Runoff data series prediction based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Radial Basis Function Neural Network extension

Abstract

This study investigated the influence of data extension on the decomposition and prediction accuracy of runoff data series. To this end, an original data series was constructed using annual runoff data from a hydrological station in China (Tang Naihai) for the period 1956–2013, and radial basis function neural network (RBFNN) extension was applied to the original data series. Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) was then applied to both data series, and their decomposition and prediction results were compared. The decomposition results indicate that the end effect significantly lowers the accuracy of low–middle frequency components. Nevertheless, the end effect could be effectively suppressed and decomposition error could be reduced by applying RBFNN extension. At the end points, the extension data series could more accurately reflect the real fluctuation characteristics of components and subsequent variation trends. Regarding component prediction, the prediction results followed the variation trend of the components themselves, with a rather large gap in the prediction results of low-frequency components between the two groups of data series. The final prediction results obtained from the reconstruction of the component prediction results suggest that the extension sequence has a clearly superior prediction accuracy than the original data series. Hence, when using the CEEMDAN method to process non-stationary hydrological data, multi-time-scale information of the data series can be obtained through reasonable extension after decomposition of the original data series. The acquired information provides evidence for the analysis and prediction of the evolution law of hydrological elements.

Keywords: Runoff forecasting; complete ensemble empirical mode decomposition (CEEMD); end effect; radial basis function neural network (RBFNN); data extension; data-driven model; combination forecast; autoregressive integrated moving average (ARIMA)

1. Introduction

Runoff modeling plays an important role in hydrological studies, with wide applications extending from water supply to disaster management. The variation of runoff has multi-time scale characteristics which are of use for runoff modeling building. To determine the internal law of the evolution of hydrological data series and predict the evolution process of regional hydrological elements, the fluctuation state and trend of each time scale of runoff data series need to be analyzed. However, under the joint influence of natural and social systems, runoff data series often show complex non-stationary and nonlinear characteristics, complicating the multi-time scale analysis and prediction of runoff sequences.

A conventional method for processing non-stationary and non-linear series is empirical mode decomposition (EMD) (Huang et al., 1998, 1999). However, EMD has limited applications in the presence of perturbations and has the disadvantages of mode splitting and mode mixing (Huang and Wu, 2008; Sang et al., 2014). These shortcomings can be overcome by the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method (Torres et al., 2011, 2014), which can also improve decomposition precision and more clearly distinguish between different scales of change patterns in complex data (Marusiak & Pekar, 2014; Zhang et al., 2017). Thus, CEEMDAN has high potential in the processing of nonlinear data series. This method has already been applied in multi-time scale analysis and prediction of hydrological systems (Adarsh et al., 2015; Liu et al., 2018; Zhang et al., 2019, Norani et al., 2014; Di et al., 2014; Zhang et al., 2016; Prasad et al., 2018; Wen et al., 2019). However, the CEEMDAN method is still based on EMD decomposition; the end effect problem in the EMD algorithm (Huang et al., 1999; Deng et al., 2001) tends to lead to the distortion of decomposition results, which is a key factor limiting decomposition accuracy. Moreover, many studies (Liu et al., 2018; Adarsh et al., 2015; Prasad et al., 2018; Wen et al., 2019) have directly applied CEEMDAN to original hydrological data series without considering the end effect. Therefore, their decomposition may be inaccurate and their analysis and prediction results may not reflect the actual situation.

Thus far, the end effect has primarily been studied in the context of the EMD method, with little focus on the CEEMDAN method. Among various extension-based methods that can solve the end effect of EMD (Zhao et al., 2001, Shu et al., 2006, Zhang et al., 2003, Deng et al., 2001; Hu et al., 2007), the radial basis function neural network (RBFNN) extension technique is highly suitable for hydrological data as it can achieve high performance for predicting nonlinear and non-stationary data series (Alizadeh et al., 2017; Elanayar and Shin, 1994; Taormina et al., 2016).

The approach of decomposition, prediction, and reconstruction has widely been applied in the prediction of hydrological time series (Norani et al., 2014; Di et al., 2014; Zhang et al., 2016; Prasad et al., 2018; Wen et al., 2019). Previous studies investigated prediction using decomposed sequences and proposed that the prediction can be improved by selecting a prediction mode of the corresponding model according to the characteristics of different components.

In this study, we explored the potential of RBFNN extension to solve the end effect. To ensure the rationality and accuracy of component prediction, RBFNN prediction was applied to the strong nonlinear component obtained by CEEMDAN decomposition and autoregressive integrated moving average (ARIMA) prediction was applied to the more stable component. CEEMDAN was applied to both the original and RBFNN series, and the results were compared to verify the effectiveness of RBFNN extension.

2. Method

2.1 CEEMDAN method

The CEEMDAN method can be used to extract relatively stable intrinsic mode functions (IMFs) and trends (Res) from original data to clearly demonstrates the fluctuation characteristics of different time scales, facilitating convenient analysis and prediction of complex data series.

The CEEMDAN algorithm is described as follows (Torres et al., 2011, 2014):

Let x denote data series to be decomposed, $E_k(\cdot)$ denote the k-order modal operators generated by EMD, and $M(\cdot)$ denote the local mean operator to generate the data series to be decomposed. Where $E_1(x)=x-M(x)$. Let $w^{(i)}$denote Gaussian white noise with mean 0, and variance 1; $x^{(i)}=x+w^{(i)}$, $\langle\cdot\rangle$ is the operator for calculating the average value. ε_0 denotes the inverse of the ideal SNR for the first added noise and the original signal

$\beta_0=\varepsilon_0 std(x)/std(E_1(w^{(i)})), \beta_k=\varepsilon_0 std(r_k), k \geq 1$.

(1) Use EMD to calculate $x^{(i)}=x+\beta_0 E_1(w^{(i)})$ (i.e. $x=1$) to obtain the first residual value:

$$r_1=¿ \tag{1}$$

(2) In the first stage ($k=1$), calculate the first order modal $\tilde{d}_1=x-r_1$.

(3) Let the average value of the local mean $r_1+\beta_1 E_2(w^{(i)})$ be the estimated value of the second residual value, then define the second order mode as follows:

$$\tilde{d}_2=r_1-r_2=r_1-\langle M(r_1+\beta_1 E_2(w^{(i)}))\rangle \tag{2}$$

(4) For $k=3,\cdots,K$, calculate the k th residual value

$$r_k=\langle M(r_{k-1}+\beta_{k-1} E_k(w^{(i)}))\rangle \tag{3}$$

(5) Calculate the k order mode

$$\tilde{d}_k=r_{k-1}-r_k=r_{k-1}-\langle M(r_{k-1}+\beta_{k-1} E_k(w^{(i)}))\rangle \tag{4}$$

(6) Return to step 4 to calculate the next k.

Repeat steps (4) to (6) until the obtained residual satisfies one of the following conditions, and the calculation can be terminated: IMF conditions met; the number of local extremum points is less than 3; no further decomposition can be done by EMD.

After reconstruction by CEEMDAN, the final residual satisfies the following equation:

$$r_k=x-\sum_{k=1}^{K}\tilde{d}_k \tag{5}$$

K is the total order of the modes. Hence, the original data series (*x*) can be described as:

$$x = x - \sum_{k=1}^{K} \tilde{d}_k \tag{6}$$

The specific algorithm of CEEMDAN demonstrates that the $E_k($¿$)$ algorithm is used repeatedly in the process of decomposing data, i.e., EMD decomposition. Therefore, the end effect also affects the decomposition accuracy of CEEMDAN.

2.2 RBFNN extension

RBFNN extension is a prediction method based on data series. It has several advantages, such as simple topological structure, fixed network structure, and fast and efficient learning, and is suitable for nonlinear data series.

The RBFNN extension method can be explained in detail as follows (Hu et al., 2007):

For a given data series *x* (let length be *n*) first generated by certain rules, its learning sample matrix is $P_{m \times k}$ and its corresponding target matrix is $T_{l \times k}$, where k is the number of samples, and m, l are data numbers. In MATLAB, the newrbe function is used to design a standard radial basis network to input the training sample (P, T) into the training network. By selecting the appropriate spread value, a trained radial basis network can be obtained.

Use this network for data extension: Determine the sample matrix p_1 of data series x at the boundary (such as the right boundary); then input it into the trained RBFNN network to obtain the extension data output a_1. Let a_1 be the new boundary for the original data series to generate the new sample matrix p_2, input p_2 into the network to gain the new extension data a_2, and so on, until data series of appropriate length is extended at the right end of the data. In addition, the data series containing the appropriate length is extended at the left end likewise. In this manner, extended data series x_1 was obtained for CEEMDAN decomposition.

The prediction length should be set within a reasonable range because the prediction accuracy of RBFNN decreases as the number of prediction steps increases. Furthermore, at least one minimum and one maximum point at each end of the extension is required to inhibit the end effect in the decomposition process (Hu et al., 2007).

3. Case study

3.1 Data selection

In this study, annual runoff data from the Tang Naihai Hydrological Station in the source area of the Yellow River covering the period from 1956 to 2013 were used for the analyses. We set up three sets of data series: (1) A dataset covering 44 years from 1963 to 2006 was intercepted as the original data series. (2) The left and right ends of the original sequence were considered for RBFNN extension, where the left end extended to 1956 and the right end extended to 2013. In this case, the obtained data series from 1956 to 2013 is extension data series. (3) The entire dataset of annual runoff from 1956 to 2013 was taken as the standard data series.

3.2 Data extension

In the original data series, the maximum interval between two adjacent maximum or minimum points is 6 years. In order to ensure that new extreme points can emerge in the extension to inhibit the end effect with certain accuracy, both left and right ends of the original data series were selected as the prediction length of the RBFNN model. The extension results and relative error obtained by comparing the measured data from 1956 to 1962 and 2007 to 2013 with the RBFNN extension results are shown in Fig. 1 and Table 1, respectively.

Fig. 1 Original data series and extension result

Table 1. RBFNN extension error

Fig. 1 and Table 1 show the occurrence of six extreme points during 2007–2012 in the right extension. The first 5 extreme points are in accordance with the change trend of the measured data, while the last extreme point (2012) shows a deviation, with a relative error reaching 36.59%. Regarding the left extension, three extreme points in 1958, 1960, and 1962 were obtained. The extreme point in 1962 is inconsistent with the trend of the measured data, with an error of 37.96%, while the other two extreme points are generally in line with the change trend of the measured data. Hence, for annual runoff data series with strong non-stationarity, their law of change is difficult to determine using the RBFNN model, and direct prediction may lead to large errors at some points. On the whole, although there are two points with large deviation, the left and right ends of the extension can basically reflect the general trend of the standard data series.

3.3 CEEMDAN decomposition

3.3.1 Comparison of decomposition results

The original data series, extension data series, and standard data series were decomposed using the CEEMDAN method; the decomposition results are presented in Fig. 2. The extension sequence and standard sequence in the 1956–1962 and 2007–2013 data series were listed as "pollution" data in CEEMDAN attributable to the end effect. Therefore, these two parts of the series were discarded. The decomposition accuracy of the original data series and the extension sequence was evaluated based on the decomposition result of the standard data series.

Fig. 2 CEEMDAN decomposition results of original data series, extension data series, and standard data series

According to Fig. 2, CEEMDAN provides five components each for the original data series, extension data series, and standard data series. We compare each layer of the three data series and discuss the effect of RBFNN extension on CEEMDAN.

Regarding the IMF1 component, the three sets of data basically overlap, and the end effect has no obvious influence on the first layer component. Thus, there is no convincing evidence that data extension can improve the decomposition accuracy.

Regarding the IMF2 component, the middle parts of the three data series basically overlap. However, at both ends of the component, although the original data series and the standard data series follow the same trend, a certain separation occurs between the lines, and errors attributable to the end effect begin to show. In contrast, the extension data series remains close to the standard value and the end effect is inhibited.

Regarding the IMF3 component, the overlapping parts between the original data series and the standard data series are dramatically reduced, and the change trend and real value at both ends are out of sync. In particular, the mismatch is more significant at and the quasi-period is significantly larger at the right end, further reflecting the influence of the end effect. The decomposition results of the extension data series still coincide with the real value to a large extent and separation only occurs at the two ends. Although the amplitude is somewhat deviated, the trend is generally synchronized.

Regarding the IMF4 component, the gap between the original data series and the standard data series is further widened, the amplitude and quasi-period are significantly deviated, and both ends show a trend of further expansion. The extension data and standard value are also widely separated, but the amplitude difference is rather small, and the variation trend, quasi-period, and other information are still consistent with the real data series.

The Res components of the last three groups all reflect a downward trend of runoff series on a long time scale, and the trend changes of the three groups are basically the same, except for the difference in amplitude. In this layer, the original data series and the extension data series are separated from the standard value, and the improvement effect of data extension on the decomposition accuracy is not obvious.

The results showed that the CEEMDAN method is susceptible to the end effect and the corresponding errors gradually spread inward from the end points along with decomposition, thus affecting the decomposition effect of the entire data series. For the IMF1 component and trend, data extension does not provide any obvious improvement in decomposition accuracy. However, for other middle- and low-frequency components, RBFNN extension can significantly inhibit the end effect, improve the decomposition accuracy, and more accurately reflect the fluctuation characteristics under multiple time scales. Particularly at the end point, the extension data series can reflect the trend of data and facilitate convenient runoff prediction.

3.3.2 Decomposition error

In this study, root-mean-square error (RMSE) and correlation coefficient (R) (Chai and Draxler, 2014; Krause et al., 2005) were calculated to quantitatively describe the decomposition accuracy of extension data series and non-continuation sequences with reference to the decomposition results of the standard data series.

RMSE can be calculated as follows:

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-X_i\right)^2}{n}} \tag{7}$$

where n denotes the number of data, x_i denotes the ith data in the real data series, and X_i denotes the ith data of the corresponding component after CEEMDAN. The smaller the value of e, the smaller the error between the data series and real data series.

R can be expressed as follows:

$$R_j=\frac{\operatorname{cov}\left(x_j,\mathrm{IMF}_j\right)}{\sqrt{\delta\left(x_j\right)}\sqrt{\delta\left(\mathrm{IMF}_j\right)}} \tag{8}$$

where x_j is the jth component of the original data series, IMF_j is the jth intrinsic mode function, δ is variance, and cov is covariance. The correlation coefficient R_j can represent the similarity between IMF and standard data series components and reflect the decomposition accuracy. The closer the value of R to 1, the more accurate the decomposition result.

Table 2. Errors in the original data series and extension data series

As shown by the RMSE and R of the IMF2, IMF3, and IMF4 components in Table 2, the decomposition accuracy of the extension series was significantly improved compared with the original data series. In the IMF1 component, both series showed relatively high accuracy, with no obvious advantage of the extension, but the error of the Res component was smaller in the extension data series. Overall, the decomposition accuracy of the extension data series is higher than that of the original data series.

3.4 Runoff prediction

3.4.1 Component prediction

Each layer component of the original data series and the extension data series was predicted. A prediction length of seven years was set, covering the period from 2007 to 2013. Although the standard series covers the period from 2007 to 2013, it is located in the right end of the sequence. Under the influence of the end effect of CEEMDAN, the decomposition results for these years may produce errors and do not accurately reflect the actual change rule of the component. Therefore, the were not considered the standard for evaluating the performance of the prediction. For each layer component, the prediction result of the extension data series is termed prediction 1, and the prediction result of the original data series is termed prediction 2. The difference between the two groups of predictions were primarily investigated.

After CEEMDAN decomposition, IMF1 of the two groups of runoff data series retained strong non-stationarity. Therefore, the RBFNN model was used to make predictions; the ARIMA model was used for other components as they appeared to be relatively stable. The prediction results of each layer component are illustrated in Figs. 3–7.

Fig. 3 Prediction results of the IMF1 component

Fig. 4 prediction results of the IMF2 component

Fig. 5 Prediction results of the IMF3 component

Fig. 6 Prediction results of the IMF4 component

Fig. 7 Prediction results of the Res component

As shown in Fig. 3, as the IMF1 components of the original data series and extension data series basically overlap, differences in their predicted values are also negligible. The first four points of prediction 1 are below those of prediction 2, which may be due to the fact that the value of the original data series at the right end of the IMF1 component is smaller than the continuation sequence, and this trend is further reflected in the prediction.

In Fig. 4, although the IMF2 components of the two sets of data series generally overlap in the middle, the inconsistencies at the right end points lead to a more pronounced deviation of their prediction results. The occurrence of the extreme point of prediction 2 lags behind that of prediction 1, and its peak value is also smaller, consistent with the variation characteristics of the right end of the IMF2 component in both groups of data series.

As illustrated in Figs. 5 and 6, the predictions of the two series reflect the variation trend of their respective components in this layer, particularly at the end points. However, the difference between them is more significant and the fluctuation rules reflected are also quite different. As for the IMF3 component, prediction 1 begins to decline after reaching the peak, while prediction 2 continues to rise from the trough to near the peak. In this component, prediction 1 first increases to the maximum value and then exhibits a downward trend, while prediction 2 maintains an upward trend from trough to peak.

Fig. 7 shows that the decline rate of the Res component of the original data series slowed down at the end. Prediction 1 maintains this trend and shows a slight increase, while prediction 2 maintains the downward trend towards the end.

In summary, the prediction results of the IMF1 component are basically consistent. The prediction gap between the original and extension data series is mainly reflected in other middle- and low-frequency components, and the change law at the end point exerts a major influence on the prediction results. In addition, after CEEMDAN decomposition, the low- and middle-frequency components were relatively stable, with rather weak nonlinearity. Therefore, the prediction results are relatively reliable. The final prediction obtained by the reconstruction of the prediction results of IMF1 with small differences and other components of the reliable prediction results can accurately and objectively reflect the prediction performance of the two groups of data series.

3.4.2 Reconstruction

Prediction 1 and prediction 2 of each component were reconstructed to obtain prediction results of the extension data series and original data series (Fig. 8 and Table 3). The previously described extension results for 2007–2013 using RBFNN (Section 2.2) correspond to the direct application of RBFNN to the original data series without any preprocessing. This represents is a general method commonly used in hydrological time series prediction. In Table 3, the relative errors of extension results from 2007 to 2013 are added for comparison with the two groups of prediction results using the "decompositions, predictions-reconstruction" mode.

Fig. 8 Prediction results of the extension data series and original data series

Table 3. Prediction results and relative errors

As shown in Fig. 8 and Table 3, the predicted fluctuation rules of the extension and original data series are consistent with the standard series, but their prediction accuracy is somewhat different. In the first five points (2007–2011), the predicted runoff of the original data series was significantly lower than that of the standard data series, with the relative error in 2008 reaching 22.88%; in contrast, the prediction result of the extension data series was more accurate, and relative error remained within 10%. Combined with the prediction results of the IMF1 and Res components in Fig. 3–7, the prediction results of the original data series are larger than those of the extension data series, whereas the opposite is true for the 2–4 components, with widely varying prediction results. In the reconstruction process, the predicted values of the original data series were relatively smaller because of the superposition of this gap.

In general, the "decomposition-prediction-reconstruction" mode improved the prediction results of both series for 2012, which had a rather large error in the RBFNN extension results (Fig. 1 and Table 1); the error of the prediction results of the original data series was smaller. However, in the prediction results of the extension data series, the largest error was still observed for 2012, which is consistent with the distribution law of the extension error. In the prediction results of the original data series, the errors from 2008 to 2010 were higher than that of 2012; even for 2008 and 2010, the prediction error was higher than that of RBFNN extension result. This may be due to the end effect and the deviation of component prediction, due to which errors in the prediction results of the non-extended original data series were transferred to other points after " decomposition-prediction-reconstruction".

Among the three methods (direct application of RBFNN on the original sequence, direct application of "decomposition-prediction-reconstruction", and direct application of "extension-decomposition-prediction-reconstruction" on prediction results), the direct use of RBFNN showed the poorest prediction performance, with an average relative error reaching 15.11%. The average relative error of the CEEMDAN decomposition-prediction-reconstruction method was 11.67%, showing that the prediction performance was improved to some extent. The average relative error of the RBFNN extension-CEEMDAN decomposition-reconstruction method was 5.23%, indicating further improvements in prediction accuracy. Therefore, when using the CEEMDAN method for prediction, the original data series can be reasonably extended and the "decomposition-prediction-reconstruction" approach can be implemented to effectively reduce prediction errors.

4. Conclusions

In this study, we addressed the issue of the end effect in the CEEMDAN method and discussed the impact of RBFNN extension on decomposition accuracy. On this basis, predictions were made to verify the effectiveness of RBFNN extension in improving prediction performance. The main conclusions can be summarized as follows:

(1) The end effect of the CEEMDAN method introduces decomposition errors. As decomposition proceeds, the errors gradually spread inward from the end points, with a relatively small impact on high-frequency components but a rather large impact on medium- and low-frequency components.

(2) RBFNN extension on original data can significantly improve the decomposition accuracy of middle- and low-frequency components, and more accurately retain the characteristics of the original data series, such as fluctuation period and amplitude at respective time scales. Thus, RBFNN extension improves the reliability of the multi-time scale analysis of hydrological series.

(3) The fluctuation characteristics at the end points of the components significantly affect the prediction results of the component. The decomposition results of the extension data series can accurately reflect variation trends at the end points of the components, providing a basis for accurate predictions.

(4) The CEEMDAN method can handle nonlinear data well and provide relatively stable components for prediction. In the practical application of CEEMDAN for prediction based on the decomposition-prediction-reconstruction approach, RBFNN extension can effectively enhance the decomposition accuracy, thus improving prediction accuracy and prediction performance.

References

Adarsh, S., Janga, M., 2015, Multiscale analysis of suspended sediment concentration data from natural channels using the Hilbert-Huang Transform. Geoderma. 4, 780-788.

Alizadeh, M. J., Kavianpour, M. R., Kisi, O., Nourani, V., 2017. A new approach for simulating and forecasting the rainfall-runoff process within the next two months. J. Hydrol. 548, 588-597.

Chai, T., Draxler, R. R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?– arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7 (3), 1247-1250.

Colominas, M. A., Schlotthauer, G., Torres, M. E., 2014. Improved complete ensemble EMD: a suitable tool for biomedical signal processing. Biomedical Signal Processing and Control 14, 19-

29.

Deng, Y. J., Wang, W., Qian, C. C., Wang, Z., Qian, D. J., 2001. Boundary processing technique in EMD method and Hilbert transform. Chinese Science Bulletin. 46(1), l-8.

Di, C., Yang, X., Wang, X., 2014. A four-stage hybrid model for hydrological time series forecasting. PLoS ONE 9 (8), e104663.

Elanayar, V. T., Shin, Y. C., 1994. Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems. IEEE Trans. Neural Networks 5 (4), 594-603.

Hu, J. S., Yang, Sh. X., 2007. Application of EMD method with data extension technique based on RBF neural network to time-frequency analysis. Journal of Mechanical Strength. 29(6), 894-899

Huang, N. E., et al., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society, pp. 903-995.

Huang, N. E., Shen, Z., Long, S. R., 1999. A new view of nonlinear water waves: the Hilbert spectrum. Annu. Rev. Fluid Mech. 31 (1), 417-457.

Huang, N. E., Wu, Z., 2008. A review on Hilbert-Huang transform: method and its applications to geophysical studies. Rev. Geophys. 46 (2), RG2006.

Karthikeyan, L., Kumar, D.N., 2013. Predictability of nonstationary time series using wavelet and EMD based ARMA models. J. Hydrol. 502 (2), 103-119.

Liu, D., et al., 2018. Multifractal detrended fluctuation analysis of regional precipitation sequences based on the CEEMDAN-WPT. Pure & Applied Geophysics. 175(8): 3069-3084.

Maheswaran, R., Khosa, R., 2012. Wavelet-Volterra coupled model for monthly stream flow forecasting. J. Hydrol. 450, 320–335.

Marusiak, O., Pekar, J., 2014. Analysis of multiannual fluctuations and long term trends of hydrological time series. Proc of Advances in Environmental Sciences, Development and Chemistry 17-21 July 2014 Greece, pp.156-159.

Norani, V., Baghanam, A. H., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet–artificial Intelligence models in hydrology: a review. J. Hydrol. 514, 358-377.

Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2018. Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition. Geoderma. 330, 136-161.

Sang, Y.F., Wang, Z., Liu, C., 2014. Comparison of the MK test and EMD method for trend identification in hydrological time series. J. Hydrol. 510 (3), 293-298.

Shu, ZH. P., Yang, ZH. CH., 2006. A better method for effectively suppressing end effect of empirical mode decomposition. Journal of Northwestern Polytechnical University. 24(5), 639-643.

Taormina, R., Galelli, S., Karakaya, G., Ahipasaoglu, S. D., 2016. An information theoretic approach to select alternate subsets of predictors for data-driven hydrological models. J. Hydrol. 542, 18-34.

Torres, M. E., Colominas, M. A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4144–4147.

Wen, X. H., et al., 2019. Two-phase extreme learning machines integrated with the complete ensemble empirical mode decomposition with adaptive noise algorithm for multi-scale runoff prediction problems. J. Hydrol. 570, 167-184.

Zhang, H., Singh, V. P., Wang, B., Yu, Y., 2016. CEREF: a hybrid data-driven model for forecasting annual streamflow from a socio-hydrological system. J. Hydrol. 540, 246-256.

Zhang, J. P., Xiao, H. L., Zhang, X., Li, F. W., 2019. Impact of reservoir operation on runoff and sediment load at multi-time scales based on entropy theory. J. Hydrol. 569, 809-815

Zhao, J. P., Huang, D. J., 2001. Mirror extending and circular spline function for empirical mode decomposition method. Journal of Zhejiang University. Science A, 2(3), 247-252.

Zhang, W., et al., 2017. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy Convers. Manage. 136, 439-451.

Zhang, Y. S., Liang, J. W., 2003. Suppression of the end effect of empirical mode decomposition by using auto-regressive model. Science Development. 13(10), 1054-1059.