Water Resources Management (2026) 40:314
https://doi.org/10.1007/s11269-026-04717-w

Streamflow Simulation Based on Pre-Processing-Based Hybrid Model: SSA-AR Model

Baydaa Abdul Kareem[1,2] · Salah L. Zubaidi[2,3] · Hussein Al-Bugharbee[4] · Ali W. Alattabi[2] · Mawada Abdellatif[5]

Received: 12 February 2025 / Accepted: 26 April 2026 / Published online: 16 May 2026

Abstract

An accurate medium-term streamflow forecast is one of the significant functions for managing and planning water resources. Considering the characteristics of trend, periodicity, and stochasticity of streamflow into account. So, this research aims to develop a new strategy, including a singular spectrum analysis (SSA) technique and a linear autoregressive (AR) model to predict monthly Tigris River streamflow data with three scenarios. The first scenario applies the SSA to decompose the normalised and cleaned time series into different signals (i.e., trend, seasonal, stochastic, and noise), then reconstruct the signals without noise and use the AR model to simulate the new time series. The second scenario employs the AR model to forecast each signal of streamflow without noise separately. The simulated time series was obtained by summing each predicted signal. The third scenario uses the AR model to simulate the raw data. Based on several statistical tests, the comparative analysis reveals that the first and second scenarios were much more accurate than the third ones. The second scenario is the best, reaching RMSE = 0.1223 (m^3/s) and MAE = 0.0913 (m^3/s) in the testing phase. The novelty of this study lies in the comparative evaluation of three SSA-AR modelling scenarios and the finding that forecasting decomposed components individually leads to superior accuracy compared to conventional approaches. The results are of substantial significance to the Ministry of Water Resources in managing and planning freshwater resources amid growing water demand.

Keywords Autoregressive model · Singular spectrum analysis · Streamflow prediction · Tigris river

1 Introduction

One of the greatest challenges many countries face is precise streamflow forecasting, which could directly impact flood early signs and control, water supply and irrigation, recreation, water quality, and agriculture (Jozaghi et al. 2021; Wen Yan Xing et al. 2022; Yilmaz et al. 2022). However, as water demand continues to rise due to economic expansion, population rise, irrigation requirement, and industrial applications, traditional water resource manage-

Extended author information available on the last page of the article

ment would urge for more effective approaches or methodologies to forecast more precisely (Sharma et al. 2022; Wang et al. 2015).

Areas worldwide have not been affected equally regarding temperature and rainfall resulting from global warming. Therefore, it can be clearly stated that one of the world's most impacted countries by climate change is Iraq (Alawsi et al. 2022). Salman et al. (2017) mentioned that Iraq's average temperature might be increased 2–7 times faster than the global rates. Additionally, the future patterns and magnitudes of precipitation in Iraq are projected to be strongly impacted by various assessments and scenarios (Osman et al. 2017). The primary water resources in Iraq are the Tigris and Euphrates Rivers, and both rivers contribute to the formation of marshes. Iraq's marshes are the greatest wetland environment in the Middle East (Abdul Kareem et al. 2022). The previous regime in Iraq conducted a drainage act in 1991 to make these marshes arid. Numerous efforts have been made after 2003 to rehabilitate these wetland regions. However, the main rivers' fluctuating water levels have severely impacted the marshes' recovery and environmental restoration efforts (Guarasci 2015). Therefore, highlighting streamflow prediction is essential for sustainable development and will assist decision-makers in improving water systems.

Hybrid approaches are widely used in various research studies fields because they allow researchers to explore different techniques, evaluate their efficacy, and ultimately combine them into novel conceptual frameworks. Therefore, a hybrid methodology is able to overcome the drawbacks of each individual approach (Sulandari et al. 2020). The preprocessing-based hybrid model (PBH) is a model that depends on integrating singular forecasting approaches, and this approach has been successfully implemented in different fields of hydrology, such as drought (Pham et al. 2021), streamflow (Siddiqi et al. 2021), river water temperature (Mirzania et al. 2024), evaporation (Katipoğlu 2023), water level (Altunkaynak and Kartal 2019; Mohamadi et al. 2022), and water quality (Hanoon et al. 2022; Zhang et al. 2020).

Recently, considerable literature has grown around the theme of data pre-processing in enhancing the quality of time series. Therefore, several approaches of signal pre-treatment, for example, singular spectrum analysis (SSA) (Halit Apaydin 2021), variational mode decomposition (VMD) (Sun et al. 2022), ensemble empirical mode decomposition (EEMD) (Yuan et al. 2021), discrete wavelet transform (DWT) (KatİPoĞLu 2022), and wavelet transform (WT) (Zhao et al. 2021b), have been utilised to denoise or analyse components of the streamflow data (i.e., trend, seasonal, and stochastic). The SSA method is an example of a time-series-based analysis technique used previously in different areas, for example, drought and water levels (Nguyen et al. 2022; Sheikholeslami et al. 2026).

Autoregressive (AR) methods, the most widely used model for updating, take advantage of the "memory" of errors in hydrological simulations (more precisely, their autocorrelation structure) (Khalifa et al. 2019; Niedzielski and Halicki 2023). The basic idea behind AR updating is to predict future errors based on the linear function of the known errors at previous time steps. These expected errors are then factored into revised forecasts. Conceptually, updating AR is straightforward, and in practise, it almost always results in vastly enhanced predictions. The results of AR updating have been shown to be comparable to those of more complex non-linear and non-parametric updating methods (Li et al. 2015). AR has been successfully used in various fields, including streamflow (Mehdizadeh and Kozekalani Sales 2018), weather variables (Karimuzzaman and Moyazzem Hossain 2020) and temperature forecasting (Möller and Groß 2019). Combining SSA with AR linear and nonlinear approaches is a relatively new option that has proven its effectiveness and applicability.

Barba and Rodríguez (2016) employed two hybrid techniques: AR-SSA and artificial neural network (ANN) combined with SSA to estimate multi-step road accidents in Santiago de Chile. Both hybrid models were compared with three single models, including Hankel singular value decomposition (HSVD), AR, and ANN. The outcomes exposed that the hybrid models outperformed the other single techniques. Vahabie et al. (2007) enhanced the performance of the AR method by combining it with SSA to forecast Short-Term Load Forecasting (STLF) in Iran. The load time series has been analysed by the SSA method and recombined then to create the new load time series. The AR–SSA model was compared with neural networks (NN) and fuzzy TSK with Levenberg-Marquardt learning models. The outcomes showed that the AR–SSA technique had a promising result. Li et al. (2014) used a hybrid SSA-AR to predict power load, their model combines an AR model with the SSA. The SSA-AR model results were compared with the backpropagation neural network (BPNN) model, SSA-based linear recurrent method (SSA-LRF), and single AR. The outcomes indicated that the hybrid technique, SSA-AR, outperformed other techniques in predicting short-term power load.

The existing studies lack an investigation of the influence of the SSA-based PBH technique of forecasting univariate monthly streamflow data. Accordingly, this research aims to assess the PBH technique's (SSA-AR) performance to simulate multistep ahead monthly streamflow data of the Tigris River at Amarah City, Iraq, with different scenarios. The first scenario uses SSA to analyse the normalised and cleaned data into various principal components (PC) (trend, seasonal, stochastic, and noise) and reconstructs all PCs without noise. Then, apply an AR model to simulate the denoise time series. The second scenario builds three AR models for each PC (without noise), and the final simulated time series can be obtained by combining the outputs of these three models. The third scenario applies the AR model to simulate the original monthly streamflow time series to evaluate and validate the first two scenarios.

This study introduces a novel SSA-based preprocessing hybrid framework (SSA-AR) for streamflow forecasting of the Tigris River at Amarah, Iraq. Unlike previous studies, it systematically compares three modelling scenarios, including direct modelling, denoised reconstruction, and component-wise forecasting. The main contribution lies in demonstrating that forecasting individual decomposed components significantly improves prediction accuracy and provides deeper insight into the underlying hydrological processes. It is expected that this research will provide an accurate methodology for forecasting streamflow time series, as it uses SSA to alleviate the influence of structureless noise and models the informative time series component using an AR model.

2 Area of Study and Dataset

Maysan Governorate is one of the Iraqi governorates bordering Iran to the east, while its capital, Amarah City, is located on the Tigris River (see Figure S1). It has a total land area of sixteen thousand seventy-two square kilometres and a population of 1,106,208 per capita (Hayder Oleiwi Shami Al Saidia and Abd 2020). The climate of Amarah is characterised by scorching summers and cold winters. Summertime temperatures frequently exceed

40 degrees Celsius. Annual precipitation averages 177 mm throughout the winter season (Abdul Kareem et al. 2022).

Maysan's Water Resources Directorate provided streamflow data (m^3/s) on a monthly basis for the period of 2010 to 2020. Based on these data, the model was built and evaluated. Figure S2 shows the Tigris River's monthly streamflow timeline at Amarah City. The city represents a critical downstream section of the Tigris River, where streamflow reflects the cumulative impacts of upstream processes, including climate variability, reservoir operations, and transboundary water management, making it highly representative of the river's overall behaviour. Furthermore, the region is essential for sustaining the Mesopotamian marshlands—an ecologically sensitive system and key indicator of Iraq's hydrological balance—thereby highlighting the importance of accurate streamflow forecasting for effective water resource management, ecological preservation, and drought assessment.

3 Methods

The proposed methodology includes data pre-processing techniques, an AR model, and a performance assessment criteria model, as shown in Figure S3.

3.1 Data Pre-Processing

Pre-processing the data is an essential step to ensure the distribution is normal or very near to it and to lessen the effect of outliers (Tabachnick and Fidell 2013). To make the data more static and less prone to multicollinearity among predictors, it will normalise the time series using the natural logarithm approach (Alawsi et al. 2022). Then, the remaining outliers (if found) will be rescaled after the transformation method. Outliers outside the 1.5 IQR timeframe were identified using the box-whisker approach (IQR = 3rd quartile (Q3)—1st quartile (Q1) (Kossieris and Makropoulos 2018). This study used the SPSS 24 statistical software to clean data from outliers. Utilising the SSA approach for decomposition and denoising will follow normalisation and cleaning of the time series.

3.2 Singular Spectrum Analysis (SSA)

SSA is generally prominent as an adaptive noise-minimising technique in time series analysis. It is an initial component analysis used for the lagged components of a time series. SSA analyses the raw signal into several individual components, including the primary components (PCs). The basic time series can be recreated via multiple computers (Al-Bugharbee and Trendafilova 2016). As demonstrated by Eq. (1), a time series (X) could be analysed into a trend (D), stochastic (C), seasonal (S), and noise (N) (Donkor et al. 2014).

$$\mathrm{X} = \mathrm{D} + \mathrm{C} + \mathrm{S} + \mathrm{N} \tag{1}$$

The SSA includes two significant phases: decomposition and rebuilding. In the decomposition, the time series is divided into several components, with the highest percentage of the original signal variance in the first component and a minor portion in the last component. During the reconstruction phase, just the initial few signal components are employed to rebuild the signal. The final components are ignored since they constitute structureless noise (Zhigljavsky 2010). In this study, the segmented signals were decomposed using SSA and reconstructed using several principal components. The SSA can be employed for linear and nonlinear time series (Hassani and Mahmoudvand 2013). The aforementioned strategy has been effectively implemented in numerous areas, including groundwater forecast (Polomčić et al. 2017) and streamflow forecasting (Halit Apaydin 2021; Liu et al. 2023).

3.3 Autoregressive Model (AR)

An AR model finds a mathematical signal by linking each data point to several previous ones. The number of previous points is called the model order, whereas the weights of linking past points are referred to as the model coefficients. The AR model's basic form is linear with time-invariant coefficients (Adamowski et al. 2012). This study employs the linear AR model of time-invariant coefficients to anticipate streamflow. A streamflow series can be noted as [X1, X2, X3….XL-2, XL-1, XL] where L is the series length. Then, the AR will represent the series as shown in Eq. (2) (Al-Bugharbee and Trendafilova 2016):

$$Xn = a0 + a1\,X\,n-1 + a2\,X\,n-2 + \dots\dots\dots + ap\,X\,n-p+ \in n \tag{2}$$

Where Xn is the predicted signal value at time n, which is linearly related to (p) previous values, p is the order of the model, a.i. = 0, 1, 2,…… p time-invariant coefficients of the autoregressive model, and €n is the error term, which indicates the difference between actual and forecasted values. In general, the order of a model should not exceed half L (i.e., half the length of the time series). The coefficients of the AR model were calculated using MATLAB's least squares approach.

As previously stated, the AR model relates a given point to several past points. This approach predicts new points using the precision AR model obtained. A section of the streamflow data (126) is utilised for training the AR model, while (26) are data tests. The testing portion is subsequently compared to the actual data. SSA is suitable as it can handle non-stationarity hydrological time series and can extract interpretable components such as trend and seasonal components. On the other hand, the AR represents a good choice for modelling the extracted components in a linear form.

3.4 Performance Assessment Criteria

The forecast models' performance was examined utilising different statistical tests because there is no globally applicable performance measure (Seo et al. 2018). Four criteria are adopted to assess the forecasting capability of the forecast techniques, categorised as absolute and relative errors. The absolute statistical tests include mean absolute error (MAE) and root mean squared error (RMSE), while the relative tests include mean absolute relative error (MARE) and mean bias error (MBE), as given by the Equations below. Because it is evaluated in the same unit as the original data, MAE and RMSE effectively depict the typical error size; in contrast, MARE is more prone to forecasting errors at low data values, making it an excellent index for such cases. Meanwhile, MBE provides complementary information by indicating the average direction of errors, revealing whether the model systematically overestimates or underestimates the observed values.

$$\mathrm{MAE} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathrm{O_i} - \mathrm{F_i}|}{\mathrm{N}} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} (O_i - \mathrm{F_i})^2}{\mathrm{N}}} \tag{4}$$

$$MARE = \frac{1}{N} \sum_{i=1}^{N} \frac{|O_i - F_i|}{O_i} \tag{5}$$

$$\mathrm{MBE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{O_i} - \mathrm{F_i}) \tag{6}$$

Where Oi and Fi are the observed and forecasted streamflow, respectively, and N is the number of data points considered.

4 Results and Discussion

4.1 Data Set Preparation and Model Configuration

The first step is to normalise the streamflow time series by adopting the natural logarithm method, and then the outliers will be examined using the box-whisker approach to treat the outliers later. Following the cleaning and normalisation of the streamflow time series, the following four components will be extracted using the SSA (see Eq. 1): trend, seasonal, stochastic, and noise (Fig. 1). The figure illustrates the normalised and cleaned streamflow time series (first line), the trend signal (second line), the seasonal signal (third line),

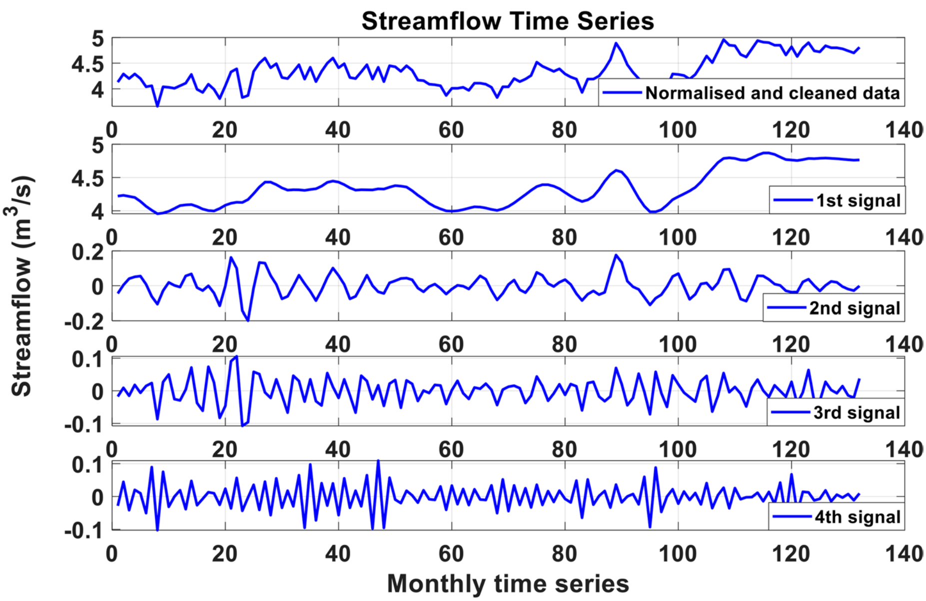

Fig. 1 The four signals provided by SSA, along with normalised and cleaned streamflow data

and stochastic and noise signals (fourth and fifth lines). After that, the first three PCs were selected and reconstructed to build a new streamflow time series without noise (i.e., denoising time series). The SSA decomposition separates the streamflow series into interpretable components (trend, seasonal, and stochastic) and noise. The first three capture meaningful dynamics, while noise is excluded due to its non-predictive nature and low contribution to the explained variance.

Both the low- and high-frequency data sets have been split into a training and testing set, with the former containing 80% (105 datasets) of the samples and the latter 20% (27 datasets). Both the low- and high-frequency data sets have been split into a training and testing set, with the former containing 80% of the samples and the latter 20%. Before utilising the AR model, it is crucial to identify the optimal model order, as improper model order might result in either over- or under-fitting, resulting in incorrect model prediction and an incorrect and/or poor diagnosis. Therefore, a suitable order choice can significantly lower the errors of over- or under-fitting (Al-Bugharbee and Trendafilova 2016).

Three orders for the AR model were selected based on the number of scenarios. The first scenario uses the AR model to simulate the denoised streamflow time series. The second scenario applies the AR model to forecast each PC (without noise) individually, and then the three forecasted PCs were reconstructed to build a new simulated streamflow time series. The last scenario utilises the AR model to predict the raw streamflow time series. Considering these scenarios, the order values for the AR model are 28, 27 and 29 for the 1 st, 2nd, and 3rd scenarios, respectively. For example, Fig. 2 illustrates the best order of the 3rd scenario, showing that errors for the testing stage increase after 29, considering the normalised mean squared error (NMSE) as a fitness function. The selection of AR model order was based on minimising the NMSE, ensuring an optimal trade-off between model complexity and predictive performance. Increasing the order beyond the selected values resulted in higher errors, indicating overfitting, while lower orders led to underfitting. This approach ensures that the selected model order captures the underlying temporal dependencies of the streamflow series without introducing unnecessary complexity.

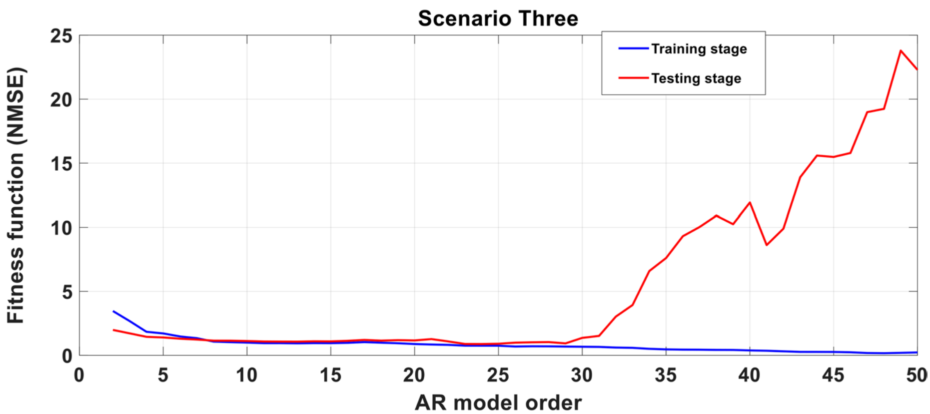

Fig. 2 The best order of the raw data

4.2 Performance Evaluation

After preparing the data sets and the order values for the AR model for the three scenarios, the AR model performs the prediction. The training data set is used to build the AR model to forecast 27 months ahead in the testing stage for each scenario. The efficiency performance of each scenario has been examined according to four error evaluation criteria, including RMSE, MAE, MARE, and NBE. Based on the four statistical criteria, the results for the comparisons between the scenarios for the training and testing stages can be seen in Table 1.

The comparative analysis reveals that the data pre-processing, particularly the SSA technique, has efficiently improved the accuracy of 1 st and 2nd scenarios in comparison to 3rd scenario for the training and testing stages, e.g., the values of MAE in the testing stage are 0.1560, 0.0913, and 12.9008 (m^3/s) for 1 st, 2nd, and 3rd scenarios, respectively. Also, according to Dawson et al. (2007), the outcomes of 2nd scenario offer much more accurate results than 1 st scenario, with MBE = 0.0549 (m^3/s) compared with MBE = −0.1545 (m^3/s) for the 1 st scenario in the testing stage. The 2nd scenario offers the lowest values of all statistical criteria in the testing stage, which indicates that predicting each PC separately offers deep insight into the characteristics of monthly streamflow data.

Moreover, a comparison of graphical plots was also used to reveal the closeness between the measured and simulated data for the three various scenarios in the testing phase, as presented in Figs. 3, 4 and 5. What stands out in Figs. 3, 4 and 5 is that scenarios 1 and 2 can mimic streamflow data closer than scenario three. Also, scenario 1 is the best based on the scale of error.

The outcomes depicted in Table 1 and in Figs. 3, 4 and 5 clearly show that the data pre-processing improved the AR model, especially the SSA technique, concerning the forecasting capability of multistep ahead monthly streamflow data in the testing stage, which confirms the capability of the 2nd scenario to generate unseen monthly streamflow data. The findings of this study highlight a key methodological contribution: modelling each decomposed component separately provides a more effective representation of streamflow dynamics than direct modelling or reconstructed series. This confirms the advantage of SSA-based hybrid frameworks in isolating meaningful structures and improving predictive performance, representing a novel contribution to streamflow forecasting of the Tigris River. To the best of the authors' knowledge, this is the first study that systematically evaluates these three SSA-AR scenarios for the Tigris River streamflow forecasting.

Table 1 Performance of SSA-AR model for:

A: Scenario 1				
Data stage	RMSE(m^3/s)	MAE(m^3/s)	MARE	MBE(m^3/s)
Training	0.1634	0.1195	0.0283	0.00002
Testing	0.21119	0.1560	0.0333	−0.1545
B: Scenario 2				
Data stage	RMSE(m^3/s)	MAE(m^3/s)	MARE	MBE(m^3/s)
Training	0.1498	0.1085	0.0257	−0.0010
Testing	0.1223	0.0913	0.0189	0.0549
C: Scenario 3 (raw data)				
Data stage	RMSE(m^3/s)	MAE(m^3/s)	MARE	MBE(m^3/s)
Training	12.9261	8.8162	0.1285	0.0587
Testing	21.2852	12.9008	0.0925	5.8851

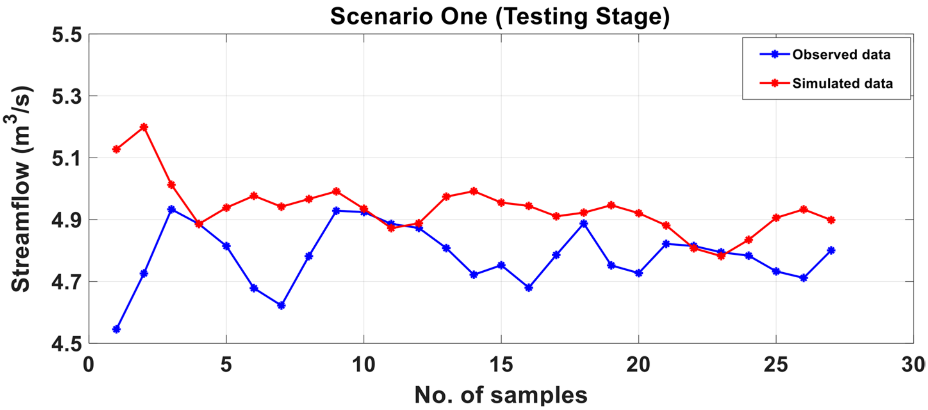

Fig. 3 Measured and predicted streamflow data for scenario 1

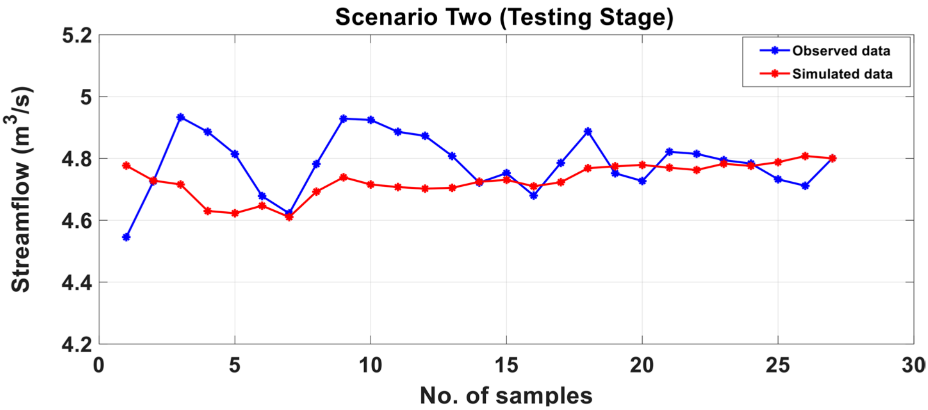

Fig. 4 Measured and predicted streamflow data for scenario 2

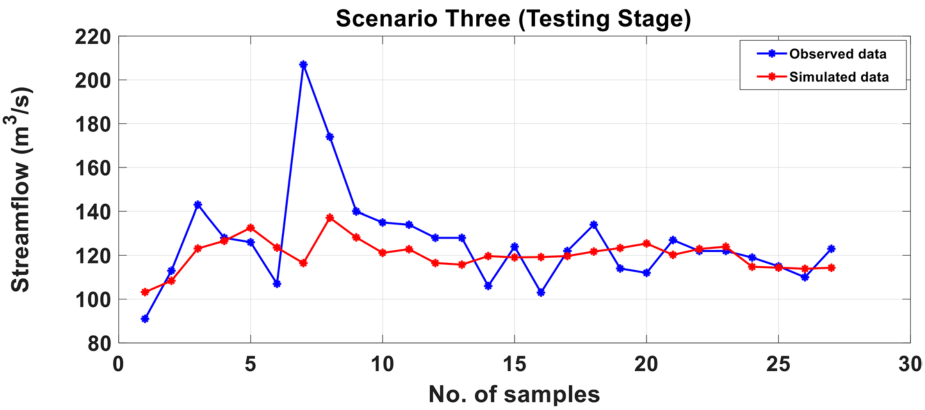

Fig. 5 Measured and predicted streamflow data for scenario 3

5 Conclusions

This research aims to investigate the PBH technique's (SSA-AR) performance to forecast the multistep streamflow data with multiple scenarios. To develop and guarantee the validity of the suggested hybrid model, the Tigris River in Amarah City's monthly streamflow data was used for this study from 2010 to 2020. In the first scenario, the normalised and cleaned streamflow time series analysis into several PCs (i.e., trend, seasonal, stochastic, and noise) utilises the SSA method and reconstructs the PCs without noise. After that, the AR model forecasts the new (denoise) streamflow time series. The second scenario employs the AR model to simulate the first three PCs separately and then reconstructs the simulated PCs to get the final forecasted streamflow time series. The final scenario uses the AR model to predict the raw streamflow time series. Five statistical and graphical performance evaluation criteria were used to assess the various built models' performances, i.e., RMSE, MAE, MARE, and MBE. The outcomes obtained in the study indicate that the SSA technique can improve prediction accuracy, and the suggested SSA-AR model can effectively enhance the AR method for monthly streamflow simulation. Additionally, the performance of the 2nd scenario is better than that of the 1 st one, which means that predicting each PC separately offers deep insight into the characteristics of monthly streamflow data. The 2nd scenario yield MBE = 0.0549 (m^3/s) compared to the 1 st and 3rd scenarios that offer MBE equal to −0.1545 (m^3/s) and 5.8851(m^3/s), respectively. The outperformance of the 2nd scenario can be attributed to the lower complexity of the individual components' structures compared to those in the 1 st and 3rd scenarios, which led to more efficient representations. Accordingly, it can be clearly stated that the SSA-AR model improves streamflow prediction, with the 2nd scenario being the most accurate.

These outcomes can help inform the decision makers in Iraqi Directorates (i.e., managers and policymakers) to better manage the existing streamflow system, which would lead to a better plan for the distribution of water rations in the Maysan Province. Additionally, it can help the Marsh Resuscitation Department restore Iraqi marshes, leading to blooming marshes and enhanced service tourism and economy in the Maysan Governorate. Consequently, considering the benefits described previously, it is recommended that further study be undertaken to mimic the streamflow data caused by weather variables due to extreme weather, which is likely to become more common (i.e., using multivariate prediction models). Also, additional PBH models should be utilised by combining the AR model with other pre-treatment signals (i.e., DWT, VMD, and EEMD techniques) to simulate streamflow time series or any other hydrological issues.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11269-026-04717-w.

Author contributions Baydaa Abdul Kareem: Formal analysis, Investigation, Methodology, Validation, Writing - original draft, Writing - review & editing. Salah L. Zubaidi: Methodology, Conceptualization, Supervision, Software, Writing - review & editing. Hussein Al-Bugharbee: Methodology, Writing - review & editing. Ali W Alattabi: Methodology, Writing - review & editing. Mawada Abdellatif: Methodology, Software, Writing - review & editing.

Funding No funds, grants, or other support were received for this stud.

Declarations

Competing Interests The authors declare that they have no competing interests.

References

Abdul Kareem B, Zubaidi SL, Ridha HM, Al-Ansari N, Al-Bdairi NSS (2022) Applicability of ANN model and CPSOCGSA algorithm for multi-time step ahead river streamflow forecasting. Hydrology Basel 9:171. https://doi.org/10.3390/hydrology9100171

Adamowski J, Fung Chan H, Prasher SO, Ozga-Zielinski B, Sliusarieva A (2012) Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour Res 48(1). https://doi.org/10.1029/2010wr009945

Al Saidia HOS, Abd NR (2020) Analysis of Human Development Indicators in the Maysan Governorate. Int J Innov Creat Chang 14(2):1296

Al-Bugharbee H, Trendafilova I (2016) A Fault Diagnosis Methodology for Rolling Element Bearings Based on Advanced Signal Pretreatment And Autoregressive Modelling. J Sound Vib 369:246–265. https://doi.org/10.1016/j.jsv.2015.12.052

Alawsi MA, Zubaidi SL, Al-Ansari N, Al-Bugharbee H, Ridha HM (2022) Tuning ANN Hyperparameters by CPSOCGSA, MPA, and SMA for Short-Term SPI Drought Forecasting. Atmosphere 13(9):1–24. https://doi.org/10.3390/atmos13091436

Altunkaynak A, Kartal E (2019) Performance comparison of continuous wavelet-fuzzy and discrete wavelet-fuzzy models for water level predictions at northern and southern boundary of Bosphorus. Ocean Eng. https://doi.org/10.1016/j.oceaneng.2019.06.002

Barba L, Rodríguez N (2016) Hybrid Models Based on Singular Values and Autoregressive Methods for Multistep Ahead Forecasting of Traffic Accidents. Math Probl Eng 2016:1–14. https://doi.org/10.1155/2016/2030647

Dawson CW, Abrahart RJ, See LM (2007) HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environ Model Softw 22:1034–1052. https://doi.org/10.1016/j.envsoft.2006.06.008

Donkor EA, Mazzuchi TA, Soyer R, Alan Roberson J (2014) Urban Water Demand Forecasting: Review of Methods and Models. J Water Resour Plan Manag 140(2):146–159. https://doi.org/10.1061/(asce)wr.1943-5452.0000314

Guarasci BL (2015) The National Park: Reviving Eden in Iraq's Marshes. Arab Stud J 23(1):128–153. https://www.jstor.org/stable/44744902

Halit Apaydin MTS, Kambiz Falsafian R Prasad (2021) Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions. J Hydrol 600:126506. https://doi.org/10.1016/j.jhydrol.2021.126506

Hanoon MS, Ammar AM, Ahmed AN, Razzaq A, Birima AH, Kumar P, Sherif M, Sefelnasr A, El-Shafie A (2022) Application of Soft Computing in Predicting Groundwater Quality Parameters. Front Environ Sci 10:1–16. https://doi.org/10.3389/fenvs.2022.828251

Hassani H, Mahmoudvand R (2013) Multivariate singular spectrum analysis: a general view and new vector forecasting approach. Int J Energy Stat 01(01):55–83. https://doi.org/10.1142/s2335680413500051

Jozaghi A, Shen H, Ghazvinian M, Seo D-J, Zhang Y, Welles E, Reed S (2021) Multi-model streamflow prediction using conditional bias-penalised multiple linear regression. Stoch Env Res Risk Assess 35(11):2355–2373. https://doi.org/10.1007/s00477-021-02048-3

Karimuzzaman M, Moyazzem Hossain M (2020) Forecasting performance of nonlinear time-series models: an application to weather variable. Model Earth Syst Environ 6(4):2451–2463. https://doi.org/10.1007/s40808-020-00826-6

KatİPoĞLu OM (2022) Monthly streamflow prediction in Amasya, Türkiye, using an integrated approach of a feedforward backpropagation neural network and discrete wavelet transform. Model Earth Syst Environ. https://doi.org/10.1007/s40808-022-01629-7

Katipoğlu OM (2023) Evaporation prediction with wavelet-based hyperparameter optimised K-nearest neighbors and extreme gradient boosting algorithms in a semi-arid environment. Environ Processes 10(4). https://doi.org/10.1007/s40710-023-00669-0

Khalifa AA, Al-Maadid A, Caporin M (2019) Water demand in Qatar: future trends and conservation scenarios. Water Util J 22:27–42

Kossieris P, Makropoulos C (2018) Exploring the statistical and distributional properties of residential water demand at fine time scales. Water Basel. https://doi.org/10.3390/w10101481

Li H, Cui L, Guo S (2014) A hybrid short-term power load forecasting model based on the singular spectrum analysis and autoregressive model. Adv Electr Eng 2014:1–7. https://doi.org/10.1155/2014/424781

Li M, Wang QJ, Bennett JC, Robertson DE (2015) A strategy to overcome adverse effects of autoregressive updating of streamflow forecasts. Hydrol Earth Syst Sci 19(1):1–15. https://doi.org/10.5194/hess-19-1-2015

Liu G, Savic D, Fu G (2023) Short-term water demand forecasting using data-centric machine learning approaches. J Hydroinform 25(3):895–911. https://doi.org/10.2166/hydro.2023.163

Mehdizadeh S, Kozekalani Sales A (2018) A comparative study of autoregressive, autoregressive moving average, gene expression programming and Bayesian networks for estimating monthly streamflow. Water Resour Manage 32(9):3001–3022. https://doi.org/10.1007/s11269-018-1970-0

Mirzania E, Roshni T, Ghorbani MA, Heddam S (2024) River water temperature prediction using a hybrid model based on variational mode decomposition (VMD) and Outlier Robust Extreme Learning Machine. Environ Process. https://doi.org/10.1007/s40710-024-00716-4

Mohamadi S, Sheikh Khozani Z, Ehteram M, Ahmed AN, El-Shafie A (2022) Rainfall prediction using multiple inclusive models and large climate indices. Environ Sci Pollut Res 29(56):85312–85349. https://doi.org/10.1007/s11356-022-21727-4

Möller A, Groß J (2019) Probabilistic temperature forecasting with a heteroscedastic autoregressive ensemble postprocessing model. Q J R Meteorol Soc 146(726):211–224. https://doi.org/10.1002/qj.3667

Nguyen AD, Le Nguyen P, Vu VH, Pham QV, Nguyen VH, Nguyen MH, Nguyen TH, Nguyen K (2022) Accurate discharge and water level forecasting using ensemble learning with genetic algorithm and singular spectrum analysis-based denoising. Sci Rep 12(1):19870. https://doi.org/10.1038/s41598-022-22057-8

Niedzielski T, Halicki M (2023) Improving linear interpolation of missing hydrological data by applying integrated autoregressive models. Water Resour Manage 37(14):5707–5724. https://doi.org/10.1007/s11269-023-03625-7

Osman Y, Abdellatif M, Al-Ansari N, Knutsson S, Jawad S (2017) Climate change and future precipitation in an arid environment of the Middle East: case study of Iraq. J Environ Hydrol 25:1–18

Pham QB, Yang T-C, Kuo C-M, Tseng H-W, Yu P-S (2021) Coupling singular spectrum analysis with Least Square Support Vector Machine to improve accuracy of SPI drought forecasting. Water Resour Manag 35(3):847–868. https://doi.org/10.1007/s11269-020-02746-7

Polomčić D, Gligorić Z, Bajić D, Cvijović Č (2017) A hybrid model for forecasting groundwater levels based on Fuzzy C-Mean Clustering and singular spectrum analysis. Water Basel. https://doi.org/10.3390/w9070541

Salman SA, Shahid S, Ismail T, Chung E-S, Al-Abadi AM (2017) Long-term trends in daily temperature extremes in Iraq. Atmos Res 198:97–107. https://doi.org/10.1016/j.atmosres.2017.08.011

Seo Y, Kwon S, Choi Y (2018) Short-term water demand forecasting model combining variational mode decomposition and Extreme Learning Machine. Hydrology Basel. https://doi.org/10.3390/hydrology5040054

Sharma S, Raj Ghimire G, Siddique R (2022) Machine learning for postprocessing ensemble streamflow forecasts. J Hydroinformatics 25(1):126–139. https://doi.org/10.2166/hydro.2022.114

Sheikholeslami R, Khaleghi H, Karimaee A (2026) Determining nonlinear and interactive processes driving PM_{10} variability via interpretable machine learning. Environ Process. https://doi.org/10.1007/s40710-026-00813-6

Siddiqi TA, Ashraf S, Khan SA, Iqbal MJ (2021) Estimation of data-driven streamflow predicting models using machine learning methods. Arab J Geosci 14(11):1058. https://doi.org/10.1007/s12517-021-07446-z

Sulandari W, Subanar S, Lee MH, Rodrigues PC (2020) Time series forecasting using singular spectrum analysis, fuzzy systems and neural networks. MethodsX 7:101015. https://doi.org/10.1016/j.mex.2020.101015

Sun X, Zhang H, Wang J, Shi C, Hua D, Li J (2022) Ensemble streamflow forecasting based on variational mode decomposition and long short term memory. Sci Rep. https://doi.org/10.1038/s41598-021-03725-7
Tabachnick BG, Fidell LS (2013) Using Multivariate Statistics. Pearson Education, Inc, United States of America
Vahabie AH, Yousefi MMR, Araabi BN, Lucas C, Barghinia S (2007) Combination of singular spectrum analysis and autoregressive model for short term load forecasting. *Proc.*, *IEEE Lausanne Power Tech*, 1090–1093
Wang W-c, Chau K-w, Xu D-m, Chen X-Y (2015) Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. Water Resour Manage 29(8):2655–2675. https://doi.org/10.1007/s11269-015-0962-6
Xing WY, Bai YL, Ding L, Yu QH, Song W (2022) Application of a hybrid model based on GA–ELMAN neural networks and VMD double processing in water level prediction. J Hydroinform 24(4):818–837. https://doi.org/10.2166/hydro.2022.016
Yilmaz M, Tosunoğlu F, Kaplan NH, Üneş F, Hanay YS (2022) Predicting monthly streamflow using artificial neural networks and wavelet neural networks models. Model Earth Syst Environ 8(4):5547–5563. https://doi.org/10.1007/s40808-022-01403-9
Yuan R, Cai S, Liao W, Lei X, Zhang Y, Yin Z, Ding G, Wang J, Xu Y (2021) Daily runoff forecasting using ensemble empirical mode decomposition and long short-term memory. Front Earth Sci. https://doi.org/10.3389/feart.2021.621780
Zhang Y-F, Fitch P, Thorburn PJ (2020) Predicting the trend of dissolved oxygen based on the kPCA-RNN model. Water Basel. https://doi.org/10.3390/w12020585
Zhao X, Lv H, Lv S, Sang Y, Wei Y, Zhu X (2021b) Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimiser. J Hydrol 601:126607. https://doi.org/10.1016/j.jhydrol.2021b.126607
Zhigljavsky A (2010) Singular spectrum analysis for time series: introduction to this special issue. Stat Its Interface 3(3):255–258. https://doi.org/10.4310/SII.2010.v3.n3.a1

Authors and Affiliations

Baydaa Abdul Kareem[1,2] · Salah L. Zubaidi[2,3] · Hussein Al-Bugharbee[4] · Ali W. Alattabi[2] · Mawada Abdellatif[5]

✉ Salah L. Zubaidi
salahlafta@uowasit.edu.iq

✉ Mawada Abdellatif
M.E.Abdellatif@ljmu.ac.uk

Baydaa Abdul Kareem
bida.k.z@uomisan.edu.iq

Hussein Al-Bugharbee
hrazzaq@uowasit.edu.iq

Ali W. Alattabi
anakemish@uowasit.edu.iq

1 Department of Civil Engineering, University of Maysan, Maysan 57000, Iraq

2 Department of Civil Engineering, Wasit University, Wasit 52001, Iraq

3 College of Engineering, University of Warith Al-Anbiyaa, Karbala 56001, Iraq

4 Department of Mechanical Engineering, Wasit University, Wasit 52001, Iraq

5 Civil Engineering and Built Environment Department, Faculty of Engineering Technology, Liverpool John Moores University, Liverpool L3 5UX, UK