



A long lead time forecast model applying an ensemble approach for managing the great Karun multi-reservoir system

Ehsan Mostaghimzadeh¹ · Seyed Mohammad Ashrafi¹ · Arash Adib² · Zong Woo Geem³

Received: 12 October 2021 / Accepted: 19 April 2023 / Published online: 5 May 2023
© The Author(s) 2023

Abstract

Flow prediction is regarded as a major computational process in strategic water resources planning. Prediction's lead time has an inverse relationship with results' accuracy and certainty. This research studies the impact of climate-atmospheric indices on surface runoff predictions with a long lead time. To this end, the correlation of 36 long-distance climate indices with runoff was examined at 10 key nodes of the Great Karun multi-reservoir system in Iran, and indices with higher correlation are divided into 4 different groups. Then, using Artificial Neural Network (ANN) and Ensemble Learning to combine the input variables, flow is predicted in 6-month horizons, and results are compared with observed values. To assess the impact of extending the prediction lead time, results from the proposed model are compared with those of a monthly prediction model. The performed comparison shows that using an ensemble approach improves the final results significantly. Moreover, Tropical Pacific SST EOF, Caribbean SST, and Nino1 + 2 indices are found to be influential parameters to the basin's inflow. It is observed that the performance of the prediction process varies in different hydrological conditions and the best results are obtained for dry seasons.

Keywords Runoff forecast · Water resources management · Prediction lead time · Wavelet transformation · Artificial neural network

Introduction

Water is the most important element of vitality in human life, such that all civilizations were formed by rivers. In many countries around the world, rainfalls are not evenly

distributed over the year or do not conform to the demands. While climates in many parts of the Middle East are entering a critical situation. It is essential to understand that the water shortage crisis also adversely affects the qualitative characteristics of the flow (Bakhsipoor et al. 2019). Meanwhile, asymmetric distribution of drought and flood on the surface of the earth cause widespread human and financial losses (Kim et al. 2005; Moazami et al. 2016). According to the Parliament of Iran, an estimated \$2.5 billion per year in damages is caused by drought in the country. In 2019, the Inter-governmental Panel on Climate Change (IPCC) reported that over the past decades, changes in climate caused by global warming have led to alterations in the hydrological regime (Change 2019; Gholami et al. 2022). Over the years, a distorted pattern of inflow to reservoirs due to climate change and human activities such as land-use change has led to significant alterations in the amount, frequency, and timing of the inflow to the water reservoirs (Demirel et al. 2013; Halgamuge and Nirmalathas 2017; Yang et al. 2018; Mostaghimzadeh et al. 2021). Therefore, the need for the employment of strategies to reduce the damages caused by droughts and floods has become more eminent. Predicting

✉ Seyed Mohammad Ashrafi
ashrafi@scu.ac.ir; semo.ashrafi@gmail.com

Ehsan Mostaghimzadeh
e.mostaghimzadeh@gmail.com

Arash Adib
arashadib@scu.ac.ir

Zong Woo Geem
zwgeem@gmail.com

¹ Department of Civil Engineering, Faculty of Civil Engineering and Architecture, Shahid Chamran University of Ahvaz, Ahvaz, Iran

² Department of Environmental Engineering, Faculty of Civil Engineering and Architecture, Shahid Chamran University of Ahvaz, Ahvaz, Iran

³ College of IT Convergence, Gachon University, Seongnam 13120, Korea

the inflow to reservoirs is considered a fundamental solution to deduce more efficient operation policies and thus reduce damages. However, alteration of the inflow to the system affects its supply values, energy production, and quality properties in the long term (Ashrafi and Mahmoudi 2021). It is important to note that extending the prediction further into the future will provide the policy-makers and operators with more leeway in making immediate or long-term decisions (Mostaghimzadeh et al. 2022). However, longer-term predictions come with higher uncertainty and therefore make the decision-making processes more complicated. In this context, data-based models are widely used. In these models, hydrologic processes are not presented definitively, but the main focus is on finding the relationship between the variables influencing the output of the models (Shortridge et al. 2016; Peng et al. 2018). With this rationale, using these models to estimate the values outside the observation intervals must be carried out with caution (Tongal and Booij 2018).

In general, awareness of the emerging uncertain condition requires an estimation of future status (MostaghimZadeh et al. 2021; Huang et al. 2022; Beshavard et al. 2022; Luo et al. 2023). This means the prediction should present the desired accuracy and meet the evaluation indices at a significant level and, at the same time, its horizon must be as far ahead in the future as possible to paint a better picture of the situation and allow for making short-term decisions following the long-term course of events. A review of the previous research reveals that the development of the flow predictive models is summarized in several different structures. Much of previous research has developed predictive models with a one-month lead time using novel methods and smart algorithms. In such researches, a wide range of feature selection methods was employed and various learning models such as neural networks, support vector regression, etc., were used to relate inputs to outputs. Also, in studies where the forecasting lead time extended further than a month, the work was based solely on the amount of rainfall. By predicting the rainfall values and subsequently using a rainfall-runoff model, the flow was obtained in the future. Despite all capabilities of Artificial Neural Networks (ANNs) in prediction of real-world phenomenon (Geem 2011), the framework of learning processes affects the accuracy of the results considerably. Osman et al. (2020) employed Fast Orthogonal Search to investigate the nonlinear relation of the inflow to the Aswan High Dam basin. The results showed that this method outperformed the classic artificial intelligence methods, especially for larger numbers. Ni et al. (2020) proposed the GMM-XGboost hybrid model, created by coupling the Extreme Gradient Boosting and the Gaussian Mixture models, to develop a monthly flow model in the Gangtze River basin. The results showed that the proposed hybrid model was more successful than the SVM model in the detection

of hydrologic patterns. Ren et al. (2020) compared the performance of eight feature selection methods in predicting the monthly flow (CAMELS). Ashrafi et al. (2020) used the Ensemble Learning technique to compare seven different approaches to estimate the monthly flow upstream of the Dez dam in the Southwest of Iran. The results showed the superiority of coupled Discrete Wavelet Transformation-Artificial Neural Network (DWT-ANN) model with four-item groups. Quedi and Fan (2020) developed a seasonal hydrologic predictive model and extended the results for tropical weather conditions as well as large-scale basins. The results showed the desirable performance of Ensemble Prediction in the assessment of flow patterns and consequently its acceptable estimation on higher lead times. Sigaroodi et al. (2014) attempted to develop a long-term precipitation forecasting model by investigating the correlation of the oceanic-atmospheric indices on the precipitation of the Maharloo basin in Iran. The results indicated a deep correlation between the fluctuations in rainfall patterns and some of these indices. In the end, the final model was able to estimate the amount of precipitation with more than one-month lead times based on a set of influential indices. Kilinc and Haznedar (2022) developed a model to estimate future streamflow conditions considering sustainability using deep learning models. The estimation results of the models were evaluated with RMSE, MAE, MAPE, STD and R2 statistical metrics. The comparison of daily streamflow predictions results revealed that the proposed model provided promising accuracy results and mainly presented higher performance than the benchmark model and the linear regression model. Liu et al. (2022) integrate meteorological forecasts, land surface hydrological model simulations and machine learning to forecast hourly streamflow over the Yantan catchment, where the streamflow is influenced by both the upstream reservoir water release and the rainfall-runoff processes within the catchment. Evaluation of the hourly streamflow hindcasts during the rainy seasons of 2013–2017 shows that the hydrometeorological ensemble forecast approach reduces probabilistic and deterministic forecast errors by 6% compared with the traditional ensemble streamflow prediction (ESP) approach during the first 7 days.

The literature shows that most of the flow prediction models use novel feature selection methods along with learning models to assess the flow in the form of monthly models. Also in some cases, models with lead times shorter than one month have been investigated. Indirect methods were often used for runoff predictions with longer than one-month lead times. In other words, first, a model for long-term rainfall is obtained. Then, by using a rainfall-runoff model and estimating the other contributors at the basin level, flow in a long-time scale (more than one month) is estimated based on the predicted precipitation (Adamowski and Prasher 2012; Chu et al. 2017; Mosavi et al. 2018; Peng et al. 2019). In

that sense, there is a research gap in the development of models that can directly predict runoff within water resource systems with long forecasting lead times. Furthermore, comparisons of flow prediction models with short and long lead times have rarely been carried out.

Therefore, this study attempts to develop a model based on the ensemble learning approach derived from (Ashrafi et al. 2020) and the various oceanic-atmospheric indices that can estimate the inflow to a drainage basin with a six-month lead time. In an ensemble learning system, the impact of predictors on target parameters is investigated systematically in groups and the final solution can be a range of predictors. In the beginning, the flow with one-month lead time is calculated and used as a benchmark to compare the results of the short-term and long-term lead time. To do so, first, different approaches to the development of a flow prediction model with monthly time steps are examined (Ashrafi et al. 2020). Then, solutions to extend the prediction lead time are proposed and a model is deduced to provide a proper estimation of the flow for a long-time step (six-month lead time). In the end, the best long- and short-term models according to assessment indices are selected and their performances are compared to study the impact of extending the prediction horizon. Investigation of the direct impact of climate indices on inflow to a multi-reservoir system and the model development approach are innovations of this study.

Research methodology

In this study, two sub-models are presented. First, the short-term model, which is a prediction model for monthly inflow to a multi-reservoir system. Second, the long-term model, which is a flow prediction model with a six-month lead time. For the one-month lead time, the seven approaches proposed by Ashrafi et al. (2020) are extended to predict future inflows into a multi-reservoir system and the best scheme for each reservoir is selected. Then, to increase the prediction lead time, the flow time series is used in combination with various climatic-oceanic indices contributing to the flow of the multi-reservoir basin. In addition, a new structure based on ensemble learning is introduced to select proper inputs for model learning. In the end, the best short- and long-term models are selected and compared to study the impact of prediction lead time on the model's performance and the evaluation indices.

For the short-term forecast model, the wavelet components of the streamflow time series are adopted as potential predictors considering the lags of 1 to 10. In contrast, for longer forecast horizon the oceanic atmospheric teleconnection indices are used as predictors. As these indices have a seasonal behaviour, it is better not to use them for every month and sampling of them was different from monthly horizons. Moreover, the lack of enough data was a barrier

to consider their lags within the feature selection process. In order to develop a solution, the present state of each index is just considered as a potential predictor. Finally, the potential lags are identified using ACF and PACF function according to the rate of correlation between streamflow and selected indices. The length of forecast horizon is considered as the closest effective lags to the present state.

The proposed forecasting model

The proposed structure is based on the developed model by Ashrafi et al. (2020) to obtain short- and long-term prediction models. This structure consists of learning sub-models (McCulloch 1990), input selection model (Phuong et al. 2005; Wang et al. 2009), and evaluation indices (Nash and Sutcliffe 1970; Willmott 1981; Shamseldin and Press 1984; Legates and McCabe 1999; Krause et al. 2005; Chai and Draxler 2014). In this structure, seven proposed approaches are studied for each reservoir and the final model is selected based on the evaluation indices. The short-term flow prediction is generally based either on the exclusive use of time series of the inflow or the forecast deduced from satellite and radar data on precipitation and its conversion into the runoff. With the extended prediction lead time, the flow time series is not sufficient as more data are required to estimate the flow with optimal accuracy. Long lead times surpass the one-month benchmark and which fills the gap between short-term predictions and long-term climate changes (Tourigny and Jones 2009). These forecasts are often influenced by significant uncertainties. Many researchers have been attempting to propose methods for more accurate predictions (Wu et al. 2011). Many studies have concluded there are nonlinear relations between meteorological phenomena, climate indices, and accuracy of answers (Kim and Barros 2001; Tourigny and Jones 2009; Guérémy et al. 2012). According to Teschl and Randeu (2006), certain variables representing the atmospheric flow fluctuation, which are directly related to the sea surface temperature fluctuations, have a significant long-term impact on rainfall, and as a consequence, on stream flows. At the same time, atmospheric flow dynamics are influenced by solar activities including radiation intensity and sunspots which are defined as temporary phenomena of the photosphere and are seen as dark patches on the surface of the sun. Solar radiation is considered the energy source for climate systems and is often measured as a constant factor. According to Sigaroodi et al. (2014), the number of solar patches is correlated with rainfall. Wang et al. also confirmed the correlation between sunspot and runoff in river basins (Wang and Sheeley 1997). They also showed that the length of the sunspot cycles is strongly correlated with the temperature of the northern hemisphere and drought periods (Vennerstrøm et al. 1991; Weickmann et al. 2000). Meanwhile, the high thermal capacity of the oceans

has a significant impact on the atmosphere–ocean interaction, which ultimately changes the atmospheric flow patterns over the subsequent periods (Penland and Matrosova 1998). Based on the above-mentioned studies, climate indices are highly correlated with precipitation at a basin scale. In the same way, the amounts of rainfall and runoff are always correlated. Therefore, the correlation between climate indices and runoff can be directly measured concerning other contributors to runoff, such as soil conditions, vegetation, snow depth. Also, a model to predict the direct runoff can be developed based on the indices using machine learning algorithms.

To develop a prediction model, the relation between the oceanic-atmospheric currents and the time series of inflows to a multi-reservoir system is examined with the proposed approach. Then, according to the level of correlation between these indices and the inflow to each system, inputs are selected for the learning model. For input selection, various quality grades are defined based on the amount of correlation for each index, and the model's input parameters are selected by categorizing the elements at each grade. Then, after the best group in each grade is determined, combinations with the best groups from all the levels are created and the final model is chosen based on the evaluation indices. In Fig. 1 a flow chart of the model is demonstrated.

Atmospheric-oceanic indices

Atmospheric-oceanic Indices represent the physical phenomena correlated with rainfall. In this study, 36 climate indices from the NOAA's website (<http://www.esrl.noaa.gov/gmd/dv/ftpdata>) are used to evaluate existing correlations. In Table 1, a list of the indices as climate predictors along with their respective harvest periods is presented.

The study area

The Karun River in Khuzestan, Iran comprises five operational dams: the Karun4, the Karun3, the Karun1, Masjed Soleyman, and Gotvand-Olya, and the Dez River includes the Dez dam. The two rivers join at Band-e-Ghir and form the Great Karun River. The average annual precipitation in the basin range from about 150 mm per year in the lowland plains to around 1800 mm per year in the adjacent Kookhrang Mountains. The Khersaan River is one of the most water-rich tributaries of the Karun River which drains into the Karun-3 reservoir. In addition, the outflow from the Karun-4 also drains into the Karun-3 reservoir. Past the Karun-3, all the dams on the Karun basin are positioned serially behind one another where at the end of each reservoir is another dam. At the downstream of the Karun3 is the Karun1, which has been in operation for years. The Masjed Soleyman dam is located in Masjed Soleyman downstream of the Karun1 and is tasked with energy supply. The last dam on the Karun, the Gotvand-Olya, is currently in operation. Meanwhile, on the Dez basin is located the Dez storage dam whose outflow joins that of the Gotvand-Olya dam in Band-e-Ghir and form the Great Karun River (Ashrafi 2021). In the figure below, the layout of the reservoirs on the Great Karun basin, which is investigated in this research, is presented. In this study, 10 flow nodes including tributaries and sub-basin inflows are accounted for according to the following Table 2 and Fig. 2.

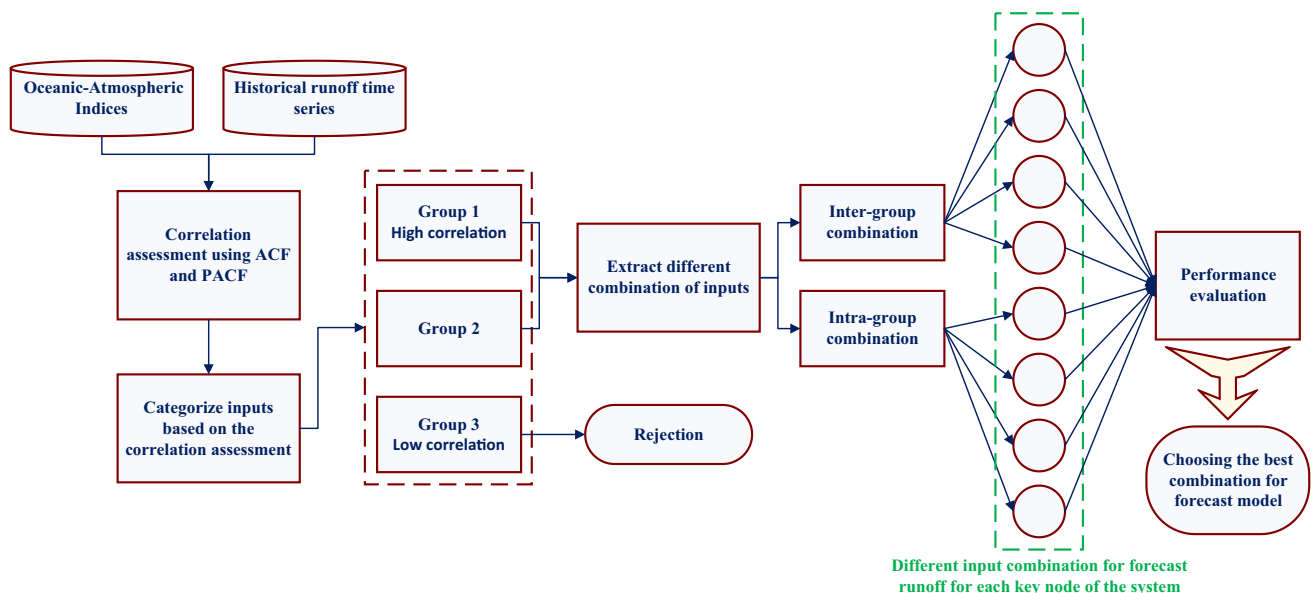


Fig. 1 Conceptual flow chart of long-term forecast model

Table 1 Primary selection of ocean-atmospheric predictors

NO	Climate index	Data period	No	Climate index	Data period
1	AAO	1979–2019	19	Northeast Brazil rainfall	1948–1999
2	AO	1950–2019	20	NP	1948–2018
3	Atlantic multidecadal oscillation smoothed	1948–2014	21	NTA	1950–2019
4	Atlantic multidecadal oscillation unsmoothed	1948–2019	22	ONI	1950–2019
5	Atlantic tropical SST EOF	1948–2007	23	Pacific warm pool	1948–2007
6	Caribbean SST)CAR(1951–2019	24	PDO	1948–2017
7	Central Indian percipitation	1948–1998	25	PNA	1950–2019
8	ENSO percipitation index	1979–2015	26	QBO	1948–2019
9	Global mean land–ocean temperature index	1948–2018	27	Sahel rainfall	1948–2017
10	Globally integrated angular momentum	1958–2013	28	SOI	1951–2019
11	Hurrican activity	1948–2015	29	Solar flux	1948–2018
12	MEI	1979–2018	30	Sunspot	1749–2018
13	NAO jones	1948–2000	31	SW mansoon region Rainfall	1948–2010
14	NAO	1950–2019	32	TNA	1948–2019
15	Nino 1 +2	1950–2019	33	TNI	1948–2018
16	Nino 3	1950–2019	34	Tropical Pacific SST EOF	1948–2008
17	Nino 4	1950–2019	35	Atlantic Tripole SST EOF	1948–2008
18	NOI	1948–2006	36	WHWP	1948–2019

Table 2 Inflow nodes of great Karun Basin

Bakhtiari head flow	Karun4 to Karun3 Sub-basin inflow
Cezar head flow	Karun3 to Karun1 Sub-basin inflow
Dez to band-e-Ghir Sub-basin inflow	Karun1 to MasjedSoleyman Sub-basin inflow
Karun4 head flow	MasjedSoleyman to Gotvand Sub-basin inflow
Khersan head flow	Gotvand to Band-e-Ghir Sub-basin inflow

Results and discussion

The achieved results are presented in two following sub-sections for short and long lead times separately, and the performance of the proposed approach is discussed.

Runoff forecasting with Short-lead time

According to the initial framework proposed by Ashrafi et al. (2020), the short-term forecast model has been developed using an approach consisting of an Artificial Neural Network (ANN) and a Discrete Wavelet Transformation (DWT) operator. This combination has been employed in order to deduce the frequency of variation dominant in the initial time series. The general equation of the DWT function is defined as follows (Polikar 1999);

$$DWT_x(s, b) = \frac{1}{\sqrt{s_0^m}} \int_{-\infty}^{+\infty} x(t) \cdot \psi\left(\frac{t - nb_0s_0^m}{s_0^m}\right) dt \quad (1)$$

Here, $\psi(\cdot)$ is the mother wavelet function, $s > 0$, is the scale parameter, b , stands for the time transfer parameter, and $DWT_x(s, b)$ indicates the discrete wavelet transform of $x(t)$. The scale parameter determines the wavelet width. The monthly runoff time series of all watershed nodes are decomposed to the fourth level. At each level of wavelet decomposition, one of the coefficients indicates low-frequency, long-term changes and the other indicates high-frequency, short-term fluctuations. Also, at other levels, the low-frequency coefficient is split and two new coefficients are obtained. Thereby, the changes hidden in the initial time series are more accurately evaluated, enabling the learning model to estimate the relationship more precisely by accurately recognizing the underlying runoff patterns. As the levels of decomposition progress, the coefficients become smoother. The process proceeds until the initial mother wavelet replaces the coefficients. In this process, at each level of decomposition, multiple parts of the time series are sampled by the mother wavelet, and the similarity is determined in the form of the internal multiplication of the two vectors. This process continues until the sampling replaces the initial time series with the mother wavelet chart, as the smoothed surface. Accordingly, the number of decomposition levels should not be too high (to inhibit fluctuation patterns from being hidden) and too low (to restrict the model from reaching the initial mother wavelet due to low oscillations). Hence, as previously mentioned, each time series is split into four levels. An “Only ANN” and six “pre-processed ANN” approaches including self-correlation and correlation with the main data are analysed to develop

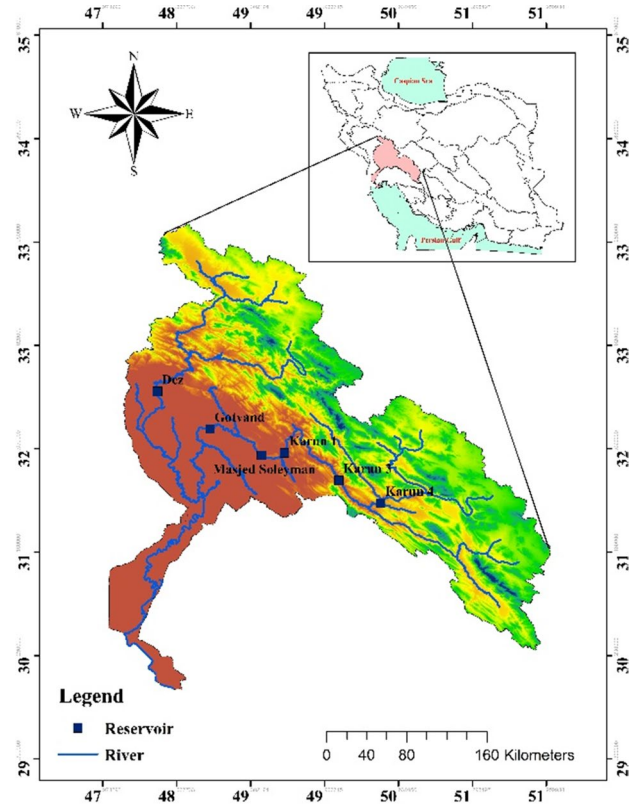
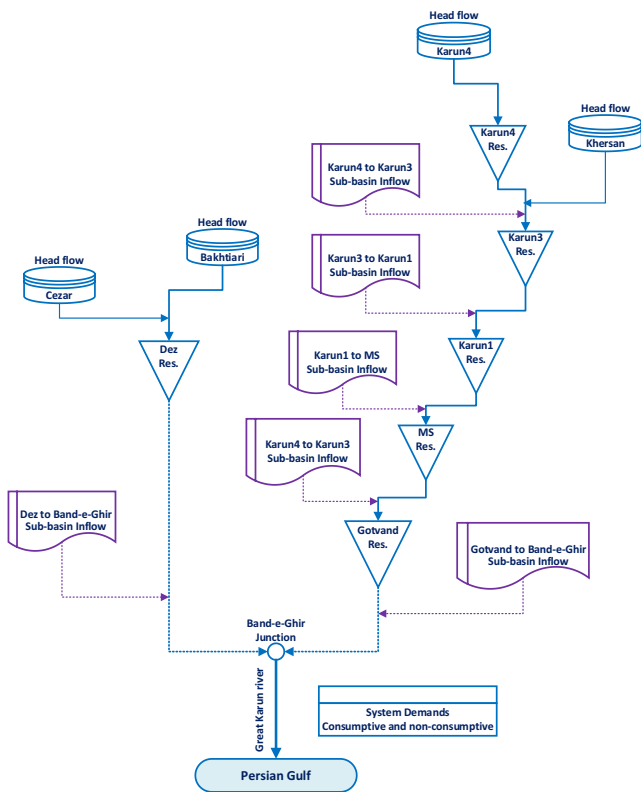


Fig. 2 Schematic of great Karun Basin

the monthly forecast model for each of the time series of reservoirs' inflows and sub-basin inflows. Several models are extracted for each time series (Table 3). Importantly, ECWT models include the best possible combination with the desired evaluation indicators.

Several neural networks have been trained to estimate the runoff into each key point of the system using 85% of the observed data. The trained networks are evaluated using 15% of the observed data and the best one is implemented for predicting the runoff of each key point. The best networks for different key nodes of the system are

presented in Table 3. Table 4 summarizes the performance evaluation indicators related to 10 independent training processes in the form of mean and standard deviation (SD).

The highest correlation coefficient (*R*) is associated with the prediction process of the Sub-basin inflow of Karun1 to MasjedSoleyman. The Willmotts index works like a correlation coefficient. In this case, the Sub-basin inflow of Karun1 to MasjedSoleyman has the best performance. Also, by analysing the Nash Sutcliffe index, and according to (Shamseldin 1997), all predictions are satisfactory, and in some

Table 3 Best predictor combination for inflow nodes (monthly forecast model)

Location	Best model	Predictor combination
Bakhtiari head flow	ECWT (Foursquare)	$(a_4, d_1, d_3, d_4)/Q_t$
Cezar head flow	ECWT (Foursquare)	$(a_4, d_1, d_3, d_4)/Q_t$
Dez to band-e-Ghir Sub-basin Inflow	ECWT (Foursquare)	$(a_4, d_1, d_3, d_4)/Q_t$
Karun4 head flow	ECWT (Quintuple)	$(a_4, d_1, d_2, d_3, d_4)/Q_t$
Kherasan head flow	ECWT (Foursquare)	$(a_4, d_1, d_3, d_4)/Q_t$
Karun4 to Karun3 Sub-basin inflow	ECWT (Foursquare)	$(a_4, d_1, d_3, d_4)/Q_t$
Karun3 to Karun1 Sub-basin inflow	ECWT (Quintuple)	$(a_4, d_1, d_2, d_3, d_4)/Q_t$
Karun1 to MasjedSoleyman sub-basin inflow	ECWT (Foursquare)	$(a_4, d_1, d_3, d_4)/Q_t$
MasjedSoleyman to Gotvand sub-basin inflow	ECWT (Foursquare)	$(a_4, d_1, d_3, d_4)/Q_t$
Gotvand to band-e-Ghir Sub-basin inflow	ECWT (Quintuple)	$(a_4, d_1, d_2, d_3, d_4)/Q_t$

Table 4 Performance evaluation indicators for the best monthly forecast models

Method	Index	R	WI	Nash	RRMSE (%)	MAPE (%)
Bakhtiari head flow	Average	0.944	0.917	0.901	24.01	32.169
	Std	0.003	0.011	0.010	1.066	0.996
Cezar head flow	Average	0.927	0.923	0.874	26.36	33.731
	Std	0.005	0.009	0.013	0.267	0.246
Dez to Band-e-Ghir Sub-basin Inflow	Average	0.917	0.932	0.896	27.51	32.23
	Std	0.007	0.004	0.008	0.279	0.156
Karun4 Head flow	Average	0.921	0.909	0.903	26.03	31.25
	Std	0.007	0.007	0.007	0.380	0.128
Khersan Head flow	Average	0.912	0.915	0.885	27.45	34.51
	Std	0.006	0.006	0.006	0.399	0.52
Karun4 to Karun3 Sub-basin Inflow	Average	0.899	0.921	0.898	29.32	35.169
	Std	0.004	0.005	0.011	0.114	0.270
Karun3 to Karun1 Sub-basin Inflow	Average	0.951	0.930	0.908	23.22	32.404
	Std	0.006	0.007	0.009	0.249	0.323
Karun1 to MasjedSoleyman Sub-basin Inflow	Average	0.963	0.946	0.891	24.44	30.107
	Std	0.021	0.014	0.012	0.441	0.435
MasjedSoleyman to Gotvand Sub-basin Inflow	Average	0.923	0.914	0.900	28.84	32.52
	Std	0.006	0.007	0.008	0.39	0.445
Gotvand to Band-e-Ghir Sub-basin Inflow	Average	0.889	0.910	0.875	28.81	31.36
	Std	0.005	0.005	0.006	0.411	0.51

cases, such as the Bakhtiari and Karun 4 Head flows, the overall performance is highly acceptable.

RMSE and MAE indicators are two scalar indexes in terms of quantity and their numerical value does not provide information on their relative utility, i.e., we cannot certainly discuss the predictive performance of these indicators. Therefore, the relative RRMSE and MAPE indicators, which measure the residual value or error of the model relative to the target value, are calculated. From Table 4, the chosen models for all runoff time series have a low relative error and are in the acceptable range. Figure 3 shows simulated runoffs compared to the observed flows in all considered nodes.

From Fig. 3, the performances of the selected methods for detecting the runoff pattern in all key nodes of the system are highly acceptable. At this step, the wavelet transformation operator has helped to accurately identify the existing fluctuations and repetitive patterns. It has also helped the learning process to perform and the corresponding coefficients to be determined following these patterns. Importantly, peak values have not been desirable in most sub-basins. Peak values are always difficult to predict in methods based on the streamflow time series. The reason is that in a long time, the maximum number of samples is very small, and if learning is done only based on the number of occurrences, these values cannot be accurately predicted. However, according to Fig. 3, in all key nodes, the pattern of peak occurrence has been correctly identified and its period obtained. Also, the evaluation of the proposed method in the study of time series

with a natural cycle is important. According to Table 4 and Fig. 3, the evaluation indicators are achieved in nearly the same range, meaning that time series derived from a natural cycle have similar patterns, and if this pattern is correctly identified, then it can be accurately predicted. However, this is expected not to change the general patterns governing the natural cycle. That's why peak events that are somehow the result of disruption of natural cycles are not properly identified.

Accordingly, if the forecast lead time increases, it is not possible to merely follow the pattern of current changes, and other influential parameters need to be identified. The following figure shows the peak estimation performance of the selected methods at all key nodes.

Figure 4 shows that the peak values are often estimated to be less than the observed values. As mentioned earlier, due to the small number of training patterns and the data-driven nature of the neural network, more samples are required to determine the exact behaviour of a real-world phenomenon. The lower peak value of runoff diminishes the rate of predictive relative error and increases the accuracy of the estimated value, as the low peak values are closer to the normal data range, and there are more samples in that range. During the learning process, the neural network has more similar samples to determine its weight coefficients for estimating normal ranges (Geem et al. 2007). A box-plot diagram is used to further examine the distribution of the predicted data. Due to its nonparametric nature, the box plot is independent of the

Fig. 3 Observed and predicted monthly streamflow in testing period **(a)** Bakhtiari Head flow, **(b)** Cezar Head flow, **(c)** Dez to Band-e-Ghir Sub-basin Inflow, **(d)** Karun4 Head flow, **(e)** Karun4 to Karun3 Sub-basin Inflow, **(f)** Karun3 to Karun1 Sub-basin Inflow, **(g)** Karun1 to Masjed-Soleyman Sub-basin Inflow, **(h)** MasjedSoleyman to Gotvand Sub-basin Inflow, **(i)** Gotvand to Band-e-Ghir Sub-basin Inflow, **(j)** Khersan Head flow)

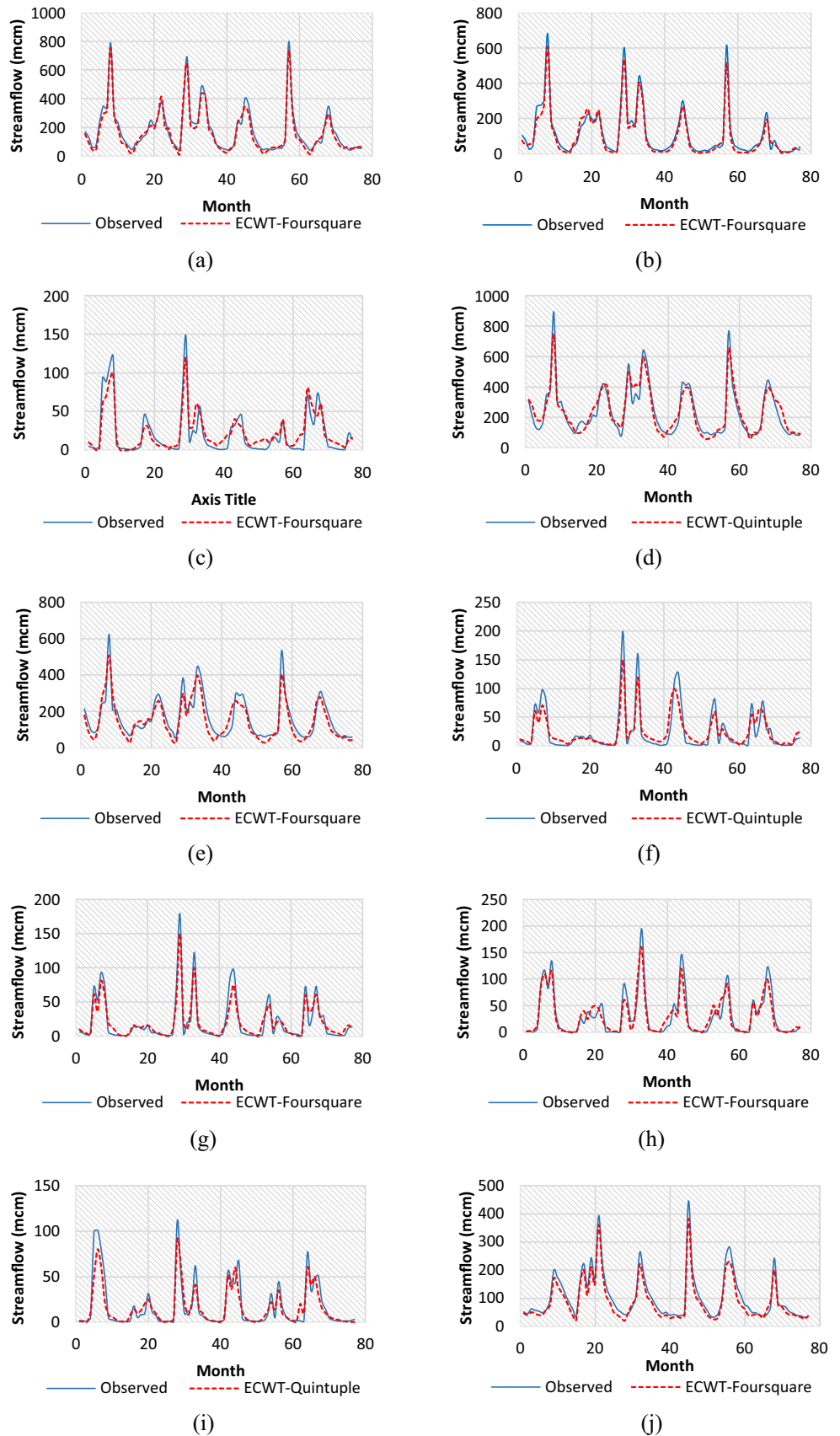
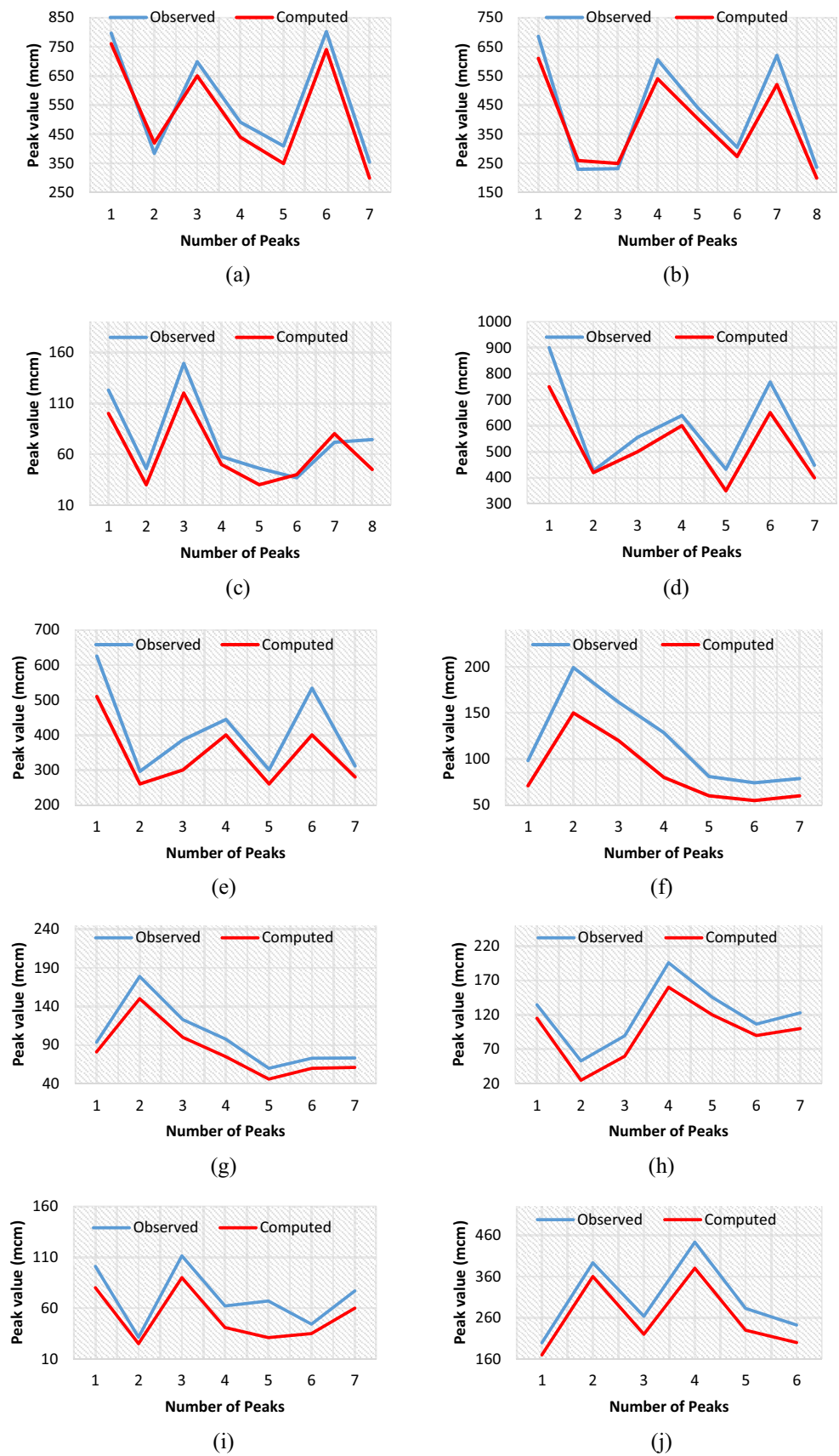


Fig. 4 Peak estimation within the test period for (a) Bakhtiari Head flow, b Cezar Head flow, c Dez to Band-e-Ghir Sub-basin Inflow, d Karun4 Head flow, e Karun4 to Karun3 Sub-basin Inflow, f Karun3 to Karun1 Sub-basin Inflow, g Karun1 to MasjedSoleyman Sub-basin Inflow, h MasjedSoleyman to Gotvand Sub-basin Inflow, i Gotvand to Band-e-Ghir Sub-basin Inflow, j Khersan Head flow)



initial assumption regarding statistical distribution, leading to a better understanding of the degree of expansion and skewness of a time series. Also, whiskers show the diversity in the data in down and upper quartiles (25% and 75%).

Figure 5 shows that the predictions of the chosen model are almost identical to observational data in all sub-basins. Besides, in estimating the minimum, maximum, and median values, the chosen models have acceptable performance with

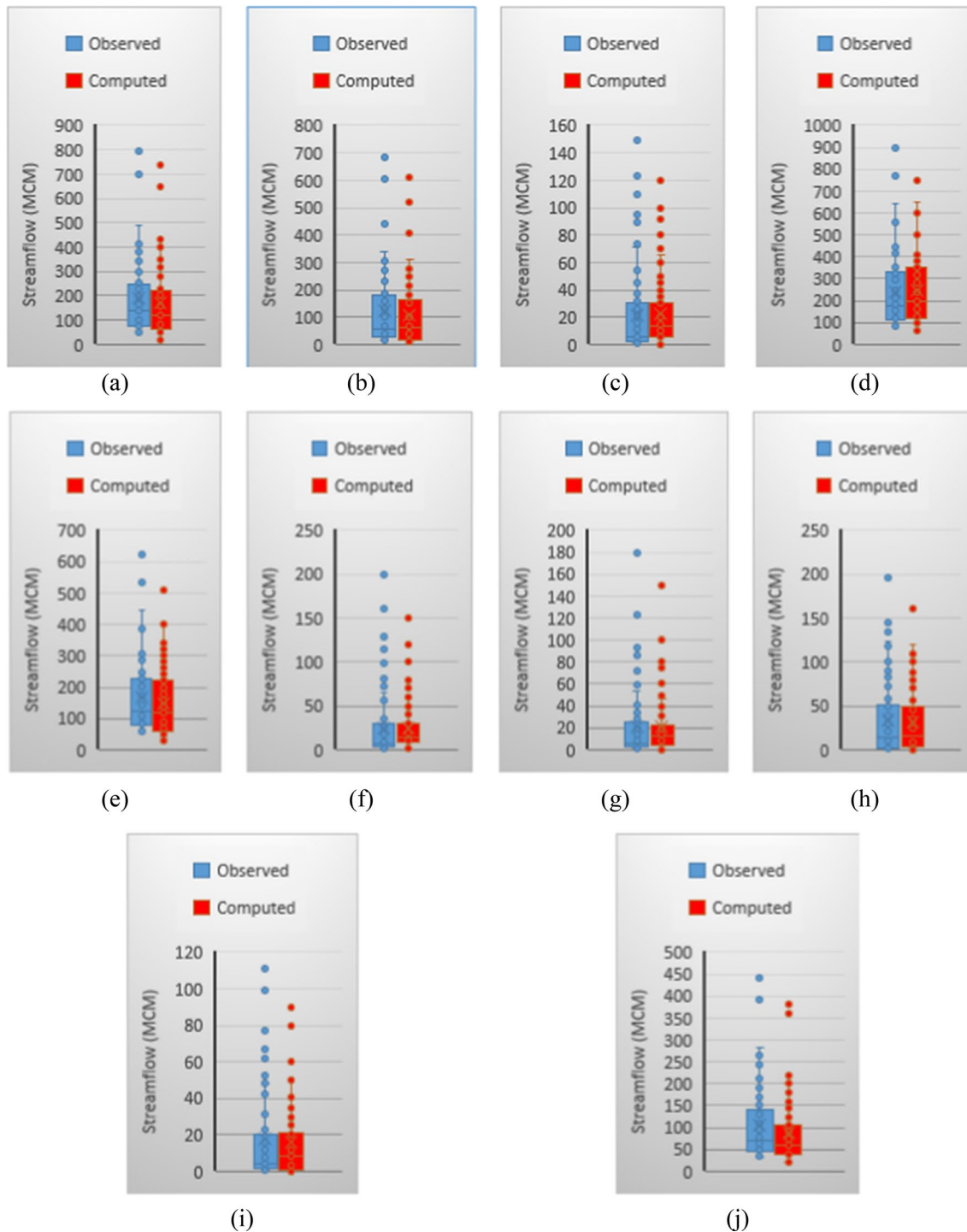


Fig. 5 Monthly Streamflow Peak estimation within the test period for **(a)** Bakhtiari Head flow, **(b)** Cezar Head flow, **(c)** Dez to Band-e-Ghir Sub-basin Inflow, **(d)** Karun4 Head flow, **(e)** Karun4 to Karun3 Sub-basin Inflow, **(f)** Karun3 to Karun1 Sub-basin Inflow, **(g)** Karun1 to Mas-

jedSoleyman Sub-basin Inflow, **(h)** MasjedSoleyman to Gotvand Sub-basin Inflow, **(i)** Gotvand to Band-e-Ghir Sub-basin Inflow, **(j)** Khersan Head flow)

a justified error level. It is also observed that the estimated values of the chosen models are skewed to the upper values as the observational values, which can be due to the presence of a larger number of data greater than the median in the observational values. In other words, there are many points outside the box range, which tends to distribute upwards. Importantly, in a limited number of nodes, such as Khersan, Karun 4 Head flows, and also Sub-basin inflow between Karun 4 and Karun 3, the distribution of data varies from the observed value, especially in small quantities. This could be due to a lack of appropriate data in this range. Also, the high difference between scattered points outside of the box in all basins is due to the inability of the proposed models to estimate peak values, which has caused a large difference between the maximum points.

Runoff forecasting with long-lead-time

As discussed in the methodology section, as the forecast lead time increases, the uncertainty of the results progresses, and the prediction accuracy decreases. Accordingly, it is no longer feasible to rely only on the current time series and even its decomposed components as a proper input for the

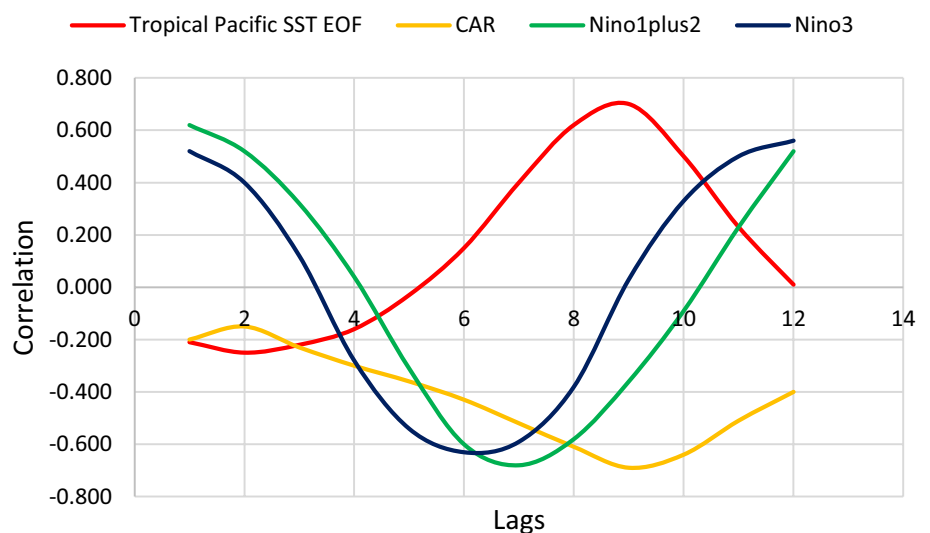
learning model. In this study, a wide range of atmospheric-climatic characteristics are introduced (Table 1). Initially, the Pearson test was performed for each of these indicators relative to all current time series in the Great Karun Basin. In other words, a linear correlation was obtained between each indicator and the amount of streamflow in each of the inlet nodes of the catchment area. Importantly, a common period was initially obtained in each pair and became the basis of the work due to the inconsistency of the observed flows and climatic characteristics. In the following, different quality levels were defined based on the correlation values. In other words, concerning the amount of correlation, the influential indicators were ranked from the highest value to the lowest level of correlation, respectively. The results of the grouping of indicators are summarized in Table 5.

According to Table 5, from 36 indicators examined in this study, CAR, Tropical Pacific SST EOF, Nino1 + 2, and Atlantic Tripole SST EOF have the highest correlation with the streamflow time series in all key nodes of the watershed. As the goal of developing a long-term model is to predict runoff values with a more than six-month lead time, in classifying the correlation of an index, values related to 7th lag onwards are considered as a basis to can estimate the runoff at the 7th month according to these values in each time step. Note that each indicator, due to its nature, has an internal relationship with the amount of flow and has a unique correlation pattern. Figure 6 shows the correlation values of the last twelve lags of 1st-group indicators with the Bakhtiari Head flow time series. From the figure, each index has its correlation pattern with different values in each lag. For example, the Tropical Pacific SST EOF (Red) index is more consistent from the sixth lag onwards, with the highest number happening in the ninth lag. Therefore, the assumption of choosing the seventh lag onwards as the inputs of the learning model to predict the runoff that leads to the

Table 5 Grouping of the climatic index based on the correlation coefficient

Group	Correlation range	Climatic indicator
1	$ R \geq 0.65$	Tropical Pacific SST EOF, CAR, Nino1 + 2, Atlantic Tripole SST EOF
2	$0.5 \leq R < 0.65$	Nino3, HA, MEI, SMRF, TNA, NBR
3	$0.4 \leq R < 0.5$	AMOS, AMOSUn, EPI, NTA, QBO, ATSE, NAOJones, NP, PDO
4	$ R < 0.4$	Other indexes

Fig. 6 Correlation patterns of the first group indicators



six-month growth of the forecast lead time is correct and can be quantified by having the input variables per month. The runoff in the seventh month can be estimated from input variables in each month. The seasonal nature of these indicators is also important. The effects of all indicators are different in seasons and according to the dominant climatic phenomena, their effects also vary. As it is clear in Fig. 6, the behaviour of some of these indicators is similar. It should be mentioned that in order to develop an efficient forecasting model, a wide range of different indicators should be used so that somehow all the dominant phenomena are well identified and used in the model. But one of the most important problems of seasonality is the small number of monthly and seasonal samples. In other words, there is a small number of indicators in each stage, which causes problems for numerical models that are dependent on data.

For longer forecast horizons effective teleconnection indices for each month were identified. Afterwards, ACF and PACF functions were drawn to identify effective lags. As we know, we must estimate streamflow for future month within a specified time interval. So, the distance between the current states and the first lag with high correlation determines forecast horizon. It is obvious that the other lags more than horizon lag also could be used as predictor but the first one is the forecast horizon. Figure 6 shows the behaviour of each index against streamflow. The length of this interval could identify the forecast lead time because there is a distance between current state and past. Indeed, the graph identifies how these indices could affect streamflow. In other words, the seasonality is gained from the fluctuations reflected in each chart.

Models are grouped after determining each of the influential indicators and qualitative grouping. In this way, single, double, triple, and quadruple categories are created at each level and the best category is chosen in each combination. In this way, the learning process is based on the similarities of a category with the main data (Ensemble Learning) and somehow distributes the effective feature and ultimately improves

the performance of the training. Importantly, due to the differences in climatic indicators and the subsequent diversity of existing correlations, in some months, the correlation rate of an indicator may be at its best but other months of the year at worst. As established in many studies (Jayawardena and Fernando 1998; Palmer et al. 2004; Guérémy et al. 2012; Sigaroodi et al. 2014), the initial correlation of the indicators with the current values is relatively low due to the long spatial-local distance of the general atmospheric indicators as well as the potential low data resolution. To solve these problems, it is recommended to use an intermediate mode with a moderate correlation in all conditions. Accordingly, the basis of grouping makes it possible to balance the remaining months of the year while taking advantage of the highest correlation rates. Also, combinations have been created between other levels to assess the possibility of the quality of other indicators and choose the best possible mode. This is because the grouping criterion is based on the maximum monthly correlation, and the maximum correlation rate may be low in one month but has a uniform distribution in other months of the year. It can increase the accuracy of the final answers. In this research, a limited number of indicators have been used in each group of the long-term forecasting model, which indicates the low correlation of other indicators in each group. It is obvious that the quality of the model should be reduced if less important indicators are included in the learning process. Accordingly, removing some of them can improve the final performance. Table 6 presents the best compounds of one, two, three, and four members and also the best interfacial compounds related to Bakhtiari Head flow as a certain case.

Eighty-five per cent of observed data are used to train neural networks and 15% to test developed networks. Table 7 presents the evaluation indicators for the 10 independent runs of the best neural network for each key point of the system in the form of mean and standard deviation.

Table 7 shows the appropriateness of the linear fit in terms of the correlation coefficient. In other words, in the

Table 6 The best combination of climatic indexes for Bakhtiari head flow

Model name	Contributed groups	Active elements of each group	Performed combination
L1_Single	1	1	(Pacific SST)/ Q_t
L1_Double	1	2	(Pacific SST, Nino1 + 2)/ Q_t
L1_Triple	1	3	(Pacific SST, CAR, Nino1 + 2)/ Q_t
L1_Foursquare	1	4	(Pacific SST, Tripole SST, CAR, Nino1 + 2, CAR)/ Q_t
L1&2_Triple	1/2	3	(Pacific SST, CAR, Nino1 + 2) and (Nino3, HA, MEI)/ Q_t
L1&2&3_Triple	1/2/3	3	(Pacific SST, CAR, Nino1 + 2) and (Nino3, HA, MEI) and (QBO, NP, PDO)/ Q_t
L1&2_Foursquare	1/2	4	(Pacific SST, Tripole SST, Nino1 + 2, CAR) and (Nino3, HA, MEI, TNA)/ Q_t
L1&2&3_Foursquare	1/2/3	4	(Pacific SST, Atlantic SST, Nino1 + 2, CAR) and (Nino3, HA, MEI, TNA) and (QBO, NP, PDO, ATSE)/ Q_t

Table 7 Evaluation Index for runoff prediction with six months lead time

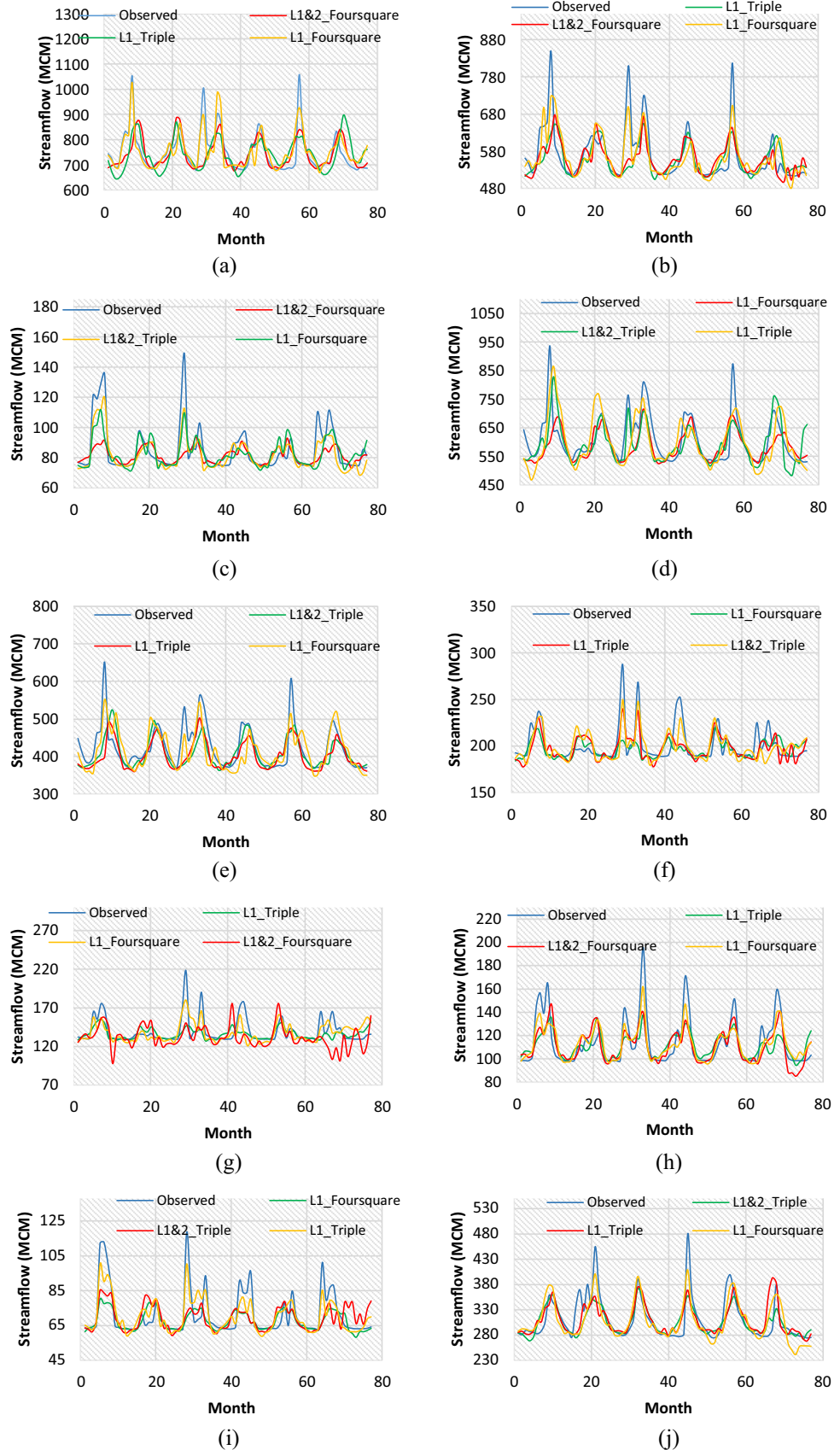
Key point/sub-basin	The best achieved model	Index	R	WI	Nash	RRMSE (%)	MAPE (%)
Bakhtiari head flow	L1_Foursquare	Average	0.803	0.794	0.810	27.06	35.169
		Std	0.000	0.014	0.016	0.541	0.98
Cezar head flow	L1_Foursquare	Average	0.796	0.781	0.803	29.86	38.731
		Std	0.006	0.006	0.010	0.293	0.221
Dez to band-e-Ghir sub-basin inflow	L1&2_Triple	Average	0.788	0.769	0.792	30.05	38.23
		Std	0.003	0.007	0.004	0.284	0.147
Karun4 head flow	L1_Triple	Average	0.801	0.799	0.803	29.12	36.374
		Std	0.002	0.009	0.006	0.375	0.118
Karun4 to Karun3 Sub-basin inflow	L1_Foursquare	Average	0.736	0.784	0.771	32.31	38.951
		Std	0.008	0.006	0.017	0.152	0.263
Karun3 to Karun1 Sub-basin inflow	L1&2_Triple	Average	0.703	0.769	0.781	32.56	39.41
		Std	0.007	0.006	0.005	0.351	0.325
Karun1 to MasjedSoleyman sub-basin Inflow	L1_Foursquare	Average	0.744	0.781	0.793	31.21	38.61
		Std	0.006	0.016	0.016	0.476	0.114
MasjedSoleyman to Gotvand sub-basin inflow	L1_Foursquare	Average	0.801	0.798	0.809	27.84	35.42
		Std	0.006	0.007	0.009	0.364	0.406
Gotvand to band-e-Ghir Sub-basin inflow	L1_Triple	Average	0.804	0.803	0.814	27.61	35.36
		Std	0.008	0.002	0.004	0.484	0.574
Khersan head flow	L1_Foursquare	Average	0.761	0.782	0.791	30.64	37.51
		Std	0.003	0.009	0.003	0.396	0.361

chosen models, a proper linear relationship has been established between the input and the output. Note that the difference in the amount of correlation coefficient in the monthly and six-month model is because the monthly model uses the time series of runoff and its decomposed components, and the one-month lag has increased the correlation to 0.9. Also, the highest correlation coefficient is related to the Gotvand to Band-e-Ghir sub-basin inflow belonging to the L1_Triple model, which is a single-group model. It is observed that the intra-group and inter-group categorization of indicators has ultimately led to the distribution of correlation values. In other words, the overlap of correlations has led a final correlation coefficient of the group to be higher than the correlation coefficient of every single component of the group compared to the original data, and in some nodes to record 0.8. The Willmotts index relatively works similar to the correlation coefficient, although it is more sensitive to differences in observational and predictive values. In this case, the Gotvand and Band-e-Ghir currents with the selected L1_Triple model have the highest value of this index. Also, it is observed that in all flow inlet nodes, this index has a value of more than 0.7. This is due to the nature of the index, which is slightly more than 0.65, even in weaker models. By examining the Nash Sutcliffe index and, according to (Shamseldin 1997) as the monthly model, all predictions are satisfactory and a uniform state is observed in all flow key nodes. The RRMSE index shows that the highest relative difference between the hypothetical fitting line and the

observed samples is 32%, which is related to the Karun3 to Karun1 sub-basin inflow with the chosen L1 & 2_Triple model. This is an acceptable value for the forecast model that estimates the value of the target parameter over the next six months. Also, in other forecasting key nodes, the index is less than this value, indicating the optimal performance of the models. The RRMSE index is more sensitive to peak values due to its structure. Part of the RRMSE error can be attributed to the fact that the performance of all chosen models in predicting peak values is not fully satisfactory. However, in the MAPE index, the remaining value is calculated by the power of one and, therefore, it is more affected by the normal data. Accordingly, it is more reliable in the current problem. Figure 7 shows comparative predicted and observed values for all key nodes. The simulated values are the results of the best-chosen models in each node (Table 7).

From Fig. 7, the effect of climatic indicators on the hydrologic processes within a watershed eventually causes changes in the surface flow pattern, and surface flows can be effectively predicted by identifying such effective indicators. However, these indicators directly affect the amount of precipitation in the basin and consequently have indirect effects on the surface flow pattern. Surface flow is not entirely due to precipitation, and the amount of base current accounts for a significant portion of the surface flow. As a result, the high correlation between climatic indicators and precipitation in all seasons does not lead to a high correlation between precipitation and surface runoff. However, it can be said that

Fig. 7 Observed and forecasted streamflow with six-month lead time within the testing period (**a**.Bakhtiari Head flow, **b** Cezar Head flow, **c** Dez to Band-e-Ghir Sub-basin Inflow, **d** Karun4 Head flow, **e** Karun4 to Karun3 Sub-basin Inflow, **f** Karun3 to Karun1 Sub-basin Inflow, **g** Karun1 to Masjed-Soleyman Sub-basin Inflow, **h** MasjedSoleyman to Gotvand Sub-basin Inflow, **i** Gotvand to Band-e-Ghir Sub-basin Inflow, **j** Kherasan Head flow)



the base current is generally due to the infiltration process, which is itself due to precipitation in previous months. Therefore, in the months when the base flow is minimal, the climatic characteristics are directly related to the amount of runoff, whilst in the months when the base flow has a significant value, the surface flow has two direct and indirect correlations and is estimated using the indicators affecting the rate of penetration and precipitation.

In general, the performance of each of the chosen methods in detecting the runoff pattern is very acceptable. For example, we can highlight the input node of Karun4 Head flow, in which the L1_Triple model has a very small relative error in appearance, besides obtaining a correlation coefficient above 0.8. Also, most of the chosen models are from the indicators available in level 1. And the L1 & 2_Triple model has been chosen as an intermediate level model only in estimating the Dez to Band-e-Ghir and Karun3 to Karun1 sub-basin inflow, where the second level indicators have been influential. The overlap of correlation values in estimating a target variable is another point to consider. Each indicator has a unique correlation pattern based on its internal nature (Fig. 6). In all key nodes, Ensemble models have been chosen from different multi-group member combinations. Also, in some nodes, such as the Karun3 to Karun1 sub-basin inflow, the effects of correlation overlap are observed. In other words, despite a lower correlation between the two groups, the final chosen model, L1 & 2_Triple, is an intermediate group combination and gives a better result due to the synergy of the correlation values. Also, the six-month forecasting model is not optimal in estimating peak values, which can be due to the low number of peak observation samples. Figure 8 shows the peak estimation results using the top three models in each node compared to the observed values.

From Fig. 8, there is a kind of underestimation with peak values. This could be due to the low number of observed data of peak events throughout the considered time horizon. For example, Karun4 to Karun3 sub-basin inflow, the L1_Foursquare model, which originated from the quadruple grouping of level one, can make the least error in peak values. However, it underestimates all peak values. Another case is the node of Dez to Band-e-Ghir sub-basin inflow, in which the chosen model L1 & 2_Triple has experienced the most relative error in the fourth peak. It is also observed that the relative error in smaller peaks is less, which is consistent with the monthly forecast performance. For example, in the node of Karun1 to MasjedSoleyman sub-basin inflow, the chosen L1_Foursquare model has a relative error of 22% in the second peak and 1% in the seventh peak. This indicates that higher peaks have a higher relative error and a higher underestimation rate. Also, the proper performance of a forecasting model in terms of statistical indicators does not mean the exact prediction value of peak values. For example, in

the node of Gotvand to Band-e-Ghir sub-basin inflow, the L1_Foursquare, and L1 & 2_Triple models, which had a high correlation with the main current, could not even identify the peak current pattern and its time leg, because statistical indicators measure the mean time series and exclude local effects such as peak values. Also, data-driven learning methods are data-orientated and can train neural networks and ultimately estimate similar values by repeating a predefined pattern (Geem, and Roper 2009). As a result, due to the lack of peak samples, it is not appropriate to use only classical learning models to accurately estimate peak values. To examine the distribution of forecast data, the box-plot diagram of the results is shown in Fig. 9.

In the key nodes of Bakhtiari Head flow, Karun4 Head flow, Karun1 to MasjedSoleyman sub-basin inflow, and MasjedSoleyman to Gotvand sub-basin inflow the distribution is such that the values outside the box are similar to the observations and the maximums are estimated more accurately. It can be attributed to the better estimation of peak values by the selected model (yellow box). The figure clearly shows that in the selected models the span, mean, and range are 25% and 75% similar to the observations, whereas in other models generally the whiskers are propagated in the opposite direction or the dispersion of the answers is not sufficient. For example, for the node of Karun1 to MasjedSoleyman sub-basin inflow, the L1&2 Foursquare model, derived from the combination of first and second four-item groups, shows a downward whisker which is contrary to the observed data. This is not just an underestimation in the peak values but indicates that the means and minimums were also underestimated. In another example, the Dez to Band-e-Ghir sub-basin inflow where the L1 & 2 Triple model, resulting from the combination of first- and second-grade three-item groups, has a very small span and in a sense formed a limited box which is usually because the pattern of the target variable is not accurately recognized in the correlation. Here, the six-item group comprised of three first- and second-grade indices has not been able to accurately identify the pattern of flow change in the node; hence, the estimated values are calculated in a limited range of original data changes. It is also seen that in some of the selected models, such as L1_Foursquare, the distribution of data is fully similar to the observed values for Dez to Band-e-Ghir sub-basin inflow while the lower whisker, which is completely skewed, is in contrast to the observation. However, the small difference between the lower whisker and the observations indicates that there is sufficient data in this interval and the difference can be due to low local correlation in some months. Because in practice calculations are based on the mean correlation and in some cases, large differences between correlation and observation in a certain month lead to such results. According to Table 5, grade-1 indices are more correlated with the overall flow in the drainage basin which indicates the flow

Fig. 8 Six-month lead time peak estimation within the test period **(a)** Bakhtiari Head flow, **(b)** Cezar Head flow, **(c)** Dez to Band-e-Ghir Sub-basin Inflow, **(d)** Karun4 Head flow, **(e)** Karun4 to Karun3 Sub-basin Inflow, **(f)** Karun3 to Karun1 Sub-basin Inflow, **(g)** Karun1 to Masjed-Soleyman Sub-basin Inflow, **(h)** MasjedSoleyman to Gotvand Sub-basin Inflow, **(i)** Gotvand to Band-e-Ghir Sub-basin Inflow, **(j)** Khersan Head flow)

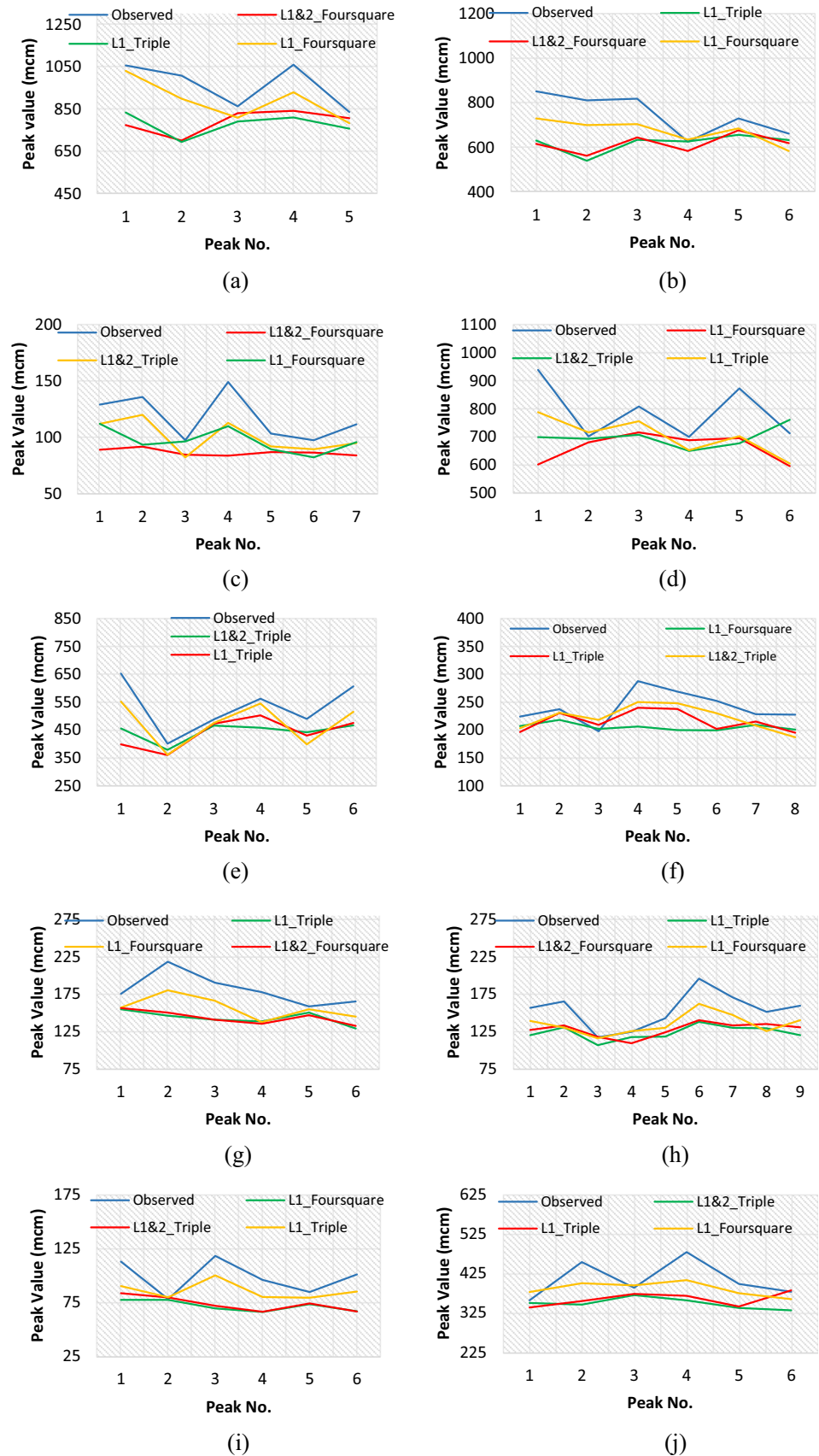
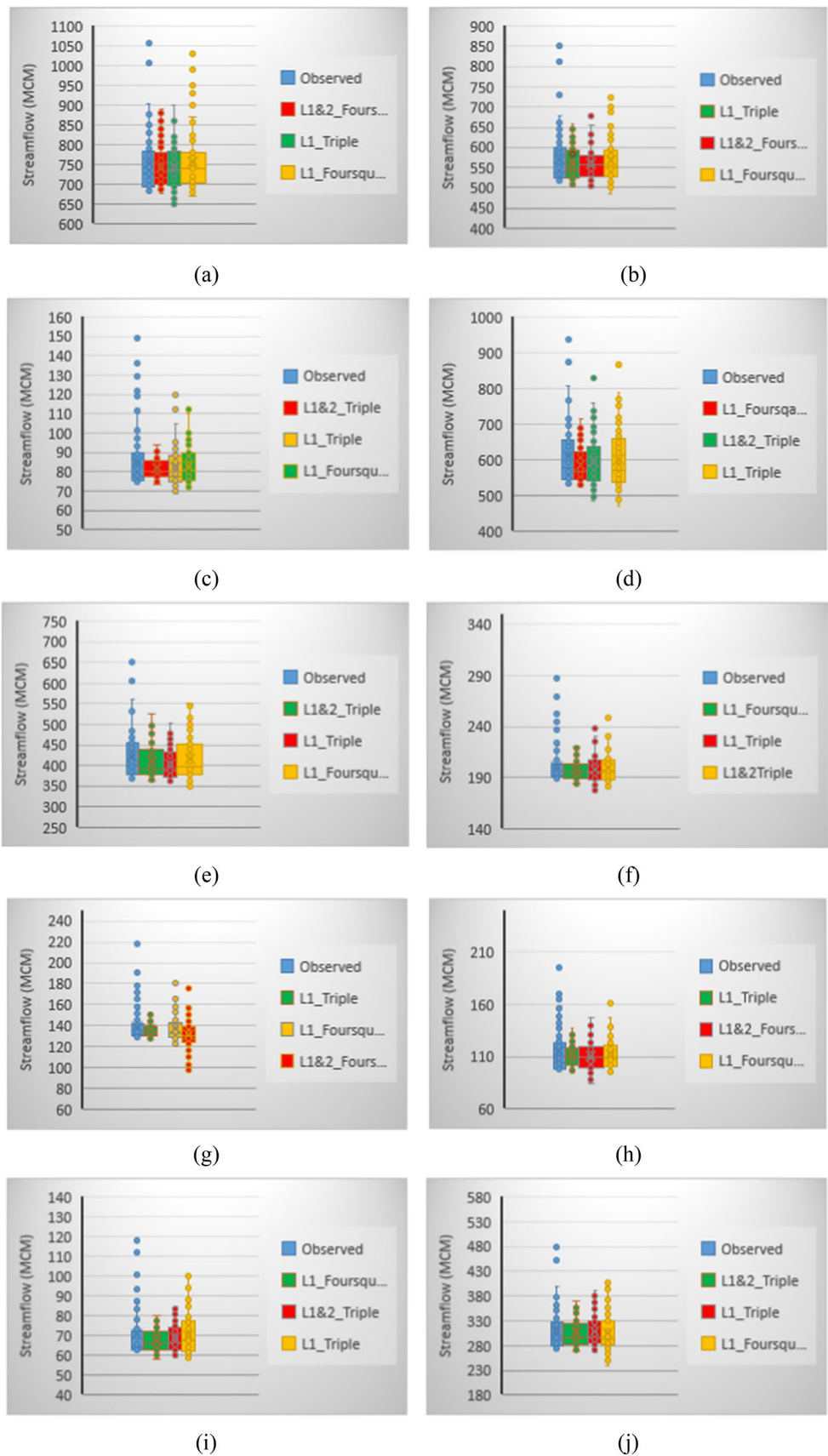


Fig. 9 Box-plots of six-month streamflow peak estimation within the test period (**a** Bakhtiari Head flow, **b** Cezar Head flow, **c** Dez to Band-e-Ghir Sub-basin Inflow, **d** Karun4 Head flow, **e** Karun4 to Karun3 Sub-basin Inflow, **f** Karun3 to Karun1 Sub-basin Inflow, **g** Karun1 to Masjed-Soleyman Sub-basin Inflow, **h** MasjedSoleyman to Gotvand Sub-basin Inflow, **i** Gotvand to Band-e-Ghir Sub-basin Inflow, **j** Kherasan Head flow)



was generally influenced by the fluctuations in Pacific SST. This is in agreement with the findings of (Gu er emy et al. 2012; Sigaroodi et al. 2014) on precipitation correlation in the Mediterranean region. It is also essential to note that the results are slightly different because the flow is partially influenced by snow melting and base flows in the Great Karun basin. For instance, the Caribbean SST is a decisive index in the first level of correlation; indicating that the flow is influenced by fluctuations in the Caribbean region. The long-distance between the studied area and the reference point of the indices such as the Pacific SST in a way represents the time lag in the correlation between precipitation and flow. In other words, while a strong correlation between an index and precipitation in a region means also a high correlation with the flow, the impact of rainfall-runoff transformation time delay, determined by snow melting, soil moisture, penetration rate, etc., influences the selection of final correlated indices. Another point is the seasonal distribution of prediction performance and the better forecasting conditions of droughts compared to wet seasons. A demonstration of the predicted flow in each season with another reference to Fig. 7 shows that the largest difference between prediction and observation occurred in spring. This indicates the maximum flow volume occurred in spring and the prediction error increased due to insufficient data on peak values. At the same time, the lowest difference was observed in autumn. In this context, for the same reason of low flow volume in autumn and adequate data distribution in this interval, the model has been well able to identify the pattern of change.

A closer view of Tables 4 and 7 reveals increasing the prediction lead time eventually led to a decrease in the values of evaluation indices. The mean correlation coefficient on the monthly lead time was estimated as 0.92 while increasing the prediction lead time lowered the mean to 0.76. The Wilmotts index was calculated as 0.92 for the monthly and 0.77 for the six-month time step. Also, the Nash Sutcliffe index was recorded as 0.9 on the monthly and as 0.76 on six-month models. For RRMSE and MAPE relative indices, 25% and 32% on the monthly calculation, as well as 29% and 36% on the six-month prediction were recorded, respectively. While the proposed technique was used for the development of six-month prediction, the results are weaker than the monthly forecasting despite the desired results in terms of assessment indices. This is due to the complex nature of flow and various parameters affecting the rainfall-runoff process as well as errors in calculating the long-distance climate indices which make a prediction with long-lead-time difficult. Figures 4 and 8 highlight the difficulty of accurately predicting peak values in short and long lead times. A shorter prediction lead time did not increase the accuracy of peak values; even at certain nodes, larger underestimations were recorded. It also showed that the quality of peak results is relatively independent of the horizon and dependent on

the frequency of occurrence. Finally, in Figs. 5 and 9, it is observed that the lead time has no bearing on the data distribution and does not follow a clear pattern.

Analysis of the results from the monthly and six-month prediction model suggests a proper ANN performance in the recognition of nonlinear patterns. The accurate identification of input patterns and their impact on output indicates the high flexibility of ANNs (Teschl and Randeu 2006; Wu and Chau 2010; Li et al. 2012). In addition, using ensemble learning in examining a group of inputs led to better recognition of patterns and proper distribution of desirable property among the members. This feature caused the overlap of the wavelet components' desired properties in the monthly model and also distributed the positive impact of each index in a group in the six-month model; improving the answers in the end.

At the end, it should be pointed that both the short- and long-term forecast models should be implemented in their suitable applications. The first and the most important function of forecast models is the water resources operating systems. The more accurate the water resources model handle the future situation, the more accurately it can allocate the available water and reduce the critical shortages. But it can be said that the main difference between short-term and long-term models in managing water resources is how they are used. Long-term models are mostly proper to extract the management strategies while, short-term models are used to modify and verify the operational actions during critical periods (e.g. within droughts or wet seasons).

Conclusion

The present study adopted the Ensemble approach based on ANN to predict the long-term inflow to the Great Karun multi-reservoir system in Iran. For this purpose, and to find the contributing input parameters, the relationship between runoff and 36 long-distance climate indices was investigated and influential indices were selected. In the input selection process, ACF and PACF functions, which calculate linear correlation as a function of time delay, were used. With a hydrological-geographic division of the drainage basin and considering the location of reservoirs, inflows to the Great Karun multi-reservoir system were modelled at 10 separate key nodes and the surface runoffs in the nodes were predicted. An examination of the correlations showed that in most of the selected indices, there is a similar behaviour in short-term (1–3 months) and long-term (6–8 months) delays. This feature was used in categorizing climate-atmospheric indices, and the resulting categories were selected to form the inputs for the learning model.

Tropical Pacific SST EOF, CAR, Nino1 + 2, Atlantic Triple SST EOF indices had a higher correlation with

inflows to the Great Karun system, indicating that runoff is primarily influenced by Pacific STT. Results showed that the fluctuation pattern of runoff can be predicted with acceptable accuracy even for six-month horizons using these indices; even if the unique correlation of each index with runoff is not high. The low correlation is due to the long-distance and time delay between the reference point of indices and the studied area. In other words, the difference between observations and forecasts suggests that other local parameters such as local wind, humidity contribute directly to precipitation and indirectly to surface runoff and therefore, must be considered in predictions. Other factors (i.e. soil moisture, slope, etc.) also contribute directly to flow rate, and excluding them increases computational error. Therefore, the use of ensemble learning resulted in the acceptable accuracy of the final model's performance.

Results also showed the nonuniform distribution of prediction performance during wet and dry seasons. The largest difference between forecast and observation values occurred in spring. This is due to an inadequate number of peak samples; because the highest runoff amount in this area occurs during spring and the learning model was unable to cover these values due to insufficient occurrence frequency. Results also show the flexibility of ANN in investigating nonlinear relationships. Finally, it was found that prediction lead time has a restrictive effect on the model's performance. The longer the prediction lead time, the more data are needed for accurate prediction, and the results' accuracy decreases proportionally. Implementing various ensemble learning methods can significantly improve the achieved results.

Acknowledgements The authors would like to thank the Shahid Chamran University of Ahvaz for financial support under the Grant number: SCU.C1400.31254.

Author contributions The authors declare that they have contributed in the preparation of this manuscript.

Funding The research leading to these results received funding from Shahid Chamran University of Ahvaz under research Grant Agreement No SCU.C1400.31254.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval The manuscript is an original work with its own merit, has not been previously published in whole or in part, and is not being considered for publication elsewhere.

Consent to participate The authors have read the final manuscript, have approved the submission to the journal, and have accepted full responsibilities pertaining to the manuscript's delivery and contents.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adamowski J, Prasher SO (2012) Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. *J Water Land Develop* 17(1):89–97
- Ashrafi SM (2021) Two-stage metaheuristic mixed integer nonlinear programming approach to extract optimum hedging rules for multireservoir systems. *J Water Resour Plan Manag* 147(10):04021070
- Ashrafi SM, Mahmoudi M (2021) Decentralized calibration process for distributed water resources systems using the self-adaptive multi-memory melody search algorithm. *J Hydroinf* 23(5):966–984
- Ashrafi SM, Mostaghimzadeh E, Adib A (2020) Applying wavelet transformation and artificial neural networks to develop forecasting-based reservoir operating rule curves. *Hydrol Sci J*. <https://doi.org/10.1080/02626667.2020.1784902>
- Bakhsipoor IE, Ashrafi SM, Adib A (2019) Water quality effects on the optimal water resources operation in great Karun River Basin. *Pertanika J Sci Technol* 27(4):1881–1900
- Beshavard M, Adib A, Ashrafi SM, Kisi O (2022) Establishing effective warning storage to derive optimal reservoir operation policy based on the drought condition. *Agric Water Manag* 274:107948
- Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. *Geosci Model Dev* 7(3):1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Change IC (2019) Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. 2019. In: The approved summary for policymakers (SPM) was presented at a press conference on, Vol 8
- Chu H, Wei J, Li J, Qiao Z, Cao J (2017) Improved medium-and long-term runoff forecasting using a multimodel approach in the Yellow River headwaters region based on large-scale and local-scale climate information. *Water* 9(8):608
- Demirel MC, Booij MJ, Hoekstra AY (2013) Impacts of climate change on the seasonality of low flows in 134 catchments in the River Rhine basin using an ensemble of bias-corrected regional climate simulations. *Hydrol Earth Syst Sci* 17(10):4241–4257. <https://doi.org/10.5194/hess-17-4241-2013>
- Geem ZW (2011) Transport energy demand modeling of South Korea using artificial neural network. *Energy Policy* 39(8):4644–4650
- Geem ZW, Roper WE (2009) Energy demand estimation of South Korea using artificial neural network. *Energy Policy* 37(10):4049–4054
- Geem ZW, Tseng CL, Kim J, Bae C (2007) Trenchless water pipe condition assessment using artificial neural network. In: *Pipelines 2007: advances and experiences with trenchless pipeline projects*, pp. 1–9
- Gholami H, Lotfifard M, Ashrafi SM, Biazar SM, Singh VP (2023) Multi-GCM ensemble model for reduction of uncertainty in

- runoff projections. *Stoch Environ Res Risk Assess* 37(3):953–964. <https://doi.org/10.1007/s00477-022-02311-1>
- Guérémy J-F, Laanaia N, Céron J-P (2012) Seasonal forecast of French mediterranean heavy precipitating events linked to weather regimes. *Nat Hazard* 12(7):2389–2398. <https://doi.org/10.5194/nhess-12-2389-2012>
- Halgamuge MN, Nirmalathas A (2017) ‘Analysis of large flood events: based on flood data during 1985–2016 in Australia and India. *Int J Disaster Risk Reduct* 24:1–11. <https://doi.org/10.1016/j.ijdr.2017.05.011>
- Huang X, Xu B, Zhong PA, Yao H, Yue H, Zhu F, Liu W (2022) Robust multiobjective reservoir operation and risk decision-making model for real-time flood control coping with forecast uncertainty. *J Hydrol* 605:127334
- Jayawardena AW, Fernando DAK (1998) Use of radial basis function type artificial neural networks for runoff simulation. *Comput-Aided Civ Infrastruct Eng* 13(2):91–99. <https://doi.org/10.1111/0885-9507.00089>
- Kilinc HC, Haznedar B (2022) A hybrid model for streamflow forecasting in the Basin of euphrates. *Water* 14:80. <https://doi.org/10.3390/w14010080>
- Kim G, Barros AP (2001) Quantitative flood forecasting using multisensor data and neural networks. *J Hydrol* 246(1–4):45–62. [https://doi.org/10.1016/S0022-1694\(01\)00353-5](https://doi.org/10.1016/S0022-1694(01)00353-5)
- Kim SK, Lee YONGDAE, Kim JAEHEE, Ko IH (2005) A multiple objective mathematical model for daily coordinated multi-reservoir operation. *Water Sci Technol Water Supply* 5(3–4):81–88
- Krause P, Boyle DP, Båse F (2005) Comparison of different efficiency criteria for hydrological model assessment to cite this version: advances in geosciences comparison of different efficiency criteria for hydrological model assessment. *Adv Geosci* 5:89–97. <https://doi.org/10.5194/adgeo-5-89-2005>
- Legates DR, McCabe GJ (1999) Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35(1):233–241. <https://doi.org/10.1029/1998WR900018>
- Li H, Xie M, Jiang S (2012) Recognition method for mid- to long-term runoff forecasting factors based on global sensitivity analysis in the Nenjiang River Basin. In: *Hydrological processes*, Wiley, vol 26, No (18), pp. 2827–2837. <https://doi.org/10.1002/hyp.9211>
- Liu J, Yuan X, Zeng J, Jiao Y, Li Y, Zhong L, Yao L (2022) Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning. *Hydrol Earth Syst Sci* 26:265–278. <https://doi.org/10.5194/hess-26-265-2022>
- Luo C, Ding W, Zhang C, Yang X (2023) Exploiting multiple hydrologic forecasts to inform real-time reservoir operation for drought mitigation. *J Hydrol* 618:129232
- Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11(7):674–693. <https://doi.org/10.1109/34.192463>
- McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 52:99–115. <https://doi.org/10.1007/BF02459570>
- Moazami S, Abdollahipour A, ZakeriNiri M, Ashrafi SM (2016) Hydrological assessment of daily satellite precipitation products over a basin in Iran. *J Hydraul Struct* 2(2):35–45
- Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: literature review. *Water* 10(11):1536
- Mostaghimzadeh E, Ashrafi SM, Adib A, Geem ZW (2021) Investigation of forecast accuracy and its impact on the efficiency of data-driven forecast-based reservoir operating rules. *Water* 13(19):2737
- Mostaghimzadeh E, Adib A, Ashrafi SM, Kisi O (2022) Investigation of a composite two-phase hedging rule policy for a multi reservoir system using streamflow forecast. *Agric Water Manag* 265:107542
- Nash JE, Sutcliffe JV (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](https://doi.org/10.1016/0022-1694(70)90255-6)
- Ni L et al (2020) Streamflow forecasting using extreme gradient boosting model coupled with gaussian mixture model. *J Hydrol* 586:124901. <https://doi.org/10.1016/j.jhydrol.2020.124901>
- Osman A, Afan HA, Allawi MF, Jaafar O, Nouredin A, Hamzah FM, El-shafie A (2020) Adaptive fast orthogonal search (FOS) algorithm for forecasting streamflow. *J Hydrol* 586:124896
- Palmer TN, Alessandri A, Andersen U, Cantelaube P, Davey M, Delécluse P, Graham R (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull Am Meteorol Soc* 85(6):853–872
- Peng Y, Zhang X, Xu W, Shi Y, Zhang Z (2018) An optimal algorithm for cascaded reservoir operation by combining the grey forecasting model with DDDP. *Water Sci Technol Water Supply* 18(1):142–150
- Peng A, Zhang X, Peng Y, Xu W, You F (2019) The application of ensemble precipitation forecasts to reservoir operation. *Water Supply* 19(2):588–595
- Penland C, Matrosova L (1998) Prediction of tropical Atlantic sea surface temperature, using linear inverse modeling. *J Clim Am Meteorol Soc* 11(3):483–496. [https://doi.org/10.1175/1520-0442\(1998\)011%3c0483:POTASS%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011%3c0483:POTASS%3e2.0.CO;2)
- Phuong TM, Lin Z, Altman RB (2005) Bioinformatics choosing SNPs using feature selection. Available at: <http://htsnp.stanford.edu/FSFS/>. (Accessed: 12 December 2019)
- Polikar R (1999) The story of wavelets. In: *Physics and Modern Topics in Mechanical and Electrical Engineering*. World Scientific and Engineering Academy and Society, pp 192–197
- Quedi ES, Fan FM (2020) Sub seasonal streamflow forecast assessment at large-scale basins. *J Hydrol* 584:124635
- Ren K, Fang W, Qu J, Zhang X, Shi X (2020) Comparison of eight filter-based feature selection methods for monthly streamflow forecasting—three case studies on CAMELS data sets. *J Hydrol* 586:124897
- Schmitt Quedi E, Mainardi Fan F (2020) Sub seasonal streamflow forecast assessment at large-scale basins. *J Hydrol* 584:124635. <https://doi.org/10.1016/j.jhydrol.2020.124635>
- Shamseldin AY (1997) Application of a neural network technique to rainfall-runoff modelling. *J Hydrol* 199(3–4):272–294. [https://doi.org/10.1016/S0022-1694\(96\)03330-6](https://doi.org/10.1016/S0022-1694(96)03330-6)
- Shamseldin AA, Press SJ (1984) Bayesian parameter and reliability estimation for a bivariate exponential distribution parallel sampling. *J Econom* 24(3):363–378. [https://doi.org/10.1016/0304-4076\(84\)90059-9](https://doi.org/10.1016/0304-4076(84)90059-9)
- Shortridge JE, Guikema SD, Zaitchik BF (2016) Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrol Earth Syst Sci* 20(7):2611–2628. <https://doi.org/10.5194/hess-20-2611-2016>
- Sigaroodi SK, Chen Q, Ebrahimi S, Nazari A, Choobin B (2014) Long-term precipitation forecast for drought relief using atmospheric circulation factors: a study on the Maharloo Basin in Iran. *Hydrol Earth Syst Sci* 18(5):1995
- Teschl R, Randeu WL (2006) ‘A neural network model for short term river flow prediction’, natural hazards and earth system sciences. *Eur Geosci Union* 6(4):629–635. <https://doi.org/10.5194/nhess-6-629-2006>
- Tongal H, Booij MJ (2018) Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. *J Hydrol* 564:266–282. <https://doi.org/10.1016/j.jhydrol.2018.07.004>

- Tourigny E, Jones C (2009) An analysis of regional climate model performance over the tropical Americas. Part I: simulating seasonal variability of precipitation associated with ENSO forcing. The Authors Journal Compilation C. Blackwell Munksgaard, vol 61, pp. 323–342. <https://doi.org/10.1111/j.1600-0870.2008.00386.x>.
- Vennerstrøm S, Friis-Christensen E, Troshichev OA, Andersen VG (1991) Comparison between the polar cap index, PC, and the auroral electrojet indices AE, AL, and AU. *J Geophys Res Space Phys* 96(A1):101–113
- Wang YM, Sheeley NR Jr (1997) The high-latitude solar wind near sunspot maximum. *Geophys Res Lett* 24(24):3141–3144
- Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J Hydrol* 374(3–4):294–306
- Weickmann KM, Robinson WA, Penland C (2000) Stochastic and oscillatory forcing of global atmospheric angular momentum. *J Geophys Res Atmos* 105(D12):15543–15557. <https://doi.org/10.1029/2000JD900198>
- Willmott CJ (1981) On the validation of models. *Phys Geogr* 2(2):184–194. <https://doi.org/10.1080/02723646.1981.10642213>
- Willmott CJ (2013) On the evaluation of model performance in physical geography. In: Willmott CJ (ed) *Spatial statistics and models*. Springer, Dordrecht, pp 443–460. https://doi.org/10.1007/978-94-017-3048-8_23
- Wu CL, Chau KW (2010) Data-driven models for monthly streamflow time series prediction. *Eng Appl Artif Intell* 23(8):1350–1367. <https://doi.org/10.1016/j.engappai.2010.04.003>
- Wu L, Seo DJ, Demargne J, Brown JD, Cong S, Schaake J (2011) Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction. *J Hydrol* 399(3–4):281–298
- Yang J, Chang J, Wang Y, Li Y, Hu H, Chen Y, Yao J (2018) Comprehensive drought characteristics analysis based on a nonlinear multivariate drought index. *J Hydrol* 557:651–667

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.