and RMSE statistics confirmed the superiority of the LSSVM model in most of the basins and sub-basins followed by MLP model. Based on R², the best models were CS, PG, UL, CP, EB, and QQ.

The calculated values of R², RMSE, and MBE between the simulations and observations for spring precipitation using different models are presented in Table 6. Considering the MBE statistics, the lowest errors in CS, PG, UL, CP, EB, and QQ basins were acquired by MLP, MLP, MLR, LSSVM, MLR, and MLP models, respectively. The analysis of models' performances based on R² and RMSE statistics revealed that the LSSVM model (in most of the basins and

Fig. 3 The percentage of significant teleconnection (PST) between teleconnection indices and seasonal precipitation of Iran’s basins and sub-basins with delays of 1–6 months (lag time 1 to 6)



sub-basins) outperformed other models followed by the MLP model. Accordingly, the rate of change in the R^2 of the best models in CS, PG, UL, CP, EB, and QQ basins were between 0.83–0.92, 0.82–0.94, 0.89, 0.86–0.93, 0.85–0.93, and 0.84, respectively, with the corresponding RMSE values between 6.4–26.6, 19.2–45.3, 7.8, 10.7–35, 13.2–15.0, and 14.4 mm.

Efficiency evaluation of the models

The models’ efficiency based on the scatter index (SI) is shown in Fig. 5 for autumn (first column), winter (second column), and spring (third column). The SI illustrated that

the LSSVM model had a good performance in most of the sub-basins in autumn: in basin 1 (CS), the ranges of change in the SI in GRNN, MLP, LSSVM, and MLR models were 0.12–0.27, 0.10–0.18, 0.06–0.14, and 0.14–0.19, respectively (excellent to good); in basin 2 (PG), the LSSVM model had good to fair performance (excluding the KAM and SBL sub-basins); and other models had fair to poor performance. Also in basin 3 (UL), LSSVM model showed good performance and other models ranked as fair. In basin 4 (CP), LSSVM model had good performance in CTD and STL sub-basins (0.19 and 0.16) and fair performance in ABS, GKH, LUD, and SKD sub-basins (with

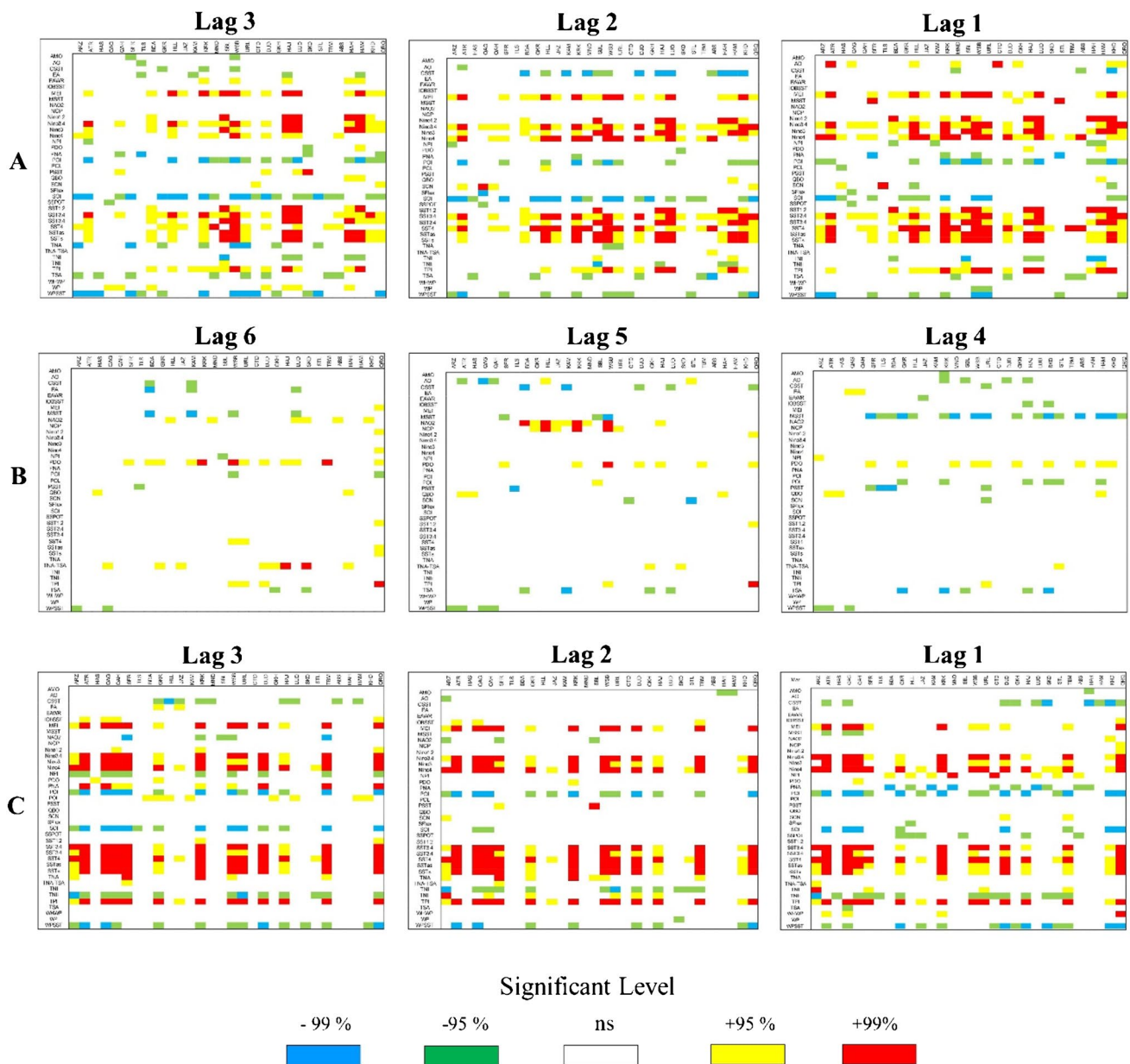


Fig. 4 The diagrams of (positive/negative) significant correlations (Pearson correlation) of indices affecting seasonal precipitation; **A**, **B**, and **C** stand for autumn, winter, and spring, respectively, in sub-basins in one- to 3-month time lags of autumn and spring precipita-

tion (lag 3 to lag 1) and in 4- to 6-month time lags of winter precipitation (lag 6 to lag 4). Black dotted line indicates the most important teleconnection indices

the values of 0.25, 0.22, 0.28, and 0.25, respectively), and MLP model had fair performance in some limited sub-basins. In basin 5 (EB), only LSSVM model performed well in KHD sub-basin (0.17), whereas it had fair performance (0.29) in HAH. The MLR model had fair performance in KHD sub-basin (0.29). In basin 6 (QQ), the MLP, LSSVM, and MLR models showed fair to good performance (0.23, 0.15, and 0.16, respectively). According to these results, the efficiency of all 4 models in modeling and predicting autumn precipitation was good only in the

CS basin, while in other sub-basins, the LSSVM model showed good and MLP appeared fair (Fig. 5, first column).

The analysis of winter precipitation forecast based on the SI showed excellent to good efficiencies using LSSVM and MLP models (Fig. 4, second column). In general, the efficiency of teleconnection indices in winter precipitation simulation was good by using LSSVM and MLP models, while it was poor by using GRNN and MLR models (Fig. 5, second column).

Table 4 The most significant teleconnection indices for seasonal precipitation at main and sub-basin of Iran

Main basin	Sub basin	Autumn			Winter			Spring		
		Var 1			Var 2			Var 3		
		Var 1	Var 2	Var 3	Var 1	Var 2	Var 3	Var 1	Var 2	Var 3
CS	ARZ	EAWR-May	TNA-Jun	WPSST-Jul	AO-Oct	SOI-Sep	WPSST-Dec	Nino4-Oct	SST4-Nov	WP-Mar
	ATR	MSST-Jun	Nino3-May	SOI-Aug	EAWR-Aug	SOI-Jul	WPSST-Aug	NAO2-Jan	NCP-Dec	POI-Mar
	HAS	POL-May	SCN-Sep	WPSST-Jul	PDO-Oct	QBO-Aug	TNA-TSA-Dec	MSST-Mar	SST4-Nov	WPSST-Feb
	QAG	CSST-May	MSST-Jun	SCN-Aug	AO-Aug	EAWR-Aug	WHWP-Aug	Nino3.4-Feb	SST3.4-Feb	NCP-Dec
	QAH	PSST-Jun	SSPOT-Jun	WHWP-May	SOI-Aug	TNA-TSA-Dec	WPSST-Aug	NCP-Dec	TPI-Mar	WPSST-Mar
	SFR	EAWR-Sep	MSST-Sep	SOI-Jul	MSST-Sep	POL-Dec	WP-Dec	SST4-Nov	NAO2-Jan	POL-Dec
	TLS	AO-Jul	PNA-Jul	SCN-Sep	NCP-Aug	PSST-Sep	QBO-Jul	SFlux-Oct	SST4-Nov	WPSST-Feb
	BDA	CSST-Aug	SCN-May	WP-Jun	EAWR-Oct	MSST-Sep	TSA-Oct	CSST-Dec	NCP-Oct	PNA-Mar
	GKR	EA-Aug	SOI-Aug	TNA-May	NAO2-Aug	TSA-Oct	WHWP-Jul	EA-Nov	NCP-Oct	POL-Jan
	HLL	SCN-May	SST3.4-Aug	WP-Aug	NCP-Aug	SCN-Aug	TNA-TSA-Dec	CSST-Dec	EA-Nov	PNA-Mar
	JAZ	EA-Aug	SOI-Aug	TNA-May	NAO2-Aug	TSA-Oct	WHWP-Jul	CSST-Jan	CSST-Nov	TSA-Jan
	KAM	AO-Sep	CSST-Aug	NPI-May	EAWR-Oct	IOBSST-Nov	TSA-Oct	CSST-Dec	NCP-Nov	PNA-Mar
PG	KRK	MSST-Sep	POI-Sep	SOI-Jul	MSST-Sep	NCP-Aug	TNA-TSA-Jul	EA-Nov	CSST-Jan	TNA-Jan
	MND	Nino4-Sep	POL-Jun	SCN-Sep	EAWR-Oct	IOBSST-Nov	TSA-Oct	PNA-Mar	TPI-Oct	EA-Mar
	SBL	Nino1.2-Sep	PDO-Sep	SCN-Jun	MSST-Sep	SSPOT-Jul	WP-Sep	PSST-Dec	PSST-Feb	PSST-Oct
	WSB	EAWR-May	SST3.4-Sep	TNA-Jun	MSST-Sep	NCP-Aug	POL-Dec	AO-Dec	EA-Nov	POL-Dec
	URL	NCP-Aug	SOI-Aug	TNA-May	NCP-Aug	SFlux-Nov	SFlux-Sep	POL-Dec	TPI-Mar	WHWP-Jan
	CTD	AO-Sep	PDO-Jul	SOI-Jul	EAWR-Oct	IOBSST-Nov	TNA-TSA-Oct	CSST-Dec	NPI-Mar	SOI-Dec
	DJD	SOI-Jul	AO-Aug	TNA-TSA-Aug	MSST-Sep	POL-Dec	TNA-TSA-Jul	Nino4-Mar	POL-Dec	SOI-Feb
	GKH	MSST-Sep	SCN-Sep	TNA-May	EAWR-Oct	IOBSST-Nov	TSA-Oct	PNA-Mar	PNA-Oct	SOI-Jan
	HAJ	CSST-Aug	SCN-May	WP-Jun	NAO2-Aug	PDO-Aug	TNA-TSA-Oct	CSST-Jan	NPI-Mar	SST4-Feb
	LUD	AO-Aug	SST3.4-Aug	SST4-Jun	EAWR-Oct	MSST-Sep	TSA-Oct	CSST-Dec	NCP-Oct	PNA-Mar
	SKD	AO-Sep	PNA-Jun	PSST-Jul	MSST-Sep	NAO2-Jul	TNA-TSA-Oct	Nino1.2-Mar	NPI-Jan	POL-Oct
	STL	AMO-Jun	MSST-Sep	SOI-Jul	PNA-Aug	SCN-Aug	TNA-TSA-Dec	EA-Mar	PNA-Mar	SST4-Oct
EB	TBM	SCN-Sep	TNA-May	TNA-TSA-Aug	MSST-Dec	MSST-Sep	TNA-TSA-Jul	NAO2-Jan	POL-Dec	SST4-Nov
	ABS	Nino3.4-May	SST4-Sep	TSA-Sep	NAO2-Jul	POL-Dec	SCN-Aug	CSST-Dec	EAWR-Jan	NPI-Mar
	HAH	CSST-Aug	PNA-Jul	WP-Jun	MSST-Sep	NCP-Sep	TNA-TSA-Jul	CSST-Mar	NCP-Feb	PDO-Mar
	HAM	CSST-Aug	Nino1.2-Sep	SCN-Jun	MSST-Sep	NCP-Sep	TNA-Aug	CSST-Dec	CSST-Nov	PNA-Dec
	KHD	CSST-Sep	SST3.4-Aug	SST4-Jun	MSST-Sep	NCP-Sep	TNA-TSA-Dec	CSST-Mar	POL-Jan	WP-Mar
	QRQ	SCN-May	WPSST-Jul	WPSST-May	MSST-Sep	EAWR-Nov	NCP-Oct	POI-Mar	QBO-Feb	QBO-Mar

Table 5 Comparison of statistical evaluation indices of the models (R^2 , RMSE, and MBE of autumn precipitation using four models in basins and sub-basins of Iran)

Main basin	Sub basin	R^2				RMSE				MBE				
		GRNN	MLP	LSSVM	MLR	GRNN	MLP	LSSVM	MLR	GRNN	MLP	LSSVM	MLR	
CS	ARZ	0.63	0.83	0.91	0.78	19.5	11.2	7.7	11.2	2.08	-5.33	1.03	0.61	
	ATR	0.82	0.83	0.92	0.69	10.5	9.7	6.9	13.5	1.29	-0.68	0.07	-0.60	
	HAS	0.47	0.62	0.83	0.55	90.1	76.0	56.2	80.1	3.69	-10.5	22.27	-2.53	
	QAG	0.52	0.61	0.82	0.66	30.7	26.7	19.0	22.3	-3.83	-0.37	5.38	-3.57	
	QAH	0.72	0.76	0.90	0.59	30.5	27.9	17.9	34.6	-6.88	0.39	1.19	6.12	
	SFR	0.77	0.86	0.87	0.71	24.9	20.8	19.8	28.2	-0.48	-5.85	-3.38	1.19	
	TLS	0.61	0.82	0.94	0.56	80.6	52.1	31.1	81.6	9.64	-3.03	-1.16	-7.45	
PG	BDA	0.81	0.89	0.86	0.75	23.9	14.0	24.1	19.8	-0.69	2.58	4.43	-3.83	
	GKR	0.75	0.86	0.92	0.64	47.0	44.3	28.2	54.1	6.44	8.92	-1.41	6.48	
	HLL	0.72	0.87	0.88	0.60	44.1	27.0	27.0	46.6	-5.96	3.56	-1.11	-1.57	
	JAZ	0.72	0.86	0.95	0.59	50.0	34.5	22.8	58.1	-12.38	3.51	-2.47	4.14	
	KAM	0.67	0.87	0.91	0.53	25.3	14.0	11.3	24.3	-1.73	-5.13	3.54	0.63	
	KRK	0.85	0.75	0.91	0.48	27.2	36.0	21.8	39.6	-4.87	-7.49	3.50	4.08	
	MND	0.61	0.79	0.92	0.57	50.5	33.4	20.3	46.8	-0.58	8.87	-0.68	6.49	
UL	SBL	0.87	0.90	0.95	0.79	17.2	14.9	11.1	21.3	-1.27	-0.70	2.65	-1.18	
	WSB	0.67	0.83	0.92	0.71	49.5	31.9	23.0	41.9	2.58	4.69	3.45	1.39	
	URL	0.76	0.75	0.90	0.74	22.6	23.5	16.1	24.2	-1.29	3.91	-4.82	-2.67	
	CP	ABS	0.37	0.68	0.90	0.73	24.5	16.7	9.5	15.5	2.98	0.81	1.49	-0.95
		CTD	0.76	0.63	0.83	0.64	11.1	13.9	9.0	13.2	-0.83	-3.37	-0.55	0.18
		DJD	0.72	0.80	0.83	0.56	11.0	9.2	8.7	13.3	-0.26	1.21	0.94	-2.69
		GKH	0.53	0.79	0.82	0.56	38.1	21.1	20.6	29.7	4.68	-0.62	5.52	-1.03
HAJ		0.83	0.89	0.90	0.79	11.4	7.6	7.1	9.9	-0.82	1.58	2.20	-1.73	
LUD		0.48	0.70	0.86	0.65	10.9	7.5	5.3	8.2	0.38	-0.84	0.05	-0.59	
SKD	0.56	0.87	0.88	0.74	15.8	7.2	7.6	10.3	-2.06	0.26	-2.20	0.17		
EB	STL	0.76	0.82	0.89	0.75	22.6	20.0	15.6	23.0	-0.44	-4.31	2.68	-1.22	
	TBM	0.44	0.57	0.86	0.51	62.5	49.1	29.6	50.5	0.09	11.29	10.12	-1.44	
	HAH	0.76	0.84	0.94	0.72	7.9	4.9	3.3	6.7	0.03	-0.17	0.83	-0.56	
	HAM	0.87	0.92	0.95	0.76	7.9	5.7	4.8	10.1	0.90	0.26	-0.07	-0.91	
	KHD	0.45	0.79	0.93	0.78	15.0	9.9	5.4	9.2	-0.31	-0.25	-1.65	-1.84	
	QQ	QRQ	0.35	0.66	0.84	0.76	14.9	10.2	6.9	7.9	-0.49	1.43	-0.70	0.09

* The best values for each sub-basin are bolded based on different indices

The results of different models efficiency in spring precipitation simulation are presented in Fig. 5 (third column). In basin 1 (CS), the rates of change of SI in GRNN, MLP, LSSVM, and MLR models were excellent to good; in basin 2 (PG), the models had poor performance in most of the sub-basins (the performance of all 4 models was good to fair except in GKR, KRK, and WSB sub-basins); in basin 3 (UL), MLP and LSSVM models had good performance and GRNN and MLR models had fair performance; in basin 4 (CP), LSSVM and MLP models had good to fair performance and GRNN and MLR models exhibited poor performance; in basin 5 (EB), the LSSVM model had fair performance (in two sub-basins of HAH and KHD) and other models had poor performance; in basin 6 (QQ), LSSVM and MLP models had good performance and MLR and

GRNN models were fair. A general review of the results shows that the simulation performance of the 4 models was good to fair only in CS basin, while in other basins and sub-basins, only LSSVM and MLP models showed fair and rarely good performance (Fig. 5, third column).

Comparison of training and testing periods

An analysis of the mean and time series of seasonal precipitation modeled (simulations) and observed (observations) in the training and testing phases in sub-basins is plotted in Fig. 6 and Fig. 7. Clearly, there was no difference between the seasonal precipitation simulations and observations in the training phase, but in the testing phase, a slight difference was observed. The precipitation in the selected sub-basins, namely

Table 6 Comparison of statistical evaluation indices of the models (R^2 , RMSE, and MBE of winter precipitation using four models in basins and sub-basins of Iran)

Main basin	Sub basin	R^2				RMSE				MBE				
		GRNN	MLP	LSSVM	MLR	GRNN	MLP	LSSVM	MLR	GRNN	MLP	LSSVM	MLR	
CS	ARZ	0.52	0.96	0.79	0.66	26.6	7.9	16.4	20.8	-2.17	-2.24	0.02	0.27	
	ATR	0.83	0.81	0.88	0.70	18.0	19.0	15.7	22.9	-1.66	-2.93	5.53	2.13	
	HAS	0.34	0.76	0.76	0.62	44.0	26.3	25.8	31.3	0.31	8.61	4.50	-1.92	
	QAG	0.63	0.84	0.95	0.60	29.2	21.2	12.0	25.5	5.39	7.91	2.73	-1.53	
	QAH	0.58	0.76	0.89	0.62	30.2	21.5	15.0	27.7	-1.91	-2.51	3.17	1.18	
	SFR	0.67	0.68	0.86	0.62	29.6	28.2	19.3	32.4	0.15	3.08	5.02	1.22	
	TLS	0.73	0.79	0.86	0.78	31.2	28.5	24.1	30.2	-0.99	-1.88	4.43	0.45	
PG	BDA	0.80	0.91	0.92	0.65	15.1	6.9	6.9	13.4	-0.75	0.15	1.95	-0.71	
	GKR	0.79	0.78	0.89	0.71	27.1	22.7	16.4	27.8	-1.40	-2.55	1.16	1.62	
	HLL	0.74	0.77	0.85	0.65	18.9	13.9	11.6	17.2	-0.36	-2.12	3.55	0.93	
	JAZ	0.48	0.74	0.79	0.49	34.3	22.0	19.6	31.2	-1.61	-5.23	5.84	-4.58	
	KAM	0.74	0.74	0.82	0.65	15.6	13.4	11.3	15.4	0.44	1.50	1.98	-0.29	
	KRK	0.67	0.88	0.88	0.68	27.4	15.2	15.7	24.3	0.92	0.32	-2.57	-3.62	
	MND	0.73	0.82	0.85	0.68	20.6	14.5	13.5	19.1	-1.33	2.82	-3.51	-1.22	
UL	SBL	0.65	0.76	0.80	0.60	15.8	11.9	10.7	15.0	0.94	3.01	-0.54	-0.06	
	WSB	0.83	0.83	0.87	0.64	22.3	22.0	19.6	31.5	-0.75	1.18	0.86	0.64	
	URL	0.55	0.82	0.73	0.62	32.9	19.1	22.9	28.1	0.08	-1.44	0.11	-2.44	
	CP	ABS	0.80	0.83	0.88	0.64	11.2	11.7	9.1	14.9	1.90	-1.70	2.31	-0.72
		CTD	0.71	0.67	0.88	0.60	15.6	16.5	10.9	18.4	-0.39	0.68	2.98	2.89
		DJD	0.36	0.85	0.91	0.62	19.3	9.6	4.4	11.0	-3.39	-0.95	0.06	0.19
		GKH	0.68	0.73	0.90	0.80	18.9	17.3	10.5	15.9	-2.97	1.10	-0.83	-3.05
HAJ		0.70	0.85	0.81	0.74	13.8	7.5	8.8	10.2	-1.93	0.46	1.53	0.42	
LUD		0.64	0.72	0.82	0.64	9.9	8.5	6.9	9.4	-1.81	-0.89	1.45	-0.34	
EB	SKD	0.56	0.84	0.84	0.60	15.3	8.7	8.7	13.5	-0.30	0.53	-0.18	0.42	
	STL	0.68	0.75	0.87	0.67	28.7	22.2	17.0	26.5	5.62	1.28	1.74	0.21	
	TBM	0.67	0.88	0.90	0.71	25.9	16	14.5	25.5	-2.49	4.98	3.20	-1.54	
	QQ	HAH	0.75	0.66	0.88	0.53	8.7	7.7	4.5	8.4	-1.13	-0.22	-0.15	-2.24
		HAM	0.62	0.78	0.86	0.71	10.3	7.4	5.8	8.6	-1.19	0.90	-0.14	0.34
		KHD	0.61	0.69	0.83	0.62	15.1	14.1	9.8	14.7	1.00	-4.97	1.07	-2.29
	QRQ	0.78	0.82	0.87	0.75	20.3	17.1	14.3	18.5	0.09	0.70	-0.34	-1.53	

* The best values for each sub-basin are bolded based on different indices

CS, PG, UL, CP, EB, and QQ, reached the values of 78.0, 12.3, 16.9, 19.3, 2.9, and 2.5 mm in autumn, respectively, the values of 4.5, 26.5, 8.1, 5.1, 4.3, and 9.3 mm in winter, and 15.8, 9.0, 0.4, 2.9, 0.5, and the values of 1.2 mm in spring. Therefore, in the selected sub-basins of each basin, modeling and precipitation forecasting were based on teleconnection indices and the models with minor errors were used in the training and testing phase. According to the previous section and the SI, the use of intelligent models (i.e., ML), especially LSSVM and MLP, might lead to higher efficiency in seasonal precipitation simulation in the main basins and sub-basin scale.

Examination of the time series results of the observed and modeled data in selected sub-basins showed an appropriate distribution of all machine learning models around the bisector axis. The distribution of the observed and modeled data

in the training phase in the LSSVM and MLP were better than the GRNN and MLR models, while in the test phase, the distribution of the MLP and GRNN were more robust (Fig. 7).

The results showed that the RMSE value of the LSSVM model had the lowest value in the autumn season in the training phase, but in the testing phase, the RMSE values of the MLP and GRN models were less than the MLR and LSSVM models. Also, the amount of RMSE values of the CTD, LUD, DJD, HAJ, SKD, HAM, HAH, KHD, and QRQ sub-basins did not substantially differ from each other (Fig. 8). The RMSE in winter rainfall in the training phase in the LSSVM and MLP models were less than the GRNN and MLR models, and also the difference between all models in the ARZ and ATR sub-basins had the lowest value.

Table 7 Comparison of statistical evaluation indices of the models (R^2 , RMSE, and MBE of spring precipitation using four models in basins and sub-basins of Iran)

Main basin	Sub basin	R^2				RMSE				MBE				
		GRNN	MLP	LSSVM	MLR	GRNN	MLP	LSSVM	MLR	GRNN	MLP	LSSVM	MLR	
CS	ARZ	0.66	0.56	0.80	0.49	8.8	10.5	6.4	10.1	0.51	-0.06	0.25	0.69	
	ATR	0.69	0.84	0.84	0.63	17.3	11.8	12.1	19.1	5.41	1.56	-2.50	-1.74	
	HAS	0.72	0.84	0.85	0.75	25.7	19.1	17.8	25.2	3.30	-4.51	1.29	-2.10	
	QAG	0.65	0.75	0.85	0.43	27.6	21.5	17.0	32.3	-2.27	-0.83	-3.62	-1.99	
	QAH	0.64	0.89	0.88	0.63	28.1	15.3	16.2	28.3	-4.51	0.11	3.18	-1.23	
	SFR	0.51	0.77	0.92	0.61	25.8	15.7	9.5	19.9	-5.03	3.94	-1.14	0.36	
	TLS	0.45	0.79	0.83	0.62	49.6	29.5	26.6	36.0	-5.39	-2.89	-1.83	3.15	
PG	BDA	0.57	0.83	0.90	0.56	76.7	40.6	30.7	55.9	2.25	2.15	-1.20	1.76	
	GKR	0.59	0.81	0.88	0.57	70.9	45.8	39.4	68.9	1.18	-0.64	11.59	-1.70	
	HLL	0.57	0.80	0.83	0.57	68.6	42.8	40.5	62.0	11.98	6.48	5.47	16.88	
	JAZ	0.68	0.74	0.82	0.70	62.4	54.3	45.3	60.9	7.63	1.46	7.96	-14.93	
	KAM	0.52	0.87	0.94	0.47	78.0	33.6	25.4	36.8	-5.05	1.22	9.70	2.02	
	KRK	0.54	0.76	0.82	0.78	39.8	27.9	23.9	27.5	-3.05	-7.48	6.19	2.59	
	MND	0.35	0.64	0.86	0.30	93.0	63.0	39.3	72.5	6.30	8.73	2.89	-3.56	
UL	SBL	0.83	0.83	0.88	0.71	24.8	21.4	19.2	29.8	11.02	-0.30	4.13	-0.68	
	WSB	0.61	0.83	0.88	0.66	35.9	22.1	19.5	31.1	4.15	3.93	1.83	4.03	
	URL	0.86	0.63	0.89	0.42	8.8	14.0	7.8	17.2	1.24	2.25	1.17	1.11	
	CP	ABS	0.28	0.66	0.91	0.63	42.4	27.0	15.0	28.7	0.65	0.69	3.96	-4.23
		CTD	0.55	0.73	0.88	0.57	24.9	15.7	10.7	20.3	1.85	1.58	1.19	-3.20
		DJD	0.45	0.74	0.91	0.63	36.6	21.3	13.6	24.8	0.78	-4.42	2.58	-1.31
		GKH	0.50	0.89	0.93	0.41	44.9	20.7	18.0	47.4	3.25	-4.79	-1.46	-5.17
HAJ		0.57	0.82	0.86	0.40	44.9	23.2	20.0	34.7	6.04	-4.41	2.61	1.59	
LUD		0.48	0.87	0.90	0.35	30.9	14.9	14.0	29.7	3.76	0.70	5.75	-0.93	
SKD		0.62	0.76	0.87	0.57	20.0	14.9	11.7	20.8	1.13	-0.56	3.40	0.80	
EB	STL	0.54	0.81	0.91	0.60	33.7	18.0	12.4	26.2	3.42	2.81	-0.51	5.31	
	TBM	0.75	0.82	0.89	0.58	57.0	41.9	35.0	56.0	0.69	6.24	-10.14	-2.96	
	HAH	0.63	0.81	0.85	0.55	32.4	17.2	15.0	25.7	5.90	-0.23	1.23	-0.64	
	HAM	0.55	0.89	0.85	0.62	32.0	13.0	15.5	23.8	2.36	-0.78	-0.72	-0.36	
	KHD	0.36	0.75	0.93	0.46	40.8	23.6	13.2	33.8	1.02	-1.95	4.83	0.41	
	QQ	0.62	0.77	0.84	0.60	24.5	16.3	14.4	20.4	-0.36	0.33	2.67	-1.66	

* The best values for each sub-basin are bolded based on different indices

However, in the testing phase, the RMSE values of the MLP and GRNN models had the lowest value. Therefore, it can be said the MLP and GRNN models had better accuracy than LSSVM and MLR models (Fig. 8). The results in the spring showed that the RMSE in the LSSVM model had the lowest value in the training phase in most sub-basins, but in the testing, the RMSE of the MLP and GRNN were lower and had better performance (Fig. 8).

Discussion

Seasonal precipitation simulation is one of the challenging subjects for climatologists and water resource specialists to provide efficient and practical solutions, forecasting,

and modeling. In this regard, using dynamic models is one of the answers to solve this problem. Due to a large amount of data, the complexity of the components of the climate system, and the short period of the simulation, statistical models are suggested (Doblas-Reyes et al. 2013; Rodwell and Palmer 2014; Qian and Xu 2020; Kim et al. 2020). Teleconnection indices as influential climate factors affect different parts of the world. Therefore, their correlations with climate variables and exploiting them as predictors were also confirmed (Ruigar and Golian 2015; Xiao 2019; Zhu et al. 2020). As mentioned before in similar studies, the relationship between teleconnection indices and precipitation was a case of point relationship using limited or linear models for forecasting. However, the teleconnection indices occur on a large scale and their

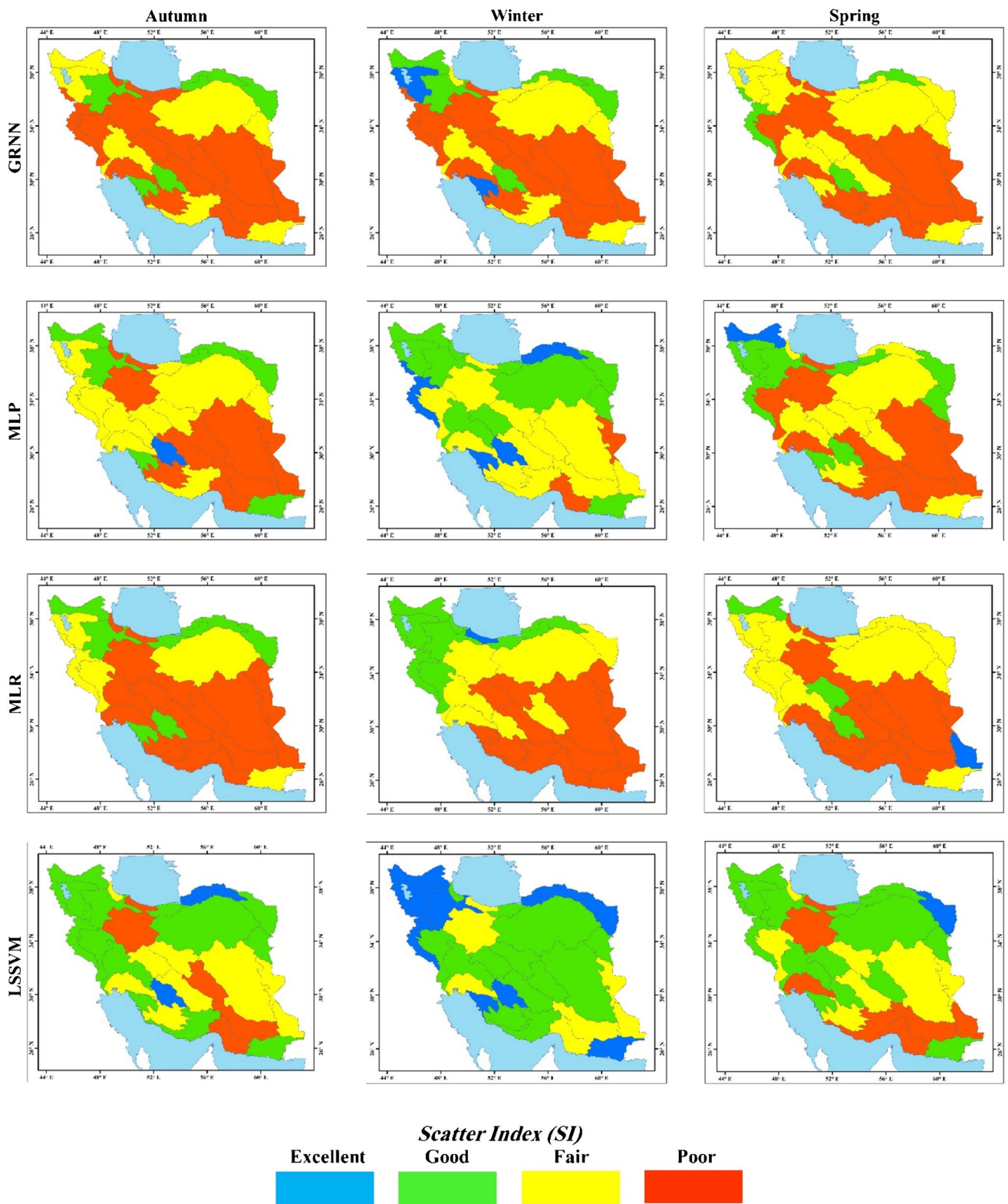


Fig. 5 Qualitative values of the scatter index (SI) in different machine learning models in autumn, winter and spring precipitation in all basins

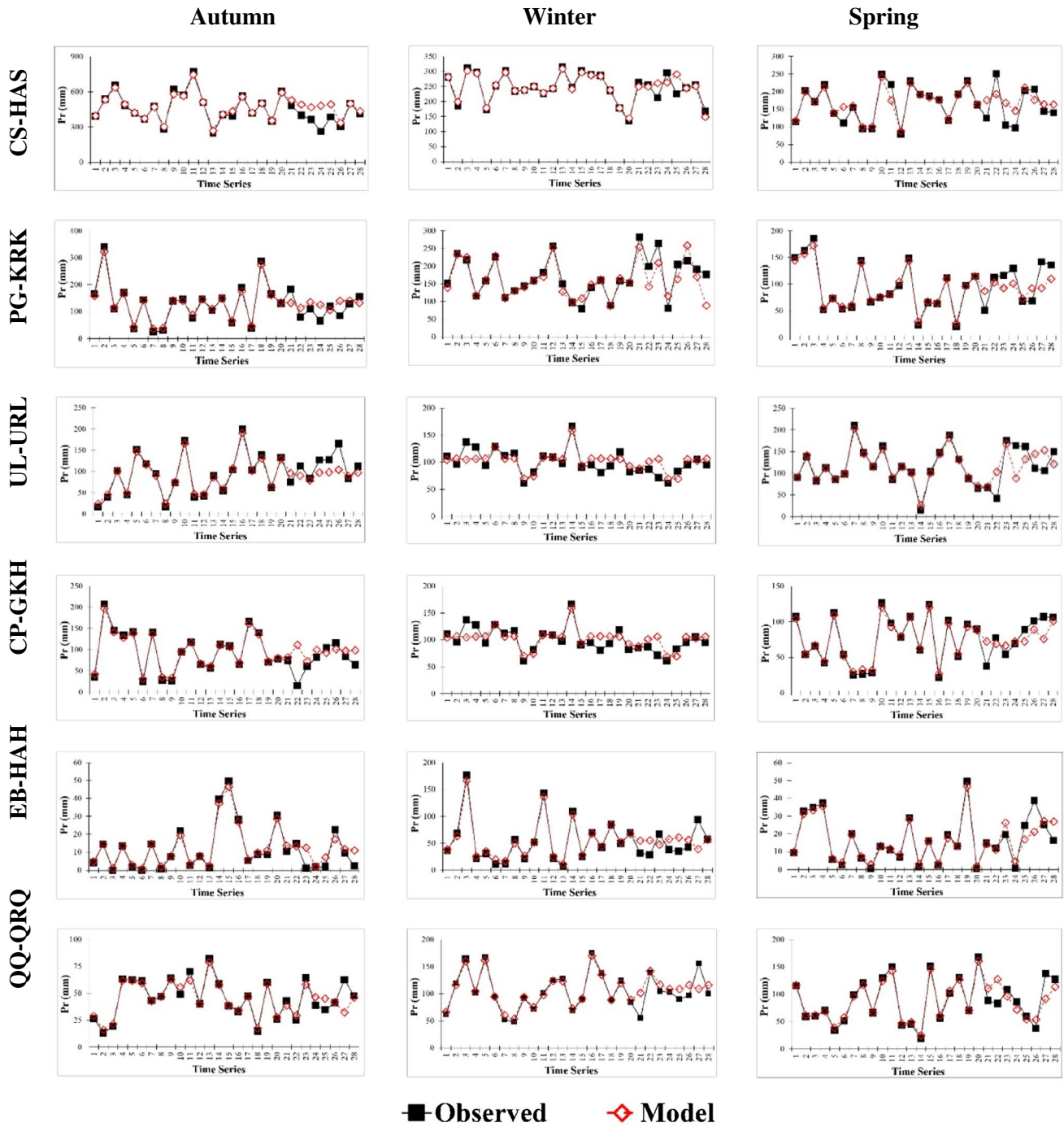


Fig. 6 Comparison between the simulations and observations in the training and testing phases in selected sub-basins of each basin (series 1 to 20 of training and series 21 to 28 of test phase)

correlation and impacts on precipitation should be correspondingly examined with the proper spatial scale of the teleconnection indices patterns (Goudarzi et al. 2017). To fill the existing gaps, the correlation between teleconnection indices and seasonal precipitation was expanded from a station or point perspective to the basin perspective and then the entire basins and sub-basins of Iran. In addition,

an attempt was made to evaluate the different predictive models' performance.

Preliminary results of the study evaluated that the highest PST was between teleconnection indices and autumn (1–4 months lag times), winter (4–6 months lag times), and spring (1–4 months lag times). In other words, seasonal precipitation of Iran's main and sub-basins were predicted

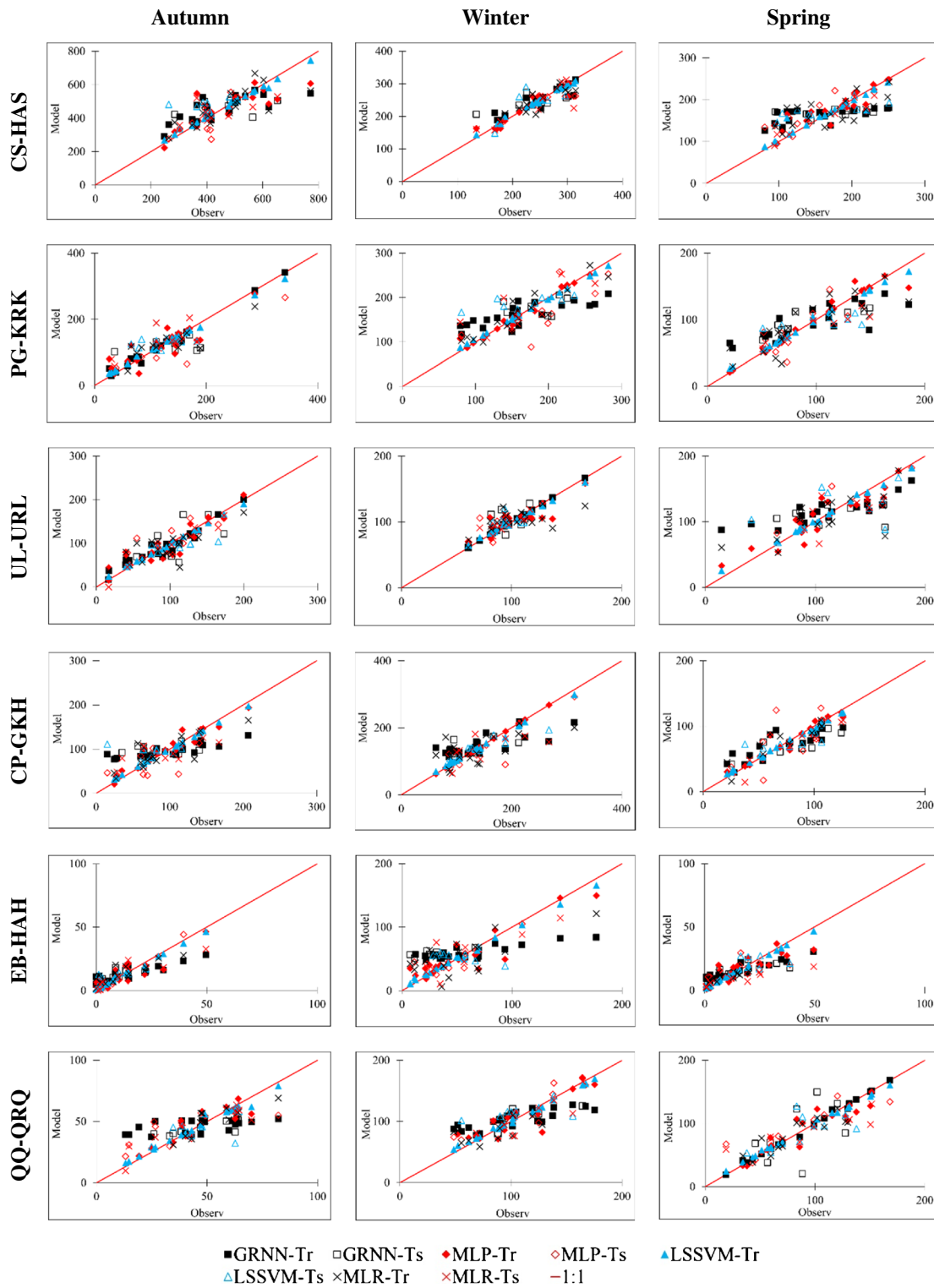


Fig. 7 Modeled and observation data in the training and testing periods in selected sub-basins on a bisector diagram

by 1–4 months lag times before the occurrence, which was in agreement with the results of Helali et al. (2020) study analyzing the correlation of autumn precipitation in basin level in Iran. The significance and pattern of indices also

varied based on the sub-basin and time lags, which was not in agreement with the results of previous studies. The current study was conducted in the sub-basin level showing the highest PST correlation between the following:

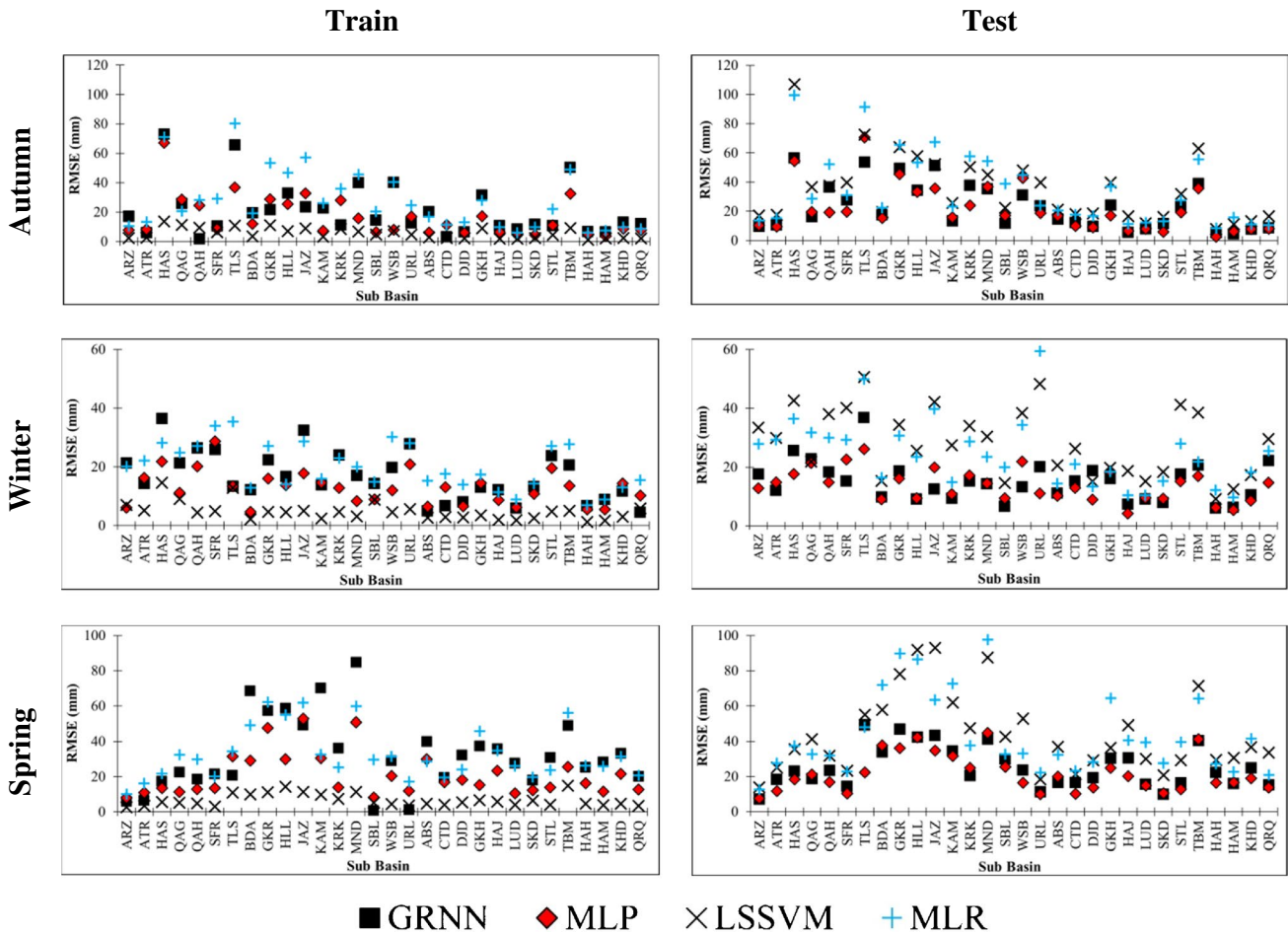


Fig. 8 The RMSE values in train and test steps of machine learning models study in the different seasons

(a) autumn precipitation and AO, SOI, SSTs, AMO, SCN, Nino 3.4, PDO, PNA, TNA, WP, and TSA indices; (b) winter precipitation and EAWR, SSTs, NAO, PNA, POL, PDO, SCN, TNA, and TSA indices; and (c) spring precipitation and SST, PNA, Nino, NAO, NPI, POL, PNA, NPI, and NCP indices. In Helali et al. (2020) study, it was found that the indices related to ENSO, SST, TNA, PDO, and AMO had the highest correlation with basin precipitation, which is consistent with the results of this research. According to Ahmadi et al. (2019), ENSO, AMO, and AO were the most related teleconnection indices affecting precipitation in Iran’s climatic regions. Moreover, Dehghani et al. (2020) concluded that SOI, NAO, and PDO indices were the most significant indices, whereas the current study showed that the type of teleconnection indices varied according to the season and sub-basin.

A detailed analysis revealed that the highest PST correlations of teleconnection indices with autumn precipitation in 6, 5, 4, 3, 2, and 1-month time lags belonged to TNA, TNA, SOI, SOI, NINO 3.4, and NINO 4, respectively. Likewise,

winter precipitation was correlated with PDO, NAO2, MSST, TSA, SOI, and POL indices, and spring precipitation was correlated with NINO 4 (in 6-, 5-, and 2-month lags), SOI (in 6-, 4-, and 3-month lags), SST4 (in 6-, 5-, and 2-month lags), POI (in 1-month lag), and WPSST (in 3- and 1-month delays). Most studies in Iran represented that teleconnection indices mainly affect autumn precipitation and partially influence winter and spring precipitation in a good correlation with some of these patterns such as the West Mediterranean Oscillation (WeMo) and Scandinavian Pattern (SP) (Ghasemi and Khalili 2008; Ghasemi 2019). While the current study showed that in addition to autumn precipitation, an appropriate and significant correlation was found between teleconnection indices and winter (NAO2, PDO, AO, CSST, MSST indices in different time lags) and spring (NINOs, SSTs, SOI, TNli, TPI, POI, and MEI in different time lags) precipitation. Hence, it is concluded that those indices are significant predictor variables. Synoptic and dynamic analysis of the relationship between autumn precipitation in Iran’s sub-basins and teleconnection indices

showed that these precipitations were partially related to the MJO index at first, which was the stimulus of ENSO phase initiators and regional indices such as NAO and AO had the same effect as ENSO through the large atmospheric bridge mechanism. In fact, regional indices are a medium for measuring the impact of global indices such as ENSO, PDO, and AMO on Iran's climate, but in general, the impact of ENSO in Iran precipitation is more evident, especially in autumn (Nazemosadat and Cordery 2000; Helali et al. 2020, 2021c).

The analysis of the ML models uncovered that LSSVM and MLP models had better performance than GRNN and MLR models in autumn, winter, and spring in most of the sub-basins of Iran between 1987 and 2015. However, the regression method (Nalley et al. 2019; Qian and Xu 2020; Kim et al. 2020), fuzzy and neural network (Karamouz et al. 2006; Choubin et al. 2014; Canchala et al. 2020; Gholami Rostam et al. 2020), and genetic algorithm method (Dariane et al. 2019) proved to be the best-known models in previous research. In essence, the efficiency of LSSVM (in most of the sub-basins) and MLP (in some of the sub-basins) was good by higher seasonal forecast accuracy in comparison with GRNN and MLR models during all times. The results of this study were consistent with other studies that proposed the LSSVM model as the best model with the lower error (Tao et al. 2017; Seyedzadeh et al. 2020). The results show that although in the training phase, the RMSE of LSSVM and MLP models were less than other models; in the testing phase, the RMSE of MLP and GRNN mostly represented the lowest values. Therefore, it is better than the MLP and GRNN to be used in forecasting and seasonal precipitation modeling seasonal precipitation. The advantages of this study were the analysis of the correlation between teleconnection indices and SP in all basins and sub-basins of Iran (spatial balance between the teleconnection indices and the study area), precipitation forecast of all seasons (unlike other studies conducted in a specific season), and using different models and evaluation of their efficiency in the seasonal precipitation forecast. This study showed the efficiency of teleconnection indices in seasonal precipitation simulation in Iran's sub-basins and suggested the appropriate models for such complex variables, to be deployed in water resource management on a larger scale as sub-basins.

Conclusion

Due to the importance of water resource management on a regional scale and the impact of large-scale teleconnection indices on it, this study attempted to model and simulate the precipitation of different main and sub-basins of Iran using ML models with different architectures. In summary, seasonal precipitation in Iran was correlated with teleconnection indices, and in most cases, their significance was higher in autumn and spring than in winter. The dynamic

and synoptic analysis of these phenomena might contribute to that and more investigation is necessary for future studies. The highest PST correlation was with autumn (in 1- to 4-month lag), winter (in 4- to 6-month lag), and spring (in 1- to 4-month lag) precipitation. The most important indices affecting autumn precipitation were MEI, NINOs, SSTs, POI, SOI, and TPI. Similarly, the winter precipitation was mainly affected by NAO2, PDO, TNA-TSA, AO, CSST, NCP, and MSST. SSTs, NINOs, MEI, NPI, PNA, POI, SOI, TNI, TPI, and TNi influenced spring precipitation, as well. These time lags were due to the distance of these indices with the climate and especially the precipitation of the study area. The analysis reflected that autumn precipitation was positively correlated with MEI, NINOs, SSTs, and TPI, and it was negatively correlated with POI, SOI, TNA, WPSST, and CSST. Winter precipitation was correlated positively with NAO2, NCP, PDO, and TNA-TSA and it was negatively affected by CSST, MSST, TSA, WPSST, EA, AO, and POL. Spring precipitation showed a positive correlation with MEI, NINOs, SST, PNA, TNA, TPI, and TNi and a negative correlation with CSST, NPI, POI, SOI, TNI, and WPSST. Therefore, teleconnection indices had the potential to simulate the seasonal precipitation of the basin and sub-basins of Iran. The seasonal precipitation forecast was more accurate by using MLP followed by the GRNN model in most basins, but in general, in sub-basins with geographical and climatic diversity, most of the models' accuracies decreased. The results of this study demonstrate the efficiency of teleconnection indices as precipitation predictor variables in Iranian sub-basins. The reason for the various results by machine learning models in different sub-basins can be the climatic characteristics and the topographic diversity of each sub-basin. It is recommended to test other algorithms to tackle the geographical and climatic diversity of such study areas. Elimination of the microclimate effect in the region restricted our study; from the basin perspective, we could no longer assess the impact of the teleconnection indices on the precipitation of a particular station in the region, but we considered the spatial fit of the teleconnection indices within the basin. Therefore, the future study will explore some other robust algorithms or integration of the above-applied ML models that consider the spatial fit of teleconnection patterns with precipitation and the microclimate effects. Moreover, it seems helpful to examine those limited stations with summer precipitation to investigate the relationship between the teleconnection indices and precipitation from seasonal and basin points of view.

Acknowledgements The authors would like to thank the Islamic Republic of Iran Meteorological Organization (IRIMO) and Ministry of Energy for providing the data needed to conduct this research.

Author contribution Author 1 and 3: conceptualized the idea and developed the theory and performed the computations. Author 1 and

2: verified the analytical methods and supervised the findings of this work. Author 4–8: wrote, edited, restructured, and optimized the manuscript. All authors discussed the results and contributed to the final manuscript.

Data availability The data used in this paper was prepared by Islamic Republic of Iran Meteorological Organization (IRIMO) and website of National Oceanic and Atmospheric Administration (NOAA) from this link: <http://www.cpc.ncep.noaa.gov/data/teledoc/teleindcalc.shtml>.

Code availability MATLAB codes used to run the machine learning models are available on request from the corresponding author.

Declarations

Ethical approval Not applicable.

Consent for publication The authors agreed to submit the article and were aware of the submission.

Conflict of interest The authors declare no competing interests.

References

- Ahmadi M, Salimi S, Hosseini SA, Poorantiyosh HA, Bayat A (2019) Iran's precipitation analysis using synoptic modeling of major teleconnection forces (MTF). *Dyn Atmos Oceans* 85:41–56
- Ahmadi K, Kalantar B, Saeidi V, Harandi EKG, Janizadeh S, Ueda N (2020) Comparison of machine learning methods for mapping the stand characteristics of temperate forests using multi-spectral sentinel-2 data. *Remote Sens* 12(18).
- Anochi JA, de Almeida VA, de Campos Velho HF (2021) Machine learning for climate precipitation prediction modeling over South America. *Remote Sens* 13:1–18
- Araghinejad S (2014) Data-driven modeling: using MATLAB in water resources and environmental engineering. Springer, Water Science and Technology Library, Volume, p 67
- Asadi Oskouei E, Helali J (2021) Correlation analysis of large-scale teleconnection indices with monthly reference evapotranspiration of Iran synoptic stations. *Iran J Soil Water Res* 66:1629–1644
- Behar O, Khellaf A, Mohammedi K (2015) Comparison of solar radiation models and their validation under Algerian climate—the case of direct irradiance. *Energy Convers Manag* 98:236–251
- Biabanaki M, Eslamian SS, Koupai JA, Cañón J, Boni G, Gheysari M (2013) A principal components/singular spectrum analysis approach to ENSO and PDO influences on rainfall in western Iran. *Hydrol Res* 45:250–262
- Canchala T, Alfonso-Morales W, Carvajal-Escobar Y, Cerón WL, Caicedo-Bravo E (2020) Monthly rainfall anomalies forecasting for Southwestern Colombia using artificial neural networks approaches. *Water* 12:2628
- Choubin B, Khalighi Sigaroodi S, Malekian A, Ahmad S, Attarod PM (2014) Drought forecasting in a semi-arid watershed using climate signals: a neuro-fuzzy modeling approach. *J Mt Sci* 11(6):1593–1605
- Cigizoglu HK (2005) Generalized regression neural network in monthly flow forecasting. *Civ Eng Environ Syst* 22(2):71–81
- Dariane BD, Ashrafi Gol M, Karami F (2019) Forecasting of rainfall using different input selection methods on climate signals for neural network inputs. *J Hydraul Struct* 5(1):42–59
- Dehghani M, Salehi S, Mosavi A, Nabipour N, Shamshirband S, Ghamisi P (2020) Spatial analysis of seasonal precipitation over Iran: co-variation with climate indices. *ISPRS Int J Geo-Inf* 9:73
- Dibike YB, Velickov S, Solomatine D, Abbott MB (2001) Model induction with support vector machines: introduction and applications. *J Comput Civ Eng* 15:208–216
- Doblas-Reyes FJ, Garcia-Serrano J, Lienert F, Biescas AP, Rodrigues LRL (2013) Seasonal climate predictability and forecasting: status and prospects. *Wires Clim Chan* 4:245–268
- Duan ZH, Kou SC, Poon CS (2013) Prediction of compressive strength of recycled aggregate concrete using artificial neural networks. *Constr Build Mater* 40:1200–1206
- Goudarzi M, Ahmadi H, Hosseini SA (2017) Examination of relationship between teleconnection indexes on temperature and precipitation components (Case Study: Karaj Synoptic Stations). *Iran J Ecohydrology* 4(3):641–651
- Ghasemi AR (2019) Influence of northwest Indian Ocean sea surface temperature and El Niño-Southern Oscillation on the winter precipitation in Iran. *J Water Clim Chang* 11(4):1481–1494
- Ghasemi AR, Khalili D (2008) The association between regional and global atmospheric patterns and winter precipitation in Iran. *Atmos Res* 88:116–133
- Gholami Rostam M, Sadatinejad SJ, Malekian A (2020) Precipitation forecasting by large-scale climate indices and machine learning techniques. *J Arid Land* 12(5):854–864
- Ghosh S, Dasgupta A, Swetapadma A (2019) A study on support vector machine based linear and non-linear pattern classification. In 2019 International Conference on Intelligent Sustainable Systems (ICISS), 24–28.
- Gong DY, Wang SW (1999) Impacts of ENSO on global precipitation changes and precipitation in China. *Chin Sci Bull* 44(9):852–857
- Gonsamo A, Chen JM, Lombardozi D (2016) Global vegetation productivity response to climatic oscillations during the satellite era. *Glob Chan Biol* 22:3414–3426
- Hartman H, Snow JA, Stein S, Su B, Zhai J, Jiang T, Valentina K, Zbigniew W (2016) Predictors of precipitation for improved water resources management in the Tarim river basin creating a seasonal forecast model. *J Arid Environ* 125:31–42
- Heino M, Guillaume JHA, Müller C, Iizumi T, Kumm M (2020) A multi-model analysis of teleconnected crop yield variability in a range of cropping systems. *Earth Syst Dynam* 11:113–128
- Helali J, Salimi S, Lotfi M, Hosseini SA, Bayat A, Ahmadi M, Naderizarneh S (2020) Investigation of the effect of large-scale atmospheric signals at different time lags on the autumn precipitation of Iran's watersheds. *Arab J Geosci* 13(18):1–24
- Helali J, Hosseinzadeh T, Cheraghalizadeh M, Mohammadi Ghalenei M (2021) Feasibility study of using Climate Teleconnection Indices in prediction of spring precipitation in Iran Basins. *Iran J Soil Water Res* 52(3):749–769
- Helali J, Momenzadeh H, Oskouei EA, Lotfi M, Hosseini SA (2021) Trend and ENSO-based analysis of last spring frost and chilling in Iran. *Meteorol Atmos Phys* 133(4):1203–1221
- Helali J, Momenzadeh H, Salimi S, Hosseini SA, Lotfi M, Mohamadi SM, MaghamiMoghim G, Pazhoh F, Ahmadi M (2021) Synoptic-dynamic analysis of precipitation anomalies over Iran in different phases of ENSO. *Arab J Geosci* 14(22):1–21
- HeydariTashehKabood S, Hosseini SA, HeydariTashehKabood A (2020) Investigating the effects of climate change on stream flows of Urmia Lake basin in Iran. *Model Earth Syst Environ* 6(1):329–339
- Hipni A, El-shafie A, Najah A, Karim OA, Hussain A, Mukhlisin M (2013) Daily forecasting of dam water levels: comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS). *Water Resour Manage* 27(10):3803–3823
- Houshyar M, Sobhani B, Hosseini SA (2018) Future projection of Maximum Temperature in Urmia through Downscaling output of CanESM2 Model. *Geogr Plan* 22(63):305–325

- Hu Y, Sun Y, Ha Y, Zhu Y, Luo Z (2020) Prediction of precipitation in the western mountainous regions of China using a statistical model. *Adv Meteorol* 4294563.
- Islam F, Imteaz MA (2020) Use of teleconnections to predict Western Australian seasonal rainfall using ARIMAX model. *Hydrology* 7:52
- Kalantar B, Ueda N, Saeidi V, Janizadeh S, Shabani F, Ahmadi K, Shabani F (2021) Deep neural network utilizing remote sensing datasets for flood hazard susceptibility mapping in Brisbane. *Australia Remote Sens* 13(13):2638
- Karamouz M, Ramazani F, Razavi S (2006) Long-term forecasting of precipitation using meteorological signals: application of artificial neural networks, 7th International Congress of Civil Engineering, Tehran, 11 p.
- Kim CG, Lee J, Lee JE, Kim NW, Kim H (2020) Monthly precipitation forecasting in the Han River Basin, South Korea, using large-scale teleconnections and multiple regression models. *Water* 12:1590
- Kingston DG, Lawler DM, McGregor GR (2006) Linkages between atmospheric circulation, climate and streamflow in the northern North Atlantic: research prospects. *Prog Phys Geogr* 30(2):143–174
- Kinouchi T, Yamamoto G, Komsai A, Liengcharernsit W (2018) Quantification of seasonal precipitation over the upper Chao Phraya river basin in the past fifty years based on monsoon and El Niño/Southern Oscillation related climate indices. *Water* 10:800
- Lee JH, Julien PY, Thornton C, Lee CH (2020) Large-scale climate teleconnections with South Korean streamflow variability. *Hydrolog Sci J* 65:57–70
- Li MF, Tang XP, Wu W, Liu HB (2013) General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Convers Manag* 70:139–148
- Loboda NS, Glushkov AV, Khokhlov VN, Lovett L (2006) Using non-decimated wavelet decomposition to analyse time variations of North Atlantic Oscillation, eddy kinetic energy, and Ukrainian precipitation. *J Hydrol* 322(1–4):14–24
- Ma C, Iqbal M (1984) Statistical comparison of solar radiation correlations monthly average global and diffuse radiation on horizontal surfaces. *Sol Energy* 33:143–148
- Masoudian SA (2005) Recognition of precipitation regimes of Iran using cluster analysis. *Geogr Res Q* 52:47–59
- Masoudian SA, Ataei H (2005) A cluster analysis of precipitation seasons of Iran. *Res Bull Isfahan Univ* 18(1):1–12
- Mekanik F, Imteaz M, Gato-Trinidad S, Elmahdi A (2013) Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. *J Hydrol* 503:11–21
- Modaresi F, Araghinejad S (2014) A comparative assessment of support vector machines, probabilistic neural networks, and K-nearest neighbor algorithms for water quality classification. *Water Resour Manag* 28(12):4095–4111
- Modaresi F, Araghinejad S, Ebrahimi K (2018) A comparative assessment of artificial neural network, generalized regression neural network, least-square support vector regression, and K-nearest neighbor regression for monthly streamflow forecasting in linear and nonlinear conditions. *Water Resour Manage* 32:243–258
- Modaresi F, Araghinejad S, Ebrahimi K (2018) Selected model fusion: an approach for improving the accuracy of monthly streamflow forecasting. *J Hydroinf* 20(4):917–933
- Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: literature review. *Water* 10(11):1536
- Mosavi A, Rabczuk T, Varkonyi-Koczy AR (2017) Reviewing the novel machine learning tools for materials design. In *International Conference on Global Research and Education* (pp. 50–58). Springer, Cham.
- Najah A, El-Shafie A, Karim OA, Jaafar O (2011) Integrated versus isolated scenario for prediction dissolved oxygen at progression of water quality monitoring stations. *Hydrol Earth Syst Sci* 15(8):2693–2708
- Nalley D, Adamowski J, Biswas A, Gharabaghi B, Hu W (2019) A multiscale and multivariate analysis of precipitation and streamflow variability in relation to ENSO, NAO and PDO. *J Hydrol* 574:288–307
- Navid MAI, Niloy NH (2018) Multiple linear regressions for predicting rainfall for Bangladesh. *Communications* 6:1–4
- Nazemosadat MJ, Cordery I (2000) On the relationships between ENSO and autumn rainfall in Iran. *Int J Climatol* 20:47–61
- Pourasghar F, Tozuka T, Jahanbakhsh S, Sari Sarraf B, Ghaemi H, Yamagata T (2012) The interannual precipitation variability in the southern part of Iran as linked to large-scale climate modes. *Clim Dyn* 39:2329–2341
- Qian H, Xu S (2020) Prediction of Autumn Precipitation over the Middle and Lower Reaches of the Yangtze River Basin Based on Climate Indices. *Climate* 8:53
- Ridwan WM, Sapitang M, Aziz A, Kushiar KF, Ahmed AN, El-Shafie A (2021) Rainfall forecasting model using machine learning methods: case study Terengganu Malaysia. *Ain Shams Eng J* 12(2):1651–1663
- Rodwell MJ, Palmer TN (2014) Using numerical weather prediction to assess climate models. *Q J Roy Meteor Soc* 133:129–146
- Ruigar H, Golian S (2015) Prediction of precipitation in Golestan dam watershed using climate signals. *Theor Appl Climatol* 123:671–682
- Salahi B, Goudarzi M, Hosseini SA (2017) Predicting the temperature and precipitation changes during the 2050s in Urmia Lake Basin. *Watershed Eng Manag* 8(4):425–438
- Salehi S, Dehghani M, Mortazavi SM, Singh VP (2019) Trend analysis and change point detection of seasonal and annual precipitation in Iran. *Int J Climatol* 40(1):308–323
- Schepen A, Wang QJ, Robertson DE (2012) Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall. *J Geophys Res* 117:D20107
- Seydi ST, Saeidi V, Kalantar B, Ueda N, van Genderen JL, Maskouni FH, Aria FA (2022) Fusion of the multisource datasets for flood extent mapping based on ensemble convolutional neural network (CNN) model. *J Sensors*, 2022.
- Seyedzadeh A, Maroufpoor S, Maroufpoor S, Shiri J, Bozorg-Haddad O, Gavazi F (2020) Artificial intelligence approach to estimate discharge of drip tape irrigation based on temperature and pressure. *Agric Water Manag* 228:105905
- Suykens JA, Osipov GV (2008) Introduction to focus issue: synchronization in complex networks. *Chaos Interdiscip J Nonlinear Sci* 18(3):037101
- Suykens JA, De Brabanter J, Lukas L, Vandewalle J (2002) Weighted least squares support vector machines: robustness and sparse approximation. *Neural Comput (spec Issue Fundam Inf Process Asp Neurocomp)* 48(1–4):85–105
- Tao L, He X, Wang R (2017) A hybrid lssvm model with empirical mode decomposition and differential evolution for forecasting monthly precipitation. *J Hydrometeorol* 18:159–176
- Widiasari IR, Nugroho LE, Widyawan (2018) Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining. *Proceedings-2017 International Conference on Innovative and Creative Information Technology: Computational Intelligence and IoT, ICITech 2017, 2018-January*, 1–5.
- Willmott CJ, Matsuura K (2006) On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *Int J Geogr Inf Sci* 20:89–102
- Woldemeskel FM, Sharma A, Sivakumar B, Mehrotra R (2014) A framework to quantify GCM uncertainties for use in impact assessment studies. *J Hydrol* 519:1453–1465

- Xiao MZ (2019) Quantifying spatiotemporal influences of climate index on seasonal extreme precipitation based on hierarchical bayesian method. *Int J Climatol* 40:3087–3098
- Yang Y, Gan TY, Tan X (2019) Spatiotemporal changes in precipitation extremes over Canada and their teleconnections to large-scale climate patterns. *J Hydrometeorol* 20:275–296
- Zhu H, He H, Fan H, Xu L, Jiang J, Jiang M, Xu Y (2020) Regional characteristics of long-term variability of summer precipitation in

the Poyang Lake basin and possible links with large-scale circulations. *Atmosphere* 11:1033

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Jalil Helali¹ · Mehdi Mohammadi Ghaleni² · Seyed Asaad Hosseini³  · Ali Lotfi Siraei⁴ · Vahideh Saeidi⁵ · Farshad Safarpour⁶ · Mojgan Mirzaei⁷ · Mohammad Lotfi⁸

Jalil Helali
jalilhelali@gmail.com

Mehdi Mohammadi Ghaleni
m.mghaleni@gmail.com

Ali Lotfi Siraei
lotfiali217@yahoo.com

Vahideh Saeidi
saeidi.va@gmail.com

Farshad Safarpour
farshadclima2008@gmail.com

Mojgan Mirzaei
mijgan_11884@yahoo.com

Mohammad Lotfi
lotfim48@yahoo.com

² Department of Water Sciences and Engineering, Arak University, Arak, Iran

³ Department of Physical Geography (Climatology), Mohaghegh Ardabili University, Ardabil, Iran

⁴ Department of Climatology, Kharazmi University, Tehran, Iran

⁵ Department of Mapping and Surveying, Tehran University, Tehran, Iran

⁶ Department of Climatology, Kurdistan University, Kurdistan, Iran

⁷ Department of Environmental Engineering, Malayer University, Malayer, Iran

⁸ Department of Meteorology, Science and Research Branch, Islamic Azad University, Tehran, Iran

¹ Department of Irrigation and Reclamation Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran