Environmental Science and Pollution Research (2023) 30:89705–89725
https://doi.org/10.1007/s11356-023-28678-4

RESEARCH ARTICLE

Application of novel artificial bee colony optimized ANN and data preprocessing techniques for monthly streamflow estimation

Okan Mert Katipoğlu[1] · Mehdi Keblouti[2] · Babak Mohammadi[3]

Received: 5 April 2023 / Accepted: 4 July 2023 / Published online: 17 July 2023

Abstract
Streamflow estimation is important in hydrology, especially in drought and flood-prone areas. Accurate estimation of streamflow values is crucial for the sustainable management of water resources, the development of early warning systems for disasters, and for various applications such as irrigation, hydropower production, dam sizing, and siltation management. This study developed the ANN algorithm by optimizing with an artificial bee colony (ABC). Then, the ABC-ANN hybrid model, which was established, was combined with different signal decomposition techniques to evaluate its performance in streamflow estimation in the East Black Sea Region, Türkiye. For this purpose, the lagged streamflow values were divided into subcomponents using the local mean decomposition (LMD) with the empirical envelope and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) signal decomposition techniques presented to the ABC-ANN algorithm. Thus, the success of the novel hybrid LMD-ABC-ANN and CEEMDAN-ABC-ANN approaches in streamflow prediction was evaluated. The outputs are reliable strategies and resources for water resource planners and policymakers.

Keywords Streamflow prediction · Artificial bee colony optimization · Empirical mode decomposition · Local mean decomposition · East Black Sea Region

Introduction

In hydrology, a watershed is an area of land that drains to a common point, such as a river, lake, or ocean. Water flow in a watershed is a critical component of the hydrological cycle, including precipitation, evapotranspiration, and runoff (Jencso et al. 2009).

The availability of a good network for measuring hydrological processes such as rainfall, river flow and groundwater levels, etc., is essential for proper management and use (Huang and Yang 1998; Zamoum and Souag-Gamane 2019), especially in areas that suffer from water scarcity on one side and the low density of the measurement network on the other side (Swain et al. 2015). Having current flow data on a spatial and temporal scale has many advantages, including a better understanding of climate change, the effects of land use change, and other environmental factors on water resources (Gudmundsson 2021; Pokhrel et al. 2017). Moreover, researchers can well assess the impact of human activities on water resources by examining changes in water availability by analyzing water stream flow data over time, (Sivapalan et al. 2012) and studying several natural phenomena such as droughts and floods (Le et al. 2022; Maity and Kashid 2011).

A good understanding of the change in streamflow records in a particular watershed is difficult because it is linked to many direct and indirect factors, such as the low density of the measuring network (Xuan Do et al. 2020) and sometimes no stream gauges in the small catchment areas, In some cases, different agencies or organizations may collect streamflow data but not shared with others (Gerlak et al. 2011; Kibler et al. 2014). Finally, changes in land use such as human activities, agriculture, urbanization, and deforestation, can significantly impact streamflow by altering the water balance of a watershed (Allan 2004; Bosch and Hewlett 1982; Zhang et al. 2001).

Responsible Editor: Marcus Schulz

✉ Okan Mert Katipoğlu
okatipoglu@erzincan.edu.tr

[1] Erzincan Binali Yıldırım University, Faculty of Engineering and Architecture, Department of Civil Engineering, Erzincan, Türkiye

[2] Abdelhafid Boussouf University Center, Institute of Sciences and Technology, Department of Civil Engineering and Hydraulic, Mila, Algeria

[3] Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden

For many decades stream flow prediction has been an important topic of hydrology and water resource engineering (Beven 2006; Gupta et al. 1999). Initially, reliance on statistical techniques to establish a relationship between historical data and future flows (Nash and Sutcliffe 1970; Razavi et al. 2012). Despite their simplicity and ease of use, these models' accuracy in prediction is related to the quality and quantity of available data. Rating curves remained the most widely used tools for stream flow estimation (Kiang et al. 2018), which relied on relating the river water level to the corresponding flow rate. The development of rating curves relies on collecting flow and stage water data across a range of flow conditions and then constructing a curve on the data using regression analysis (Sivapalan et al. 2012). The development of rating curves allows estimating the flow rate at any point in time based on the current phase measurement.

Despite their potential advantages, streamflow models require facing some challenges. One of the most important challenges that can make it difficult to develop accurate and effective models is the issue of data availability, both in terms of quality and quantity, (Ghimire et al. 2021). This is often found in remote or developing areas where data may be little or completely absent and of poor quality. In addition, we may find streamflow models are more complex, requiring great expertise in their creation and interpretation (Kavetski et al. 2006; Wagener et al. 2003).

This complexity may limit its scope of use, especially in regions with limited technical expertise. Finally, streamflow models differ concerning the level of uncertainty, the latter of which can be influenced by a range of factors, which would particularly affect flow estimation. Uncertainties can be resolved through sensitivity analyses and model calibration if additional data and resources are available.

However, In recent years, a growing body of scientific research has used machine learning techniques for streamflow estimation because of its ability to capture complex non-linear relationships between different hydrological variables and streamflow and use them to estimate streamflow at ungauged or poorly measured sites. Some popular machine-learning methods for streamflow estimation include artificial neural networks, decision trees, support vector machines, and random forests. These models have been successfully applied in a range of studies. Wang et al. (2006) used three types of hybrid artificial neural network (ANN) models, namely the threshold-based ANN (TANN), cluster-based ANN (CANN), and periodic ANN (PANN). The latest hybrid model gave better prediction results than other models for daily discharge forecasting of the headwater region of the Yellow River northeast of the Tibet Plateau in China. Rahsepar and Mahmoodi (2014) proposed an algorithm combining ANN and ABC to predict the future discharge of the Tang-e Karzin hydrometric station of Salman Farsi Dam, South Iran, with good results. The ABC significantly improved the performance of the ANN, thus improving the prediction of future discharge in the study area. Adnan et al. (2017) used three measures, the coefficient of determination (R2), the root mean square error (RMSE) and the mean absolute error (MAE), to evaluate the accuracy of the ANN and the support vector machines (SVM) in predicting monthly flow in the upper Indus basin, north of India. These measures showed that SVM has the best accuracy in predicting monthly flow. Katipoğlu and Can (2018) used the Auto-Regressive (AR) model to model monthly streamflow in Karasu River in the Euphrates Basin. Cheng et al. (2020) used ANN and long and short-term memory (LSTM) to forecast streamflow using precipitation and runoff datasets in the Nan River Basin and Ping River Basin, Thailand. The results showed that the LSTM model is superior to the ANN model in daily prediction. For multi-month prediction, the LSTM model showed less satisfactory results, and this is due to the limited availability of monthly training data. Katipoğlu (2020) employed Extreme gradient boosting (XGBoost) and K-Nearest Neighbors (KNN) to predict monthly streamflows in the lower Euphrates basin. According to the analysis, Xgboost was found to be superior to KNN. Siddiqi et al. (2021) used regression extreme learning machines (ELM) and ANN to estimate mean monthly upstream flow for the Tarbela dam in the Indus River basin. Ghimire et al. (2021) have developed a new deep-learning model called CNN-LSTM based on integrating CNN and LSTM to predict the hourly Qflow at Brisbane River and Teewah Creek, Australia, using deep neural networks. Pini et al. (2020) have used different machine learning algorithms such as ANN, support vector machine (SVM), and random forest (RF). Ha et al. (2021) used the monthly streamflow data of the Yangtze River from 1952 to 2018 to predict the monthly streamflow of the Yangtze River to estimate the streamflow in Lake Como (Italy). Le et al. (2022) have tried to estimate the monthly streamflow over several areas in the world, such as North America, South America, and Western Europe using three machine learning such as SVM, RF, and gradient-boosted trees, Akbarian et al. (2023) investigated the effect of these variables on the accuracy of streamflow forecasting using learning models, multiple linear regression (MLR), ANN, SVM, RF, and XGBoost, concerning the results, it showed a significant effect of surface runoff on the accuracy of flow forecasting, followed by precipitation and temperature, with regard to the performance of the models, the results showed that machine learning models, especially ANN, XGBoost and RF, can provide accurate predictions of surface runoff compared to other used models, to improve surface water management through accurate prediction of discharge in drought-prone areas. If we compile the results of these studies, we find that the use of artificial intelligence models for the estimation of flow offers many advantages, especially in terms of accuracy, and it helps to make informed decisions about the hydrological domain. However, the availability of high-quality data remains the main responsible for the accuracy of these models.

This study combines the ANN algorithm and artificial bee colony (ABC) to make a new integrated hybrid model called ABC-ANN for estimation streamflow time series. Furthermore, this model was combined with various signal decomposition techniques to assess its efficiency in estimating streamflow time series in the East Black Sea Region (Türkiye).

Material and method

The East Black Sea Region

The study area includes three cities located in the northeast of the country, on the shores of the East Black Sea region, Rize, Ordu, Trabzon, which are located 41°01′29″N 40°31′20″E, 40°77′45″N 37° 44′08″ and 41°00′18″N 39°43′21″E respectively. The climate of the study area is the same as the climate of the east Black Sea region, a humid subtropical climate with warm and sometimes cold summers due to the direct influence of the Black Sea, where the average temperature is about 26.5 C. It is characterized by mild to cold winters, sometimes due to snowfall, especially in the mountains, with the average temperature reaching up to 5.7 C for the period (1991–2020). Precipitation in the study area is relatively equal and moderate to high, especially in the late autumn season from (October to December), especially in the mountainous areas, where they receive large amounts of rain, with an average of 178 mm. Most of the streams in the area flow vertically to the sea in narrow and deep valleys. (Turkish State Meteorological Service 2021). The location information of the stream gauging station (SGS) for which the current is estimated is shown in Fig. 1. In Table 1, the locations of SGS are addressed.

Table 1 Information on SGS used in the current study

River name	SGS name	SGS no	City name
Melet	Arıcılar	2238	Ordu
Karadere	Değirmencik köyü	2202	Trabzon
Çamlıdere	Dereköy	2215	Rize

Statistical coefficients of monthly average flow data from 3 SGS in the Eastern Black Sea Region are given in Table 2. The data structure can be considered and model assumptions can be tested by evaluating the mean values, standard deviations and distributions of the stream flows according to these statistical parameters.

Methods

Artificial neural networks

ANNs are among the most preferred AI models for predicting incomplete hydrological data such as precipitation and river flow (Dawson et al. 2005; Kueh and Kuok 2018). These models can quickly model the relationship between variables. ANNs consist of input, hidden and output layers. The layers learn by changing the information between them. The training process is based on reducing the error between the expected and actual output values by adjusting the model parameters. The model estimates streamflow data based on training of historical data. The computational steps of the ANN model are presented in Fig. 2. The mathematical formula of the ANN is given in Eq. 1.

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

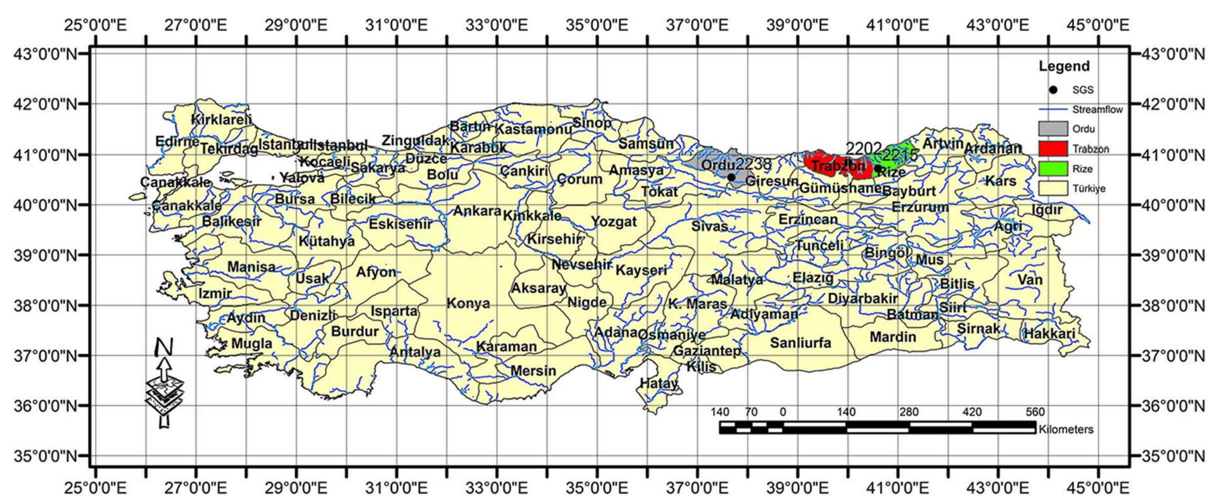


Fig. 1 Geographical coordinates of SGSs used in the study

Table 2 Statistical characteristics of streamflow data

	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec
	Ordu											
Mean	5.2	7.2	20.7	39.8	30.5	12.1	3.3	1.7	1.8	2.9	4.9	6.0
Max	18.1	25.4	51.1	76.6	58.7	35.3	6.8	3.1	3.6	8.1	22.1	18.7
Min	1.3	2.3	5.7	5.3	2.6	3.1	1.4	1.0	1.0	1.1	1.5	1.9
Std	3.7	4.9	9.6	15.6	12.3	6.5	1.3	0.5	0.5	1.7	4.1	4.0
Skewness	2.2	2.1	1.1	0.4	0.2	1.4	0.7	0.6	1.3	1.7	2.8	1.7
Kurtosis	4.9	4.5	1.8	0.1	0.1	2.9	0.2	0.3	2.8	2.4	8.5	2.3
	Trabzon											
Mean	4.3	5.6	11.1	28.9	33.5	16.2	6.8	4.7	4.4	6.6	7.5	5.4
Max	9.2	11.3	26.2	43.8	58.0	40.7	15.6	12.0	7.5	14.1	17.7	10.0
Min	1.6	2.6	5.4	17.1	14.2	6.4	2.8	2.5	1.7	1.8	2.0	2.2
Std	1.6	1.8	4.5	7.7	10.0	6.7	2.9	1.8	1.5	3.2	4.2	1.9
Skewness	0.8	0.9	2.0	0.4	0.3	1.2	1.0	2.0	0.1	0.6	1.0	0.3
Kurtosis	1.0	1.3	4.6	-0.9	-0.3	2.8	0.8	6.3	-0.7	-0.1	0.1	-0.5
	Rize											
Mean	4.3	4.4	7.1	19.6	38.9	37.4	19.0	8.5	5.9	7.1	6.8	5.2
Max	6.7	8.1	18.1	36.7	64.1	56.9	34.1	14.7	10.7	14.6	13.2	8.9
Min	2.4	2.7	3.3	8.8	27.1	20.4	9.6	4.8	3.3	3.5	3.0	2.6
Std	0.9	1.1	3.0	6.9	7.4	7.9	5.9	2.5	1.5	2.8	2.6	1.4
Skewness	0.6	1.1	2.4	0.6	0.8	0.1	0.4	0.7	1.0	1.2	1.1	0.9
Kurtosis	0.5	1.6	6.4	-0.3	1.4	-0.1	-0.3	0.0	2.0	1.2	0.6	1.0

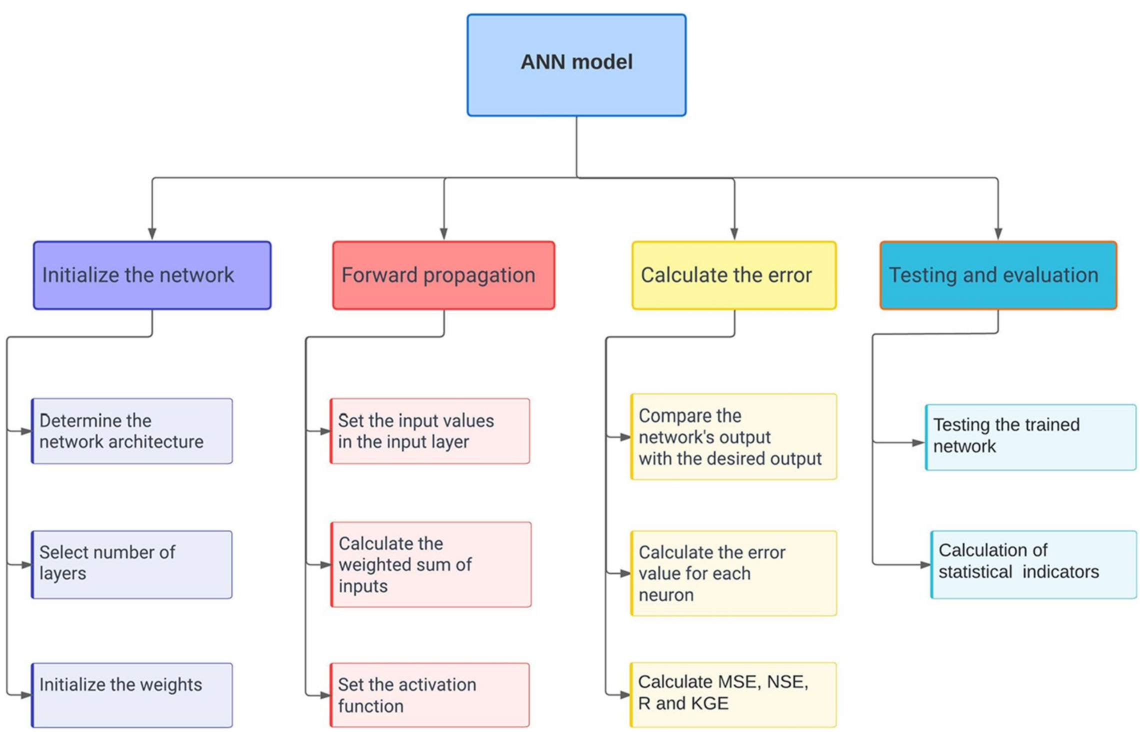


Fig. 2 Algorithmic program showing the setup steps of the ANN model

where y indicates the output, f shows the transfer function, w_i is the weight vector, x_i is the input vector, and b is the bias (Katipoğlu 2022).

Artificial bee colony (ABC)

ABC algorithm is one of the artificial intelligence techniques inspired by the intelligent behavior of bees in their search for food. Which focuses on studying the collective behavior of decentralized systems, represented by groupings of simple elements that interact locally with each other and with the surrounding environment. This algorithm is used to solve many optimization problems, i.e. issues that require reaching the optimal solution from a set of proposed solutions. (Karaboga and Basturk 2007). The ABC system has three types of bees: worker, observer, and explorer. After that, the bees disperse to search for food sources and if there is an ample nectar source compared to other sources in their search area. This determines the most abundant source among all the abundant sources (Karaboga et al. 2014) and this is exactly what optimization issues require. It is accepted that streamflow estimation is a challenging scenario that requires the utilization of a non-linear estimator to accurately reflect the correlating associations and dynamics. The current research used the ABC-ANN model as a non-linear estimator, taking advantage of the capabilities of the ANN to accurately depict the intricate and non-linear relations found in streamflow data. The ABC optimization algorithm was employed to optimize the ANN's parameters for enhancing the ability of ordinary ANN to capture the complex patterns in streamflow time series datasets. Figure 3 shows the application steps of ABC optimization technique.

The ABC algorithm conceptualizes natural processes and activities by representing them as algorithmic components and functionalities. In this representation, the concept of a "food source" is transformed into a "feasible solution" denoted as x_i, while the "nectar amount" corresponds to the fitness of a solution indicated by $F(x_i)$ as described in Eq. (2).

$$F\left(x_i\right)\begin{cases} \frac{1}{1+\mathrm{f}(\mathrm{x_i})}\mathrm{f}(\mathrm{x_i}) \geq 0 \\ \left|1+\mathrm{f}(\mathrm{x_i})\right| otherwise \end{cases} \tag{2}$$

$$\mathrm{v}_i = \mathrm{x}_i + \varphi i\left(\mathrm{x}_i - \mathrm{x}_n\right) \tag{3}$$

In the given equations, x_i represents the current solution, x_n represents the neighboring solution, and vi represents the candidate solution. The variable φi is a randomly generated number within the [-1, 1] range. The index i takes on values from 1 to N, indicating the index of the food source, where N represents the total number of food sources. Additionally, when the onlooker bees fail to find an improved solution, the scout bees can be generated using Eq. (4).

$$x_{ij} = \mathrm{LB}_j + rand(0,1) \times \left(UB_j - LB_j\right) \tag{4}$$

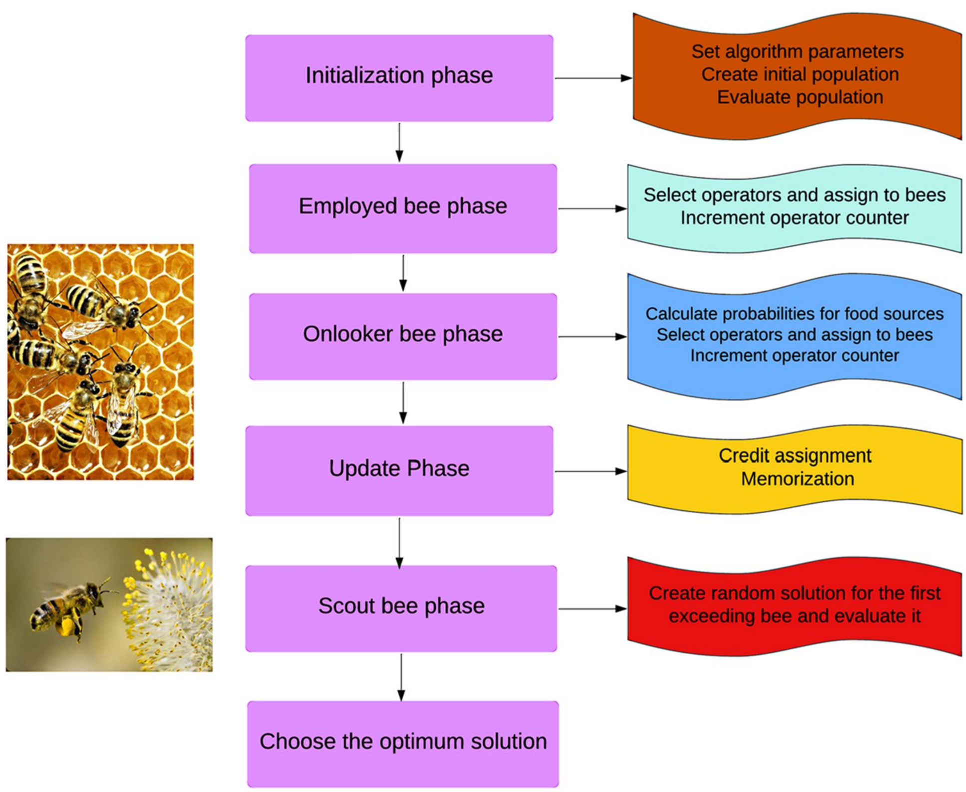


Fig. 3 Algorithmic program showing the setup steps of the ABC optimization (Durgut and Aydin 2021)

In the given context, $x_{i,j}$ represents the jth decision variable within the solution vector x_i. The index j ranges from 1 to D, meaning the total decision variables. Additionally, LB and UB denote the lower and upper boundary values specified for the decision variable (Durgut and Aydin 2021).

Local mean decomposition (LMD)

Applicable in various fields and of numerous applications because of its power (Lei et al. 2013), this signal processing technique makes complex signals into simpler components according to their local average frequencies. The LMD algorithm can iteratively extract a series of component functions from the input signal, each representing a different frequency range (Huang et al. 1998). Then, these component functions are combined to form the original signal, which can be reconstructed by summing all the extracted components (Huang et al. 1998). Then, these component functions are combined to form the original signal, which can be reconstructed by summing all the extracted components.

Application steps of the LMD technique consist of tree step: (i) Determining the input time series to the model, (ii) Determining the parsing level, (iii) Obtaining subcomponents. In LMD, the process of smoothing a signal involves applying moving averaging, while the weighting is determined by examining the gap between consecutive extrema. To begin the decomposition, the first step entails computing each half-wave oscillation's maximum and minimum points. In this scenario, the mean value "m_i" for the "ith" oscillation, positioned between two successive extrema "n_i" and "$n_i + 1$," is calculated according to the following method:

$$m_i = \frac{n_i + n_{i+1}}{2} \tag{5}$$

A uniformly varying continuous local mean function m(t) is obtained. Half-wave oscillations are expressed as follows:

$$a_i = \frac{|n_i - n_{i+1}|}{2} \tag{6}$$

For most natural data, LMD follows an iterative process that effectively captures a positive instantaneous frequency from a purely frequency-modulated signal with a constant envelope (Smith 2005).

Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)

Flandrin et al. (2011) proposed CEEMDAN, which separates complex signals into their basic components using IMFs. The CEEMDAN algorithm uses an adaptive sifting process that dynamically adjusts the bandwidth of the sifting window for each IMF to avoid mode mixing and improve the decomposition accuracy (Li et al. 2015). In addition, CEEMDAN reduces the effects of noise on the separation results. It then calculates the final decomposition result of each community member's IMFs based on the average. EEMD results are affected by residual noise that causes problems in the opposite direction. The number of trials can be improved with this technique by increasing the number of eliminations. To reduce mode mixing and the number of trials, the CEEMDAN technique is utilized (Torres et al. 2011). Signal processing techniques such as LMD and CEEMDAN provide advantages in reducing the noise in the data by separating the various subbands of the input time series, modeling the subcomponents at different frequencies and better understanding the structure of the data.

The steps for conducting a CEEMDAN analysis are as follows:

Apply the identical EEMD technique to compute the first modal function.

$$\overline{IMF_1(t)} = \frac{1}{N}\sum\nolimits_{j=1}^{N} IMF_1^j(t) \tag{7}$$

Additionally, a distinctive initial residue is determined by performing the following calculation:

$$r_1(t) = x(t) - \overline{IMF_1(t)} \tag{8}$$

The kth Intrinsic Mode Function (IMF) component, denoted as emd(t), is defined by applying Empirical Mode Decomposition (EMD). Next, the sequence sequencer1(t) + p1 * emd1(nj(t)) is decomposed to obtain the second IMF component.

$$\overline{IMF_2(t)} = \frac{1}{N}\sum\nolimits_{j=1}^{N} emd_1[r_1 t + p_1 emd_1(n_j(t)] \tag{9}$$

The residual signal is indicated following:

$$r_2(t) = r_1(t) - \overline{IMF_2(t)} \tag{10}$$

Similar to the procedure outlined in steps 1 and 2, the kth residual signal can be represented using the given equation.

$$r_k(t) = r_{k-1}(t) - IMF_k(t) \tag{11}$$

Similarly, the k + 1th Intrinsic Mode Function (IMF) component can be expressed using the provided equation.

$$\overline{IMF_{k+1}(t)} = \frac{1}{N}\sum\nolimits_{j=1}^{N} emd_1[r_k t + p_k emd_k(n_j(t)] \tag{12}$$

Continue iterating steps 1 to 3 until the residual signal satisfies the specified termination criterion. Assuming there are L IMF components, the original sequence can be expressed using the mentioned equation (Torres et al. 2011; Rezaie-Balf et al. 2019).

$$y(t) = \sum_{i=1}^{L} \overline{IMF_i(t)} + r(t) \tag{13}$$

Comparison of methods

For our case and to study the performance of the methods used to estimate the monthly streamflow, four indicators are used:

Mean squared error (MSE)

One of the common metrics used to evaluate the performance of regression models is the MSE scale. Where measures the average squared differences between the predicted (ŷ) values and the actual values (y). In other words, it measures the average amount by which forecasts deviate from actual values. The smaller the MSE value, the higher the accuracy of the model

$$MSE = \left(\frac{1}{n}\right) \sum (y_i - \hat{y}_i)^2 \tag{14}$$

Mean absolute percentage error (MAPE)

MAPE is a commonly used metric, as it measures the average percentage difference between the expected and actual values to evaluate the performance and calculate the prediction accuracy in terms of the error rate for the prediction model.

MAPE is calculated by Eq. 15 where (ŷ) are the predicted values and (y) are the actual values

$$MAPE = \left(\frac{1}{n}\right) \sum \left| \frac{(y_i - \hat{y}_i)}{y_i} \right| \tag{15}$$

Correlation and determination coefficient

Sometimes we try to find if there is a relationship or a connection between two or more variables. The correlation coefficient can answer this question graphically or numerically by calculating the correlation coefficient (R). The square of the correlation coefficient is the determination coefficient (R^2) and is calculated as follows:

$$R^2 = \left(\frac{\sum_{i=1}^{N} (x_i - \overline{x}) * (y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \overline{x})^2 * \sum_{i=1}^{N} (y_i - \overline{y})^2}} \right)^2 \tag{16}$$

Nash–sutcliffe efficiency (NSE)

NSE is an efficiency parameter that shows the fit and relationship between two-time series. NSE is a preferred indicator mostly to show the accuracy of the model. This indicator shows the variation of the observed data with the predicted data. NSE values range from -∞ to 1. A value of 1 for NSE indicates that the estimated dataset perfectly matches the actual data. It can be said that the closer the NSE value is to 1, the higher the model performance.

$$NSE = 1 - \frac{\left(\sum_{i=1}^{N} (y_i - \hat{y}_i)\right)^2}{\sum_{i=1}^{N} \left(y_i - \overline{\hat{y}}_i\right)^2} \tag{17}$$

Kling-gupta efficiency (KGE)

Developed by Gupta et al. (2009), the KGE is an indicator of goodness of fit for hydrological modeling to enable its different components to decompose correlation, variability bias, and mean bias properties. Kling et al. (2012) developed the NSE indicator to ensure that the rates of bias and variability were not cross-correlated and recommended this indicator.

$$KGE = \sqrt{(R-1) + (a-1)^2\, (b-1)^2} \tag{18}$$

$$a = \frac{S_{\hat{y}}}{S_y}, b = \frac{\overline{\hat{y}_i}}{\overline{y}} \tag{19}$$

$S_{\hat{y}}$ shows standard deviation of predictions, S_y indicates standard deviation of observations, $\overline{\hat{y}}_i$ isaverage of predictions, $\overline{y}$ is averageof observations. If the KGE value is close to 1, it means that the prediction results of the model have perfect agreement with the real values. The fact that the KGE's values are equal to zero indicates no relationship between the model estimates and the actual data.

Taylor diagrams

Taylor diagrams are used in various fields, such as meteorology and oceanography. It is a graphical method for comparing similarity and statistical parameters between two or more data sets. For this, the closeness of the prediction models to the reference point and the values of the statistical parameters are evaluated. In addition, the most suitable model can be decided according to the correlation coefficient between the prediction models and the real data set. The Taylor diagram provides a clear and concise way to visualize the similarities and differences between multiple datasets in a single graph (Taylor 2001).

Results and discussion

Streamflow estimation is vital for the risk management of floods, the supply of water resources, the construction of water structures and irrigation planning. Streamflow prediction accuracy has improved with the development of artificial intelligence technologies. In addition, ANN techniques are strengthened with various signal separation and bio-inspired optimization techniques. This study evaluated the one-month lead-time streamflow prediction success of the ANN model combined with LMD, CEEEMDAN and ABC algorithms. Furthermore, the performance of the created hybrid ANN model was evaluated according to various statistical and graphical approaches.

Choosing the input combination is critical to determine the best streamflow prediction model. Within the scope of the study, a suitable model structure was established using partial autocorrelation function (PACF) graphics. PACF graphs are shown in Fig. 4. Accordingly, lagged values exceeding 95% confidence limits were used in the modeling. Lagged values that exceed the confidence limits and have a high correlation in the PAC graphs are presented as input to the proposed model for streamflow estimation. For the analysis of the streamflows in Ordu, Trabzon and Rize according to PAC charts, 1, 2, 10, 11, 12-month delayed values, 1, 2, 11, 12-month delayed values and 1, 2, 11, 12-month lagged values were selected as an input, respectively (Table 3). Also, Fig. 5 shows Taylor diagrams of selected inputs based on PACF in each station, which were used for the input of estimator models. The t-11 datasets (green point) are most closet point to the target dataset (t + 1 which is black point) and t dataset (purple point) and t10 dataset (yellow point) are located at the second and third closet points (respectively) to the target dataset in all stations.

Figure 6 displays the subcomponents of delayed streamflow values obtained through the CEEEMDAN and LMD techniques in Ordu-2238 no SGS. The objective is to

Table 3 Selected model combinations for Q (t + 1) streamflow estimation

Station	Input	Target
Ordu- 2238 no SGS	t, t-1, t-9, t-10, t-11	t + 1
Trabzon 2202 no SGS	t, t-1, t-10, t-11	t + 1
Rize 2215 no SGS	t, t-1, t-10, t-11	t + 1

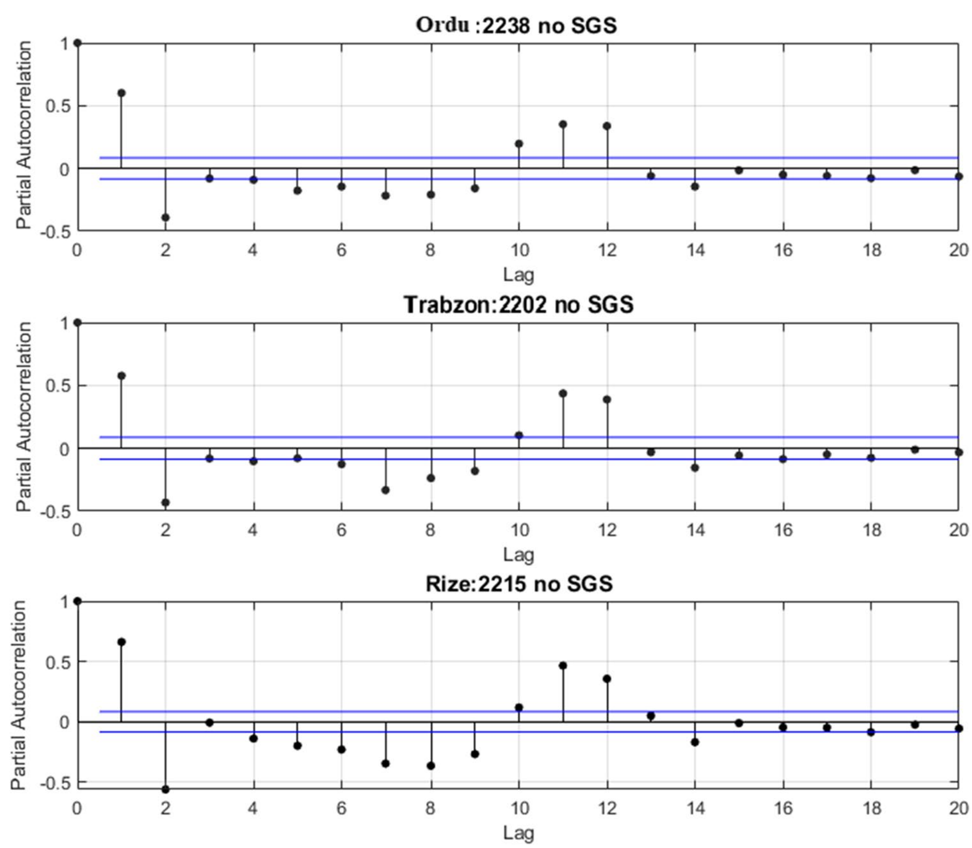


Fig. 4 PACF of streamflows of used SGS

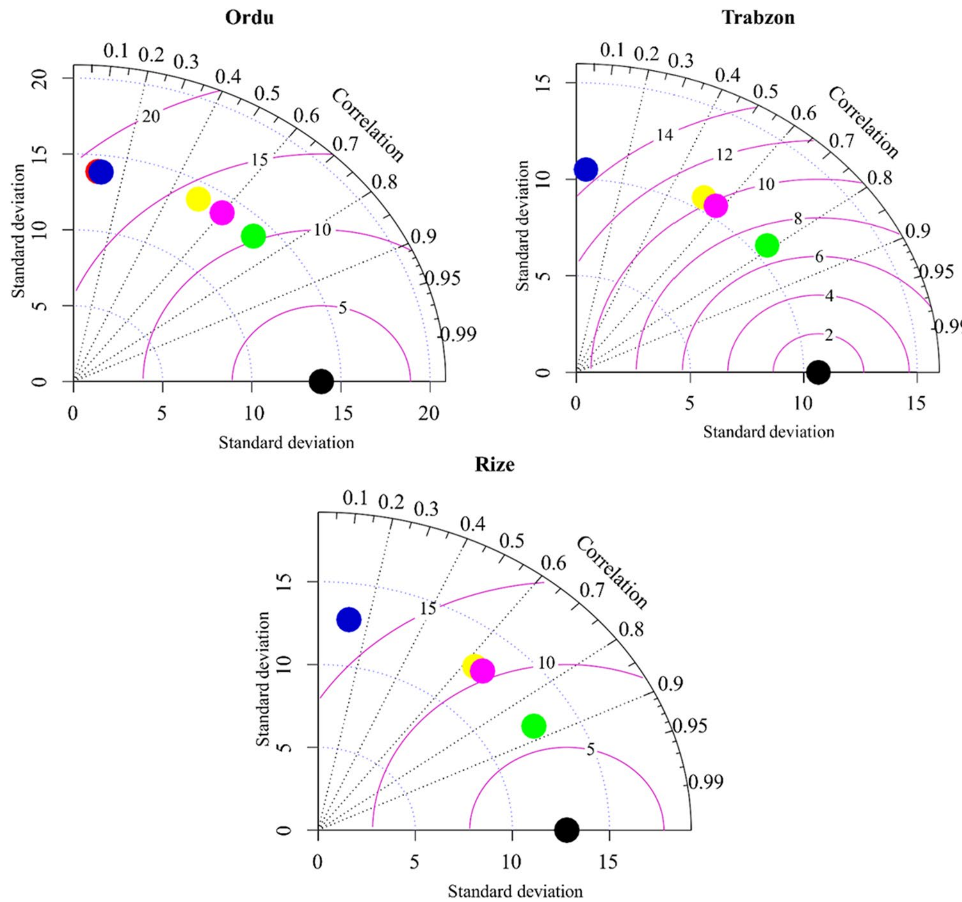


Fig. 5 Taylor diagrams of selected input data for estimator models based on PACF in each station; t-11: green point, t-10: yellow point, t-9: red point, t-1: blue point, t: purple point, t+1: black point

enhance the model's performance by presenting the obtained subcomponents to the ABC-ANN model. The CEEEMDAN algorithm separates lagged current values into various numbers of IMFs and residuals, enabling the intelligent model better to evaluate the input values' fluctuations and trends. With the LMD algorithm, the streamflow values are decomposed into sub-components, or product functions (PF), considering the impact of noise and outliers on the modeling. The goal is to strengthen the ABC-ANN model by incorporating these sub-components.

In Table 4, the performance evaluation of the models used in streamflow estimation has been made. Accordingly, the accuracy of the models was compared according to the KGE, MSE, NSE and R^2 statistical indicators. Therefore, the training and test results of the CEEEMDAN-ABC-ANN model in SGS no. 2238 in Ordu and SGS no. 2202 in Trabzon showed more successful results than ABC-ANN and LMD-ABC-ANN models in monthly streamflow estimation. The hybrid model CEEEMDAN-ABC-ANN has the highest prediction accuracy at station 2238 with the following values for training (MSE: 43.58, R^2: 0.79, NSE: 0.79, KGE: 0.84) and testing (MSE: 54.44, R^2: 0.67, NSE: 0.67, KGE: 0.75). At station 2202, the CEEEMDAN-ABC-ANN hybrid model has the highest prediction accuracy with the following values for training (MSE: 19.15, R^2: 0.83, NSE: 0.83, KGE: 0.84) and testing (MSE: 41.86, R^2: 0.66, NSE: 0.66, KGE: 0.76). In addition, the ABC-ANN model showed higher success than LMD-ABC-ANN and CEEEMDAN-ABC-ANN hybrid models during training and testing stages in Rize-2215 with no SGS. At station 2215, the ABC-ANN model has the highest prediction performance with the following statistical values for training (MSE: 17.88, R^2: 0.89, NSE: 0.88, KGE: 0.91) and testing (MSE: 36.92, R^2: 0.80, NSE: 0.80, KGE: 0.86).

In Fig. 7, the accuracy of the models used to estimate streamflow in SGS 2238 was evaluated according to the scatter diagrams. A scatter plot is a graphical indicator plotted along the X and Y axis to visualize the relationship between two variables and reveal correlations and outliers. In order to determine the appropriate model according to the scatterplots, an evaluation is made according to the distribution of the points around the 45-degree line. When the scatter diagrams are compared, it can be said that the performances of the established models in the streamflow estimation are close. Still, it can be said that the CEEEMDAN-ABC-ANN hybrid model gives more successful results than the other models. Figure 7 shows strong correlations between observed streamflow and predicted streamflow values for

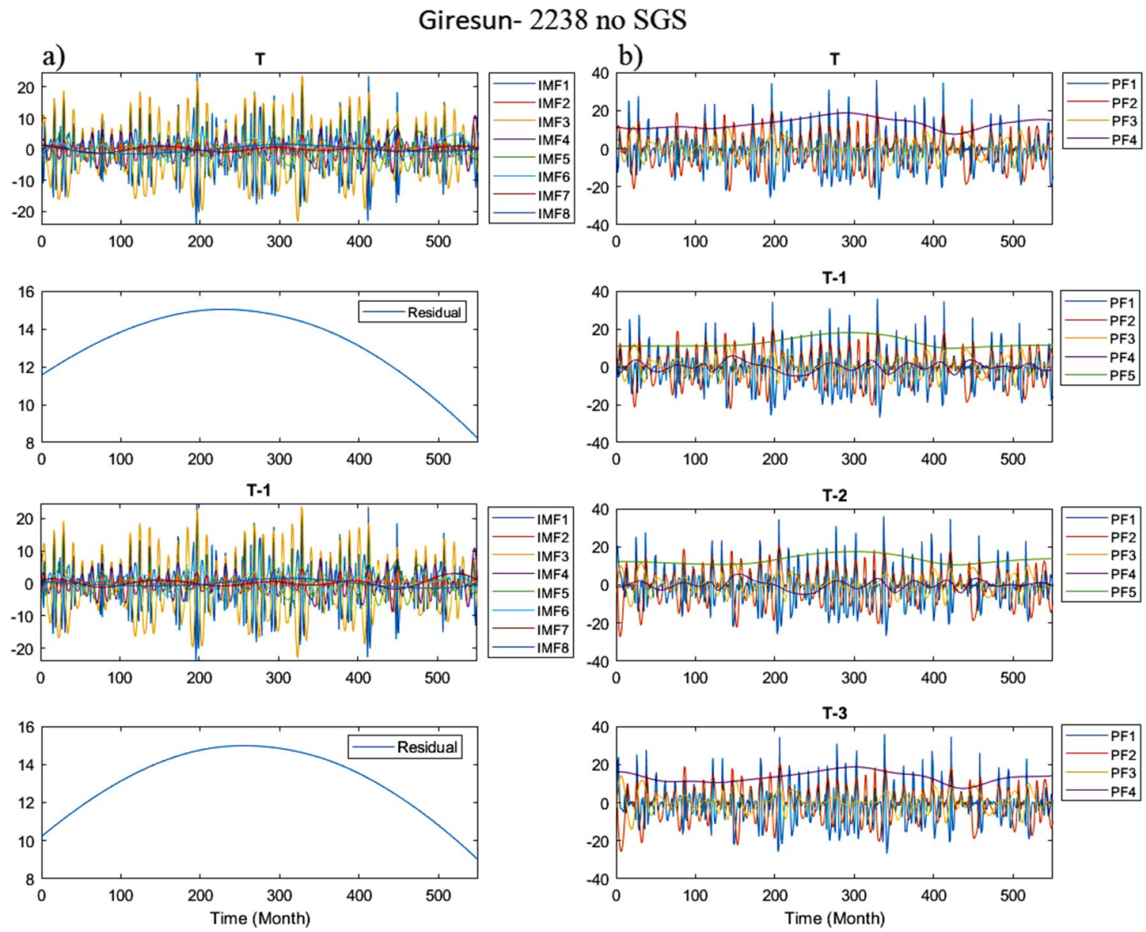


Fig. 6 Streamflow values decomposed to subcomponents in Ordu- 2238 no SGS **a**) CEEEMDAN algorithm **b**) LMD algorithm

the training and testing phases at Ordu station. Generally, correlations between predicted and observed streamflow in the training phase are higher than corresponding values in the testing phases.

Figure 8 analyzes the scattering diagrams of the models used to estimate streamflow in SGS 2202. When the scatter diagrams were evaluated, it was determined that the performances of the established models in the streamflow prediction were close to each other. However, it is seen that the CEEEMDAN-ABC-ANN hybrid model gives slightly more accurate results than other models. This issue indicates the higher performance of CEEEMDAN-ABC-ANN in both training and testing phases than other applied models. Also, it shows the ability of the ABC algorithm as a boosting tool to optimize the performance of the ANN model for streamflow precision.

In Fig. 9, the streamflow prediction performances of the model established in SGS 2215 are compared according to the scatter diagrams. According to these diagrams, the ABC-ANN model has higher accuracy than CEEEMDAN-ABC-ANN and LMD-ABC-ANN models in the training and testing stages. This higher performance of the ABC-ANN model is demonstrated in the scatter plot, where the points for the ABC-ANN model are more closely grouped around the actual target values.

Figure 10 shows time series plots of predicted and actual values in SGS 2238. These plots evaluated the relationship and spread between the actual and estimated streamflow values. According to Fig. 10, it can be said that the estimation results of the CEEEMDAN-ABC-ANN hybrid model during the training and testing phase are superior to other models since they spread following the real values. In addition, the LMD-ABC-ANN model is the weakest since it predicts the maximum current values with less accuracy than other models. The distribution of the predicted streamflow values around the observed streamflow values indicates that the CEEEMDAN-ABC-ANN model successfully captures the complexity and variability of streamflow in SGS 2238 with a high accuracy in the training and testing phases.

Figure 11 presents time series plots of estimated and actual streamflow values in SGS 2202. These plots analyzed the relationship and spread between the actual and

Table 4 Performance analysis of streamflow prediction via ABC-ANN, LMD-ABC-ANN, and CEEEMDAN-ABC-ANN models

Ordu—2238 no SGS				
Model phase	Statistic	ABC-ANN	LMD-ABC-ANN	**CEEEMDAN-ABC-ANN**
Train	MSE	57.66	55.57	**43.58**
	R^2	0.72	0.73	**0.79**
	NSE	0.72	0.73	**0.79**
	KGE	0.78	0.78	**0.84**
Test	MSE	67.27	62.65	**54.44**
	R^2	0.60	0.62	**0.67**
	NSE	0.59	0.62	**0.67**
	KGE	0.71	0.72	**0.75**
Trabzon—2202 no SGS				
Train		ABC-ANN	LMD-ABC-ANN	**CEEEMDAN-ABC-ANN**
	MSE	21.45	23.10	**19.15**
	R^2	0.81	0.79	**0.83**
	NSE	0.81	0.79	**0.83**
	KGE	0.83	0.84	**0.84**
Test	MSE	46.32	50.45	**41.86**
	R^2	0.63	0.60	**0.66**
	NSE	0.62	0.59	**0.66**
	KGE	0.76	0.72	**0.76**
Rize—2215 no SGS				
Train		**ABC-ANN**	LMD-ABC-ANN	CEEEMDAN-ABC-ANN
	MSE	**17.88**	18.83	21.67
	R^2	**0.89**	0.879	0.86
	NSE	**0.88**	0.85	0.86
	KGE	**0.91**	0.88	0.87
Test	MSE	**36.92**	56.79	45.57
	R^2	**0.80**	0.69	0.75
	NSE	**0.80**	0.67	0.75
	KGE	**0.86**	0.76	0.84

The best models are identified by bold characters representing statistical criteria

estimated streamflow values. The estimation results of the CEEEMDAN-ABC-ANN hybrid model during the training and testing phase are superior to other models since they best match the actual values. In addition, it is emphasized that the LMD-ABC-ANN model is the weakest because it predicts the maximum streamflow values with less accuracy than other models and has less overlap with the actual values.

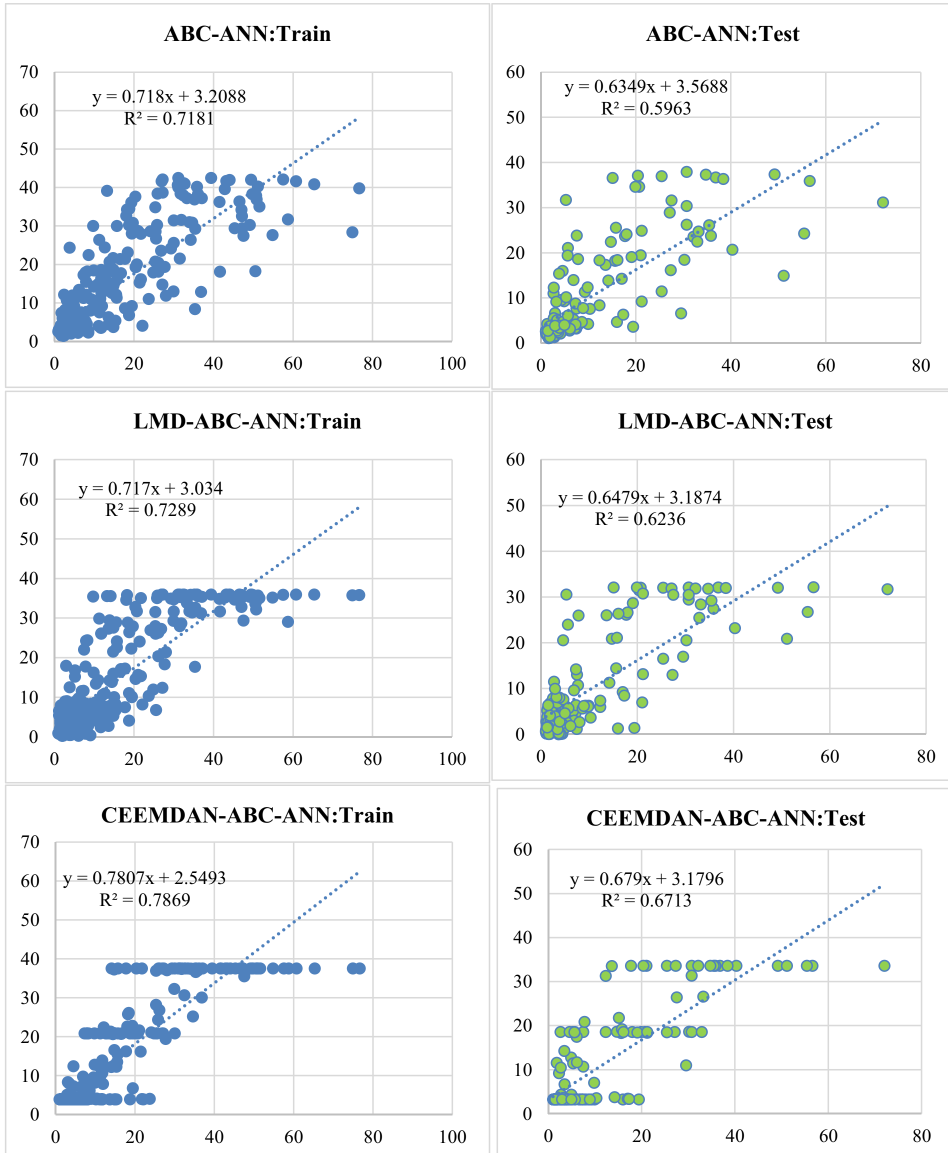
Figure 12 shows time series plots of streamflow values in SGS 2215. According to these plots, it is noteworthy that the actual and the estimated streamflow values overlap to a large extent. In addition, when the time series plots are examined in detail, it is seen that the ABC-ANN model represents the actual streamflow values better than the other models.

In Fig. 13, the potentials of the streamflow prediction models are compared with the Taylor diagrams. These graphs compared the estimated currents during the training and testing phases with the actual values. According to the statistical properties of the prediction model, which is close to the reference point, the most superior model was decided. According to these diagrams, it has been determined that the CEEEMDAN-ABC-ANN model in SGS 2238 and 2202 has the highest accuracy since it is closest to the reference point and has low RMSE and high R^2 values. Accordingly, it can be deduced that the CEEEMDAN technique is superior to the ABC-ANN model with its noise reduction in the input stream data, solving the mode mixing problem and time-varying structure. In addition, since the ABC-ANN model is closest to the reference point in SGS 2215, it is deduced that the estimations are the most realistic. In addition, all models showed satisfactory results in the flow estimation with values in the range of 0.80 to 0.95.

Discussion

The current study's main goal is to propose a new AI-based model coupled with data preprocessing approaches to enhance the accuracy of ABC-ANN for streamflow

Fig. 7 Evaluation of model performances in streamflow prediction at 2238 no SGS with scatter diagrams



simulation. The overall results showed that both preprocessing approaches (LMD and CEEEMDAN) increased the capability of ABC-ANN in the Ordu station, and CEEEMDAN-ABC-ANN only performed better than the ABC-ANN in Trabzon. Also, both LMD and CEEEMDAN acted worse than the ABC-ANN model in Rize station for streamflow simulation. Various results are reported based on different inputs to each station's different time series behavior. The key advantage of the ABC-ANN approach is that the ANN's parameters can be tuned via an optimized framework (ABC) to reach the highest accuracy of time series simulation via ABC-ANN mode. Furthermore, adding preprocessing techniques (e.g., CEEEMDAN) can help the model to detect the non-linear behavior streamflow, and therefore CEEEMDAN-ABC-ANN utilized both advantages of bio-inspired and preprocessing methods.

It can be observed that all applied models are more capable in streamflow simulation of the training phase. This issue can be justified due to the existence of a large amount of peak flow in the testing phase, which was not learned by models in the training phase and can affect the final result of the testing period. Although CEEEMDAN-ABC-ANN reported better streamflow simulation results for streamflow forecasting, some limitations to implanting CEEEMDAN-ABC-ANN include (1) combining CEEEMDAN with the ABC-ANN adds more complexity to the final model (CEEEMDAN-ABC-ANN). Furthermore, (2) although CEEEMDAN can decompose the original time series of streamflow data into several signals, ABC-ANN needs more effort and process and make more relationships between all these separated signals during the learning phase of the ANN model, which can increase model running time. (3) using the original time

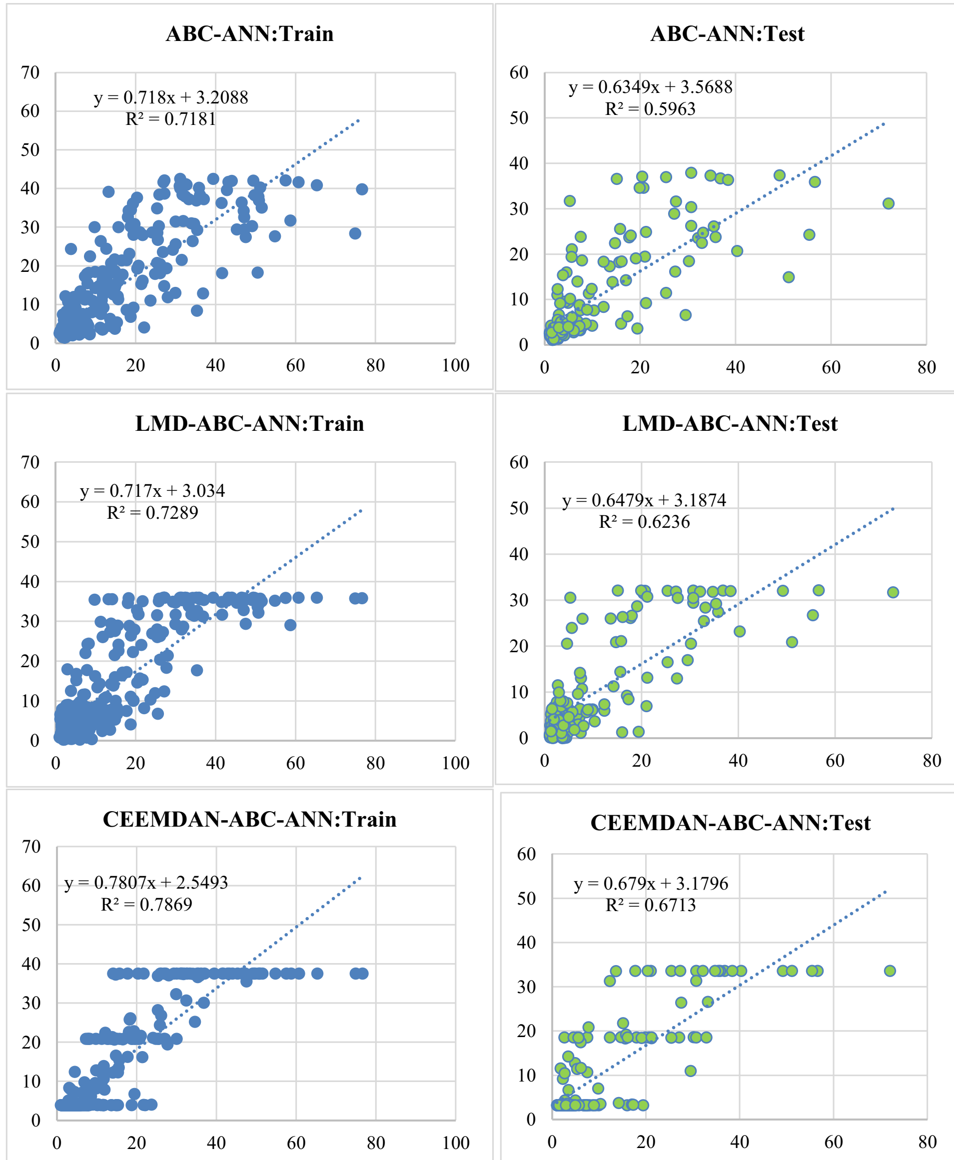


Fig. 8 Evaluation of model performances in streamflow prediction at 2202 no SGS with scatter diagrams

series of streamflow data is always understandable in the hydrological view, but using decomposed signals could not be understandable in the hydrological view. The modeling process could be a (deeper) black box with less explanation. (4) The ABC-ANN uses the artificial bee colony optimization algorithm in the ANN learning rate for tuning the ANN model's hyperparameters. Sometimes the algorithm could be trapped in the local minimize and also it is sensitive to choosing algorithm initial parameters and; therefore, reaching optimal results by ABC-ANN has more challenges compared with the ordinary ANN model.

In recent years, metaheuristic optimization algorithms have been successfully coupled with AI models as optimizer tools in solving complex non-linear issues for hydrological modeling tasks (Mahmoudi et al. 2022; Maroufpoor et al. 2019; 2020). Due to the high computational performance of AI-based models in solving non-linear problems, these models have been used for streamflow simulation worldwide. However, due to AI models' lack of hydrological terms, they fail to interpret hydrological processes (Mohammadi et al. 2022) physically. Previous studies such as Cheng et al. (2020), Difi et al. (2022), Wang et al. (2022), and Ayana et al. (2023) recommended AI techniques as powerful tools for capturing streamflow time series, while they also mentioned these AI techniques are sensitive to use data. Therefore, they should be trained well with enough time series data. However, both hydrological physically based models and AI-based models have some advantages and disadvantages in their application, while the type of case study can decide which type of model could be suitable. The ABC algorithm showed it can find the optimal solution with a high probability (Wang et al.

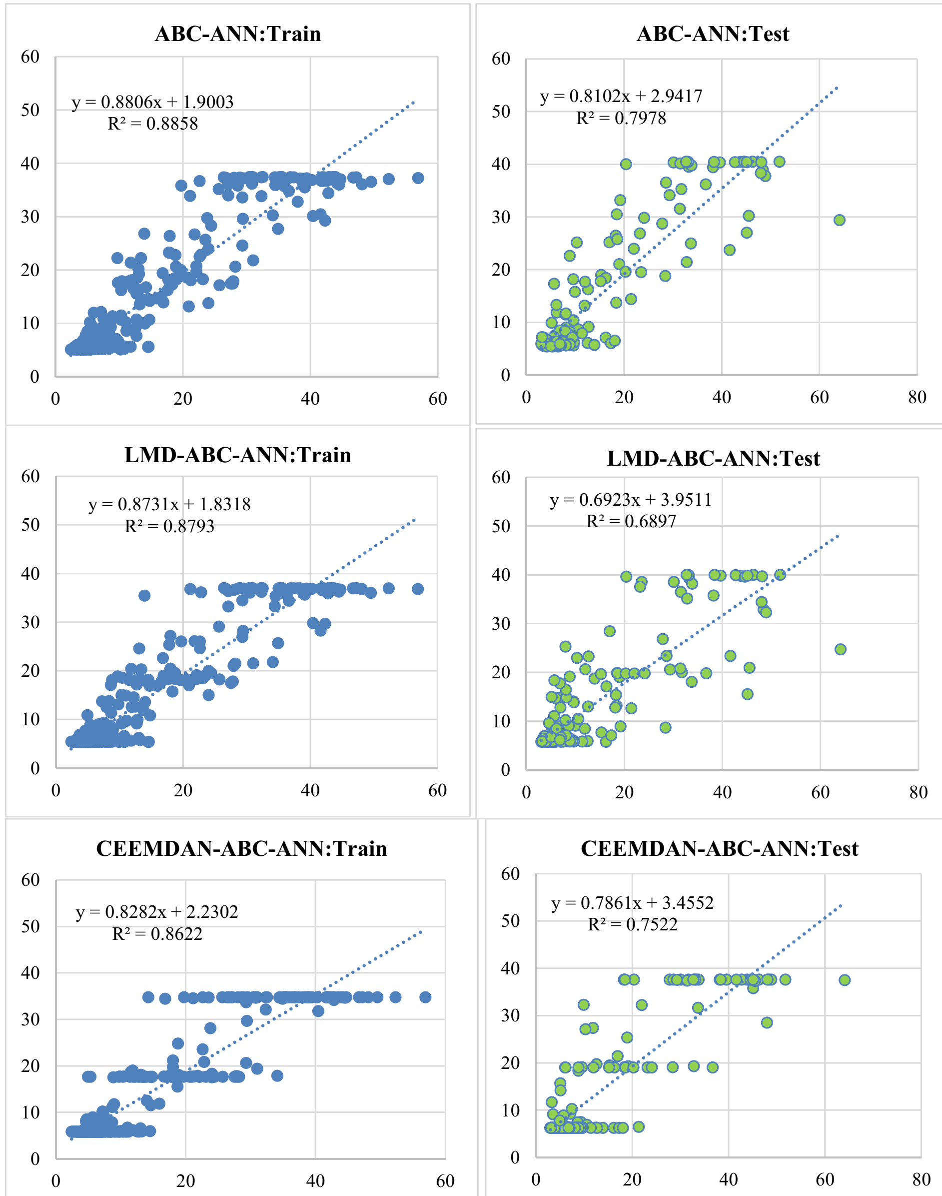


Fig. 9 Evaluation of model performances in streamflow prediction at 2215 no SGS with scatter diagrams

2020). Although it uses fewer control parameters, the ABC model can effectively solve multidimensional multimodal optimization, giving better results than other models (Karaboga and Akay 2009). Thanks to their ability to analyze non-linear and non-stationary data, the two signal processing techniques, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Local Mean Decomposition (LMD) have become popular and the most used in many fields. However, both CEEMDAN and LMD have some disadvantages as well. CEEMDAN can be computationally expensive, especially for large datasets, and may require significant computational resources. On the other hand, LMD may not perform well for signals with sharp transitions or discontinuities, as it relies on a smooth signal assumption.

Conclusion

Accurate streamflow simulation is vital for water resources management and environmental planning. The current study proposed a novel artificial intelligence technique based on an artificial bee colony combined with ANN (ABC-ANN) and the local mean decomposition

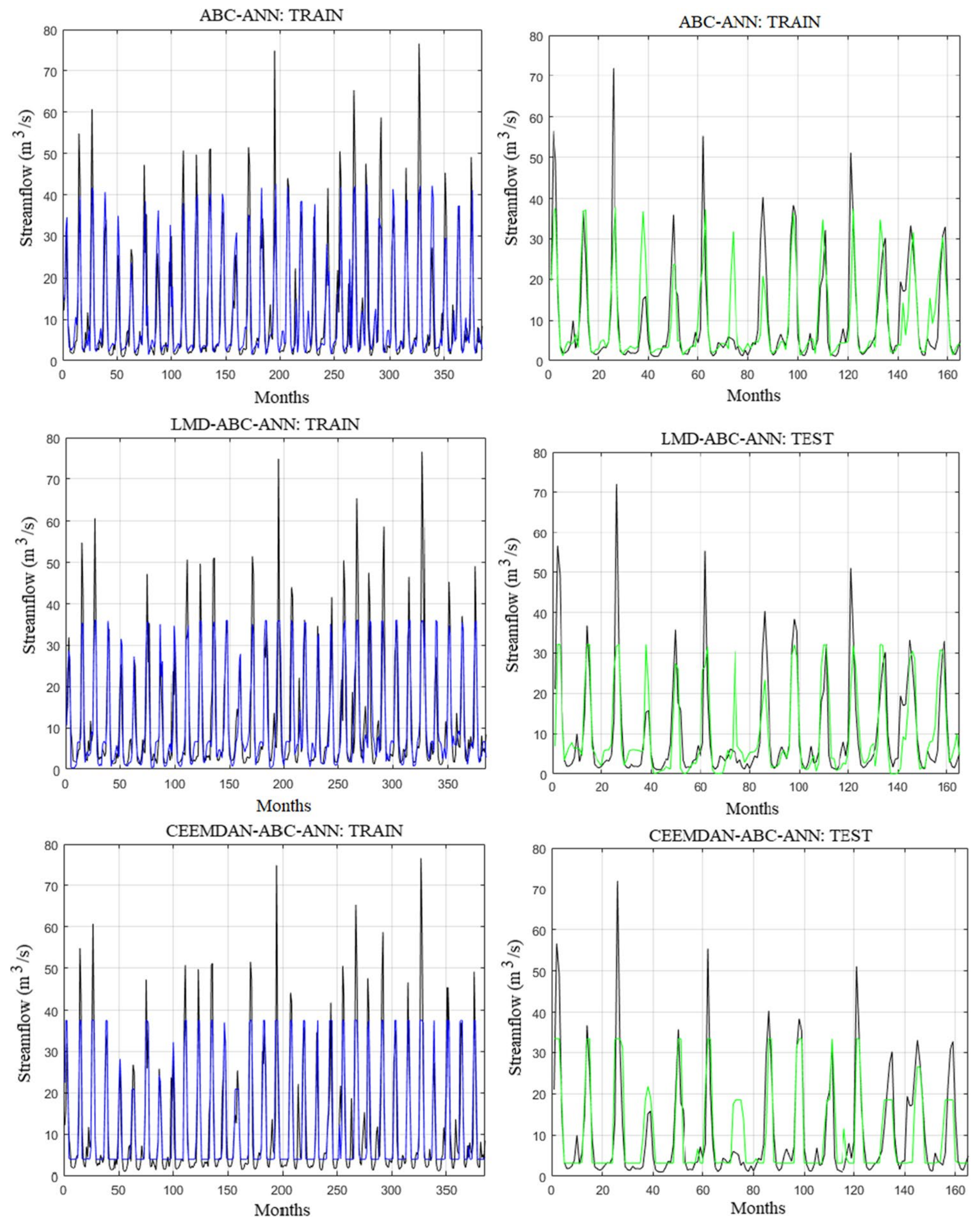


Fig. 10 Comparison of model performances with time series plot in 2238 no SGS

(LMD-ABC-ANN) for the monthly streamflow time series prediction in Ordu, Trabzon and Rize hydrometric stations (in the East Black Sea Region Türkiye). The model was enhanced by applying different signal decomposition techniques, including Local Mean Decomposition (LMD) and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). A partial autocorrelation function (PACF) was used to detect effective lag times of streamflow as input of estimator models. The lag times including *{t, t-1, t-9, t-10, t-11}* and *{t, t-1, t-10, t-11}* and *{t, t-1, t-10, t-11}* were selected as effective monthly lag times for predicting streamflow in $t+1$ in Ordu, Trabzon, and Rize stations, respectively. It can be concluded that all applied models simulated monthly streamflow with reliable performances in this study. The results showed that empirical envelope and complete ensemble empirical mode decomposition with adaptive noise combined with ANN-ABC (CEEEMDAN-ABC-ANN) outperformed in Ordu and Trabzon stations and ABC-ANN simulated streamflow by a higher accuracy compared with other applied models in Rize station. The CEEEMDAN method improved the capability of ABC-ANN in the Ordu and Trabzon stations with $R^2=0.67$ and 0.66 for the test section in the Ordu and

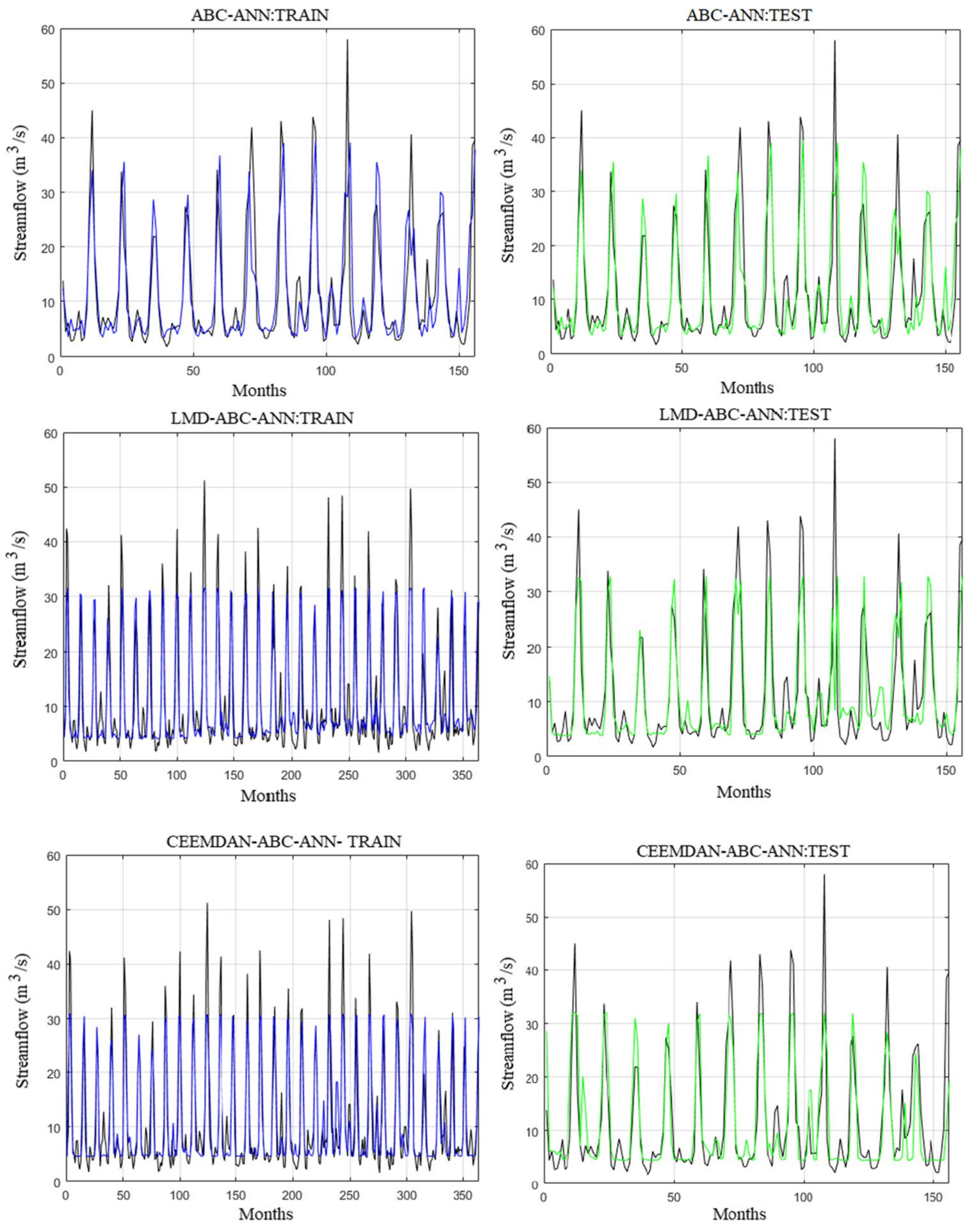


Fig. 11 Comparison of model performances with time series plot in 2202 no SGS

Trabzon stations, respectively. The results showed that the coupled ABC-ANN model, optimized via the ABC algorithm, achieved accurate streamflow estimation in the studied regions. The fluctuations and noise of streamflow time series data were captured by coupling the LMD and CEEMDAN techniques into estimator models, which helped to enhance streamflow time series prediction. The selection of lagged streamflow values by PACF analysis led to identifying the most effective input variables for each station, which also contributed to improving the accuracy of estimator models. The hybrid CEEEMDAN-ABC-ANN model resulted in lower MSE values and higher R^2 values compared to the ABC-ANN model in Ordu and Trabzon stations. This shows that integration of the CEEMDAN technique improved the accuracy of streamflow estimation during the training and testing phases in Ordu and Trabzon stations. This issue shows this coupled model can follow the dynamic pattern of streamflow time series during the training and testing phase. In Rize station, the ABC-ANN model resulted in a lower MSE and a higher R^2 values for the training and testing phases compared to the LMD-ABC-ANN and CEEEMDAN-ABC-ANN models. These findings suggest that the efficiency of the hybrid models might be contingent upon the particular characteristics and patterns of streamflow data at each station. It can be concluded that the CEEEMDAN technique can increase the accuracy of streamflow simulation compared with the LMD method and the CEEEMDAN-based estimators could be tested and generalized for streamflow simulation in various

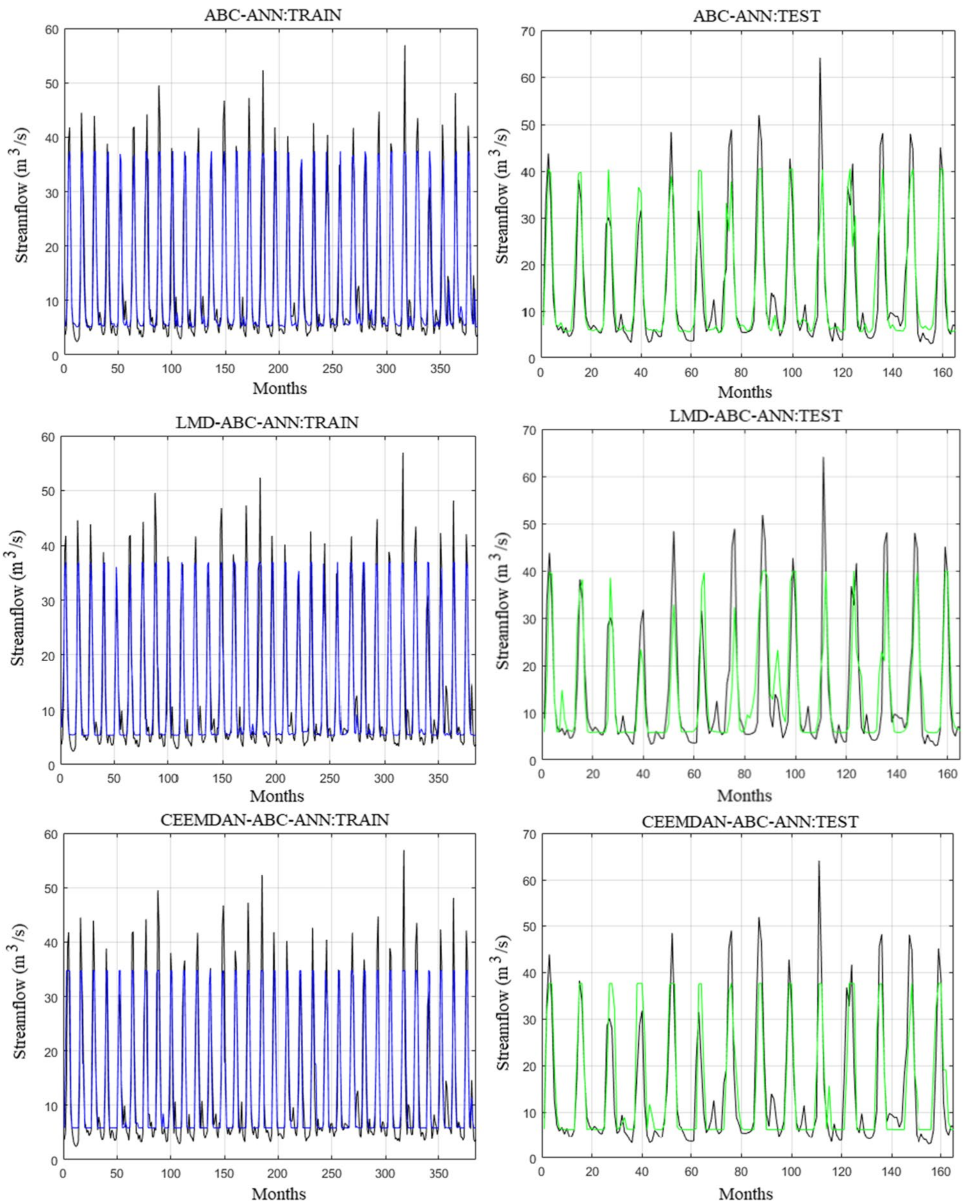


Fig. 12 Comparison of model performances with time series plot in 2215 no SGS

climates. Streamflow time series behavior and its prediction have been better understood using several AI-based methods, such as presented in this study. Streamflow simulation via these methods can be useful to increase our knowledge of the mechanisms that drive streamflow behavior in which these processes may be altering in response to climate change impacts.

Despite the good results that the artificial intelligence models give us, there are some limitations that may decrease the accuracy of the models. One of the basic limitations is the availability of data in terms of quality and quantity and the availability of other parameter data such as evaporation, evapotranspiration, and information on soil properties, and land cover changes. This last one, using it, gives us an evaluation and a better analysis of the effect of other relevant variables on the process flow simulation. Despite all these limitations, the use of artificial intelligence models in forecasting the streamflow has many advantages, which contribute to the good management of water resources, planning for drought, and solving the problem of floods. Finally, the CEEEMDAN-ABC-ANN model can be applied in other hydro-climatic contexts similar to those studied, also integrating the proposed model and comparing the proposed method with other types of metaheuristics optimization algorithms and some newly developed machine learning techniques such as deep learning models, thus providing a useful solution possible management of water resources in different regions of the world.

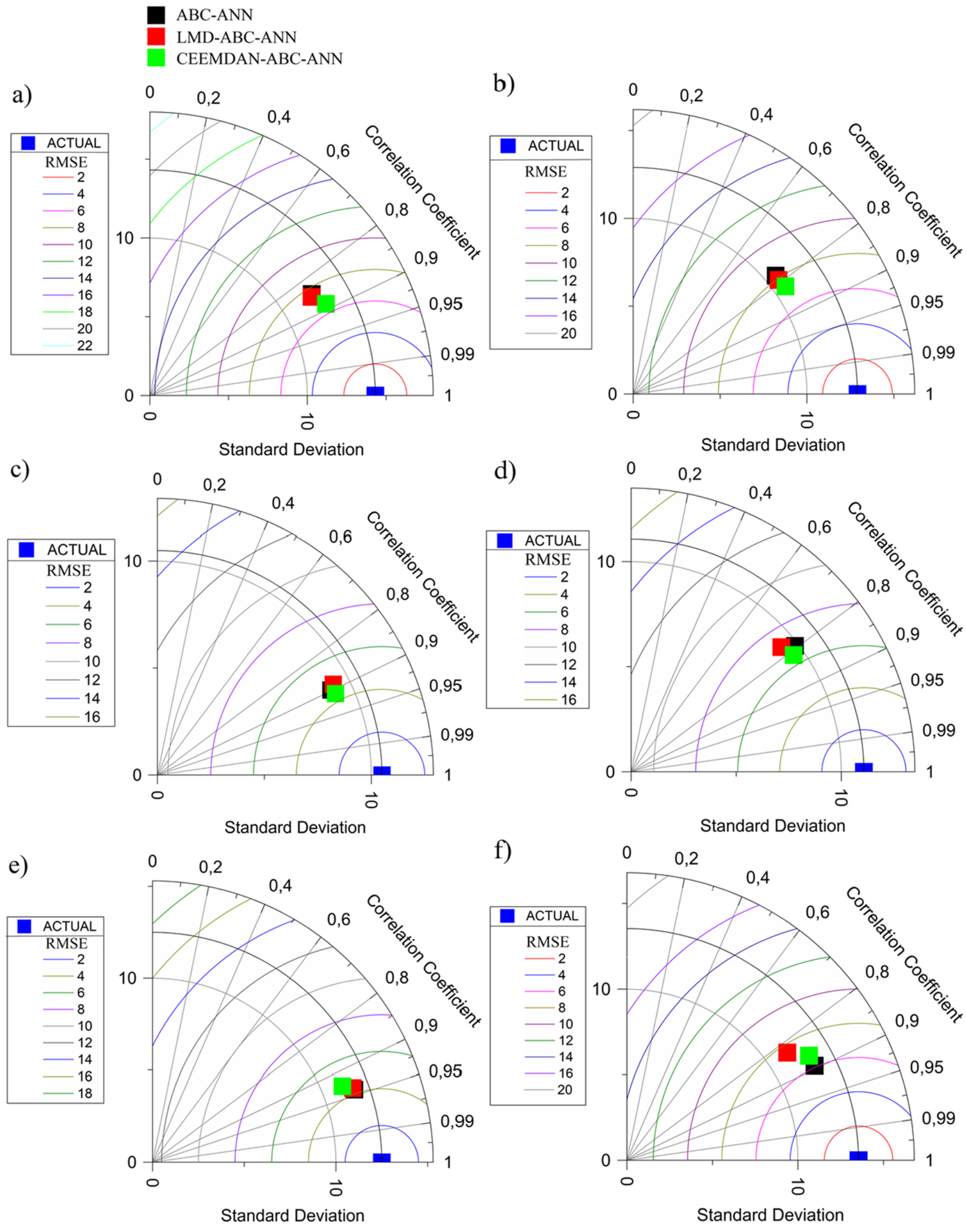


Fig. 13 Comparison of model performances with Taylor diagrams: **a**) 2238 no SGS training, **b**) 2238 no SGS test, **c**) 2202 no SGS training, **d**) 2202 no SGS test, **e**) 2215 no SGS training, **f**) 2215 no SGS test

Appendix

Table 5 List of abbreviations used in the study

Abbreviation	Explanation
ANN	Artificial Neural Network
ABC	Artificial Bee Colony
CEEMDAN	Complete Ensemble Empirical Mode Decomposition With Adaptive Noise
CNN	Convolutional Neural Network
ELM	Extreme Learning Machines
IMF	Intrinsic Mode Function
KGE	Kling-Gupta efficiency
LSTM	Long and Short Term Memory
LMD	Local Mean Decomposition
MSE	Mean Squared Error
MAPE	Mean Absolute Percentage Error
NSE	Nash–Sutcliffe Efficiency
R	Correlation Coefficient
R^2	Determination Coefficient
SGS	Stream Gauging Station

Acknowledgements The authors thank the general directorate of electric power resources survey and development administration for the monthly streamflow data provided.

Author contribution O. M. Katipoğlu contributed to the data analysis, Results and interpretation. M. Keblouti contributed to writing the literature review, Introduction and Material and Methods, B. Mohammadi wrote the Discussion and Conclusion sections and reviewed the paper. All authors read and approved the final manuscript.

Data availability Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval The manuscript complies with all the ethical requirements. The paper was not published in any journal.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflicts of interest The author declares no conflict of interest.

References

Adnan RM, Yuan X, Kisi O, Yuan Y (2017) Streamflow forecasting using artificial neural network and support vector machine models. Am Sci Res J Eng Technol Sci (ASRJETS) 29(1):286–294
Allan JD (2004) Landscapes and riverscapes: the influence of land use on stream ecosystems. Annu Rev Ecol Evol Syst 35(1):257–284. https://doi.org/10.1146/ecolsys.2004.35.issue-1. https://doi.org/10.1146/annurev.ecolsys.35.120202.110122
Akbarian M, Saghafian B, Golian S (2023) Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran. J Hydrol 620:129480. https://doi.org/10.1016/j.jhydrol.2023.129480
Ayana Ö, Kanbak DF, Kaya Keleş M, Turhan E (2023) Monthly streamflow prediction and performance comparison of machine learning and deep learning methods. Acta Geophysica 1–18. https://doi.org/10.1007/s11600-023-01023-6
Beven K (2006) A manifesto for the equifinality thesis. J Hydrol 320(1–2):18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007
Bosch JM, Hewlett JD (1982) A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. J Hydrol 55(1–4):3–23. https://doi.org/10.1016/0022-1694(82)90117-2
Cheng M, Fang F, Kinouchi T, Navon IM, Pain CC (2020) Long lead-time daily and monthly streamflow forecasting using machine learning methods. J Hydrol 590:125376. https://doi.org/10.1016/j.jhydrol.2020.125376
Dawson CW, See LM, Abrahart RJ, Wilby RL, Shamseldin AY, Anctil F, Belbachir AN, Bowden G, Dandy G, Lauzon N, Maier H, Mason G (2005) A comparative study of artificial neural network techniques for river stage forecasting. Proc Int Jt Conf Neural Networks 4:2666–2670. https://doi.org/10.1109/IJCNN.2005.1556324
Difi S, Elmeddahi Y, Hebal A, Singh VP, Heddam S, Kim S, Kisi O (2022) Monthly streamflow prediction using hybrid extreme learning machine optimized by bat algorithm: a case study of Cheliff watershed, Algeria. Hydrol Sci J 1–20. https://doi.org/10.1080/02626667.2022.2149334
Durgut R, Aydin ME (2021) Adaptive binary artificial bee colony algorithm. Appl Soft Comput 101:107054. https://doi.org/10.1016/j.asoc.2020.107054
Flandrin P, Torres E, Colominas MA (2011) A complete ensemble empirical mode decomposition Laboratorio de Senales y Dinamicas no Lineales, Universidad Nacional de Entre R Laboratoire de Physique (UMR CNRS 5672). Ecole Normale Superieure de Lyon, France, pp 4144–4147
Gerlak AK, Lautze J, Giordano M (2011) Water resources data and information exchange in transboundary water treaties. Int Environ Agreements Polit Law Econ 11:179–199. https://doi.org/10.1007/s10784-010-9144-4
Ghimire S, Yaseen ZM, Farooque AA, Deo RC, Zhang J, Tao X (2021) Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. Sci Rep 11:1–26. https://doi.org/10.1038/s41598-021-96751-4
Gudmundsson L (2021) Globally observed trends in mean and extreme river flow attributed to climate change. Science 371(6534):1159–1162. https://doi.org/10.1126/science.aba3996
Gupta HV, Sorooshian S, Yapo PO (1999) Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. J Hydraul Eng 4(2):135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)
Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J Hydrol 377(1–2):80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003
Ha S, Liu D, Mu L (2021) Prediction of Yangtze River streamflow based on deep learning neural network with El Niño-Southern Oscillation. Sci Rep 11:1–23. https://doi.org/10.1038/s41598-021-90964-3
Huang WC, Yang FT (1998) Streamflow estimation using Kriging. Water Resour Res 34:1599–1608. https://doi.org/10.1029/98WR00555
Huang NE, Shen Z, Long SR, Wu MC, Snin HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and

the Hubert spectrum for non-linear and non-stationary time series analysis. Proc R Soc A Math Phys Eng Sci 454:903–995. https://doi.org/10.1098/rspa.1998.0193

Jencso KG, McGlynn BL, Gooseff MN, Wondzell SM, Bencala KE, Marshall LA (2009) Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale. Water Resour Res 45:1–16. https://doi.org/10.1029/2008WR007225

Karaboga D, Akay B (2009) A comparative study of Artificial Bee Colony algorithm. Appl Math Comput 214:108–132. https://doi.org/10.1016/j.amc.2009.03.090

Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. J Glob Optim 39:459–471. https://doi.org/10.1007/s10898-007-9149-x

Karaboga D, Gorkemli B, Ozturk C, Karaboga N (2014) A comprehensive survey: Artificial bee colony (ABC) algorithm and applications. Artif Intell Rev 42:21–57. https://doi.org/10.1007/s10462-012-9328-0

Katipoğlu OM (2020) Data division effect on machine learning performance for prediction of streamflow. Dicle Univ Eng Fac J Eng 13(4):653–660. https://doi.org/10.24012/dumf.1158748

Katipoğlu OM (2022) Evaluation of the performance of data-driven approaches for filling monthly precipitation gaps in a semi-arid climate conditions. Acta Geophysica 1–21. https://doi.org/10.1007/s11600-022-00963-9

Katipoğlu OM, Can I (2018) Determining the lengths of dry periods in annual and monthly stream flows using runs analysis at Karasu River, in Turkey. Water Sci Technol: Water Supply 18(4):1329–1338

Kavetski D, Kuczera G, Franks SW (2006) Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory Water Resour Res 42:1–9. https://doi.org/10.1029/2005WR004368

Kiang JE, Gazoorian C, McMillan H, Coxon G, Le Coz J, Westerberg IK, Belleville A, Sevrez D, Sikorska AE, Petersen-Øverleir A, Reitan T, Freer J, Renard B, Mansanarez V, Mason R (2018) A Comparison of Methods for Streamflow Uncertainty Estimation. Water Resour Res 54:7149–7176. https://doi.org/10.1029/2018WR022708

Kibler KM, Biswas RK, Lucas AMJ (2014) Hydrologic data as a human right? Equitable access to information as a resource for disaster risk reduction in transboundary river basins. Water Policy 16:36–58. https://doi.org/10.2166/wp.2014.307

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

Kueh SM, Kuok KK (2018) Forecasting long term precipitation using cuckoo search optimization neural network models. Environ Eng Manag J 17:1283–1291. https://doi.org/10.30638/eemj.2018.127

Le M, Kim H, Adam S, Do HX, Beling PA (2022) Streamflow estimation in ungauged regions using machine learning : quantifying uncertainties in geographic extrapolation. Hydrol Earth Syst Sci Discuss 1–24

Lei Y, Lin J, He Z, Zuo MJ (2013) A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mech Syst Signal Process 35:108–126. https://doi.org/10.1016/j.ymssp.2012.09.015

Li C, Zhan L, Shen L (2015) Friction signal denoising using complete ensemble EMD with adaptive noise and mutual information. Entropy 17:5965–5979. https://doi.org/10.3390/e17095965

Mahmoudi N, Majidi A, Jamei M, Jalali M, Maroufpoor S, Shiri J, Yaseen ZM (2022) Mutating fuzzy logic model with various rigorous meta-heuristic algorithms for soil moisture content estimation. Agric Water Manag 261:107342. https://doi.org/10.1016/j.agwat.2021.107342

Maity R, Kashid SS (2011) Importance analysis of local and global climate inputs for basin-scale streamflow prediction. Water Resour Res 47(11). https://doi.org/10.1029/2010WR009742

Maroufpoor S, Maroufpoor E, Bozorg-Haddad O, Shiri J, Yaseen ZM (2019) Soil moisture simulation using hybrid artificial intelligent model: Hybridization of adaptive neuro fuzzy inference system with grey wolf optimizer algorithm. J Hydrol 575:544–556. https://doi.org/10.1016/j.jhydrol.2019.05.045

Maroufpoor S, Bozorg-Haddad O, Maroufpoor E (2020) Reference evapotranspiration estimating based on optimal input combination and hybrid artificial intelligent model: Hybridization of artificial neural network with grey wolf optimizer algorithm. J Hydrol 588:125060. https://doi.org/10.1016/j.jhydrol.2020.125060

Mohammadi B, Safari MJS, Vazifehkhah S (2022) IHACRES, GR4J and MISD-based multi conceptual-machine learning approach for rainfall-runoff modeling. Sci Rep 12(1):12096. https://doi.org/10.1038/s41598-022-16215-1

Nash JE, Sutcliffe JV (1970) River Flow Forecasting Through Conceptual Models - Part I - A Discussion of Principles. J Hydrol 10:282–290

Pini M, Scalvini A, Liaqat MU, Ranzi R, Serina I, Mehmood T (2020) Evaluation of machine learning techniques for inflow prediction in Lake Como. Italy Procedia Comput Sci 176:918–927. https://doi.org/10.1016/j.procs.2020.09.087

Pokhrel Y, Satoh Y, Kim H, Ward PJ, Ostberg S (2017) Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century. Nat Commun 8(1):15697. https://doi.org/10.1038/ncomms15697

Rahsepar M, Mahmoodi H (2014) Predicting weekly discharge using artificial neural network (ANN) optimized by Artificial Bee Colony (ABC) algorithm: a case study. Civil Engineering and Urban Planning: An International Journal (CiVEJ) 1(1)

Razavi S, Tolson BA, Burn DH (2012) Review of surrogate modeling in water resources. Water Resour Res 48. https://doi.org/10.1029/2011WR011527

Rezaie-Balf M, Naganna SR, Kisi O, El-Shafie A (2019) Enhancing streamflow forecasting using the augmenting ensemble procedure coupled machine learning models: case study of Aswan High Dam. Hydrol Sci J 64(13):1629–1646. https://doi.org/10.1080/02626667.2019.1661417

Siddiqi TA, Ashraf S, Khan SA, Iqbal MJ (2021) Estimation of data-driven streamflow predicting models using machine learning methods. Arab J Geosci. https://doi.org/10.1007/s12517-021-07446-z

Sivapalan M, Takeuchi K, Franks SW, Gupta VK, Karambiri H, Liang X, Mcdonnell JJ, Mendiondo EM, Connell PEO, Oki T, Pomeroy JW, Schertzer D, Uhlenbrook S, Zehe E (2012) IAHS Decade on Predictions in Ungauged Basins (PUB), 2003 – 2012 : Shaping an exciting future for the hydrological sciences IAHS Decade on Predictions in Ungauged Basins (PUB), 2003 – 2012 : Shaping an exciting future for the hydrological sciences 6667:2003–2012. https://doi.org/10.1623/hysj.48.6.857.51421

Smith JS (2005) The local mean decomposition and its application to EEG perception data. J R Soc Interface 2(5):443–454. https://doi.org/10.1098/rsif.2005.0058

Swain JB, Jha R, Patra KC (2015) Stream Flow Prediction in a Typical Ungauged Catchment Using GIUH Approach. Aquat Procedia 4:993–1000. https://doi.org/10.1016/j.aqpro.2015.02.125

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res: Atmospheres 106(D7):7183–7192

Torres ME, Colominas MA, Schlotthauer G, Flandrin P (2011, May) A complete ensemble empirical mode decomposition with adaptive noise. In 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4144–4147). IEEE. https://doi.org/10.1109/ICASSP.2011.5947265

Turkish State Meteorological Service (2021) Resmi İstatistikler: İllerimize Ait Mevism Normalleri (1991–2020)

Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV (2003) Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrol Process 17:455–476. https://doi.org/10.1002/hyp.1135

Wang W, Van Gelder PH, Vrijling JK, Ma J (2006) Forecasting daily streamflow using hybrid ANN models. J Hydrol 324(1–4):383–399. https://doi.org/10.1016/j.jhydrol.2005.09.032

Wang Y, Liu J, Li R, Suo X, Lu E (2020) Precipitation forecast of the Wujiang River Basin based on artificial bee colony algorithm and backpropagation neural network. Alexandria Eng J 59:1473–1483. https://doi.org/10.1016/j.aej.2020.04.035

Wang K, Band SS, Ameri R, Biyari M, Hai T, Hsu CC, Hadjouni M, Elmannai H, Chau KW, Mosavi A (2022) Performance improvement of machine learning models via wavelet theory in estimating monthly river streamflow. Eng Appl Comput Fluid Mech 16(1):1833–1848. https://doi.org/10.1080/19942060.2022.2119281

Xuan Do H, Zhao F, Westra S, Leonard M, Gudmundsson L, Eric Stanislas Boulange J, Chang J, Ciais P, Gerten D, Gosling SN, Müller Schmied H, Stacke T, Telteu CE, Wada Y (2020) Historical and future changes in global flood magnitude - evidence from a model-observation investigation. Hydrol Earth Syst Sci 24:1543–1564. https://doi.org/10.5194/hess-24-1543-2020

Zamoum S, Souag-Gamane D (2019) Monthly streamflow estimation in ungauged catchments of northern Algeria using regionalization of conceptual model parameters. Arab J Geosci 12:1–14. https://doi.org/10.1007/s12517-019-4487-9

Zhang L, Dawes WR, Walker GR (2001) Response of mean annual evapotranspiration to vegetation changes at catchment scale. Water Resour Res 37:701–708. https://doi.org/10.1029/2000WR900325