



Application of the novel circulant singular spectrum analysis ensemble model for forecasting of streamflow data

Levent Latifoğlu¹

Received: 22 November 2021 / Accepted: 27 April 2022 / Published online: 12 May 2022
© Saudi Society for Geosciences 2022

Abstract

Accurate forecasting of streamflow is an important part of water resource management and a difficult job for engineers. In this study, a circulant singular spectrum analysis (CiSSA)-based novel approach for forecasting daily streamflow data is proposed. Obtained features using CiSSA methods are applied to support vector regression (SVR), random forest (RF), and artificial neural network (ANN) models. In addition, the variational mode decomposition (VMD) method was used for preprocessing the streamflow data and to compare the performance of the CiSSA method.

One to five-day ahead forecasting performance was investigated and compared with single SVR, RF, ANN single models. VMD-SVR, VMD-RF, VMD-ANN, CiSSA-SVR, CiSSA-RF, CiSSA-ANN, CiSSA-VMD-SVR, CiSSA-VMD-RF, and CiSSA-VMD-ANN ensemble models and Kruskal–Wallis test was used to show whether the forecasting results are statistically significant. The analysis of the obtained results showed that the performance of the models (CiSSA-SVR, CiSSA-RF, and CiSSA-ANN) that use CiSSA features as input for far-day ahead predictions were significantly better than the other models. Taylor diagrams and mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (R), Nash–Sutcliffe efficiency coefficient (NSE), and Willmott's refined index (WI) performance parameters revealed that the performance of the CiSSA-RF model was better. Furthermore, the CiSSA-RF model proved to be significantly much reliable for longer predicting periods.

Keywords Streamflow forecasting · Circulant singular spectrum analysis · Variational mode decomposition · Support vector regression · Random forest · Artificial neural network

Introduction

Hydrology is a basic applied interdisciplinary science that studies the cycle, distribution, formation, physical, chemical, and biological properties of water in the earth (i.e., on the surface, underground, and in the atmosphere). Also, hydrological studies are critical during the planning, design, development, and service phases of the facilities that are used for water-related studies such as water usage and water quantity management. In water engineering research, hydrological data (such as streamflow and precipitation) play an important role in understanding and predicting nature's character and behavior (Sivapalan

2018). The nature of such data is nonlinear and non-stationary. To estimate and simulate such data, numerous studies were performed in the literature. Studies on forecasting for hydrological parameters and time series models began with linear methods, such as the parametric auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA) (Box and Jenkins 1976). Depending on the underlying physics, most deterministic processes require large amount of data for modeling. Without considering the underlying physical processes, black-box models assist in establishing relationships between input and output variables. These models are data-driven models which connect the system state variables (input, internal, and output variables) and need only limited information about the details of the physical behavior of the system. Due to the availability of the data, these methods have been gaining popularity over the past two decades. Machine learning (decision trees, Bayesian and instance-based methods, neural networks, reinforcement learning) methods, fuzzy rule-based systems, support vector regression (SVR) data-driven techniques can be

Responsible Editor: Broder J. Merkel

✉ Levent Latifoğlu
latifoglu@erciyes.edu.tr

¹ Department of Civil Engineering, Faculty of Engineering, Erciyes University, Kayseri, Türkiye

given as examples (Akçay and Manisali 2018; Akyuncu et al. 2018; Acar Yildirim and Akçay 2019; Mosaffa et al. 2022). According to literature, approaches based on artificial intelligence techniques such as ANN, SVR, random forest (RF), and adaptive network-based fuzzy inference system (ANFIS), which are nonlinear models, have recently been used in the forecasting of hydrological variables (Rasouli et al. 2012; Özger 2009; Luo et al. 2019; Sabzi et al. 2018; Yaseen et al. 2017, 2015; Oyebo and Stretch 2019). However, hybrid approaches have been developed to increase the performance of these approaches.

Some examples of the widely used time series pre-processing techniques can be given as wavelet transform (WT), Fourier transform (FT), singular spectrum analysis (SSA), empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and variational mode decomposition (VMD) (Liu et al. 2014; Kisi et al. 2014; Latifoğlu et al. 2015; Zhu et al. 2019; Lahmiri 2016; He et al. 2019; Wang et al. 2017; Zhang et al. 2018; Unnikrishnan and Jothiprakash 2018; Zounemat-Kermani et al. 2021a; Hassani 2007). These techniques split the original non-stationary signals into multiple smaller subband signals. The reconstructed time series are then used as partial or complete inputs to machine learning models.

Recently, literature has started being enriched with the empirical mode decomposition (EMD) technique in which empirical approach is used for the analysis of nonlinear and non-stationary data in addition to the WT transform (Liu et al. 2014; Kisi et al. 2014). But WT-based, EMD-based models along with an extension of these methods are employed to restrict the boundary effect. This causes the prediction performance of hybrid models to deteriorate (Zhang et al. 2015). These are the main disadvantages of mentioned methods. Another important method that separates the signal according to time information is the SSA technique, also considered the most powerful techniques for multivariate analysis (Hassani 2007). This technique is a non-parametric model and does not require prior knowledge about the data and it does not suffer the boundary effect as is the case for the SSA-based forecasting models (Zhu et al. 2019; Zhang et al. 2015).

Recently, the variational mode decomposition (VMD) method has been used for the forecasting of economic and financial time series, carbon price, stock price, daily runoff, wind speed, and streamflow data (Lahmiri 2016; He et al. 2019; Wang et al. 2017; Zhang et al. 2018; Zuo et al. 2020). The subband signals of VMD are more uncorrelated than those of EEMD, according to Jungang Luo's study (Zuo et al. 2020). As a result, VMD outperforms EEMD and DWT in terms of decomposition.

One of the newest methods, the CiSSA algorithm has been proposed as a nonparametric, nonlinear method for the analysis of non-stationary and nonlinear time series data (Bógalo et al. 2021). Bógalo showed the performance of the CiSSA algorithm

compared to the basic SSA and EMD methods (Bógalo et al. 2021) and revealed that the decomposed signals using the CiSSA were strongly separable. In this study, a reliable and efficient CiSSA-based ensemble model is proposed, as a novel approach, for the forecasting of streamflow data. Additionally, the VMD-based ensemble model is applied for the forecasting study. This is the first study in the literature which analyses the prediction performance of the CiSSA model and compares it with the VMD model.

In the hybrid approaches used for prediction, the signals are decomposed into subband data and the subband using preprocessing techniques such as EMD, WT, and VMD. Separate models have been developed for the estimation of each of these subband signals. The total estimation data is obtained with the sum of the estimated subband signals. In this study, the streamflow data was decomposed into subband signals using the CiSSA and VMD methods, and the obtained data was used as features for the prediction. Therefore, instead of applying machine learning techniques separately to predict each subband signal, a machine learning algorithm was run once in the estimation study using the extracted features.

RF, ANN, and SVR methods were performed for forecasting daily streamflow data using extracted features, and forecasting performances of CiSSA-RF, CiSSA-ANN, and CiSSA-SVR were compared with VMD-RF, VMD-ANN, and VMD-SVR. An ensemble of CiSSA-VMD models such as CiSSA-VMD-RF, CiSSA-VMD-ANN, and CiSSA-VMD-SVR and the Kruskal–Wallis test was used to determine if the predicted series satisfied the statistical significance criteria in each model. The flow of the proposed study is shown in Fig. 1.

Materials and methods

Study area and data

The daily streamflow (mm) data was measured over 61 years, from 1951 to 2012, resulting in a total of 22,646 data points for this analysis. The data was collected from the Upsalquitch river in the Upsalquitch basin, which is located at 47° 49' 56" N, 66° 53' 13" W (latitude: 47.83247, longitude: -66.8871) in New Brunswick, Canada. The data is obtained from the CANOPEX database for the site with a drainage area of 2305.77 km² (<http://canopex.etsmtl.net/>; Arsenault et al. 2016).

The daily streamflow data was split into a 70:30 ratio, where 70% of data was employed for training the model, and the remaining 30% data was tested to show the effectiveness of the model. As a result, 15,852 streamflow data points were used in the training process, and the remaining 6794 data points were used at the testing stage. Figure 2 shows the normalized training and testing data.

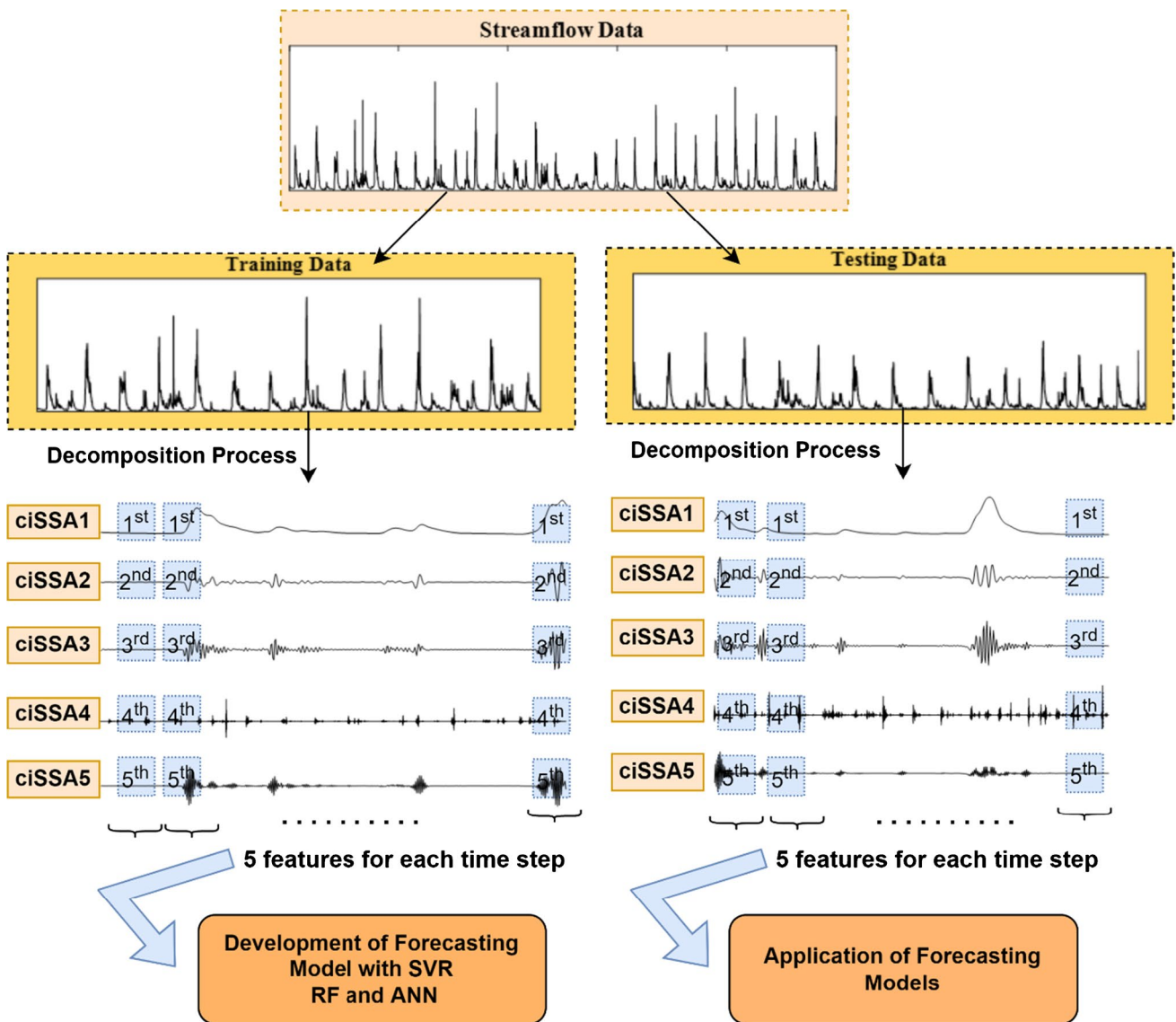
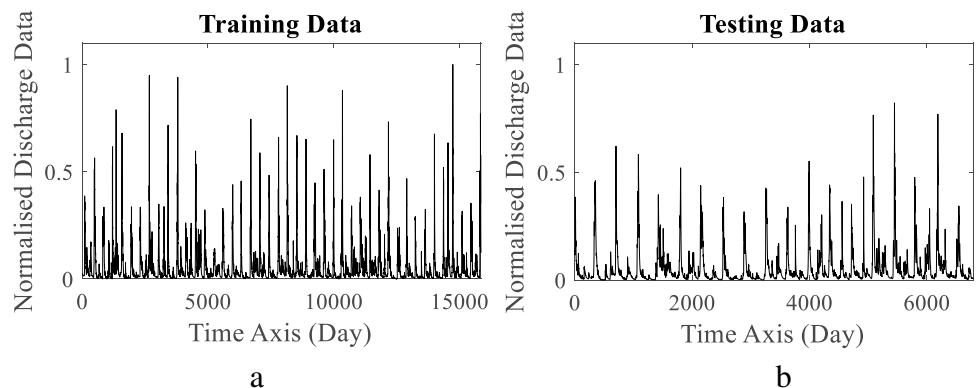


Fig. 1 The proposed forecasting model for daily streamflow data

By using preprocessing techniques on the entire streamflow data, some information from the validation phase may

be sent into the training process of data-driven models, resulting in “hindcasting experiments” (HE). Thus, in a real-world

Fig. 2 Daily streamflow data measured on Upsalquitch river. **a** Training data, **b** testing data



application of streamflow forecasting known as “forecasting experiments,” the outcomes of streamflow prediction at a certain time are calculated using some future knowledge that would not be accessible at that exact moment (Zhang et al. 2015).

To create models, in some studies, first full-time series were decomposed into sub-signals and then divided into training and testing samples. This is an incorrect or unrealistic strategy because some data from the testing step is mostly missing which necessitates the separation of the streamflow data into training and testing sections. After separation, the CiSSA and VMD techniques were applied to the normalized training data for the developing of proposed model.

Therefore, the streamflow data in this study was initially separated into two parts: training and testing data.

Circulant singular spectrum analysis

Singular spectrum analysis (SSA) is a nonparametric signal extraction approach based on subspace algorithms. The basic goal of SSA is to extract the time series’ underlying signals, such as trend, cycle, seasonal, and irregular components. The main disadvantage of the SSA approach is that it requires identifying the frequency of the decomposed components after they have been extracted.

To eliminate this disadvantage, the new circulant singular spectrum analysis was proposed by Bógalo in 2021 (Bógalo et al. 2021). Circulant singular spectrum analysis is a nonparametric signal decomposition approach that may rebuild a time series as the sum of orthogonal components of known frequencies (Bógalo et al. 2021). The main advantage of the CiSSA method is that users can group the extracted components according to their needs, as the components are defined precisely by frequency.

The circulant SSA method consists of embedding, decomposition, grouping, and reconstruction steps that are shown in Fig. 3.

Building the trajectory matrix stage:

A trajectory matrix is defined as shown in Eq. 1.

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ x_2 & x_3 & \cdots & x_{N+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_L & x_{L+1} & \cdots & x_T \end{bmatrix} \quad (1)$$

where T is the length of time series x_t , L is the window length that is $L < T/2$, and X is the trajectory matrix.

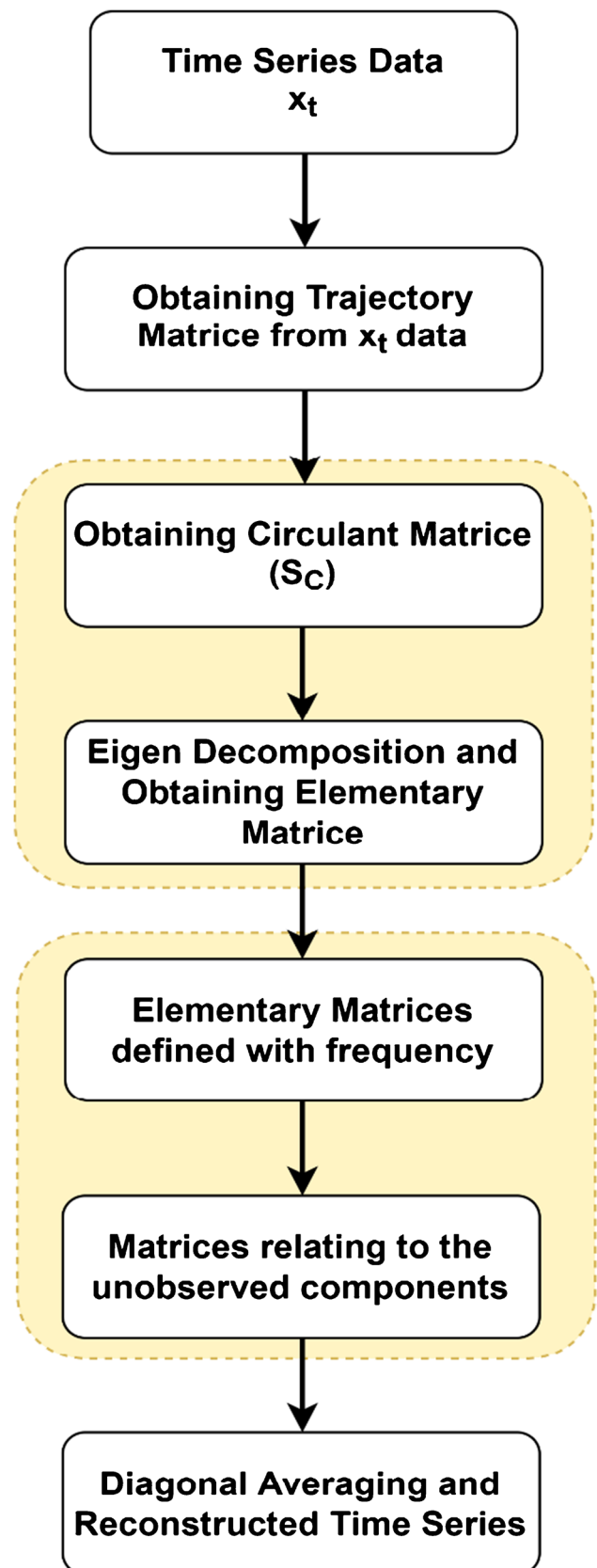


Fig. 3 Flow chart of the CiSSA algorithm

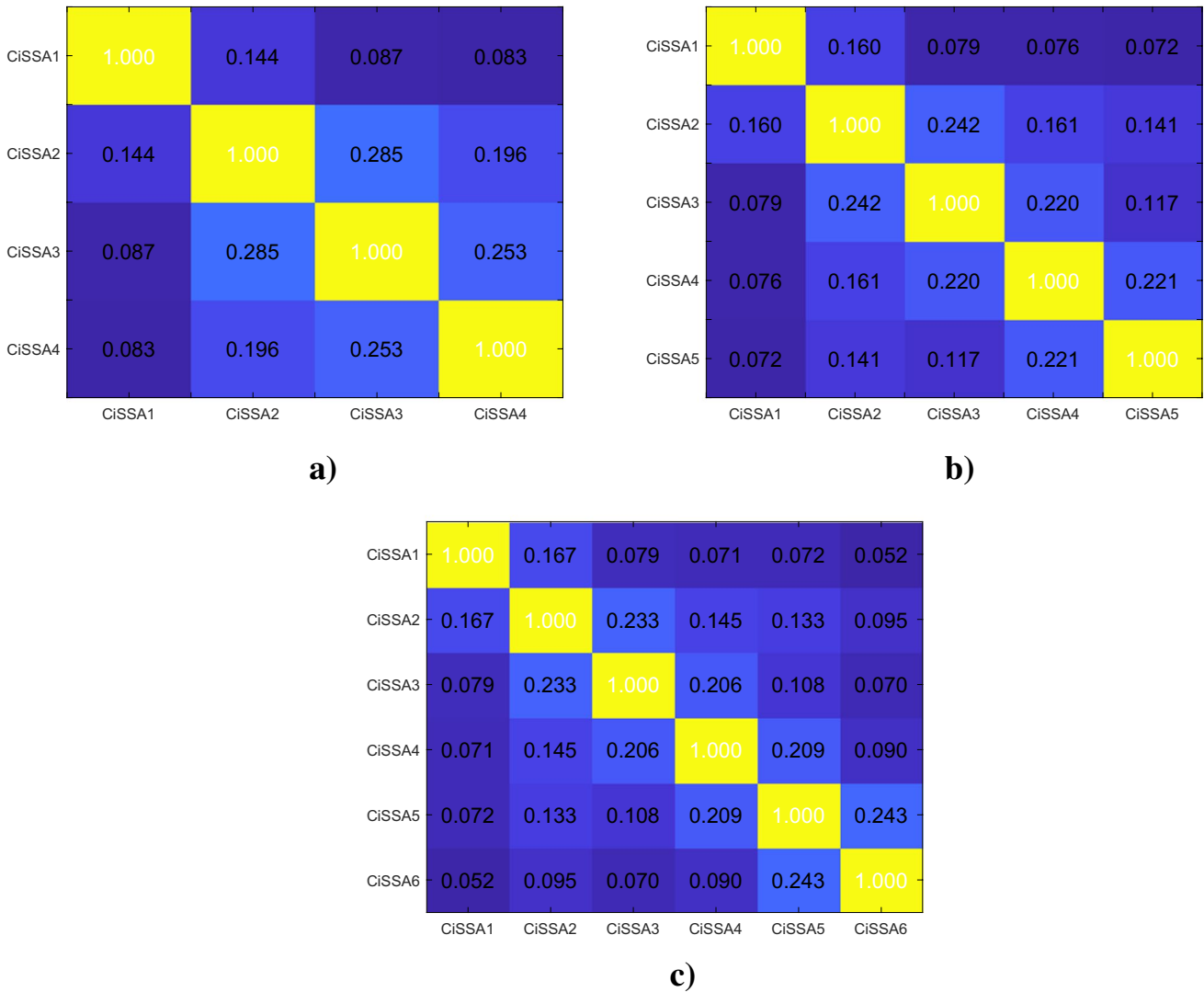


Fig. 4 Pearson correlation coefficients of the sub-signals of **a** four decomposition levels of CiSSA, **b** five decomposition levels of CiSSA, **c** six decomposition levels of CiSSA

Decomposition stage:

To extract the unobserved components, the trajectory matrix is projected to the space spanned by a set of eigenvectors. For this, the circulant matrix S_C is created using the sample second moments and elements of S_C which are defined according to the following equations;

$$s_m = \frac{1}{T - m} \sum_{t=1}^{T-m} x_t x_{t-m}, m = 0, 1, \dots, L - 1 \tag{2}$$

First row components of S_C are given in the Eq. 3 and the eigenvectors of S_C are given by Eq. 4.

$$\alpha_m = \frac{L - m}{L} s_m + \frac{m}{L} s_{L-m}, m = 0, 1, \dots, L - 1 \tag{3}$$

$$\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_L) = U^* S_C U \tag{4}$$

where U shows the Fourier matrix. Unnikrishnan and Jothiprakash (2018) can be referred for detailed explanations. The eigenvalues λ_k are related directly to the spectral density of the x_t time series for the frequency $w_k = \frac{k-1}{L} k = 1, \dots, L$ as shown in Eq. 5

$$\lambda_k \cong f\left(\frac{k-1}{L}\right) = \sum_{m=-\infty}^{\infty} s_m \exp\left(i2\pi m \frac{k-1}{L}\right) \tag{5}$$

In this way, the X trajectory matrix can be decomposed with X_k elementary matrices and this is defined as Eq. 6.

$$X = \sum_{k=1}^L x_k = \sum_{k=1}^L U_k U_k^* X \tag{6}$$

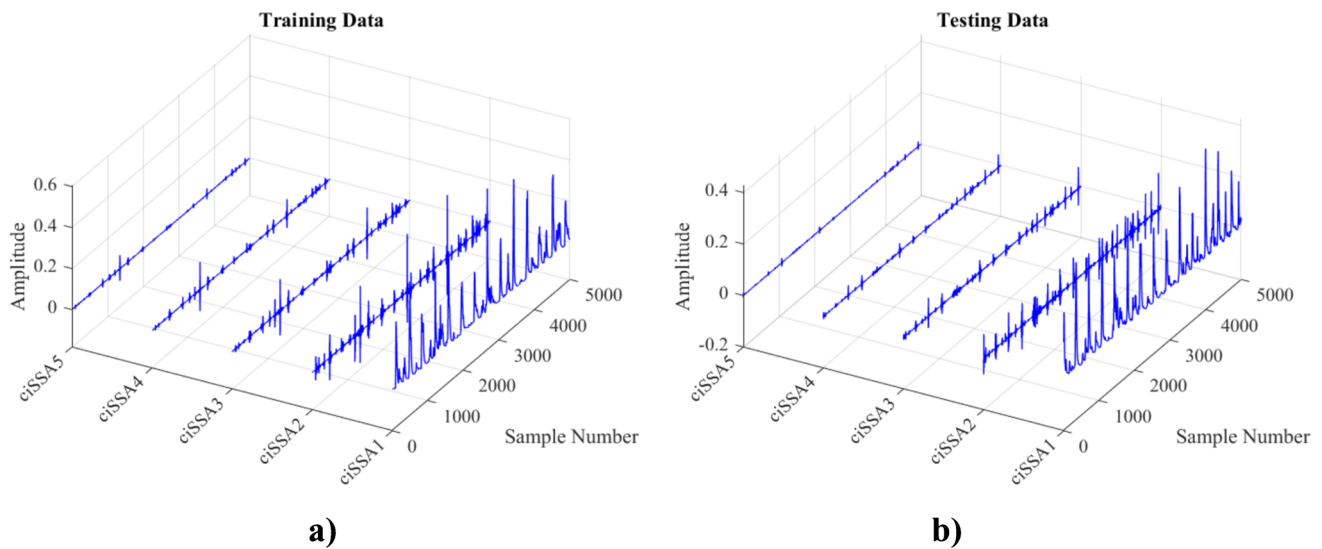


Fig. 5 CiSSA subband signals of **a** training and **b** testing data

Grouping stage

Because the spectral density function is symmetric, corresponding eigenvectors are complex conjugate. Therefore, the X_k elementary matrices and detailed in Unnikrishnan and Jothiprakash (2018).

Reconstruction stage

In this stage, time series can be obtained from transformation matrices using the same approach as SSA method (Bógalo et al. 2021).

Subband features obtained using the CiSSA method are named as $CiSSA_i$ in this study. Training and testing data were decomposed into five subbands using the CiSSA method.

The reason why five subbands were chosen is that the number of five decomposition levels is the most appropriate value in terms of computational cost and forecasting performance in different trials. The Pearson correlation coefficient of subband signals is shown in Fig. 4. According to Fig. 4, five-level subband signals of CiSSA are more uncorrelated than the four or six-level decomposition of subband signals. For each t time (the forecast day), 5 features were used which were obtained from each subband signals. The CiSSA subband signals of training and testing data are shown in Fig. 5.

The CiSSA toolbox written in MATLAB was used to decompose the streamflow data using CiSSA method (Juan Bógalo Román 2021).

Variational mode decomposition

The VMD method is used as an alternative to the EMD method. The purpose of the VMD method is to decompose an input signal $x(t)$, a real value, into a certain number of subband signals (modes, u_k). In the VMD method, each mode (u_k) is clustered around a central frequency (w_k). Each of the K modes u_k ($k = 1, 2, 3, \dots, K$) of a signal $x(t)$ has a central frequency and a limited bandwidth. To obtain the unilateral spectrum, the Hilbert transform is used to get the analytical signal of each mode $((\delta(t) + (j/\pi t)) * u_k(t))$. Each mode's analysis signal is multiplied by an exponential term to adjust its approximate center frequency $((\delta(t) + (j/\pi t)) * u_k(t)e^{-jw_k t})$. Finally, the demodulation signal's gradient squared L2 bound norm is measured, and the signal bandwidth of each mode is estimated (Dragomiretskiy and Zosso 2013). Constrained variational problem is constructed as Eqs. 7 and 8:

$$\min_{\{u_k\}, \{w_k\}} \sum_k \|\partial_t [(\delta(t) + (j/\pi t)) * u_k(t)] e^{-jw_k t}\|^2 \tag{7}$$

$$\sum_k u_k = x(t) \tag{8}$$

where $\{u_k\} = \{u_1, u_2, \dots, u_k\}$ is the set of all modes, $\{w_k\} = \{w_1, w_2, \dots, w_k\}$ is the central frequency set of all modes. Detailed information about the VMD can be found in ref. (Dragomiretskiy and Zosso 2013). Subband features

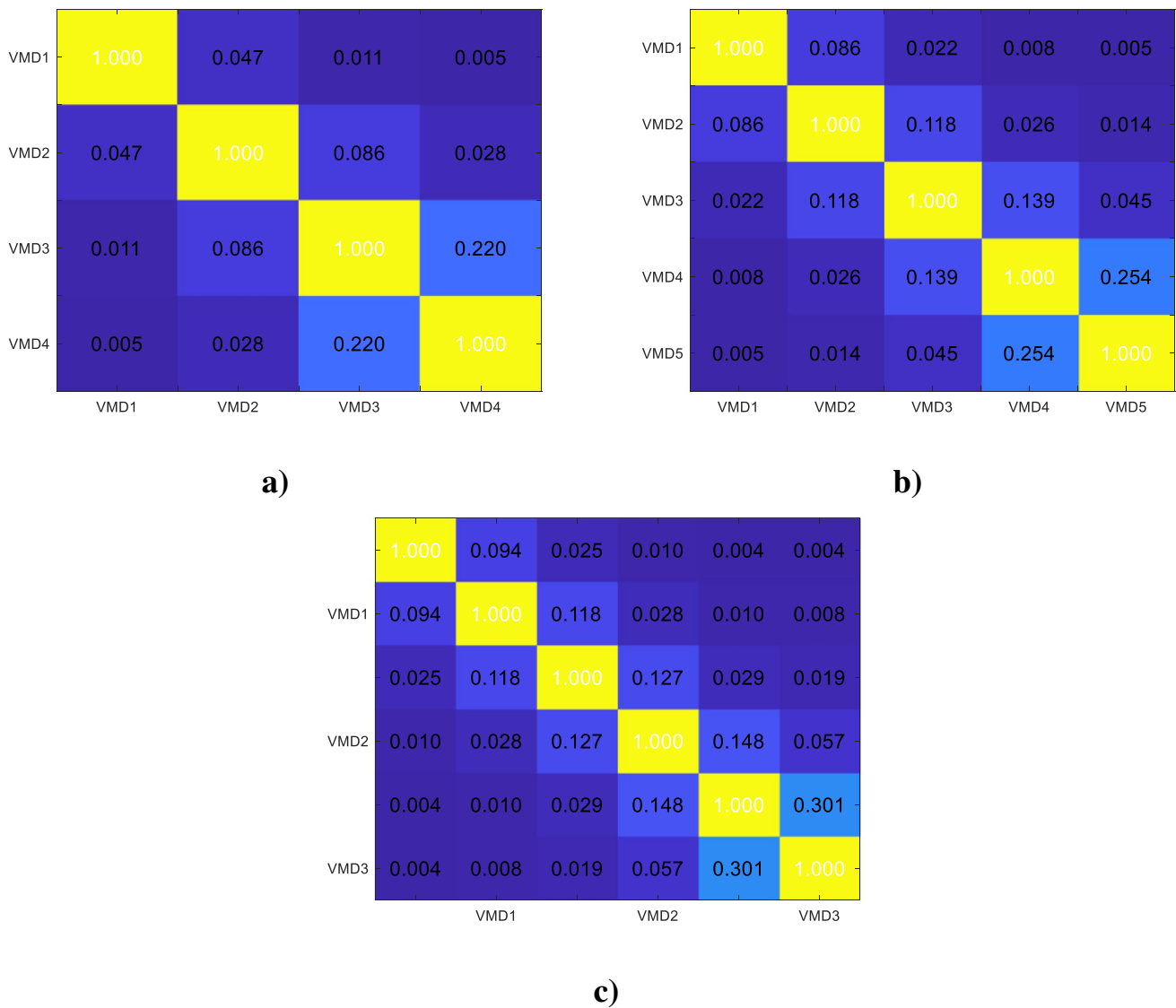


Fig. 6 Pearson correlation coefficients of the sub-signals of **a** four decomposition levels of VMD; **b** five decomposition levels of VMD; **c** six decomposition levels of VMD

obtained using VMD method are named as VMD_i in this study. Training and testing data were decomposed into five subbands using VMD method as in the CiSSA method. The Pearson correlation coefficient of VMD subband signals are shown in Fig. 6. According to Fig. 6, increasing decomposition level does not increase the uncorrelatedness. The VMD subband signals of training and testing data are shown in Fig. 7.

Support vector regression

Support vector regression (SVR) is a type of supervised learning method that is commonly used in the

classification problems. It was established by Vladimir N. Vapnik and Alexey Y. Chervonenkis (Cortes and Vapnik 1995) in 1963 in which a line is drawn to separate the points placed on a plane. While the basic purpose of this line is to connect the points of both groups at the greatest possible distance, it is significantly useful for datasets that are complex but small to medium in size.

SVR is similar to linear regression in which the equation of the line is $\vec{w} \cdot \vec{x} + b = y$. This straight line is known as a hyperplane in SVR. The nearest data points on either side of the hyperplane are called support vectors which are used to map the boundary line. Detailed information about SVR can be found in the ref. (Cortes and Vapnik 1995). In this

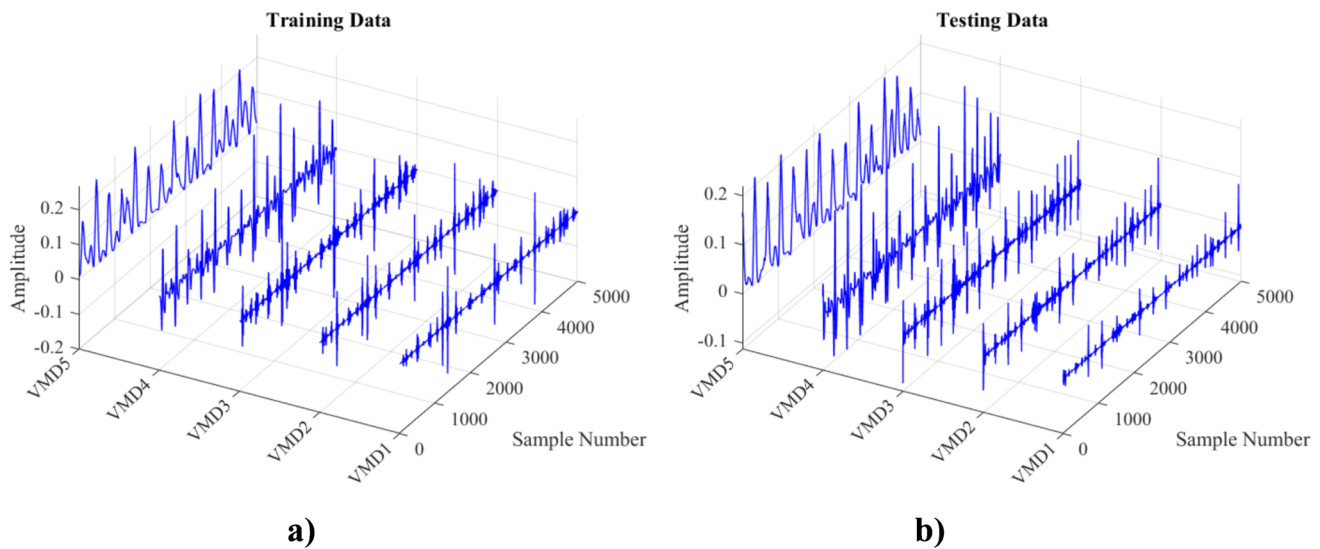


Fig. 7 VMD subband signals of **a** training and **b** testing data

Table 1 Random forest parameters used in this study

Iteration number	300
Folds number for back fitting	0 (no back fitting)
Bag size, as a percentage of the training set size	100
Used or not seed for random number generator	Yes
for batch prediction, the desired batch size	100

study, Smola and Schölkopf's sequential minimal optimization (SMO) algorithm was used for SVR parameter optimization of Platt (1998).

Random forest algorithm

The random forest algorithm is a supervised classification algorithm and can be used to solve regression problems. This algorithm aims to increase the classification performance by generating more than one decision tree during the classification process. The algorithm randomly creates a forest. There is a direct relationship between the number of trees in the algorithm and the result. As the number of trees increases, a precise result is obtained. The main steps of the random forest algorithm are given as follows (Breiman 2001).

Step 1: From the given dataset, the algorithm selects random samples.

Step 2: For each chosen sample, the algorithm will generate a decision tree. Then, for each decision tree it will get a prediction result.

Step 3: For each forecasted result, voting will be performed. It will use mode to solve a classification problem and mean to solve a regression problem.

Step 4: The algorithm will choose the prediction with the most votes as the final prediction.

During construction of the model, random forest parameters listed in the Table 1

Artificial neural networks

Artificial neural networks parallel and distributed information processing structures, inspired by the human brain, are composed of processing elements that are connected to each other through weighted connections and each having its own memory. In other words, they are computer programs that mimic the biological neural networks.

Just as biological neural networks have nerve cells, artificial neural networks also have artificial neural cells. Artificial nerve cells are also called process elements.

The inputs (x_1, x_2, \dots, x_n) transmit the information received from the environment to the nerve. Inputs can be applied to the neural network from previous nerves or from the outside. Weights (w_1, w_2, \dots, w_i) determine the effect of inputs to ANN on the nerve. Each input has its weight. The large value of weight means that it is strongly connected to the artificial nerve of that input or that it is important, and a small one means that it is weakly connected or it is not important. The result of the addition function is applied to the activation function. As is the case for the summation function, different formulas are used to calculate the output in the activation function. Artificial nerve cells come together to form the ANN. Nerve cells do not come together

Table 2 Input features applied the models in this study

	Single model inputs	Single output	VMD ensemble model inputs	VMD ensemble model output
N-day ahead forecast				
One-day ahead forecast	Q(t-1)	Q(t)	Q(t-1), VMD1(t-1), VMD2(t-1), VMD3(t-1), VMD4(t-1), VMD5(t-1)	Q(t)
Two-day ahead forecast	Q(t-2)	Q(t)	Q(t-2), VMD1(t-2), VMD2(t-2), VMD3(t-2), VMD4(t-2), VMD5(t-2)	Q(t)
Three-day ahead forecast	Q(t-3)	Q(t)	Q(t-2), VMD1(t-3), VMD2(t-3), VMD3(t-3), VMD4(t-3), VMD5(t-3)	Q(t)
Four-day ahead forecast	Q(t-4)	Q(t)	Q(t-4), VMD1(t-4), VMD2(t-4), VMD3(t-4), VMD4(t-4), VMD5(t-4)	Q(t)
Five-day ahead forecast	Q(t-5)	Q(t)	Q(t-5), VMD1(t-5), VMD2(t-5), VMD3(t-5), VMD4(t-5), VMD5(t-5)	Q(t)
N-day ahead forecast	CiSSA Ensemble ModelInputs	CiSSA ensemble	VMD-CiSSA ensemble model inputs	VMD-CiSSA ensemble model output
One-day ahead forecast	Q(t-1), CiSSA1(t-1), CiSSA2(t-1), CiSSA3(t-1), CiSSA4(t-1), CiSSA5(t-1)	Q(t)	Q(t-1), VMD1-CiSSA1(t-1), VMD2-CiSSA2(t-1), VMD3-CiSSA3(t-1), VMD4-CiSSA4(t-1), VMD5-CiSSA5(t-1)	Q(t)
Two-day ahead forecast	Q(t-2), CiSSA1(t-2), CiSSA2(t-2), CiSSA3(t-2), CiSSA4(t-2), CiSSA5(t-2)	Q(t)	Q(t-2), VMD1-CiSSA1(t-2), VMD2-CiSSA2(t-2), VMD3-CiSSA3(t-2), VMD4-CiSSA4(t-2), VMD5-CiSSA5(t-2)	Q(t)
Three-day ahead forecast	Q(t-3), CiSSA1(t-3), CiSSA2(t-3), CiSSA3(t-3), CiSSA4(t-3), CiSSA5(t-3)	Q(t)	Q(t-3), VMD1-CiSSA1(t-3), VMD2-CiSSA2(t-3), VMD3-CiSSA3(t-3), VMD4-CiSSA4(t-3), VMD5-CiSSA5(t-3)	Q(t)
Four-day ahead forecast	Q(t-4), CiSSA1(t-4), CiSSA2(t-4), CiSSA3(t-4), CiSSA4(t-4), CiSSA5(t-4)	Q(t)	Q(t-4), VMD1-CiSSA1(t-4), VMD2-CiSSA2(t-4), VMD3-CiSSA3(t-4), VMD4-CiSSA4(t-4), VMD5-CiSSA5(t-4)	Q(t)
Five-day ahead forecast	Q(t-5), CiSSA1(t-5), CiSSA2(t-5), CiSSA3(t-5), CiSSA4(t-5), CiSSA5(t-5)	Q(t)	Q(t-5), VMD1-CiSSA1(t-5), VMD2-CiSSA2(t-5), VMD3-CiSSA3(t-5), VMD4-CiSSA4(t-5), VMD5-CiSSA5(t-5)	Q(t)

randomly. Generally, cells come together parallel to form a network in three layers which are named as the Input layer, hidden layer, and output layer. Detailed information about ANN can be found in ref Livingstone (2009). In the ANN model, a multi-layer perceptron (MLP) is used, and the network structure consists of an input, an output, and a hidden layer. To create the network structure with the highest predictive performance, the number of interlayer neurons was tried by trial and error. As a result of the forecasting, the ANN model was performed with the number of neurons with the highest performance and the lowest error. In this structure, the number of iterations was determined as 100 and the sigmoidal function was used as the activation function. Levenberg Marquardt backpropagation algorithm is used as a learning algorithm in the MLP structure.

Performance evaluation

To evaluate the performance of the proposed model, MAE, RMSE, R, NSE, and WI parameters were calculated (Nou-rani et al. 2019; Ali and Abustan 2014). The MAE is a parameter for determining how close a prediction is to the actual result which is defined as Eq. 9.

$$MAE = \sum_{t=1}^N |y_t - y| / N \tag{9}$$

In this equation, y_t is the forecasted data, y is the observed data, and N represents the number of data.

The sample standard deviation of the differences between forecasted (y_t) and observed values (y) is represented by RMSE, where N is the number of observations. RMSE is defined as Eq. 10.

$$RMSE = \sqrt{\sum_{t=1}^n (y_t - y) / N} \tag{10}$$

R is the correlation coefficient, which is used to determine the strength and direction of a linear relationship between y_t and y . The value of R is always between + 1 and - 1, and the closer the value of R is to + 1, higher is the optimality of the model. The value of R is determined as shown in Eq. 11.

$$R = \frac{\sum_{t=1}^N (y_t - \bar{y}_t)(y - \bar{y})}{\left\{ \sum_{t=1}^N (y_t - \bar{y}_t)^2 \sum_{t=1}^N (y - \bar{y})^2 \right\}^{1/2}} \tag{11}$$

where \bar{y}_t is the mean of forecasted data, and \bar{y} is the mean of observed data.

The NSE parameter ranges from -∞ to 1.0, with NSE = 1 being the best value. Unsatisfactory performance is indicated by the NSE ≤ 0, while the acceptable range is 0 < NSE < 1. The NSE parameter can be written as shown in Eq. 12.

Table 3 Forecasting performance of SVR, RF, and ANN model

SVR model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9737	0.9387	0.9385	0.0062	0.0191
Two-day ahead forecast	0.9305	0.8212	0.8940	0.0107	0.0307
Three-day ahead forecast	0.8898	0.6964	0.8613	0.014	0.0382
Four-day ahead forecast	0.85	0.5521	0.8345	0.0167	0.0442
Five-day ahead forecast	0.8135	0.4087	0.8128	0.0189	0.0488
RF model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.968	0.9337	0.9219	0.0079	0.0206
Two-day ahead forecast	0.9191	0.8289	0.8689	0.0132	0.0325
Three-day ahead forecast	0.8751	0.7238	0.8284	0.0173	0.0399
Four-day ahead forecast	0.8278	0.6054	0.7944	0.0208	0.0464
Five-day ahead forecast	0.7878	0.4911	0.7662	0.0236	0.051
ANN model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9738	0.9447	0.9321	0.007	0.0189
Two-day ahead forecast	0.9327	0.8068	0.8739	0.0198	0.0337
Three-day ahead forecast	0.894	0.7803	0.8405	0.0161	0.0375
Four-day ahead forecast	0.8564	0.6194	0.8250	0.0177	0.0428
Five-day ahead forecast	0.8213	0.4739	0.7784	0.0224	0.0485

Table 4 Forecasting performance of VMD-SVR, VMD-RF, and VMD-ANN model

VMD-SVR model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.979	0.9502	0.9415	0.0062	0.0172
Two-day ahead forecast	0.9522	0.8800	0.9047	0.0096	0.0256
Three-day ahead forecast	0.9247	0.8050	0.8770	0.0124	0.0317
Four-day ahead forecast	0.9033	0.7367	0.8557	0.0146	0.0358
Five-day ahead Fforecast	0.8701	0.6372	0.8329	0.0169	0.0409
VMD-RF model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9775	0.9538	0.9355	0.0065	0.0173
Two-day ahead forecast	0.9526	0.8984	0.9022	0.0101	0.0246
Three-day ahead forecast	0.9295	0.8446	0.8789	0.0122	0.0303
Four-day ahead forecast	0.9161	0.8089	0.8636	0.0138	0.0329
Five-day ahead forecast	0.8834	0.7283	0.8392	0.0162	0.0385
VMD-ANN model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9785	0.9559	0.9306	0.007	0.0169
Two-day ahead forecast	0.954	0.8695	0.9029	0.0199	0.0278
Three-day ahead forecast	0.935	0.8231	0.8816	0.0119	0.0299
Four-day ahead forecast	0.9086	0.7814	0.8499	0.0151	0.0344
Five-day ahead forecast	0.8711	0.6284	0.7650	0.0212	0.0422

$$NSE = 1 - \frac{\sum_{i=1}^n [y_i - y_{t,i}]^2}{\sum_{i=1}^n [y_i - \bar{y}_i]^2} \tag{12}$$

where n is the number of samples.

The WI index has a range of -1 to 1 and is defined as given in Eq. 13 and Eq. 14 (Willmott et al. 2012).

Table 5 Forecasting performance of CiSSA-SVR, CiSSA-RF, and CiSSA-ANN model

CiSSA-SVR model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9806	0.9534	0.9434	0.0057	0.0166
Two-day ahead forecast	0.9565	0.8827	0.9080	0.0093	0.0249
Three-day ahead forecast	0.9363	0.8233	0.8831	0.0118	0.0297
Four-day ahead forecast	0.9213	0.7856	0.8630	0.0138	0.0325
Five-day ahead forecast	0.891	0.6801	0.8410	0.016	0.0381
CiSSA-RF model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9842	0.9673	0.9449	0.0056	0.0145
Two-day ahead forecast	0.968	0.9319	0.9189	0.0083	0.0206
Three-day ahead forecast	0.9547	0.9022	0.9024	0.0098	0.0244
Four-day ahead forecast	0.9479	0.8882	0.8948	0.0106	0.0261
Five-day ahead forecast	0.9225	0.8312	0.8664	0.0135	0.0317
CiSSA-ANN model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.984	0.9659	0.9420	0.0059	0.0148
Two-day ahead forecast	0.9584	0.9008	0.8541	0.0147	0.0245
Three-day ahead forecast	0.9386	0.8669	0.8727	0.0129	0.0284
Four-day ahead forecast	0.9235	0.7936	0.8593	0.0142	0.0321
Five-day ahead forecast	0.8933	0.6344	0.7469	0.0255	0.0419

Table 6 Forecasting Performance of VMD-CiSSA-SVR, VMD-CiSSA-RF, and VMD-CiSSA-ANN model

VMD-CiSSA-SVR model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9809	0.9544	0.9435	0.0057	0.0164
Two-day ahead forecast	0.9562	0.8863	0.9076	0.0093	0.0247
Three-day ahead forecast	0.9328	0.8187	0.8820	0.0119	0.0303
Four-day ahead forecast	0.9213	0.7860	0.8628	0.0138	0.0325
Five-day ahead forecast	0.8798	0.6237	0.8408	0.0169	0.0425
VMD-CiSSA-RF model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9833	0.9657	0.9439	0.0057	0.0149
Two-day ahead forecast	0.9662	0.9278	0.9169	0.0084	0.0212
Three-day ahead forecast	0.9476	0.8862	0.8977	0.0103	0.0262
Four-day ahead forecast	0.9475	0.8879	0.8940	0.0107	0.0262
Five-day ahead forecast	0.9196	0.8239	0.8667	0.0134	0.0323
VMD-CiSSA-ANN model	R	NSE	WI	MAE	RMSE
One-day ahead forecast	0.9837	0.9645	0.9358	0.0065	0.0149
Two-day ahead forecast	0.9663	0.9203	0.8844	0.0117	0.0217
Three-day ahead forecast	0.946	0.8567	0.8820	0.0119	0.0272
Four-day ahead forecast	0.9405	0.8070	0.8757	0.0125	0.0297
Five-day ahead forecast	0.8995	0.6936	0.7810	0.0221	0.0393

$$WI = 1 - \frac{\sum_{i=1}^n |y_{t,i} - y_i|}{c \times \sum_{i=1}^n |y_i - \bar{y}_i|} \tag{13}$$

Table 7 *p*-values of Kruskal–Wallis test at 95% significance level for 1–5-day ahead forecast for all folds

Models	<i>p</i> value	*H ₀	<i>p</i> value	*H ₀	<i>p</i> value	*H ₀	<i>p</i> value	*H ₀	<i>p</i> value	*H ₀
	One-day ahead		Two-day ahead		Three-day ahead		Four-day ahead		Five -day ahead	
SVR	0.28728	<i>Reject</i>	0.6949	<i>Reject</i>	05,497	<i>Reject</i>	0.5807	<i>Reject</i>	0.4142	<i>Reject</i>
RF	0.2958	<i>Reject</i>	0.6007	<i>Reject</i>	0.1881	<i>Reject</i>	0.2853	<i>Reject</i>	0.1141	<i>Reject</i>
ANN	0.0836	<i>Reject</i>	0.7484	<i>Reject</i>	0.5529	<i>Reject</i>	0.529	<i>Reject</i>	0.4065	<i>Reject</i>
VMD-SVR	0.0898	<i>Reject</i>	0.7289	<i>Reject</i>	0.3270	<i>Reject</i>	0.1901	<i>Reject</i>	0.2940	<i>Reject</i>
VMD-RF	0.338	<i>Reject</i>	0.7605	<i>Reject</i>	0.5807	<i>Reject</i>	0.3474	<i>Reject</i>	0.2209	<i>Reject</i>
VMD-ANN	0.2940	<i>Reject</i>	0.1737	<i>Reject</i>	0.3525	<i>Reject</i>	0.1901	<i>Reject</i>	0.2674	<i>Reject</i>
CiSSA-SVR	0.1552	<i>Reject</i>	0.3270	<i>Reject</i>	0.1264	<i>Reject</i>	0.1228	<i>Reject</i>	0.2384	<i>Reject</i>
CiSSA-RF	0.2555	<i>Reject</i>	0.2243	<i>Reject</i>	0.3270	<i>Reject</i>	0.1901	<i>Reject</i>	0.2396	<i>Reject</i>
CiSSA-ANN	0.3644	<i>Reject</i>	0.0791	<i>Reject</i>	0.1901	<i>Reject</i>	0.0658	<i>Reject</i>	0.2322	<i>Reject</i>
VMD-CiSSA-SVR	0.1812	<i>Reject</i>	0.1812	<i>Reject</i>	0.0777	<i>Reject</i>	0.0846	<i>Reject</i>	0.4405	<i>Reject</i>
VMD-CiSSA-RF	0.3338	<i>Reject</i>	0.2144	<i>Reject</i>	0.3525	<i>Reject</i>	0.3270	<i>Reject</i>	0.0186	<i>Reject</i>
VMD-CiSSA-ANN	0.5046	<i>Reject</i>	0.1507	<i>Reject</i>	0.1473	<i>Reject</i>	0.0968	<i>Reject</i>	0.4731	<i>Reject</i>

*H₀ represents a hypothesis based on the significant difference between mean predicted and observed values

$$\text{when } \sum_{i=1}^n |y_{t,i} - y_i| \leq c \times \sum_{i=1}^n |y_i - \bar{y}_i|, \text{ with } c = 2$$

$$WI = \frac{c \times \sum_{i=1}^n |y_i - \bar{y}_i|}{\sum_{i=1}^n |y_{t,i} - y_i|} - 1 \tag{14}$$

$$\text{when } \sum_{i=1}^n |y_{t,i} - y_i| > c \times \sum_{i=1}^n |y_i - \bar{y}_i|, \text{ with } c = 2$$

Results

In this study, 1–5-day ahead forecasting of the streamflow data has been performed using SVR, RF, ANN, CiSSA-SVR, CiSSA-RF, CiSSA-ANN, VMD-SVR, VMD-RF, VMD-ANN, and combination of CiSSA-VMD method. Table 2 contains the input features used for forecasting model.

In the first part of the study, the forecasting performance of SVR, RF, ANN methods and in the second part of the study, the forecasting performance of the hybrid models were analyzed without using data preprocessing methods. While the calculated parameters from SVR, RF, and ANN models are listed in Table 3, Table 4 contains the forecasting performance of VMD-SVR, VMD-RF, and VMD-ANN model. Forecasting performances of CiSSA-SVR, CiSSA-RF, and CiSSA-ANN model are given in Table 5. Moreover, the individual forecasting performances of CiSSA-VMD-SVR, CiSSA-VMD-RF, and CiSSA-VMD-ANN models are given in Table 6.

In this study, Kruskal–Wallis test was used to analyze whether the predicted and observed time series have similar distributions or not. Table 7 shows the Kruskal–Wallis test results for 1 to 5 days ahead. If there is no significant difference between the mean and variance of the predicted and observed time series, the H₀ hypothesis is rejected for which the *p*-values are above 0.05 (Başakın et al. 2021; Hussain and Khan 2020).

In addition, in terms of visual evaluation of the performance of the models, the Taylor diagrams are shown in Fig. 8. The forecasting and scatter plots of the models in which the best estimations are obtained for one to five day ahead estimations are shown in Fig. 9.

Discussion and conclusion

In literature, many studies are carried out for forecasting of streamflow data. Machine learning algorithms (such as ANN, SVM, M5 tree model, and RF) were compared to evaluate their forecasting performance using the specified parameters (Nourani et al. 2019; Hussain and Khan 2020; Zounemat-Kermani et al. 2021b). For the training and testing phases, each algorithm is used with different time-lagged streamflow data for modeling (Nourani et al. 2019; Hussain and Khan 2020; Zounemat-Kermani et al. 2021b; Alizadeh et al. 2021). In addition, in literature, wavelet, empirical mode decomposition, singular spectrum analysis, and the VMD subband signals have been used for forecasting studies (Wang et al. 2017; Nourani et al. 2019; Hussain and Khan 2020; Zounemat-Kermani et al. 2021b; Alizadeh et al. 2021). The VMD is a new decomposition approach that can decompose original streamflow data into

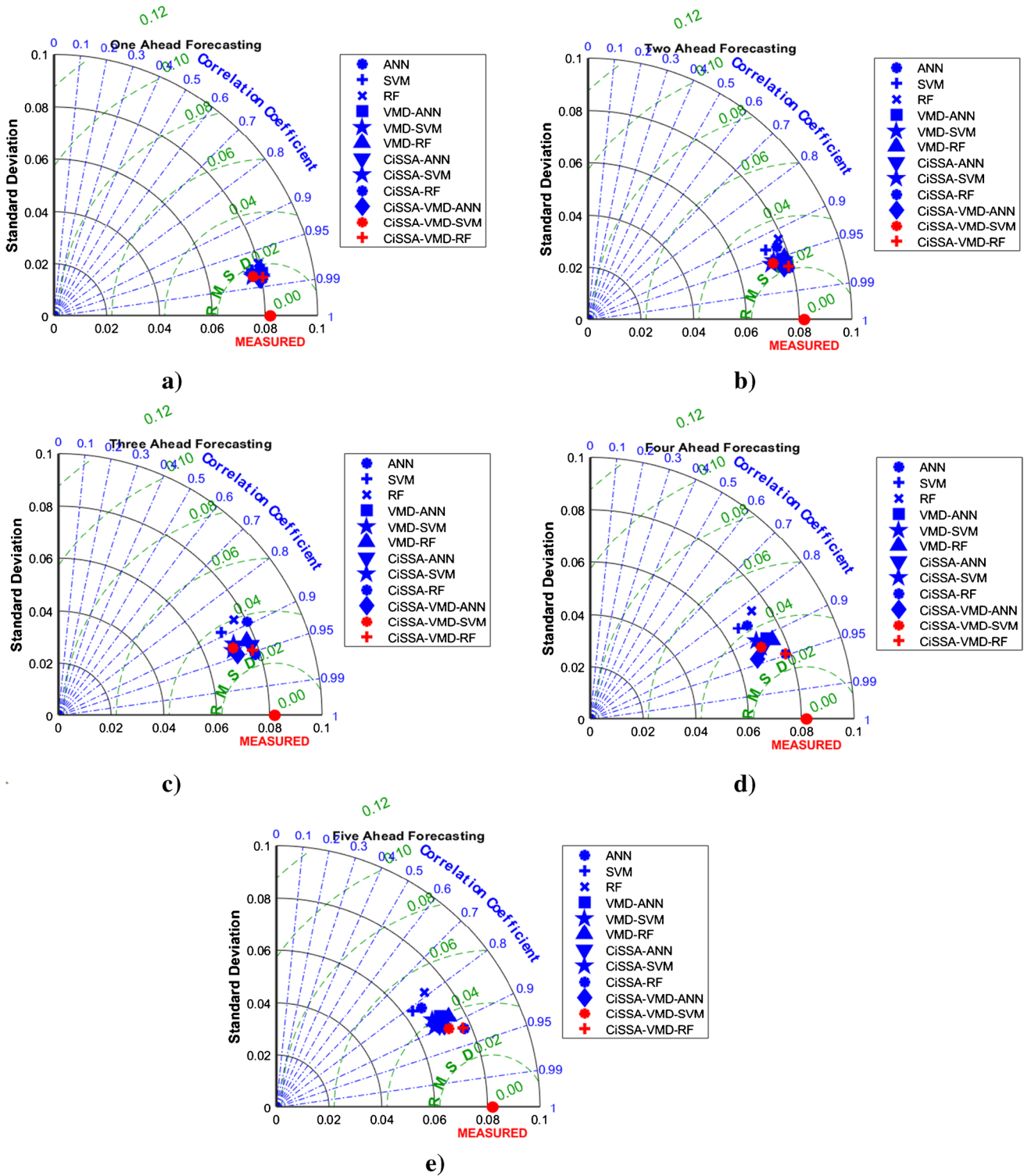


Fig. 8 Taylor diagram **a** for 1-day ahead, **b** for 2-day ahead, **c** for 3-day ahead, **d** for 4-day ahead, **e** for 5-day ahead forecasting to evaluate the forecasting performance

several subseries, removing noise and revealing the original data's key characteristics. The WT-based and EMD-based techniques are often used as decomposition methods. WT-based techniques, for example, offer good localization

properties in both the time and frequency domains, but their effectiveness is highly dependent on the wavelet function and decomposition level used. The EMD-based techniques are adaptive and have few hyperparameters,

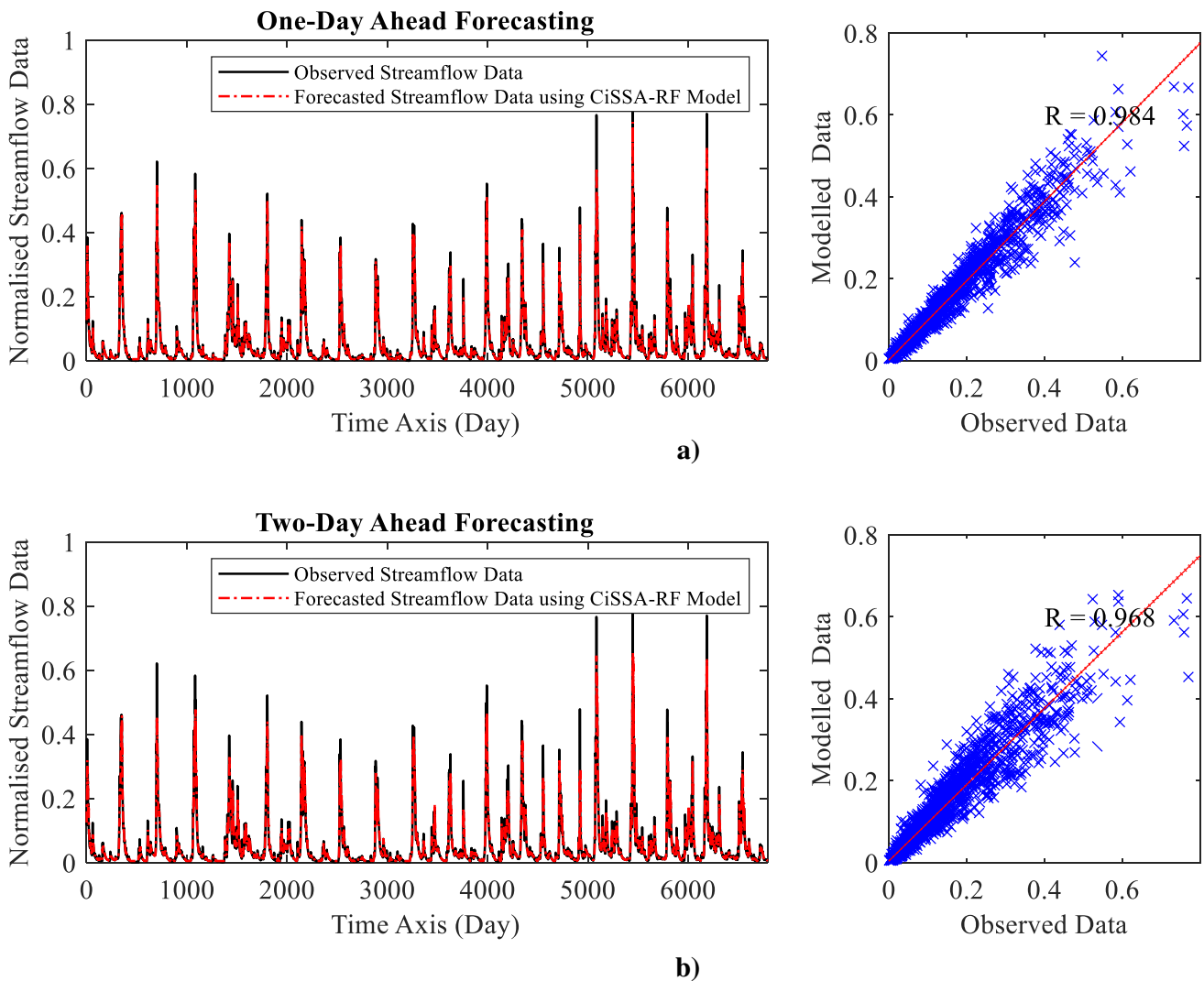


Fig. 9 Forecasting of daily streamflow data using **a** CiSSA-ANN model with scatter plot for 1-day ahead forecasting, **b** for 2-day ahead forecasting, **c** for 3-day ahead forecasting, **d** for 4-day ahead forecasting, **e** for 5-day ahead forecasting

but they lack mathematical theory and are susceptible to noise and sampling. The VMD, in comparison with the other decomposition methods, has a powerful mathematical theory, which can achieve precise signal separation, and has a high operational efficiency. The VMD-based models perform better than the WT-based and EMD-based models (Wang et al. 2018).

CiSSA was recently introduced by Bógalo et al. (Bógalo et al. 2021), which links the eigen decomposition to the frequency of the collected signals and proves the three variations' asymptotic equivalence (basic, toeplitz, and circulant SSA). The CiSSA technique eliminates the end effect that occurs in many preprocessing techniques such as EMD and DWT techniques.

In this study, 1–5-day ahead forecasting of daily river flow data was performed using CiSSA-SVR, CiSSA-RF,

CiSSA-ANN models with a novel approach. The proposed approach uses the CiSSA method for feature extraction and CiSSA features were applied to SVR, RF, and ANN models. Therefore, there is no need to use the models for forecasting of CiSSA subband signals separately and consequently, the computational complexity is reduced. For the comparison of proposed CiSSA-based models, VMD ensemble, CiSSA-VMD ensemble, and single models were constructed for forecasting of streamflow data. Therefore, a comparative analysis was drawn between the forecasting capabilities of SVR, RF, ANN, VMD-SVR, VMD-ANN, CiSSA-SVR, CiSSA-RF, CiSSA-ANN, and CiSSA-VMD-SVR, CiSSA-VMD-RF, CiSSA-VMD-ANN combined models.

Firstly, normalized daily river flow data is divided into training and testing data. The forecasting models are applied

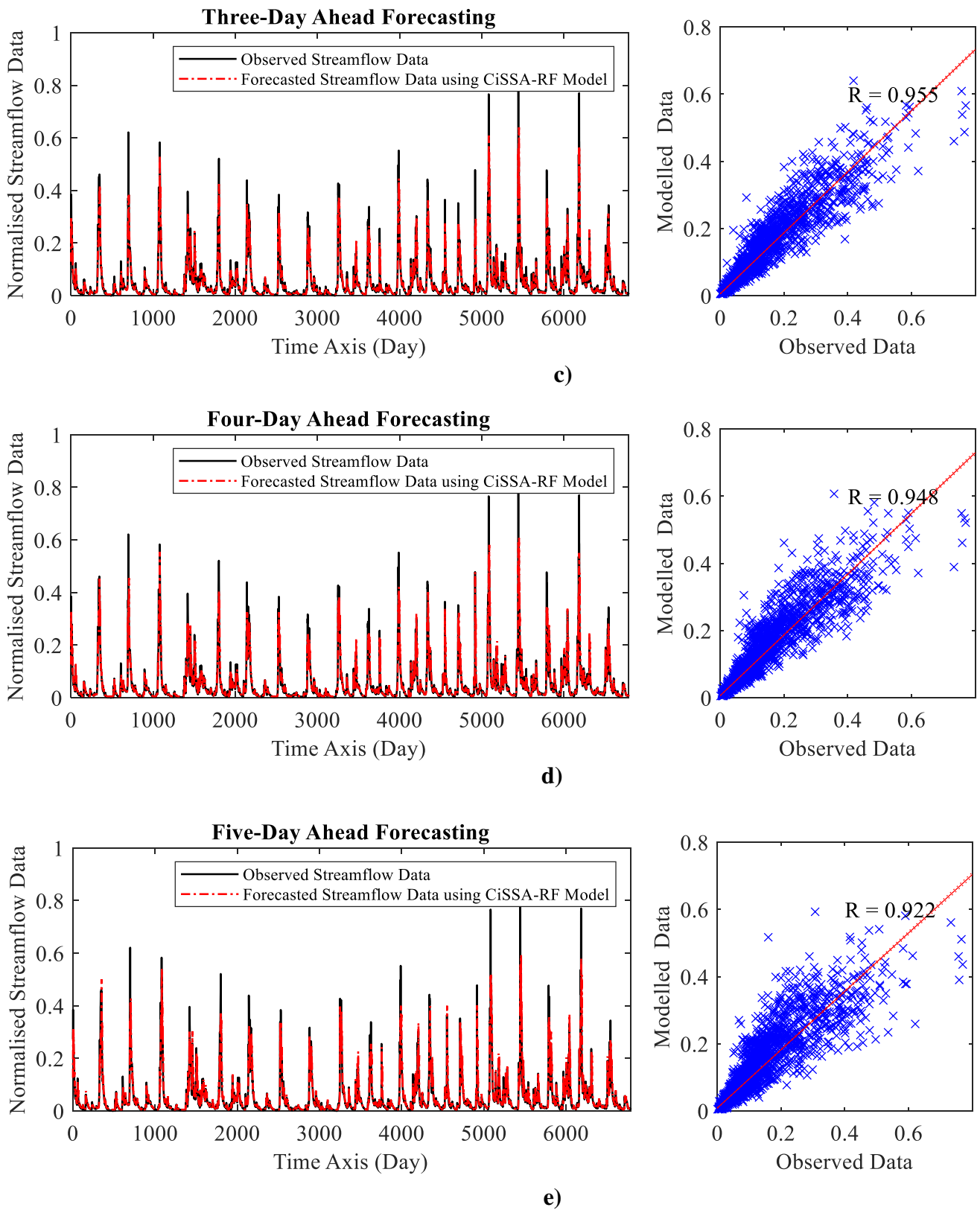


Fig. 9 (continued)

to training data to construct the model. After the development of the models, the forecasting performances of the models were tested using testing data. Forecasting performances are seen in Tables 3, 4, 5, and 6.

Performances obtained from the SVR, RF, ANN, VMD-SVR, VMD-RF, VMD-ANN, CiSSA-SVR, CiSSA-RF, CiSSA-ANN, CiSSA-VMD-SVR, CiSSA-VMD-RF, CiSSA-VMD-ANN models are close to each other for 1-day ahead forecasting. But it is seen that forecasting performance is improving significantly for far-day ahead forecasting obtained from CiSSA-SVR, CiSSA-RF, and CiSSA-ANN models.

As an example, when the 5-day ahead estimation results are analyzed, R, NSE, WI, MAE, and RMSE parameters were calculated as 0.8135, 0.4087, 0.8128, 0.0189, and 0.0488 respectively for SVR model, 0.7878, 0.4911, 0.7662, 0.0236, and 0.051 respectively for RF model and 0.8213, 0.4739, 0.7784, 0.0224, and 0.0485 respectively for ANN model. It can be seen from these numerical values, that the NSE parameter was significantly worse in 5-day ahead estimations according to single models.

If the effect of VMD features on forecast performance is analyzed, R, NSE, WI, MAE, and RMSE parameters were calculated as 0.87, 0.6372, 0.8329, 0.0169, and 0.0409 respectively for VMD-SVR model, 0.8834, 0.7283, 0.8392, 0.0162, and 0.0385 respectively for VMD-RF model and 0.8711, 0.6284, 0.7650, 0.0212, and 0.0422 respectively for VMD-ANN model for 5-day ahead forecasting. According to these results, it was seen that the performance of VMD-SVR, VMD-RF, and VMD-ANN models was better than the single SVR, RF, and ANN single models.

The forecasting study was additionally carried out using the features obtained with the CiSSA model and the R, NSE, WI, MAE, and RMSE values were calculated as 0.891, 0.6801, 0.8410, 0.016, and 0.0381 respectively for CiSSA-SVR model, 0.9225, 0.8312, 0.8664, 0.0135, and 0.0317 respectively for CiSSA-RF model and 0.8933, 0.6344, 0.7469, 0.0255, and 0.0419 respectively for CiSSA-ANN model for 5-day ahead forecasting. According to these results, it was seen that the performance of the models combined with CiSSA was better than the models combined with VMD.

In this study, the performance of the estimation study was also analyzed by using the features obtained with VMD and CiSSA together. However, it was observed that using the features obtained from the VMD and CiSSA algorithms together did not improve the prediction performance more than the CiSSA model.

Based on the obtained results, the prediction performance of the models established by using the features obtained from the CiSSA method has improved significantly. The performance of the CiSSA-RF model was comparatively

better than the prediction performance of all other models as shown in the Taylor diagrams.

It can be concluded from this study that the CiSSA features as input to forecasting models not only improve the forecasting performance but also can prove to be an effective tool for the forecasting studies in various fields.

Declarations

Conflict of interest The author declares no competing interests.

References

- Acar Yildirim H, Akcay C (2019) Time-cost optimization model proposal for construction projects with genetic algorithm and fuzzy logic approach. *Revista De La Construcción* 18(3):554–567
- Akcay C, Manisali E (2018) Fuzzy decision support model for the selection of contractor in construction works. *Revista de la Construcción. J Constr* 17(2):258–266
- Akyuncu V, Uysal M, Tanyildizi H, Sumer M (2018) Modeling the weight and length changes of the concrete exposed to sulfate using artificial neural network. *Revista de la Construcción. J Constr* 17(3):337–353
- Ali MH, Abustan I (2014) A new novel index for evaluating model performance. *Journal of Natural Resources and Development (JNRD)* 4:1–9
- Alizadeh A, Rajabi A, Shabanlou S, Yaghoubi B, Yosefvand F (2021) Modeling long-term rainfall-runoff time series through wavelet-weighted regularization extreme learning machine. *Earth Sci Inform*, 1–17
- Arsenault R, Bazile R, Ouellet Dallaire C, Brissette F (2016) CAN-OPEX: a Canadian hydrometeorological watershed database. *Hydrol Process* 30(15):2734–2736
- Başakın EE, Ekmekcioğlu Ö, Çıtakoğlu H, Özger M (2021) A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. *Neural Comput Appl*, 1–30.
- Bógalo J, Poncela P, Senra E (2021) Circulant singular spectrum analysis: a new automated procedure for signal extraction. *Signal Process* 179:107824
- Box GEP, Jenkins GM (1976) *Time series analysis forecasting and control*. 1st Holden Day Inc
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Dragomiretskiy K, Zosso D (2013) Variational mode decomposition. *IEEE Trans Signal Process* 62(3):531–544
- Hassani H (2007) Singular spectrum analysis: methodology and comparison. *J Data Sci* 5:239–257
- He X, Luo J, Zuo G, Xie J (2019) Daily runoff forecasting using a hybrid model based on variational mode decomposition and deep neural networks. *Water Resour Manage* 33(4):1571–1590
- Hussain D, Khan AA (2020) Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. *Earth Science Informatics*, 1–11.
- Juan Bógalo Román (2021). CiSSA: circulant SSA under Matlab, GitHub <https://github.com/jbogalo/CiSSA/releases/tag/2.1.2>. Accessed 20 May 2021

- Kisi O, Latifoğlu L, Latifoğlu F (2014) Investigation of empirical mode decomposition in forecasting of hydrological time series. *Water Resour Manage* 28(12):4045–4057
- Lahmiri S (2016) A variational mode decomposition approach for analysis and forecasting of economic and financial time series. *Expert Syst Appl* 55:268–273
- Latifoğlu L, Kişi Ö, Latifoğlu F (2015) Importance of hybrid models for forecasting of hydrological variable. *Neural Comput Appl* 26(7):1669–1680
- Liu Z, Zhou P, Chen G, Guo L (2014) Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. *J Hydrol* 519:2822–2831
- Livingstone DJ (2009) *Artificial neural networks* (Vol. 458). Springer
- Luo X, Yuan X, Zhu S, Xu Z, Meng L, Peng J (2019) A hybrid support vector regression framework for streamflow forecast. *J Hydrol* 568:184–193
- Mosaffa H, Sadeghi M, Mallakpour I, Jahromi MN, Pourghasemi HR (2022) Application of machine learning algorithms in hydrology. In *Computers in Earth and Environmental Sciences* (pp. 585–591). Elsevier
- Nourani V, Davanlou Tajbakhsh A, Molajou A, Gokcekus H (2019) Hybrid wavelet-M5 model tree for rainfall-runoff modeling. *J Hydrol Eng* 24(5).
- Oyebode O, Stretch D (2019) Neural network modeling of hydrological systems: a review of implementation techniques. *Nat Resour Model* 32(1):e12189
- Özger M (2009) Comparison of fuzzy inference systems for streamflow prediction. *Hydrol Sci J* 54(2):261–273
- Platt J (1998) Sequential minimal optimization: a fast algorithm for training support vector machines
- Rasouli K, William WH, Cannon AJ (2012) Daily streamflow forecasting by machine learning methods with weather and climate inputs. *J Hydrol* 414:284–293
- Sabzi HZ, King JP, Dilekli N, Shoghli B, Abudu S (2018) Developing an ANN based streamflow forecast model utilizing data-mining techniques to improve reservoir streamflow prediction accuracy: A case study. *Civ Eng J*, 4(5)
- Sivapalan M (2018) From engineering hydrology to Earth system science: milestones in the transformation of hydrologic science. *Hydrol Earth Syst Sci* 22(3):1665–1693
- Unnikrishnan P, Jothiprakash V (2018) Daily rainfall forecasting for one year in a single run using singular spectrum analysis. *J Hydrol* 561:609–621
- Wang D, Luo H, Grunder O, Lin Y (2017) Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. *Renew Energy* 113:1345–1358
- Wang X, Yu Q, Yang Y (2018) Short-term wind speed forecasting using variational mode decomposition
- Willmott CJ, Robeson SM, Matsuura K (2012) A refined index of model performance. *Int J Climatol* 32(13):2088–2094
- Yaseen ZM, El-Shafie A, Jaafar O, Afan HA, Sayl KN (2015) Artificial intelligence based models for stream-flow forecasting: 2000–2015. *J Hydrol* 530:829–844
- Yaseen, Z. M., Ebtahaj, I., Bonakdari, H., Deo, R. C., Mehr, A. D., Mohtar, W. H. M. W., ... & Singh, V. P. (2017). Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. *J Hydrol*, 554:263-276
- Zhang X, Peng Y, Zhang C, Wang B (2015) Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. *J Hydrol* 530:137–152
- Zhang J, Zhu Y, Zhang X, Ye M, Yang J (2018) Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. *J Hydrol* 561:918–929
- Zhu J, Wu P, Chen H, Liu J, Zhou L (2019) Carbon price forecasting with variational mode decomposition and optimal combined model. *Physica A* 519:140–158
- Zounemat-Kermani M, Batelaan O, Fadaee M, Hinkelmann R (2021a) Ensemble machine learning paradigms in hydrology: a review. *J Hydrol* 598:126266
- Zounemat-Kermani M, Mahdavi-Meymand A, Hinkelmann R (2021b) A comprehensive survey on conventional and modern neural networks: application to river flow forecasting. *Earth Sci Inform*, 1–19.
- Zuo G, Luo J, Wang N, Lian Y, He X (2020) Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. *J Hydrol* 585:124776