Hydrological Sciences Journal

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/thsj20

A novel approach to predictor selection among large-scale climate indices for seasonal rainfall forecasting in small catchments

Fereshteh Modaresi, Kumars Ebrahimi & Ali Danandeh Mehr

To cite this article: Fereshteh Modaresi, Kumars Ebrahimi & Ali Danandeh Mehr (20 Mar 2024): A novel approach to predictor selection among large-scale climate indices for seasonal rainfall forecasting in small catchments, Hydrological Sciences Journal, DOI: 10.1080/02626667.2024.2313572

To link to this article: https://doi.org/10.1080/02626667.2024.2313572

HYDROLOGICAL SCIENCES JOURNAL
https://doi.org/10.1080/02626667.2024.2313572

A novel approach to predictor selection among large-scale climate indices for seasonal rainfall forecasting in small catchments

Fereshteh Modaresi [a], Kumars Ebrahimi [b] and Ali Danandeh Mehr [c,d]

[a]Department of Water Science and Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad (FUM), Mashhad, Iran; [b]Department of Renewable Energies and Environmental Engineering, University of Tehran, Tehran, Iran; [c]Civil Engineering Department, Antalya Bilim University, Antalya, Turkey; [d]MEU Research Unit, Middle East University, Amman, Jordan

ABSTRACT

Utilizing identical climate indices as predictors for all climate divisions within large basins may result in unreliable rainfall forecasts at the sub-basin scale. This study aimed to develop a new approach to identify the most effective predictors among large-scale climate indices for seasonal rainfall forecasting in small areas. The proposed approach combines a selective singular value decomposition method (SSVD) with a non-linear sequential forward selection method (NLSFS). Applying the new algorithm for seasonal rainfall forecasting within two climate divisions in Karkheh basin, Iran, indicated that the climate indices identified by the SSVD differed between the study areas. The combination of these indices exhibited a correlation with seasonal rainfall approximately 11% higher than those derived from the SVD method. Moreover, NLSFS significantly enhanced the forecast accuracy compared to the frequently employed linear sequential forward selection (SFS) method, and the optimal predictors chosen by the two methods differed across all seasons.

ARTICLE HISTORY
Received 29 January 2023
Accepted 16 January 2024

EDITOR
S. Archfield

ASSOCIATE EDITOR
A. Agarwal

KEYWORDS
hybrid model; non-linear sequential forward selection; selective singular value decomposition; input selection; Karkheh; Seimareh

1 Introduction

Rainfall is one of the most important components of the hydrological cycle and can impact catchment runoff significantly. It is well-documented that accumulated rainfall amount over a long time horizon (i.e. monthly, seasonal, and annual rainfall) may be influenced by the interaction between oceanic and atmospheric cycles, aka climatic oscillations (Mohammadi *et al.* 2020, Xie *et al.* 2021, Wu *et al.* 2022). For example, the effect of El Niño Southern Oscillation (ENSO) on seasonal rainfall in Eastern Africa (Vashisht *et al.* 2021), Southern America (Cai *et al.* 2020), Australia (Ma *et al.* 2023), Eastern Asia (Ng *et al.* 2019), Western Europe (Martija-Díez *et al.* 2021), and Iran (Dehghani *et al.* 2020, Bahrami *et al.* 2021) has been reported in the relevant literature. Similarly, the influence of the Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), and Atlantic Multi-decadal Oscillation (AMO) on long-term rainfall variation over the US (Liu *et al.* 2018, Zhao and Brissette 2022), Western Europe (Fernández-González *et al.* 2012, Tabari and Willems 2018), Morocco (Marchane *et al.* 2015), and Iran (Ruigar and Golian 2015, Dehghani *et al.* 2020) has been shown. Several studies have also proved the effect of sea surface temperature (SST) on local and regional rainfall events. Examples include but are not limited to the studies demonstrating the effect of Pacific and Indian Ocean SST on the Australian monsoon (Heidemann *et al.* 2022); the Indian Ocean SST on summer rainfall across India (Vibhute *et al.* 2020); the Atlantic Ocean SST on winter rainfall over Northern Europe, West Africa, and South America (Vallès-Casanova *et al.* 2020, Börgel *et al.* 2022); the Mediterranean Sea SST on rainfall across Turkey (Desbiolles *et al.* 2021), the Sahel (Park *et al.* 2016), Greece (Kassomenos and Mcgregor 2006), and Iran (Meidani and Araghinejad 2014); and the Persian Gulf SST on rainfall in the west and south of Iran (Nasemosadat 1998). Relying on teleconnections between climate oscillations/SSTs and rainfall patterns, many recent studies have shown that large-scale climate indices can be used as predictors for long-term rainfall forecasting. For instance, Kalra and Ahmad (2012) utilized the PDO, NAO, AMO, and ENSO series to forecast annual rainfall in the Colorado River basin. Schepen *et al.* (2012) pointed out that PDO and the South Oscillation Index (SOI) can be used for seasonal rainfall forecasting across Australia.

Undoubtedly, rainfall forecasting is one of the most challenging tasks in the hydrological forecasting community, as the level of uncertainty in rainfall patterns is significantly higher than those of other variables such as temperature and streamflow (Danandeh Mehr 2021, Fayaz *et al.* 2022). For appropriate watershed management, it is important to forecast the amount of monthly, seasonal, and even annual rainfall as it benefits the authorities in the decision-making process concerning water deficit, droughts, and the optimal operation of hydro systems, such as dams and reservoirs, and change in cultivation/land use pattern for a certain period so that precious resources can be saved or allocated for high-priority demands (Li *et al.* 2020, Zhu *et al.* 2022, Asghari Saraskanrood *et al.* 2023). In addition, flooding is a major challenge in watershed management that emphasizes the need for accurate rainfall forecasting models (Hosseini *et al.* 2021, Oborie and Rowland 2023). Since a large

CONTACT Fereshteh Modaresi fmodaresi@um.ac.ir Department of Water Science and Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad (FUM), Mashhad 9177948974, Iran

portion of Iran is in an arid and semi-arid climate facing both surface and groundwater scarcity, medium- to long-term forecasts of rainfall amount are also crucial for food production, water allocation, and developing an optimal irrigation system.

Focusing on Iran's climate, only a few studies have used large-scale climate indices for rainfall prediction across the country; however, the results were promising. For instance, Fallah-Ghalhary *et al.* (2009) showed that pressure differences between the Persian Gulf and the Adriatic Sea, the Oman Sea, and the Red Sea could be efficiently used for six-month rainfall forecasting in eastern Iran. Nazemosadat *et al.* (2010) reported that Persian Gulf SST is a dominant predictor for autumn and winter rainfall in the southwest of Iran. In another study, NAO and SOI, as well as the SST and pressure variation across the Persian Gulf, were reported as suitable predictors for long-lead rainfall forecasting in the Karoon River basin (Sarah *et al.* 2011). More recently, Danandeh Mehr *et al.* (2017) demonstrated that SST of the Black Sea, Mediterranean Sea, and Red Sea can be successfully used for maximum monthly rainfall prediction in the northwest of Iran. Also, Choubin *et al.* (2017) revealed that AMO, SOI, and East Central Tropical Pacific Sea surface temperature (NINO3.4) are the most effective climate indices for the prediction of winter and spring rainfall in Maharloo-Bakhtegan basin in the southwest of Iran.

Since the underlying patterns linking oceanic–atmospheric oscillations and rainfall features are often complex and inherently non-linear, artificial intelligence (AI) methods are commonly implemented to develop rainfall forecasting models based on climate oscillations (e.g. Danandeh Mehr *et al.* 2017). Artificial neural networks (ANNs) and support vector regression (SVR) are some of the notable AI methods commonly implemented for this task. The structure of ANN models is based on empirical risk minimization, which minimizes total error (Modaresi and Araghinejad 2014). Since ANNs have various architectures, selecting an optimum architecture involves a time-consuming trial-and-error process (Samadi *et al.* 2015, Abdullah *et al.* 2023). In contrast, SVR models are based on structural risk minimization in which overfitting is avoided via minimizing empirical risk. Thus, they are more suitable for generalization tasks (Modaresi *et al.* 2018). In addition, the architecture of any SVR model is expressed by a quadratic optimization problem which can be solved by standard programming algorithms (Pradhan 2012, Li and Sun 2020). Therefore, the SVR model was found to be more efficient than ANNs for hydrological and meteorological forecasting (Modaresi *et al.* 2018, Ghasemi *et al.* 2023).

Choosing proper inputs for AI models is a crucial task as they significantly affect the model's accuracy. In previous studies, different input selection criteria such as Pearson correlation coefficients (PCC) (Modaresi *et al.* 2016, Mohammadi *et al.* 2020), principal component analysis (PCA) (Noori *et al.* 2011, Choubin *et al.* 2017), singular value decomposition (SVD) analysis (Lin *et al.* 2019, Ali *et al.* 2020, Bhanu 2021), the forward selection method (Noori *et al.* 2011, Dehghani *et al.* 2014, Selvi and Huseyinov 2020) and the mutual information (MI) index (Modaresi *et al.* 2016, Danandeh Mehr *et al.* 2022) were used to reduce the number of predictors and detect more effective inputs. In a recent study, Ren *et al.* (2020) compared the impact of PCC, MI, and gamma test criteria on input selection for three streamflow forecasting models, namely multiple linear regression, K-nearest neighbour, and ensemble extreme learning machine. The authors demonstrated that there was no dominant selection method for ensemble extreme learning machine or K-nearest neighbour models; however, the partial Pearson coefficient index outperforms its counterparts when resampling is used.

Considering the previous studies related to seasonal rainfall forecasting using climate indices, this study addressed the following identified gaps:

- Climate indices are commonly used for rainfall forecasting at regional scale/large basins. However, applying the same indices for small catchments within a basin may lead to unreliable rainfall forecasts.
- Despite the critical role of spring rainfalls in providing the water needed for rainfed cultivation in Iran, especially in the Khuzestan Plain, the previous studies have instead investigated the predictability of autumn and winter rainfall using climate indices.
- So far, the SVD method has been used to combine features, not to detect the most effective ones, while the type of input variables fed to the SVD model affects the correlation between its output series.
- The sequential forward selection (SFS) method is based on linear modelling. When it is employed as a predictor selection method for a predominantly non-linear process, it could result in incorrect predictors.

To address the identified limitations in predicting seasonal rainfall in localized regions and to enhance forecast accuracy, this study was crafted to formulate a resilient method for selecting optimal predictors in seasonal rainfall forecasting through the incorporation of extensive climate indices. We developed a novel hybrid AI method that effectively synergizes a Selective Singular Value Decomposition technique (SSVD) with a Non-linear SFS (NLSFS) approach. The SSVD identifies the most influential climate indices, while the NLSFS establishes the optimal combination of these indices. The proposed algorithm was applied to forecast autumn, winter, and spring rainfall in two sub-basins of a large basin in Iran, characterized by different climatic conditions.

2 Study area and data

Karkheh basin is one of the 30 main hydrological basins of Iran, located in southwestern Iran and included in the Persian Gulf basin (Fig. 1). It lies between latitudes 33°40' and 35°00'N and longitudes 46°23' and 49°12'E, and comprises five provinces: Hamedan, Kermanshah, Ilam, Lorestan, and Khuzestan. The Khuzestan Plain, which is the most important agricultural core of Iran, is situated in the south of this basin.

The weather of the Karkheh basin ranges from mountainous (very cold) with cold winters and mild summers in the northern regions to hot with mild winters and long and warm summers in the southern regions, while its humidity is diverse, from arid to humid (Fig. 1). The average annual rainfall over this basin varies from 300 to 800 mm. The major sources of rainfall in this basin are the moist air flows originating from

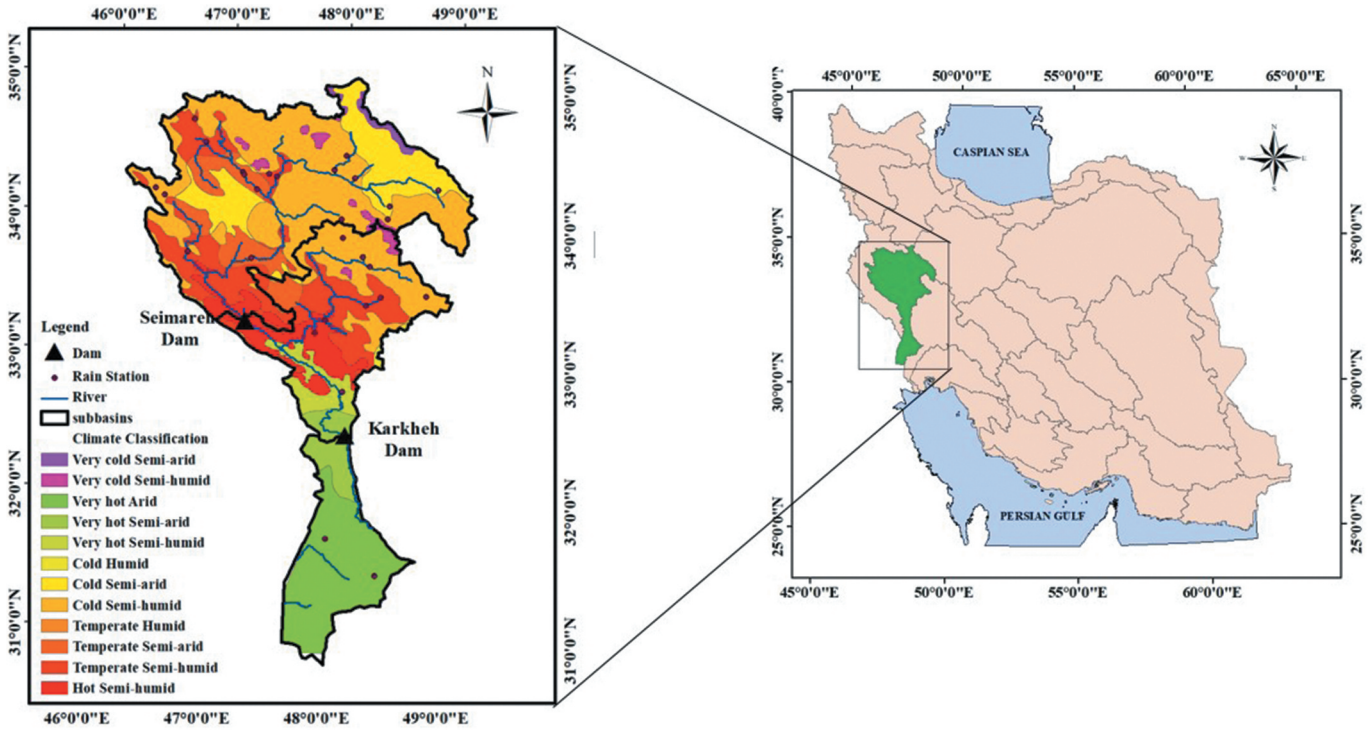

Figure 1. Location map of the study areas in Iran, illustrating the location and climatic variations (Asadi Oskouei *et al.* 2022) of Karkheh and Seimareh sub-basins at the upstream of the respective dams, and the location of the studied meteorological stations.

the Mediterranean Sea (30°15'–36°53'N, 10°01'–36°10'E) and the Persian Gulf (24°00'–30°00'N, 47°56'–56°30'E), with an annual frequency of 64.5% and 22.9%, respectively (Modaresi *et al.* 2016). Owing to the presence of different climates across the basin (Fig. 1), and the effect of diverse climate indices on the weather of the Karkheh basin (Dezfuli *et al.* 2010, Tabari *et al.* 2014, Modaresi *et al.* 2016), this proposed model was applied in the upstream Seimareh and Karkheh sub-basins, individually, to assess its performance in different conditions. As illustrated in Fig. 1, a total of 27 raingauge stations (18 at Seimareh and nine at Karkheh) are available in the study area.

In addition, the seasonal average of the climate indices including ENSO (i.e. SOI and NINO3.4; SOI data is available at: http://www.bom.gov.au/climate/enso/ and NINO 3.4 data is available at: https://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/sstoi.indices), NAO (available at: http://www.cgd.ucar.edu/cas/jhurrell/indices.html), PDO (available at: http://research.jisao.washington.edu/pdo/PDO.latest), and AMO (available at: http://www.esrl.noaa.gov/psd/data/correlation/amon.sm.long.data), as well as the seasonal averages of the SST time series of the Mediterranean Sea and the Persian Gulf (available at: http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html) were used to determine the optimum combinations of large-scale predictors. It is worth remembering that the monthly time series of all indices are available in the given links, while monthly SST series are measured at a spatial resolution of 1° × 1° (i.e. 101 and 26 nodes over the Mediterranean Sea and the Persian Gulf, respectively). The period of available data from the climate indices also differs; The shortest length belongs to the SST series of the Mediterranean Sea and Persian Gulf as well as the time series of Nino 3.4, which are available since 1982. But reliable local rainfall data for both studied areas were available from 1987. Therefore, data for the period 1987–2019 were used in this study, and all the data were normalized and de-trended. The first to fourth lags of the seasonal SST series (of all nodes) and the climate indices during this period were used as initial predictors.

3 Methodology

3.1 Selective singular value decomposition (SSVD) method

SVD is a data mining technique based on the theory of linear algebra. It is a powerful statistical method that can produce the most correlated series from a large amount of data. Due to the use of a cross-covariance matrix, this method can extract the most correlated data series from spatiotemporal variables like SST (Lin *et al.* 2019). Also, in the outputs of the SVD method, the weights of the time series in the process of combination and creating the correlated data series are recognizable. With respect to this ability, the SVD method was applied in the current research for two purposes: first, to detect the most appropriate climate indices and SST nodes of the Mediterranean Sea and Persian Gulf for each study area; and, second, to make an ensemble of the chosen indices or SST nodes to create time series that exhibits the highest correlation with seasonal rainfall.

SSVD is an advanced SVD method in which the most effective predictors are detected and combined through the following algorithm. Based on the SVD method, the cross-covariance matrix ($COVAR_{gs}$) of the standardized values of predictors (such as the seasonal SST of the Mediterranean Sea

at 101 points) and predicted variables (such as the seasonal rainfall from 18 rain stations in the Seimareh sub-basin) can be broken down as follows:

$$SVD(COVAR_{gs}) = U_{gg}S_{gs}V_{ss}^{T} \quad (1)$$

where U_{gg} and V_{ss} are two orthogonal matrices, containing singular vectors, and S_{gs} is a diagonal matrix with the singular values arranged in descending order.

In order to compare the relative importance of modes based on the singular values, the squared covariance fraction (*SCF*) is calculated as a useful measurement as follows (Bretherton *et al.* 1992):

$$SCF_j = \frac{C_j^2}{\sum_{j=1}^{\min(g,s)} C_j^2} \quad (2)$$

where C_j represents each of the singular values placed on the main diameter of the S_{gs} matrix. If the *SCF* value of the first mode is too high relative to the other modes, the combined series of SST values as well as the rainfall series which have the highest correlation with each other can be achieved using the coefficients in the first column of the singular vector matrices – U_{gg} and V_{ss}, respectively. For example, the combined series of SST (SSTP) and rainfall (R) for the first mode is calculated with $i = 1$ in Equation (3):

$$\begin{aligned} SSTP(i) &= \begin{bmatrix} sst_{1,1} & \cdots & sst_{1,g} \\ \vdots & \ddots & \vdots \\ sst_{y,1} & \cdots & sst_{y,g} \end{bmatrix} U(:,i) \\ R(i) &= \begin{bmatrix} r_{1,1} & \cdots & r_{1,s} \\ \vdots & \ddots & \vdots \\ r_{y,1} & \cdots & r_{y,s} \end{bmatrix} V(:,i) \end{aligned} \quad (3)$$

where g is the number of SST grid cells, y is the number of years and s represents the number of rain stations, while i can range from 1 up to the maximum number of significant modes.

The SVD output value placed in each column of U matrix plays the role of weights for predictors (like climate indices); so that a higher value represents a higher effect of that index. On this basis, in the SSVD method proposed in this research, the more effective predictors were identified based on their weights, and several combinations of them were tested and evaluated so that the best combination series having the most correlation with the predicted variable was achieved. The combined series produced by this method had a higher correlation than that obtained from SVD method.

3.2 Non-linear sequential forward selection (NLSFS) method

The NLSFS is a new version of the traditional SFS method in which a non-linear model is utilized for predictor selection. The SFS is originally a linear technique in which the most appropriate subset of predictors, as model input, are chosen to forecast or model another variable, as model output. In this method, first, each of the predictors is given as input to the linear model and the best of them (S1) is selected based on the model results with the least error (S1 = [Pi]). In the second step, other predictors are, one by one, paired with the best predictor selected in the previous step (S1), and each of the pairs is used as input to the linear model. If the model error in this step is lower than in the step before, the corresponding pair is selected as the best feature (S2 = [Pi,Pj]). In the next step, the remaining predictors are added to S2 and the best set of three predictors having the minimum error as compared to the previous steps is chosen (S3 = [Pi,Pj,Pk]). This process is continued until a non-predefined number of predictors that lead to the best linear modelling result are selected as the final best subset of predictors.

Since the relationship between climate features and seasonal rainfall could have a non-linear form, in the current study, an NLSFS method based on the least squares–support vector regression model (LS-SVR), which has several kernel functions ranging from linear to polynomial and radial basis functions, was applied in order to detect the best subset of predictors and the results were compared to those from the SFS method.

3.3 Least squares-support vector regression (LS-SVR)

LS-SVR is a type of SVR model in which the least squares method is implemented to find the hyper planes at the the maximum distance from the support vectors, nearest the observed data, on both sides (Modaresi and Araghinejad 2014).

In this method, a non-linear mapping of ϕ in the trait space for $X_t \in R^m$ as the input data and $Y(X_t) \in R$ as the output data is computed as follows (Suykens *et al.* 2002):

$$Y(X_t) = w^T.\phi(X_t) + b \quad (4)$$

where w and b are the value of weights and biases of the regression function, respectively, calculated by minimizing the following objective function:

$$\begin{aligned} \underset{w,b,e_i}{Min}\, j(w,e) &= \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{t=1}^{n} e_t^2 \\ S.t: T_t &= w^T\phi(X_t) + b + e_t t = 1,2,\ldots,n \end{aligned} \quad (5)$$

where e is the amount of the errors, and γ is the regularization parameter of the model that controls the flatness of the approximation function, while its optimum amount should be determined by the user.

Solving the objective function using the Lagrangian function based on the Karush-Kuhn-Tucker condition leads to the following relation:

$$Y(X) = \sum_{t=1}^{n} \alpha_t K(X, X_t) + b \quad (6)$$

where α_t (t = 1,2, ..., n) are Lagrange multipliers or support values, and $K(X,X_t)$ is a kernel function, having three types of functions – linear function, polynomial function and radial basis function (RBF) – as follows:

$$\text{Linear}: K(x_i, x_j) = x_i^T x_j \tag{7}$$

$$\text{Polynomial}: K(x_i, x_j) = (x_i^T x_j + t)^d \tag{8}$$

$$\text{RBF}: K(x_i, x_j) = exp(-\|x_i - x_j\|^2 / S^2) \tag{9}$$

where t and d are the polynomial parameters, and S is the RBF parameter. In the current paper, all three kernel functions were applied and assessed to find the best combination of the predictors. Regarding the amount of data for model calibration, the optimum values of parameters were calculated via the leave-one-out cross-validation (LOOCV) method (Bramer 2020) to achieve a reliable structure for the model.

3.4 Assessment criteria

To recognize the best subset of predictors, the results of the LS-SVR model using the NLSFS method were evaluated using Nash-Sutcliffe efficiency (NSE), root mean square error (RMSE) and PCC assessment criteria:

$$NSE = 1 - \frac{\sum_{t=1}^{n}(T_t - Y_t)^2}{\sum_{t=1}^{n}(T_t - \bar{T})^2} \tag{10}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(T_t - Y_t)^2}{n}} \tag{11}$$

$$PCC = \frac{\sum_{t=1}^{n}(T_t - \bar{T})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^{n}(T_t - \bar{T})^2}.\sqrt{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}} \tag{12}$$

where Y_t and T_t are forecasted and observed values of rainfall for t^{th} data, $\bar{T}$ and $\bar{Y}$ are the average of observed and forecasted values of rainfall, respectively, and n is the number of data.

NSE ranges from $-\infty$ to 1.0, indicating the worst and best model performance, respectively. Values from 0.0 to 1.0 indicate an acceptable performance whilst values lower than zero reveal that the average of the observed data is a better estimation than the forecasted values, which indicates unacceptable performance (Moriasi *et al.* 2007). The value of RMSE varies from 0.0 to $+\infty$, while PCC ranges from -1.0 to 1.0. Lower RMSE values and higher PCC values indicate greater model reliability.

3.5 The proposed input selection approach

In this study, a hybrid process (shown in Fig. 2) was introduced to detect the optimal predictors for improving the accuracy of seasonal rainfall forecasting in small catchments. As illustrated in the figure, first, the collected climate indices and SST series were processed by the SSVD method. In this process, the most effective climate indices and SST nodes are identified and an ensemble series of them is produced using the weights calculated with the SSVD method, so that the new time series provides the highest correlation with the historical seasonal rainfall.

In the next step, the individual climate/SST and ensemble series which had a reliable correlation at the confidence level of 98%, based on the statistical t-test method, with seasonal rainfall for one to four seasons ahead were chosen as the initial appropriate predictors. Finally, the initially selected predictors resulting from the previous steps were assessed one by one through the NLSFS method performed by an LS-SVR model which has an optimized structure according to three types of kernel functions. The best predictors for each season are eventually determined with respect to NSE, RMSE, and PCC.

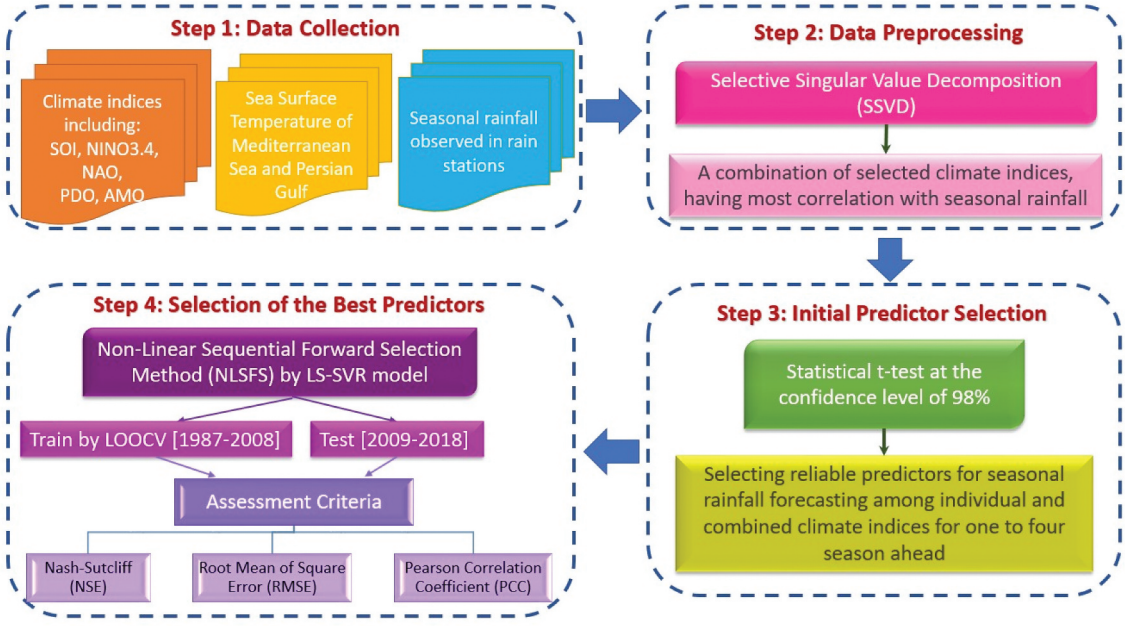

Figure 2. Flowchart of the research steps. These four steps were performed for selection of the optimal predictors for each of the autumn, winter and spring rainfalls in each of the studied sub-basins.

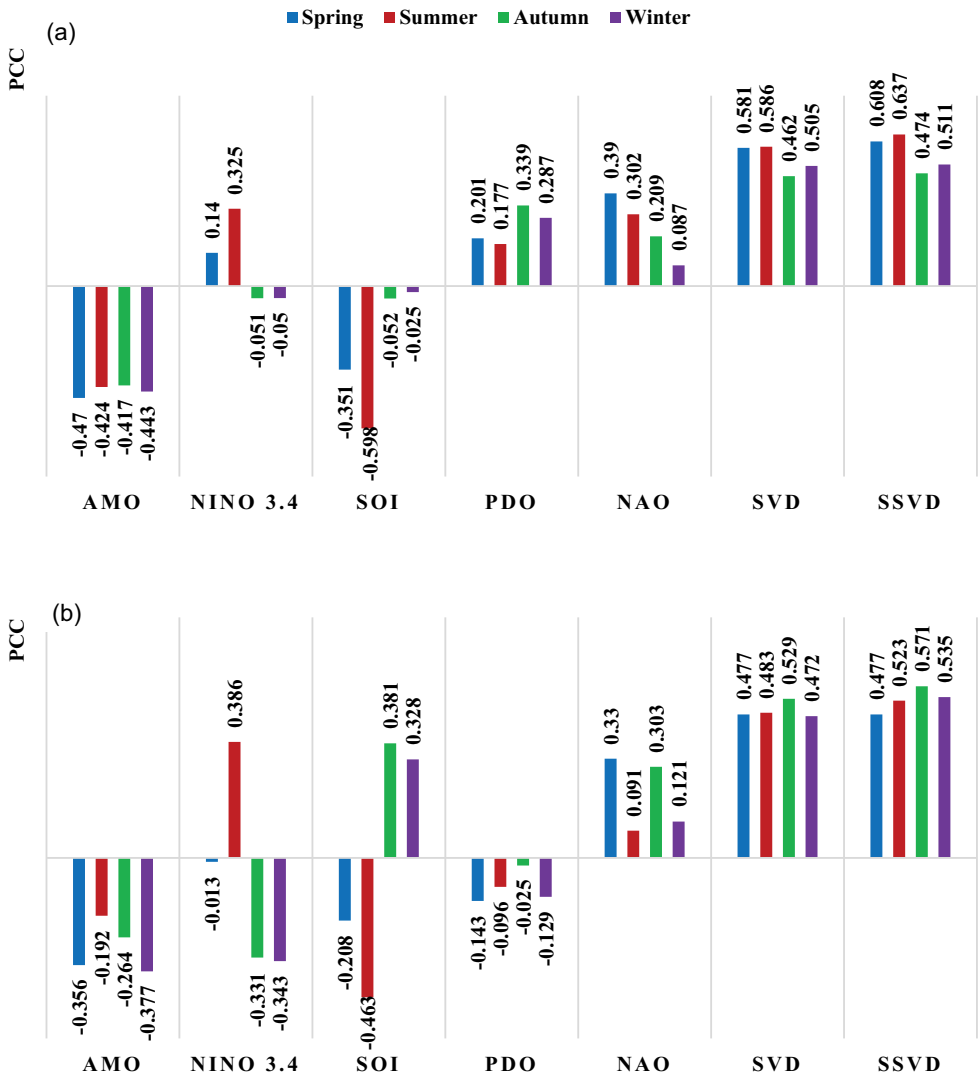

Figure 3. The Pearson correlation coefficient (PCC) of single and combined climate indices with autumn rainfall of (a) Seimareh and (b) Karkheh sub-basins.

4 Results

4.1 Results of SSVD analysis

4.1.1 Results of assessing climate indices as rainfall predictors

To assess the effect of each climate index and their combinations obtained via the SVD and SSVD methods, the PCC was calculated for delays of one to four seasons at each sub-basin (see Figs 3–5 for the autumn, winter, and spring rainfall, respectively). It is worth noting that the results of the SVD and SSVD methods are based on the first mode SCF which was above 90% in all cases. It can be seen from Figs 3–5 that the effects of the seasonal fluctuations of the studied climate indices on the seasonal rainfall of the sub-basins were generally different in two aspects. There are different positive and negative correlations between the seasonal climate indices and seasonal rainfall in different seasons in each sub-basin. Besides, the sign and value of the correlations between rainfall and climate indices vary spatially in each season.

According to Fig. 3, the effect of the seasonal average of the AMO index in all seasons on the autumn rainfall was similar in both sub-basins in terms of negative correlation; such conditions also existed for the NAO index but in terms of positive correlation. However, according to the other indices, there were no similar conditions in both sub-basins in that the correlation of PDO fluctuations in all seasons with the autumn rainfall was positive in Seimareh while it was negative in Karkheh. Besides, the fluctuations of the SOI in all seasons had negative correlations with autumn rainfall in Seimareh while there were positive correlations between autumn and winter averages of this index and autumn rainfall in Karkheh. In addition, there was a dissimilarity between the behaviour of the spring average of NINO3.4 and the autumn rainfall in the studied areas.

Figure 4 demonstrates positive correlations between the winter rainfall in both sub-basins and the seasonal average of AMO and NINO3.4, while the ratio of their correlation coefficients in different seasons was not similar. The type of correlation between the winter rainfall and the other three climate indices in some seasons differed in terms of the sign of the correlations. This was recognizable in the winter fluctuation of SOI and NAO as well as the spring fluctuation of PDO.

In addition, Fig. 5 shows that the seasonal average of SOI and AMO indices in all seasons other than winter for AMO

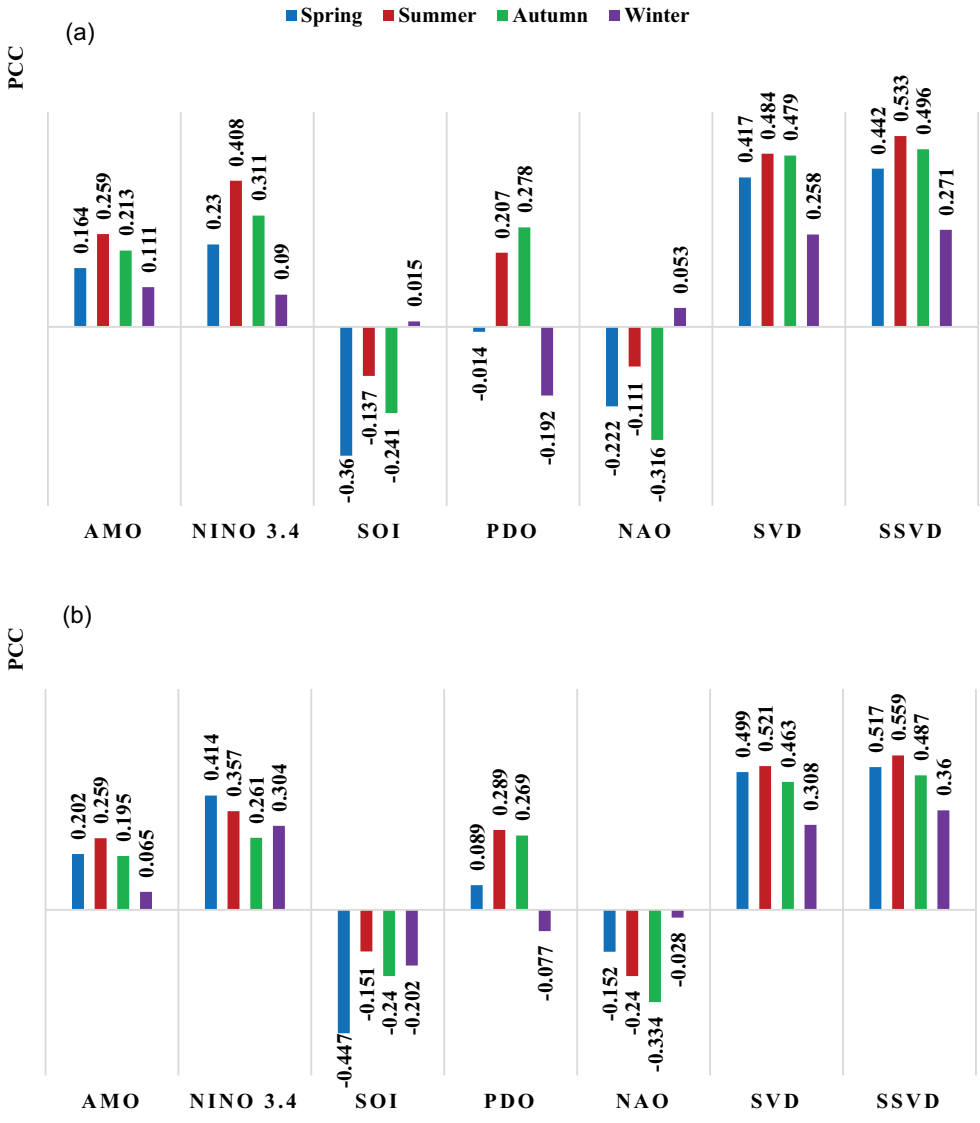

Figure 4. The Pearson correlation coefficient (PCC) of single and combined climate indices with winter rainfall of (a) Seimareh and (b) Karkheh sub-basins.

had a negative correlation with the spring rainfall in both studied areas. The same condition occurred for the PDO index in terms of positive correlation. The seasonal fluctuations of NINO 3.4 also had a positive correlation with spring rainfall in both sub-basins from summer to winter, whereas its correlation type in the spring was different in the studied areas. The type of correlation between the seasonal average of NAO and the spring rainfall was different in various seasons: there was a positive correlation in the summer and winter, and a negative one in the spring in both sub-basins. On the other hand, the autumn average of NAO had a negative correlation with Karkheh rainfall, but there was a positive correlation between it and Seimareh rainfall.

With respect to Figs 3–5, there were positive and negative correlations at different confidence levels between the seasonal average of climate indices and seasonal rainfall in both studied areas. However, the combined time series resulting from the SSVD method had dramatically positive correlations with all seasonal rainfall in both sub-basins, and they were mostly reliable at the significance level of 98% (PCC > 0.41) for the autumn and winter rainfall. Moreover, the results indicated that in almost all cases, the time series obtained from the SSVD method were on average 11% more correlated to the seasonal rainfall as compared to the SVD series. Table 1 presents the climate indices included in the combined series of the SSVD method.

With respect to Table 1, it can be inferred that the climate indices chosen by the SSVD method as the most effective ones on the seasonal rainfall were not the same for the three seasons and for the studied sub-basins, which may be due to the different climate of the studied areas. Furthermore, an assessment of the selected indices regarding their correlation with the seasonal rainfall revealed that they individually do not necessarily have a reliable correlation with a high confidence level. This could be observed, for instance, in the spring, summer, and autumn combinations of climate indices for winter rainfall in both sub-basins. In addition, the spring series of the NINO 3.4 had a reliable correlation with the winter rainfall of the Karkheh sub-basin while it was not included in the SSVD series. This is due to the data analysis method in the

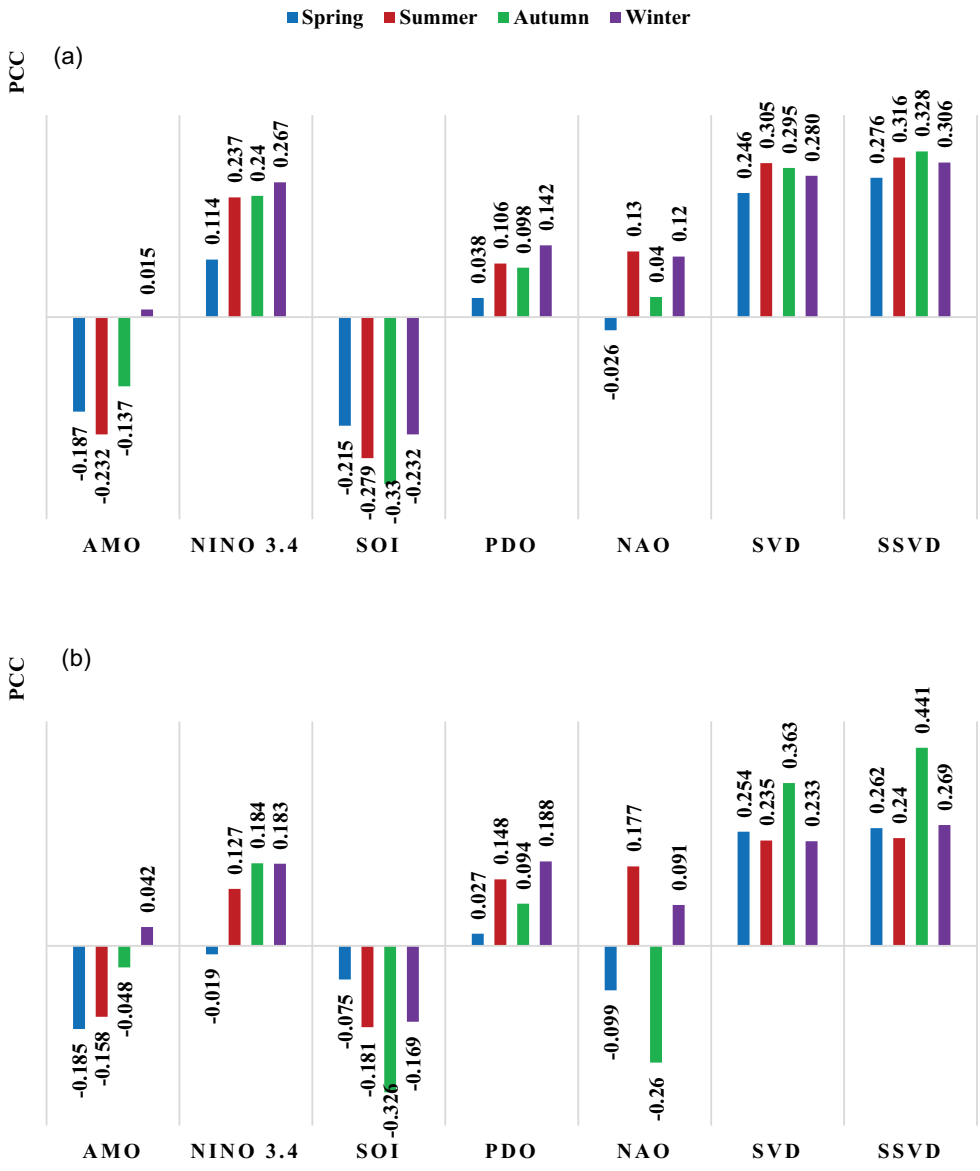

Figure 5. The Pearson correlation coefficient (PCC) of single and combined climate indices with spring rainfall of (a) Seimareh and (b) Karkheh sub-basins.

SSVD, which is based on the relationship among both the predictors and predictands. Therefore, only the correlation coefficient between climate indices and the seasonal rainfall cannot be considered as the index's selection criterion.

4.1.2 Results of assessing SSTs as rainfall predictors

To assess the effect of the SST of each sea on the seasonal rainfall of the sub-basins, the correlation coefficient was calculated between seasonal rainfall and the average SST series as well as the SSVD series of the nodes of the Mediterranean Sea and the Persian Gulf, separately with one to four seasons' delay. For instance, the results for the spring season are presented in Fig. 6 (see the Appendix, Figs A1 and A2 for autumn and winter rainfall, respectively). Figure 6 shows that the SST of the Mediterranean Sea in all seasons had a similar effect on the spring rainfall of both studied areas, although the values of its correlations in all seasons were not equal. However, the effect of the average SST of the Mediterranean Sea in all seasons on the autumn and winter rainfall of the Seimareh sub-basin was almost contrary to that of the Karkheh basin in terms of sign of correlation (see the Appendix, Figs A1 and A2). Thus, the associated PCC values are not reliable at the confidence level of 98%. Evaluation of the Persian Gulf average SST revealed that it had a similar effect on the winter and spring rainfall of both sub-basins, while there was no such condition for the autumn rainfall.

Despite the existence of positive and negative correlations, the implementation of the SSVD method for SST of the seas resulted in creating seasonal combined series of each sea SST that all had positive correlations with seasonal rainfall while their values were more than the absolute value of the correlations between average SST series and rainfall. However, only the spring rainfall of both studied areas (Fig. 6) was affected by the SST of the two seas in winter (Mediterranean Sea) and autumn (Persian Gulf) at the 98% confidence level.

Furthermore, the evaluation of the weights of SST nodes in the series resulting from the SVD method for the spring rainfall revealed those in the Mediterranean Sea between latitudes

Table 1. The most effective combination of seasonal climate indices attained for seasonal rainfall forecasting in each sub-basin.

		Seimareh sub-basin			Karkheh sub-basin		
		Seasonal rainfall			Seasonal rainfall		
Seasonal average of climate indices		Spring	Autumn	Winter	Spring	Autumn	Winter
Spring	AMO	✓	✓	✓	✓	✓	✓
	Nino 3.4				✓	✓	
	SOI	✓	✓	✓	✓	✓	✓
	PDO		✓	✓		✓	
	NAO	✓	✓	✓	✓	✓	
Summer	AMO	✓	✓	✓			✓
	Nino 3.4	✓		✓			✓
	SOI	✓	✓		✓	✓	
	PDO			✓	✓	✓	✓
	NAO	✓	✓		✓		✓
Autumn	AMO		✓	✓	✓	✓	✓
	Nino 3.4		✓	✓			
	SOI	✓			✓	✓	✓
	PDO		✓	✓			✓
	NAO	✓	✓	✓	✓	✓	✓
Winter	AMO	✓	✓		✓	✓	
	Nino 3.4	✓	✓	✓			✓
	SOI					✓	
	PDO		✓	✓	✓		✓
	NAO	✓	✓		✓		✓

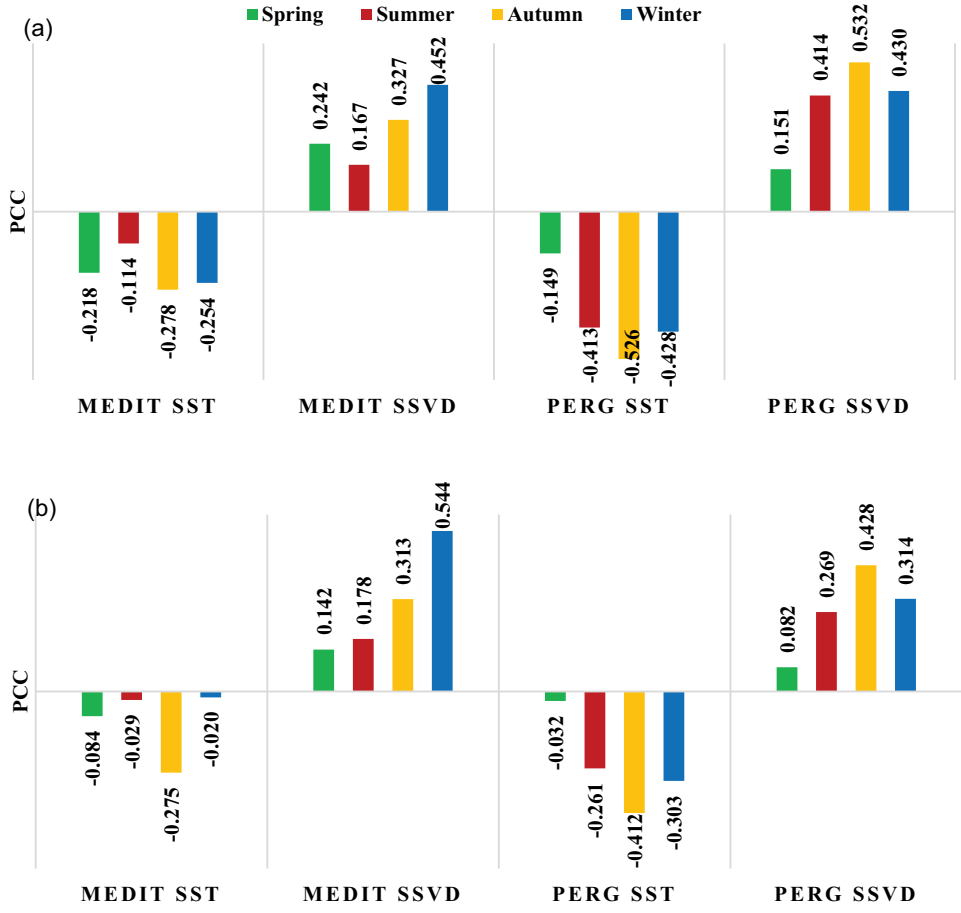

Figure 6. The Pearson correlation coefficient (PCC) of spring rainfall with seasonal average SST of Mediterranean Sea (MEDIT SST) and Persian Gulf (PERG SST) and their combination series by SSVD method (MEDI SSVD and PERG SSVD) for (a) Seimareh and (b) Karkheh sub-basins.

32°30' and 34°30'N, and longitudes 29°30' and 34°30'E were the highest (>0.4) which were included in the combined series of the winter SST by the SSVD method for both studied sub-basins. The identified section of the Mediterranean Sea is the easternmost section and the nearest one to Iran. But the weights of all nodes of the Persian Gulf were almost the same in the autumn SST series resulting from the SVD method for both sub-basins. As a result, all of them were used in the SSVD method.

Table 2. Initial selected predictors for seasonal rainfall forecasting of the Seimareh sub-basin.

	Autumn rainfall		Winter rainfall		Spring rainfall	
Predictor ID	Predictor name	*PCC*	Predictor name	*PCC*	Predictor name	*PCC*
P1	Summer SSVD	0.637	Summer SSVD	0.533	Autumn PERG SSVD	0.532
P2	Spring SSVD	0.607	Autumn SSVD	0.496	Winter MEDIT SSVD	0.452
P3	Winter SSVD	0.511	Spring SSVD	0.442	Winter PERG SSVD	0.430
P4	Autumn SSVD	0.474			Summer PERG SSVD	0.414

Table 3. Initial selected predictors for seasonal rainfall forecasting of the Karkheh sub-basin.

	Autumn rainfall		Winter rainfall		Spring rainfall	
Predictor ID	Predictor name	*PCC*	Predictor name	*PCC*	Predictor name	*PCC*
P1	Autumn SSVD	0.571	Summer SSVD	0.559	Winter MEDIT SSVD	0.555
P2	Winter SSVD	0.535	Spring SSVD	0.517	Autumn PERG SSVD	0.430
P3	Summer SSVD	0.523	Autumn SSVD	0.487		
P4	Spring SSVD	0.477	Spring Nino 3.4	0.414		

4.2 Results of initial selection of the predictors

The initial selection of the appropriate predictors was done from: (1) all the SSVD combined series of climate indices, (2) the single seasonal average series of each climate index that was not included in the SSVD combined series, and (3) the SSVD combined series of the Mediterranean Sea and the Persian Gulf SST. The selection process was based on the Student's t-test of the PCC at the confidence level of 98%. On this basis, the predictors having PCC > 0.41 were selected as the initially appropriate predictors. Tables 2 and 3 display the initial selected predictors.

According to Tables 2 and 3, it can be suggested that all the reliable predictors for seasonal rainfall forecasting of both sub-basins were the combined series of climate indices for autumn and winter rainfall, and the SST of the Mediterranean Sea and the Persian Gulf for spring rainfall while they resulted from the SSVD method. In addition, the spring series of the NINO 3.4, which had a reliable correlation with the winter rainfall of the Karkheh sub-basin and was not included in the SSVD series, was considered individually as a winter rainfall predictor for the Karkheh sub-basin.

4.3 Results of the NLSFS method

Although the predictor selection was based on the value of PCC, the difference of the PCC values from ±1 suggested that the relation between the predictors and predictand was not completely linear. So, the NLSFS method based on LS-SVR with three types of kernel functions (linear, polynomial, and RBF) was run using the initially selected predictors (see Section 3.3) to detect the best predictors for seasonal rainfall prediction in each sub-basin. To this end, the LS-SVR model was calibrated (trained) based on the LOOCV method using 22 years of data and then validated (tested) using 10 years of data (70% and 30% of seasonal data) in each of the NLSFS steps. For example, results for the spring season in the calibration and validation phases including three assessment criteria (NSE, RMSE, and PCC), as well as the best kernel type and the value of its parameters and the gamma parameter, are presented in Tables 4 and 5, for Seimareh and Karkheh sub-basins, respectively (see the Appendix, Tables A1–A4 for the results of the autumn and winter). The results indicated that the connection between all predictors and seasonal rainfall in both examined regions was non-linear, as optimal kernel func-

Table 4. NLSFS results for spring rainfall forecasting in the Seimareh sub-basin.

				Calibration			Validation		
Predictors	Best kernel	Best gamma	Best parameters	NSE	RMSE	PCC	NSE	RMSE	PCC
P1*	**RBF**	**0.4**	**S^2 = 40**	**0.411**	**0.782**	**0.698**	**0.278**	**0.791**	**0.654**
P2	RBF	2.1	S^2 = 65	0.220	0.900	0.493	0.219	0.823	0.654
P3	RBF	8.5	S^2 = 10	0.279	0.866	0.533	0.186	0.840	0.438
P4	Polynomial	0.2	t = 0.1 d = 5	0.402	0.788	0.635	0.004	0.929	0.185
P1,P2	**RBF**	**10**	**S^2 = 15**	**0.777**	**0.481**	**0.885**	**0.494**	**0.662**	**0.746**
P1,P3	RBF	0.6	S^2 = 100	0.377	0.804	0.681	0.235	0.815	0.487
P1,P4	RBF	0.2	S^2 = 100	0.320	0.841	0.697	0.027	0.918	0.183
P1,P2,P3	RBF	1	S^2 = 100	0.449	0.756	0.726	0.315	0.771	0.679
P1,P2,P4	RBF	1.8	S^2 = 50	0.654	0.600	0.838	0.152	0.858	0.415

*bold rows reveal the best predictor in each phase of the NLSFS method.

Table 5. The NLSFS results for spring rainfall forecasting in the Karkheh sub-basin.

				Calibration			Validation		
Predictors	best Kernel	best gamma	best Parameters	NSE	RMSE	PCC	NSE	RMSE	PCC
P1*	**Polynomial**	**1.1**	**t = 0.1 d = 3**	**0.254**	**0.927**	**0.505**	**0.251**	**0.703**	**0.573**
P2	RBF	0.2	S^2 = 90	0.190	0.965	0.570	0.049	0.793	0.239
P1, P2	**Polynomial**	**0.1**	**t = 0.1 d = 3**	**0.534**	**0.732**	**0.733**	**0.538**	**0.552**	**0.853**

*bold rows reveal the best predictor in each phase of the NLSFS method.

tions were identified as polynomial or RBF types, despite considering the linear kernel function as well. The basis for choosing the best combination of predictors for each season was the least error in the validation phase according to the assessment criteria.

The best result of the NLSFS method for each season of the studied areas showed that the best predictors for the autumn rainfall in both sub-basins were the autumn, summer, and winter SSVD time series of climate indices; nevertheless, the types of climate indices included in the SSVD series were different for each sub-basin. The best predictors for spring rainfall forecasting in both sub-basins were the SSVD time series of autumn SST of the Mediterranean Sea and winter SST of the Persian Gulf. However, the best predictors of the winter rainfall were different in the two sub-basins; the summer and spring SSVD time series of climate indices were the best predictors for Seimareh while the summer and autumn SSVD time series of climate indices as well as the spring average of Nino 3.4 were the best combination of predictors for Karkheh sub-basin.

To evaluate the performance of the traditional SFS method in the determination of the best predictors, the SFS method was also applied to the initially selected predictors as shown in Tables 2 and 3. The step-by-step results obtained from this method for the spring in the calibration and validation phases are illustrated in Tables 6 and 7 for Seimareh and Karkheh sub-basins, respectively.

Comparing the results of spring rainfall forecasting for Seimareh in Table 4 (NLSFS) and Table 6 (SFS) indicates that the accuracy of the NLSFS results for each of the predictors was greater than that of the SFS method, which occurred because there was a non-linear relationship between autumn rainfall and each of the predictors as demonstrated by the type of optimum kernel functions. On this basis, the best combination of the predictors identified by these methods were different. Such a condition can also be observed for the Karkheh sub-basin (Tables 5 and 7).

Comparing the best results of the NLSFS and SFS methods (Table 8) revealed that, except for winter in Seimareh, the best predictors determined by the SFS for both studied areas in all seasons differed from those identified by NLSFS. Furthermore, the accuracy of the results achieved from the NLSFS in the calibration and validation phases in almost all cases was greater

Table 6. The step-by-step results of SFS method for spring rainfall forecasting in the Seimareh sub-basin.

	Calibration			Validation		
Predictors	NSE	RMSE	PCC	NSE	RMSE	PCC
P1	0.349	0.822	0.592	0.083	0.892	0.381
P2*	**0.154**	**0.937**	**0.417**	**0.326**	**0.765**	**0.695**
P3	0.135	0.948	0.368	0.295	0.782	0.592
P4	0.306	0.849	0.558	−0.222	1.030	−0.241
P2,P1	0.380	0.803	0.623	0.267	0.798	0.544
P2,P3	**0.186**	**0.920**	**0.439**	**0.439**	**0.698**	**0.750**
P2,P4	0.308	0.848	0.563	−0.126	0.988	−0.076
P2,P3,P1	0.405	0.786	0.652	0.034	0.915	0.425
P2,P3,P4	0.328	0.835	0.573	−0.023	0.942	0.170

*bold rows reveal the best predictor in each phase of the SFS method.

Table 7. Step-by-step results of the SFS method for spring rainfall forecasting in the Karkheh sub-basin.

	Calibration			Validation		
Predictors	NSE	RMSE	PCC	NSE	RMSE	PCC
P1*	**0.205**	**0.956**	**0.453**	**0.673**	**0.464**	**0.856**
P2	0.251	0.929	0.501	−0.125	0.862	0.213
P1,P2	0.397	0.833	0.631	0.465	0.594	0.697

*bold rows reveal the best predictor in each phase of the SFS method.

Table 8. The best predictors identified by the NLSFS and SFS methods for the studied areas in all seasons.

Study area			*Seimareh*			*Karkheh*		
Season			Autumn	Winter	Spring	Autumn	Winter	Spring
NLSFS method	**Best predictors**		**P1,P3,P4**	**P1,P3**	**P1,P2**	**P1,P3,P2**	**P3,P1,P4**	**P1,P2**
	Best kernel function		**RBF**	**Polynomial**	**RBF**	**Polynomial**	**Polynomial**	**Polynomial**
	Calibration	***NSE***	0.527	0.490	0.777	0.657	0.938	0.534
		RMSE	0.749	0.679	0.481	0.634	0.253	0.732
		PCC	0.735	0.700	0.885	0.84	0.972	0.733
	Validation	***NSE***	0.334	0.28	0.494	0.621	0.425	0.538
		RMSE	0.415	0.805	0.662	0.477	0.630	0.552
		PCC	0.647	0.767	0.746	0.860	0.749	0.853
SFS method	**Best predictors**		**P3**	**P1,P3**	**P2,P3**	**P1,P2**	**P3,P2**	**P1**
	Calibration	***NSE***	0.205	0.253	0.186	0.297	0.255	0.205
		RMSE	0.970	0.822	0.920	0.908	0.877	0.956
		PCC	0.457	0.530	0.439	0.548	0.532	0.453
	Validation	***NSE***	0.009	0.293	0.439	0.567	0.396	0.673
		RMSE	0.507	0.798	0.698	0.510	0.646	0.464
		PCC	0.184	0.746	0.750	0.799	0.856	0.856

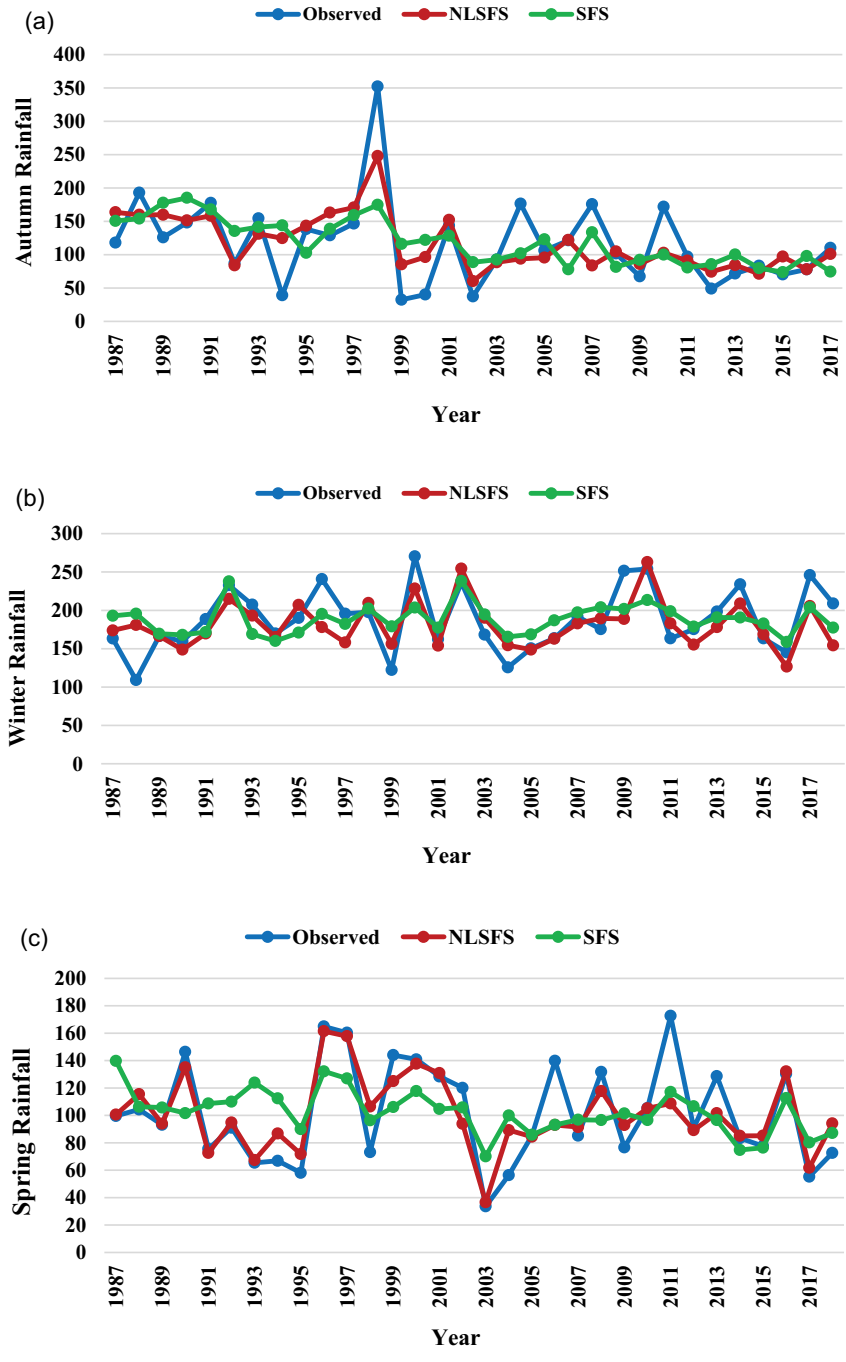

Figure 7. The time series plots of observed and forecasted rainfalls of Seimareh by NLSFS and SFS methods for (a) autumn, (b) winter, and (c) spring.

than that of the SFS. According to the best predictors characterized by the NLSFS, the best set of predictors for autumn and spring rainfall forecasting were similar for the two sub-basins, although the climate indices included in the SSVD series of autumn rainfall predictors were different. But the predictors selected by the SFS varied in both sub-basins for all seasons. The seasonal rainfall forecasted by the best set of predictors identified by the NLSFS and the SFS methods in the calibration and validation phases are displayed in Figs 7 and 8 for Seimareh and Karkheh, respectively. These graphs also confirm that applying the NLSFS method for predictor recognition resulted in more accurate forecasts, especially for peak values, as compared to the SFS method for both studied areas.

5 Discussion

The results from both case studies demonstrated non-significant correlations between the individual climate indices and local seasonal rainfall. However, trustworthy correlations at the 98% confidence level were achieved by producing ensemble series by means of SSVD. Comparing the correlation values attained by SVD, our results showed that SSVD is superior to SVD by 11% (on average). Our findings agree with the results of Nazemosadat and Cordery (2000), Araghinejad *et al.* (2006), and Dezfuli *et al.* (2010) who reported the negative correlation between summer average of SOI and autumn rainfall in the west of Iran. They are also consistent with the findings of

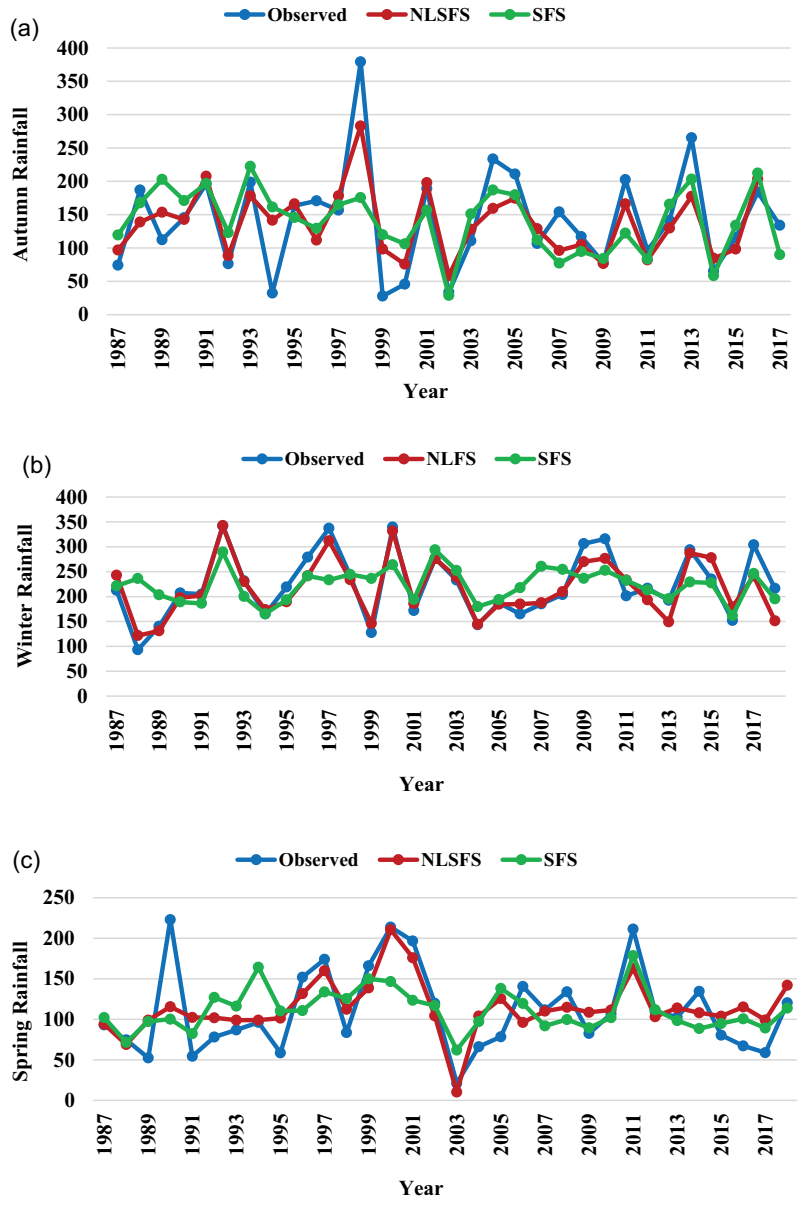

Figure 8. Time series plots of observed and forecasted rainfalls of Karkheh by NLSFS and SFS methods for (a) autumn, (b) winter, and (c) spring.

Tabari *et al.* (2014) demonstrating the effect of spring average NAO fluctuations on the autumn rainfall across the northern Karkheh basin (Seimareh sub-basin). Additionally, our findings are consistent with the results of Modaresi *et al.* (2016) and Sarah *et al.* (2011) in terms of the effect of autumn SST of the Persian Gulf on spring rainfall and streamflow in the Karoon and Karkheh basins, respectively. On the other hand, our results disagree with Meidani and Araghinejad's findings, in which autumn SST of the Mediterranean Sea was introduced as the main winter rainfall trigger in the west of Iran (Meidani and Araghinejad 2014). Such an inconsistency could be owing to the differences in (i) the duration of observed data and (ii) the extent of the studied areas. Thus, it is concluded that a change in the temporal and spatial scales of the data and study area may result in identifying different predictors.

The results also revealed that the best predictors determined using the NLSFS method were different from those selected by the SFS in all seasons. The seasonal rainfall forecasted by the NLSFS method was dramatically more accurate compared to the SFS. This is because of the dominant non-linear relationships that exist between the climate indicators and seasonal rainfall in small areas. Our NLSFS implementation results confirmed the results of Choubin *et al.* (2017) and Kalra and Ahmad (2012) regarding the higher performance of non-linear methods in comparison with the linear ones for rainfall forecasting. Furthermore, the results agreed with those of Selvi and Huseyinov (2020), who indicated that simultaneous implementation of data pre-processing and SFS methods with non-linear modelling algorithms can improve feature selection for hydrological forecasting.

The total selected predictors indicated that the SSVD ensemble series of seasonal climate indices were the best predictors for autumn and winter rainfall, while the combined series of the SST of the Mediterranean Sea and the Persian Gulf were the best for spring rainfall forecasting. According to the best predictors identified by the NLSFS method, the best set of predictors for autumn and spring rainfall forecasting were similar for the two

sub-basins, although the climate indices included in the SSVD series of autumn rainfall predictors were different. By contrast, the predictors selected by the SFS varied in both sub-basins for all seasons.

Based on the results, we conclude that the effects of SSTs of the Persian Gulf and the Mediterranean Sea on the sub-basins' rainfall are almost the same even though they are at different distances from the sub-basins. By contrast, the influences of the ocean–atmospheric climate indices far from the sub-basins were found to be different. The reason for such a varied effect could be the fluctuations of the climate indices that affect the climate variables related to rainfall, such as air pressure, air temperature, and wind speed (Aravena *et al.* 2009, Xie *et al.* 2019).

The physical mechanism of teleconnections between large-scale phenomena and climate variables is not yet well understood for different regions in Iran (Choubin *et al.* 2017). The results obtained in the present study, which is based on 32 years of observed data (1987–2019) from 27 stations, can help researchers recognize the physical relationship between large-scale climate fluctuations and local variations of the climate variables. Because of stochastic patterns of rainfall events and random characteristics of rainfall data, the patterns identified in this study must not be considered deterministic. However, as we evaluated and verified the efficiency of the proposed model using data from two case study regions with different climate conditions, and the models for each season were calibrated by the LOOCV method using the calibration data, we strongly believe that the outcomes are trustworthy.

6 Conclusion

This article introduced a new approach for predictor selection among large-scale climate indices when they are used for long-term rainfall forecasting in small catchments with different climate conditions. The new approach was based on identifying and coupling the most effective seasonal climate indices via advanced statistical and data-mining methods. The proposed algorithm was a new hybrid AI approach that effectively combines a selective singular value decomposition method (SSVD) with a non-linear SFS method (NLSFS). The new method was demonstrated and verified using data from the two sub-basins of the Karkheh basin, Iran.

The results showed that the climate indices individually had no significant correlation with the seasonal rainfall of the studied sub-basins; but applying the SSVD method resulted in recognition of the most effective of the climate indices on the seasonal rainfall of each studied area, and we were able to produce an ensemble series of them with a trustworthy correlation at the 98% confidence level. The results revealed that the effective climate indices selected by the SSVD method in the combined series were different seasonally and spatially. However, the total predictors selected by the NLSFS method indicated that the SSVD combined series of seasonal climate indices were the best predictors for autumn and winter rainfall, while the combined series of the SST of the Mediterranean Sea and the Persian Gulf were the best for spring rainfall forecasting. The results of this study suggest that when there are diverse weather conditions in a vast area, selecting rainfall predictors of climate indices for each division with a specific weather condition using the proposed algorithm results in more accurate forecasts.

In addition, our results demonstrated the superiority of the proposed hybrid algorithm compared to the use of SVD and SFS methods in terms of the type of selected predictors and the accuracy of the forecast results. Although the proposed method was used for input selection for seasonal rainfall forecasting at the catchment scale, it can be employed for optimum predictor selection in any hydrological forecasting task. As the SSVD method combines the data series possessing spatiotemporal variation, the application of this method would be limited to hydrological predictions in which at least one of the predictors or predictand variables have a spatiotemporally varying feature.

Acknowledgements

The authors are grateful to the Vice Chancellor for Research and Technology of Ferdowsi University of Mashhad (FUM) for supporting this research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the funding associated [NA].

ORCID

Fereshteh Modaresi http://orcid.org/0000-0001-7033-5402
Kumars Ebrahimi http://orcid.org/0000-0002-9914-4383
Ali Danandeh Mehr http://orcid.org/0000-0003-2769-106X

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Abdullah, D., *et al.*, 2023. An artificial neural networks approach and hybrid method with wavelet transform to investigate the quality of Tallo River, Indonesia. *Caspian Journal of Environmental Sciences*, 21 (3), 647–656.

Ali, M., *et al.* 2020. Forecasting long-term precipitation for water resource management: a new multi-step data-intelligent modelling approach. *Hydrological Sciences Journal*, 65 (16), 2693–2708. doi:10.1080/02626667.2020.1808219

Araghinejad, S., Burn, D.H., and Karamouz, M., 2006. Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors. *Water Resources Research*, 42 (3). doi:10.1029/2004WR003853

Aravena, G., *et al.* 2009. Influence of the North Atlantic Oscillation (NAO) on climatic factors and estuarine water temperature on the Basque coast (Bay of Biscay): comparative analysis of three seasonal NAO indices. *Continental Shelf Research*, 29 (4), 750–758. doi:10.1016/j.csr.2008.12.001

Asadi Oskouei, E., *et al.* 2022. Mapping climate zones of Iran using hybrid interpolation methods. *Remote Sensing*, 14 (11), 2632. doi:10.3390/rs14112632

Asghari Saraskanrood, S., Asadi, B., and Ghale, E., 2023. Land surface temperature assessment in relation to land-use/land-cover (A case study: isfahan city, Central Iran). *Caspian Journal of Environmental Sciences*, 21 (3), 725–735.

Bahrami, F., *et al.* 2021. Synoptic–dynamic patterns affecting Iran's autumn precipitation during ENSO phase transitions. *Climate*, 9 (7), 106. doi:10.3390/cli9070106

Bhanu, P.S., 2021. Seasonal prediction of rainfall over Cherthala region of Kerala through stepwise singular value decomposition. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12 (13), 2005–2013.

Börgel, F., *et al.* 2022. Atlantic multidecadal variability and the implications for North European precipitation. *Environmental Research Letters*, 17 (4), 044040. doi:10.1088/1748-9326/ac5ca1

Bramer, M., 2020. *Principles of data mining*. Springer-Verlag London Ltd, United Kingdom.

Bretherton, C.S., Smith, C., and Wallace, J.M., 1992. An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, 5 (6), 541–560. doi:10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2

Cai, W., *et al.* 2020. Climate impacts of the El Niño–Southern oscillation on South America. *Nature Reviews Earth & Environment*, 1 (4), 215–231. doi:10.1038/s43017-020-0040-3

Choubin, B., *et al.* 2017. An ensemble forecast of semi-arid rainfall using large-scale climate predictors. *Meteorological Applications*, 24 (3), 376–386. doi:10.1002/met.1635

Danandeh Mehr, A., *et al.*, 2017. A binary genetic programing model for teleconnection identification between global sea surface temperature and local maximum monthly rainfall events. *Journal of Hydrology*, 555, 397–406. doi:10.1016/j.jhydrol.2017.10.039

Danandeh Mehr, A., 2021. Seasonal rainfall hindcasting using ensemble multi-stage genetic programming. *Theoretical and Applied Climatology*, 143 (1–2), 461–472. doi:10.1007/s00704-020-03438-3

Danandeh Mehr, A., *et al.* 2022. A new evolutionary time series model for streamflow forecasting in boreal lake-river systems. *Theoretical and Applied Climatology*, 148 (1–2), 255–268. doi:10.1007/s00704-022-03939-3

Dehghani, M., *et al.* 2014. Uncertainty analysis of streamflow drought forecast using artificial neural networks and monte-carlo simulation. *International Journal of Climatology*, 34 (4), 1169–1180. doi:10.1002/joc.3754

Dehghani, M., *et al.* 2020. Spatial analysis of seasonal precipitation over Iran: co-variation with climate indices. *ISPRS International Journal of Geo-Information*, 9 (2), 73. doi:10.3390/ijgi9020073

Desbiolles, F., *et al.* 2021. Links between sea surface temperature structures, clouds and rainfall: study case of the Mediterranean Sea. *Geophysical Research Letters*, 48 (10), e2020GL091839. doi:10.1029/2020GL091839

Dezfuli, A.K., Karamouz, M., and Araghinejad, S., 2010. On the relationship of regional meteorological drought with SOI and NAO over southwest Iran. *Theoretical and Applied Climatology*, 100 (1), 57–66. doi:10.1007/s00704-009-0157-2

Fallah-Ghalhary, G.A., Mousavi-Baygi, M., and Nokhandan, M.H., 2009. Annual rainfall forecasting by using mamdani fuzzy inference system. *Research Journal of Environmental Sciences*, 3 (4), 400–413. doi:10.3923/rjes.2009.400.413

Fayaz, S.A., Zaman, M., and Butt, M.A., 2022. Numerical and experimental investigation of meteorological data using adaptive linear m5 model tree for the prediction of rainfall. *Review of Computer Engineering Research*, 9 (1), 1–12. doi:10.18488/76.v9i1.2961

Fernández-González, S., *et al.* 2012. Connection between NAO, weather types and precipitation in León, Spain (1948–2008). *International Journal of Climatology*, 32 (14), 2181–2196. doi:10.1002/joc.2431

Ghasemi, M., *et al.*, 2023. A comparative study of black-box and white-box data-driven methods to predict landfill leachate permeability. *Environmental Monitoring and Assessment*, 195 (7), 862. doi:10.1007/s10661-023-11462-9

Heidemann, H., *et al.*, 2022. The influence of interannual and decadal indo-pacific sea surface temperature variability on Australian monsoon rainfall. *Journal of Climate*, 35 (1), 425–444.

Hosseini, F.S., *et al.* 2021. Towards a flood vulnerability assessment of watershed using integration of decision-making trial and evaluation laboratory, analytical network process, and fuzzy theories. *Environmental Science and Pollution Research*, 28 (44), 62487–62498. doi:10.1007/s11356-021-14534-w

Kalra, A. and Ahmad, S., 2012. Estimating annual precipitation for the Colorado River Basin using oceanic-atmospheric oscillations. *Water Resources Research*, 48 (6). doi:10.1029/2011WR010667

Kassomenos, P. and Mcgregor, G., 2006. The interannual variability and trend of precipitable water over southern Greece. *Journal of Hydrometeorology*, 7 (2), 271–284.

Li, J., *et al.*, 2020. Toward monitoring short-term droughts using a novel daily scale, standardized antecedent precipitation evapotranspiration index. *Journal of Hydrometeorology*, 21 (5), 891–908. doi:10.1175/JHM-D-19-0298.1

Li, X. and Sun, Y., 2020. Stock intelligent investment strategy based on support vector machine parameter optimization algorithm. *Neural Computing & Applications*, 32 (6), 1765–1775. doi:10.1007/s00521-019-04566-2.

Lin, R., Zhu, J., and Zheng, F., 2019. The application of the SVD method to reduce coupled model biases in seasonal predictions of rainfall. *Journal of Geophysical Research: Atmospheres*, 124 (22), 11837–11849. doi:10.1029/2018JD029927

Liu, Y.-C., *et al.* 2018. Relationships of rainy season precipitation and temperature to climate indices in California: long-term variability and extreme events. *Journal of Climate*, 31 (5), 1921–1942. doi:10.1175/JCLI-D-17-0376.1

Ma, Y., *et al.*, 2023. More profound impact of CP ENSO on Australian spring rainfall in recent decades. *Climate Dynamics*, 60, 3065–3079. doi:10.1007/s00382-022-06485-w

Marchane, A., *et al.*, 2015. Relationship between snow, temperature, rainfall and the North Atlantic Oscillation on the Moroccan Atlas. ed. EGU General Assembly Conference Abstracts, Vienna, Austria, 14083.

Martija-Díez, M., Rodríguez-Fonseca, B., and López-Parages, J., 2021. ENSO Influence on Western European summer and fall temperatures. *Journal of Climate*, 34 (19), 8013–8031.

Meidani, E. and Araghinejad, S., 2014. Long-lead streamflow forecasting in the southwest of Iran by sea surface temperature of the Mediterranean Sea. *Journal of Hydrologic Engineering*, 19 (8), 05014005. doi:10.1061/(ASCE)HE.1943-5584.0000965

Modaresi, F. and Araghinejad, S., 2014. A comparative assessment of support vector machines, probabilistic neural networks, and K-nearest neighbor algorithms for water quality classification. *Water Resources Management*, 28 (12), 4095–4111. doi:10.1007/s11269-014-0730-z

Modaresi, F., Araghinejad, S., and Ebrahimi, K., 2016. The combined effect of persian gulf and Mediterranean Sea surface temperature on operational forecast of spring streamflow for Karkheh basin, Iran. *Sustainable Water Resources Management*, 2 (4), 387–403. doi:10.1007/s40899-016-0068-1

Modaresi, F., Araghinejad, S., and 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 Resources Management*, 32 (1), 243–258. doi:10.1007/s11269-017-1807-2

Mohammadi, B., Vaheddoost, B., and Mehr, A.D., 2020. A spatiotemporal teleconnection study between Peruvian precipitation and oceanic oscillations. *Quaternary International*, 565, 1–11. doi:10.1016/j.quaint.2020.09.042

Moriasi, D.N., *et al.* 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50 (3), 885–900. doi:10.13031/2013.23153

Nasemosadat, M.J., 1998. Persian Gulf sea surface temperature as a drought diagnostic for southern Iran. *Drought Network News (1994–2001)*, 55.

Nazemosadat, M. and Cordery, I., 2000. On the relationships between ENSO and autumn rainfall in Iran. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 20 (1), 47–61. doi:10.1002/(SICI)1097-0088(200001)20:1<47::AID-JOC461>3.0.CO;2-P

Nazemosadat, M.J., Setoodeh, P., and Safavi, A.A., 2010. Improving neural network models for forecasting seasonal precipitation in southwestern Iran: the evaluation of oceanic–atmospheric indices. *Advances in Geosciences, Volume 16: Atmospheric Science (AS)*, 16.

Ng, C.H.J., *et al.* 2019. An asymmetric rainfall response to ENSO in East Asia. *Climate Dynamics*, 52 (3–4), 2303–2318. doi:10.1007/s00382-018-4253-9

Noori, R., *et al.* 2011. Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology*, 401 (3), 177–189. doi:10.1016/j.jhydrol.2011.02.021

Oborie, E. and Rowland, E.D., 2023. Flood influence using GIS and remote sensing based morphometric parameters: a case study in Niger delta region. *Journal of Asian Scientific Research*, 13 (1), 1–15. doi:10.55493/5003.v13i1.4719

Park, J.-Y., Bader, J., and Matei, D., 2016. Anthropogenic Mediterranean warming essential driver for present and future Sahel rainfall. *Nature Climate Change*, 6 (10), 941–945. doi:10.1038/nclimate3065

Pradhan, A., 2012. Support vector machine-a survey. *International Journal of Emerging Technology and Advanced Engineering*, 2 (8), 82–85.

Ren, K., *et al.*, 2020. Comparison of eight filter-based feature selection methods for monthly streamflow forecasting–three case studies on CAMELS data sets. *Journal of Hydrology*, 586, 124897. doi:10.1016/j.jhydrol.2020.124897

Ruigar, H. and Golian, S., 2015. Assessing the correlation between climate signals and monthly mean and extreme precipitation and discharge of Golestan Dam Watershed. *Earth Sciences Research Journal*, 19 (1), 65–72. doi:10.15446/esrj.v19n1.40996

Samadi, M., *et al.*, 2015. Estimation of scour depth below free overfall spillways using multivariate adaptive regression splines and artificial neural networks. *Engineering Applications of Computational Fluid Mechanics*, 9 (1), 291–300. doi:10.1080/19942060.2015.1011826

Sarah, A., *et al.*, 2011. Long term rainfall forecasting by integrated artificial neural network-fuzzy logic-wavelet model in Karoon basin. *Scientific Research and Essays*, 6 (6), 1200–1208.

Schepen, A., Wang, Q., and Robertson, D., 2012. Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *Journal of Climate*, 25 (4), 1230–1246. doi:10.1175/JCLI-D-11-00156.1

Selvi, O. and Huseyinov, İ., 2020. A novel algorithm for feature selection based on geographic distance metric: a case study of streamflow forecasting of Austria's water resources. *International Journal of Environmental Science & Technology*, 17 (1), 295–308. doi:10.1007/s13762-019-02485-2

Suykens, J.A., *et al.* 2002. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, 48 (1–4), 85–105. doi:10.1016/S0925-2312(01)00644-0

Tabari, H., Abghari, H., and Hosseinzadeh Talaee, P., 2014. Impact of the North Atlantic Oscillation on streamflow in western Iran. *Hydrological Processes*, 28 (15), 4411–4418. doi:10.1002/hyp.9960

Tabari, H. and Willems, P., 2018. Lagged influence of Atlantic and Pacific climate patterns on European extreme precipitation. *Scientific Reports*, 8 (1), 5748. doi:10.1038/s41598-018-24069-9

Vallès-Casanova, I., *et al.* 2020. On the spatiotemporal diversity of Atlantic Niño and associated rainfall variability over West Africa and South America. *Geophysical Research Letters*, 47 (8), e2020GL087108. doi:10.1029/2020GL087108

Vashisht, A., Zaitchik, B., and Gnanadesikan, A., 2021. ENSO teleconnection to Eastern African summer rainfall in global climate models: role of the tropical easterly jet. *Journal of Climate*, 34 (1), 293–312. doi:10.1175/JCLI-D-20-0222.1

Vibhute, A., *et al.* 2020. Decadal variability of tropical Indian Ocean sea surface temperature and its impact on the Indian summer monsoon. *Theoretical and Applied Climatology*, 141 (1–2), 551–566. doi:10.1007/s00704-020-03216-1

Wu, X., *et al.* 2022. Long-range precipitation forecast based on multipole and preceding fluctuations of sea surface temperature. *International Journal of Climatology*, 42 (15), 8024–8039. doi:10.1002/joc.7690

Xie, T., *et al.* 2019. NAO implicated as a predictor of the surface air temperature multidecadal variability over East Asia. *Climate Dynamics*, 53 (1–2), 895–905. doi:10.1007/s00382-019-04624-4

Xie, X., *et al.*, 2021. A simple Monte Carlo method for estimating the chance of a cyclone impact. *Natural Hazards*, 107 (3), 2573–2582. doi:10.1007/s11069-021-04505-2

Zhao, C. and Brissette, F., 2022. Impacts of large-scale oscillations on climate variability over North America. *Climatic Change*, 173 (1–2), 4. doi:10.1007/s10584-022-03383-2

Zhu, X., *et al.*, 2022. Impact of dam construction on precipitation: a regional perspective. *Marine and Freshwater Research*, 74 (10), 877–890. doi:10.1071/MF22135

Appendix A

Table A1. The Non-linear sequential forward selection (NLSFS) results for autumn rainfall forecasting in Seimareh sub-basin.

				Calibration			Validation		
Predictors	Best kernel	Best gamma	Best parameters	NSE	RMSE	PCC	NSE	RMSE	PCC
P1*	**RBF**	**5.7**	**S^2 = 10**	**0.392**	**0.837**	**0.629**	**0.372**	**0.403**	**0.624**
P2	RBF	8.1	S^2 = 5	0.449	0.796	0.675	−0.100	0.534	0.072
P3	RBF	10	S^2 = 20	0.226	0.958	0.480	0.042	0.498	0.209
P4	Polynomial	10	t = 2 d = 4	0.261	0.935	0.511	0.113	0.479	0.622
P1,P2	RBF	4.4	S^2 = 5	0.554	0.717	0.758	0.217	0.451	0.477
P1,P3	**RBF**	**7.6**	**S^2 = 20**	**0.470**	**0.782**	**0.696**	**0.360**	**0.407**	**0.645**
P1,P4	RBF	9.4	S^2 = 20	0.477	0.777	0.699	0.337	0.414	0.639
P1,P3,P2	RBF	9.5	S^2 = 10	0.599	0.690	0.782	0.259	0.438	0.538
P1,P3,P4	**RBF**	**10**	**S^2 = 25**	**0.527**	**0.749**	**0.735**	**0.334**	**0.415**	**0.647**

*bold rows reveal the best predictor in each phase of the NLSFS method.

Table A2. The Non-linear sequential forward selection (NLSFS) results for winter rainfall forecasting in Seimareh sub-basin.

				Calibration			Validation		
Predictors	Best kernel	Best gamma	Best parameters	NSE	RMSE	PCC	NSE	RMSE	PCC
P1*	**Polynomial**	**10**	**t = 2 d = 3**	**0.270**	**0.812**	**0.519**	**0.198**	**0.849**	**0.677**
P2	RBF	10	S^2 = 5	0.280	0.806	0.530	−0.039	0.967	0.445
P3	Polynomial	10	t = 1.5 d = 1	0.199	0.850	0.447	−0.104	0.996	0.461
P1,P2	RBF	2.1	S^2 = 5	0.297	0.797	0.551	0.055	0.922	0.558
P1,P3	**Polynomial**	**10**	**t = 2 d = 3**	**0.490**	**0.679**	**0.700**	**0.280**	**0.805**	**0.767**
P1,P3,P2	Polynomial	0.3	t = 1.1 d = 1	0.304	0.793	0.552	0.074	0.913	0.664

*bold rows reveal the best predictor in each phase of the NLSFS method.

Table A3. The Non-linear sequential forward selection (NLSFS) results for autumn rainfall forecasting in Karkheh sub-basin.

					Calibration			Validation		
Predictors	Best kernel	Best gamma	Best parameters		NSE	RMSE	PCC	NSE	RMSE	PCC
P1*	**Polynomial**	**0.2**	**t = 1.1**	**d = 5**	**0.335**	**0.863**	**0.585**	**0.332**	**0.633**	0.667
P2	Polynomial	0.2	t = 1.8	d = 5	0.492	0.754	0.708	0.487	0.555	0.749
P3	Polynomial	0.6	t = 0.1	d = 2	0.261	0.916	0.541	0.258	0.668	0.532
P4	RBF	3.5	$S^2 = 100$		0.177	0.967	0.508	0.177	0.703	0.564
P1,P2	Polynomial	0.4	t = 0.9	d = 3	0.373	0.857	0.613	0.370	0.615	0.741
P1,P3	**Polynomial**	**0.2**	**t = 0.3**	**d = 2**	**0.592**	**0.692**	**0.804**	**0.577**	**0.504**	**0.790**
P1,P4	RBF	2.6	$S^2 = 15$		0.386	0.848	0.636	0.386	0.607	0.789
P1,P3,P2	**Polynomial**	**0.1**	**t = 0.7**	**d = 2**	**0.657**	**0.634**	**0.840**	**0.621**	**0.477**	**0.860**
P1,P3,P4	RBF	2.8	$S^2 = 15$		0.525	0.746	0.744	0.524	0.535	0.835
P1,P3,P2,P4	RBF	2.9	$S^2 = 20$		0.535	0.738	0.752	0.535	0.529	0.879

*bold rows reveal the best predictor in each phase of the NLSFS method.

Table A4. The Non-linear sequential forward selection (NLSFS) results for winter rainfall forecasting in Karkheh sub-basin.

					Calibration			Validation		
Predictors	Best kernel	Best gamma	Best parameters		NSE	RMSE	PCC	NSE	RMSE	PCC
P1	Polynomial	10	t = 1.7	d = 1	0.288	0.857	0.537	0.177	0.753	0.679
P2	Polynomial	10	t = 1.5	d = 1	0.259	0.875	0.509	0.015	0.824	0.724
P3*	**RBF**	**10**	**$S^2 = 5$**		**0.212**	**0.902**	**0.462**	**0.224**	**0.732**	**0.691**
P4	Polynomial	10	t = 2	d = 4	0.204	0.907	0.451	−0.061	0.855	0.756
P3,P1	**Polynomial**	**3.4**	**t = 2**	**d = 3**	**0.351**	**0.819**	**0.593**	**0.289**	**0.700**	**0.759**
P3,P2	Polynomial	10	t = 1.2	d = 1	0.283	0.861	0.532	0.180	0.752	0.856
P3,P4	Polynomial	0.4	t = 1.3	d = 5	0.843	0.403	0.923	0.248	0.720	0.691
P3,P1,P2	Polynomial	0.5	t = 0.9	d = 1	0.321	0.837	0.567	0.217	0.735	0.819
P3, P1, P4	**Polynomial**	**0.4**	**t = 2**	**d = 4**	**0.938**	**0.253**	**0.972**	**0.425**	**0.630**	**0.749**
P3,P1,P4,P2	Polynomial	2	t = 1.2	d = 1	0.322	0.837	0.567	0.213	0.737	0.810

*bold rows reveal the best predictor in each phase of the NLSFS method.

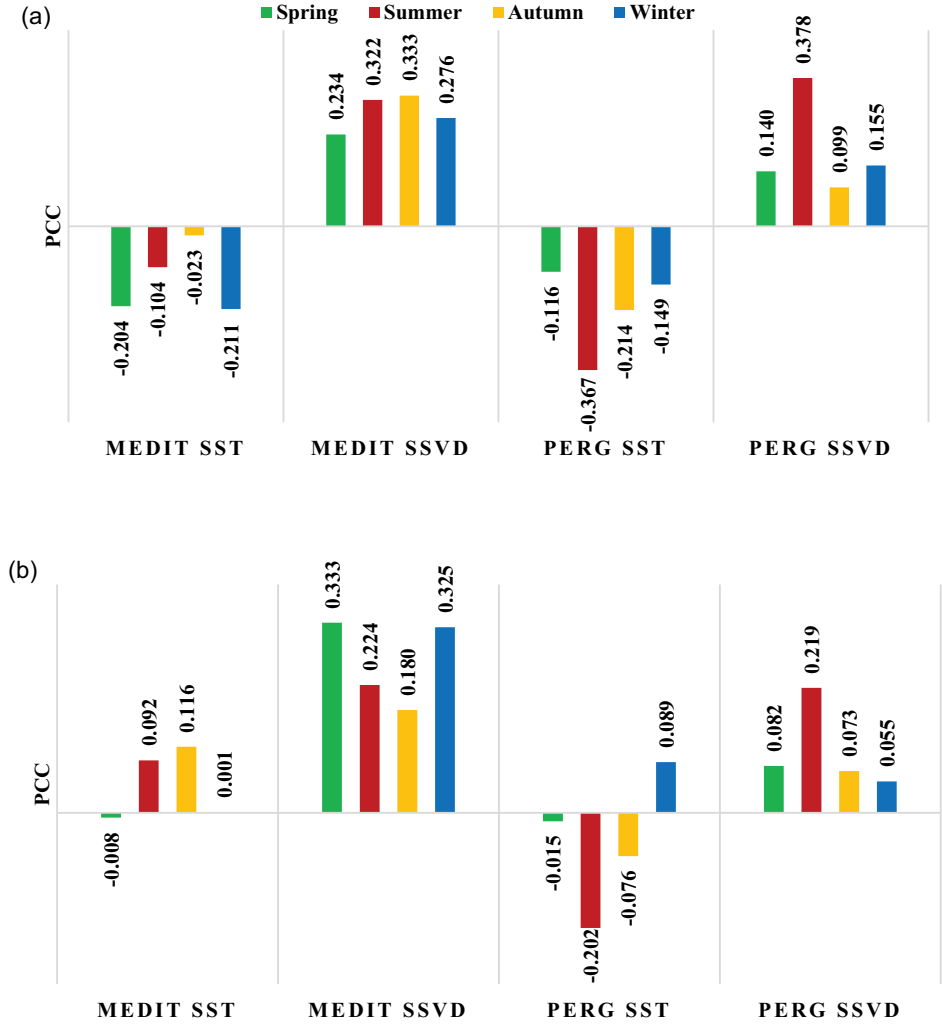

Figure A1. The Pearson correlation coefficient (PCC) of autumn rainfall with seasonal average Sea Surface Temperature (SST) of Mediterranean Sea (MEDIT SST) and Persian Gulf (PERG SST) and their combination series by Selective singular value decomposition (SSVD) method (MEDI SSVD and PERG SSVD) for (a) Seimareh and (b) Karkheh sub-basins.

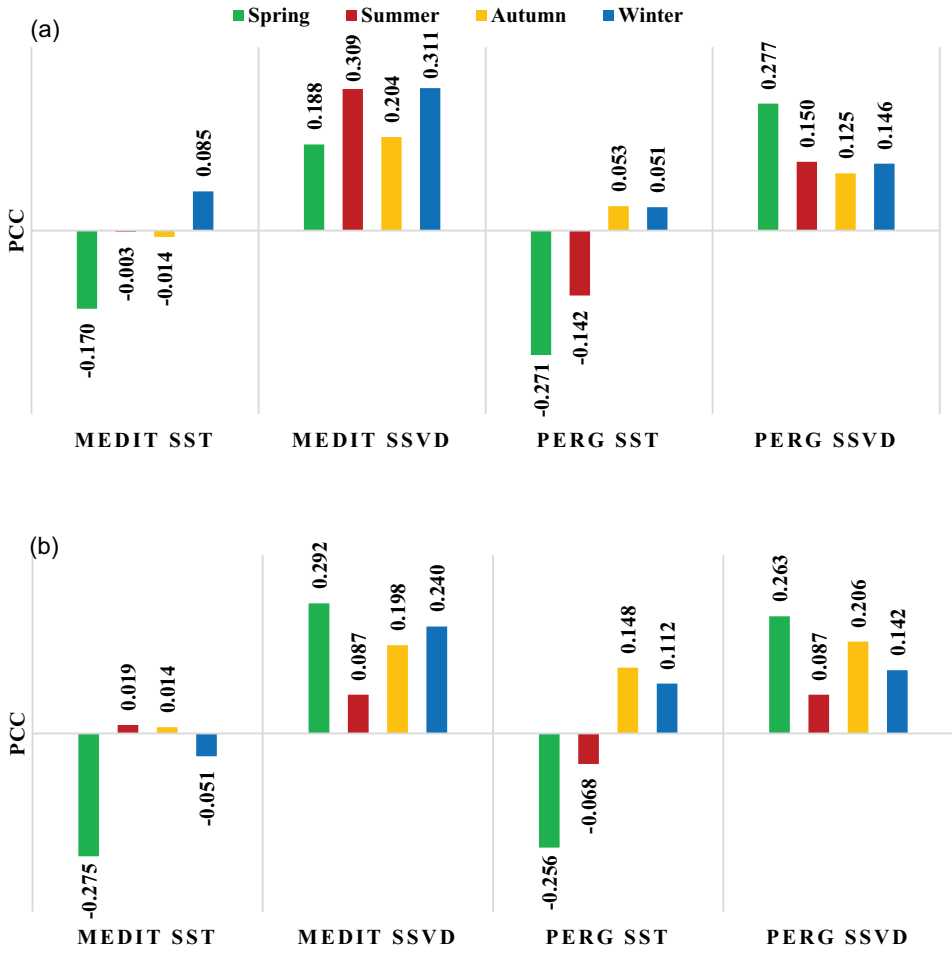

Figure A2. The Pearson correlation coefficient (PCC) of winter rainfall with seasonal average SST of Mediterranean Sea (MEDIT SST) and Persian Gulf (PERG SST) and their combination series by SSVD method (MEDI SSVD and PERG SSVD) for (a) Seimareh and (b) Karkheh sub-basins.