IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 12, 2020, accepted December 22, 2020, date of publication January 1, 2021, date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048745

Drought Prediction Method Based on an Improved CEEMDAN-QR-BL Model

YANG LIU AND LIHU WANG
School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450046, China
Corresponding author: Lihu Wang (hellowanglihu@126.com)

This work was supported in part by the National Key Research and Development Project under Grant Strategic Research Projects in Key Areas 16, and in part by the Water Conservancy Science and Technology Research Project in Henan Province under Grant GG202042.

ABSTRACT In this study, a new broad learning (BL) model based on an improved complete ensemble empirical mode decomposition adaptive noise (CEEMDAN) is proposed to resolve the low accuracy, poor robustness, and long delay problems that are present in current drought assessments. First, the extreme delay method was applied to improve the CEEMDAN end effect. The improved CEEMDAN method was then used to decompose a series of non-steady-state signals from drought monitoring into multiple steady-state components. A BL model based on orthogonal trigonometry (QR) was then used to predict these multiple steady-state components, and the predicted components were further reorganised to obtain a high-precision drought sequence. On this basis, CEEMDAN was introduced into the orthogonal triangular broad learning (QR-BL), and a drought prediction model (CEEMDAN-QR-BL) combining CEEMDAN and QR-BL was proposed. Finally, the De Martonne aridity index was used to calculate the drought sequence results and determine the drought grades. To meet the real-time requirements of drought prediction, parallel computing was introduced into the CEEMDAN-QR-BL model, and a drought prediction method based on parallel CEEMDAN-QR-BL was constructed. The experimental results show that, when compared with a support vector regression model combined with an empirical mode decomposition, the reliability and accuracy of the CEEMDAN-QR-BL increases by 29.57% and 11.84%, respectively. In addition, when compared with only BL, the prediction efficiency of QR-BL improved by 62.29%.

INDEX TERMS Broad learning, drought assessment, empirical mode decomposition, orthogonal triangular matrix decomposition.

I. INTRODUCTION

Drought is a complex interdisciplinary issue involving meteorology, hydrology, geology, ecology, agriculture, the social economy, and other multi-disciplinary and multi-sectoral subjects. However, the intermittency, uncertainty, and randomness of weather signals has brought great challenges to weather and drought forecasting.

According to different prediction principles, drought prediction can be divided into mechanism-driven and data-driven models. Mechanism-driven models are often realised by modelling the relationship between drought indicators and potential drought factors [1]. For example, the Multivariate Ensemble Streamflow Prediction model predicts future drought states by analysing the relationship between precipitation, soil moisture, and drought [2]. By studying the time change of soil moisture in combination with the standardised precipitation evapotranspiration index, the model is able to identify drought conditions in different seasons [3]. However, mechanism-driven models often involve problems such as numerous external interference factors and a poor understanding of the mechanism. Data-driven models include both single models and hybrid models. Monomer models include the linear regression [4], random forest [5], support vector regression [6], and the neural network [7] models. Monomer models often have problems such as poor model generalisation and weak noise processing abilities. Using a hybrid model to predict drought can effectively solve the shortcomings of a single model. Both genetic algorithms [8] and particle swarm optimisation algorithms [9] can effectively overcome the shortcomings of artificial neural networks (ANN) that can easily fall into local optimums. The combination of wavelet transform and support vector regression (SVR) can effectively improve the prediction

The associate editor coordinating the review of this manuscript and approving it for publication was Yu-Da Lin.

accuracy of SVR [10]. However, the wavelet transform cannot describe the local characteristics of the signal in the time domain and is not suitable for the decomposition of non-stationary signals.

Empirical Mode Decomposition (EMD) performs signal decomposition based on the time scale characteristics of the data itself, which is fundamentally different from the Fourier and wavelet decomposition methods based on a priori wavelet basis function [11], [12]. Through the combination of ANN and EMD to achieve drought prediction, the coefficient of determination is increased from 0.83 to 0.95 [13]. Through the integration of EMD and Deep Confidence Network (DBN) to predict the time series and obtain the future drought situation, the accuracy of the prediction is effectively improved [14]. To exert the advantages of EMD further, polynomial fitting can be used to correct the error caused by the end effect of EMD before applying it to time series forecasting [15]. Combining the ensemble empirical mode decomposition (EEMD) with white noise and an artificial neural network, and subsequently applying it to drought prediction can effectively avoid the modal aliasing problem of the EMD method and further improve the accuracy of drought prediction [16].

As the white noise of EEMD will not be completely cancelled in practice, complete ensemble empirical mode decomposition adaptive noise (CEEMDAN) is proposed, and the reconstructed signal of CEEMDAN is almost identical to the original signal [17]. This research first uses the extreme value delay method to improve the CEEMDAN end effect; subsequently, orthogonal triangular decomposition (QR) is used to redefine the broad learning (BL) output matrix solution method to improve the efficiency of the BL calculation. On this basis, CEEMDAN is introduced into orthogonal triangular broad learning (QR-BL) and combined with parallel computing ideas, resulting in the proposal of a drought prediction model combining CEEMDAN and QR-BL (CEEMDAN-QR-BL). The drought prediction model (EMD-SVR) after combining EMD and SVR is used as a comparison model. The experimental results show that, when compared with EMD-SVR, CEEMDAN-QR-BL has a 29.57% increase in reliability and an 11.84% increase in accuracy. In addition, when compared with only BL, the prediction efficiency of QR-BL is improved by 62.29%.

II. CORRELATION THEORY

A. BROAD LEARNING

Broad learning (BL) is a random-vector single-layer neural network learning system that links the neural network as a carrier and expands the designed network through the increment of neural nodes [18]. As it ensures that the function approximation method has a good generalisation ability, the BL method effectively eliminates the shortcomings of long training processes [19]–[22], while exhibiting a stronger generalisation ability and efficiency than those of the neural networks. The basic BL principle is shown in Figure 1.

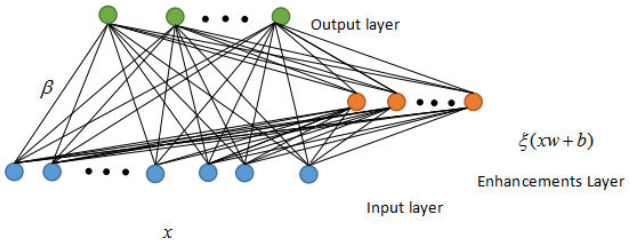


FIGURE 1. Topological structure of the broad learning model.

With given N arbitrarily different training samples $\{(x_i, t_i)\}_{i=1}^{N}$, x_i is multiplied by a set of random weights, and the random deviation is added to the enhancement layer. The weights will not change in the subsequent process. The matrix H is obtained after the data passed to the enhancement layer undergo the activation function. Finally, the original input data matrix of the input layer is defined as the combination of X and the output matrix H of the enhancement layer. The mathematical model definition of a BL network with N input neurons and M enhancement layer neurons is shown in (1):

$$[X|H]\,\beta = T \tag{1}$$

where T is the training set and matrix H is defined according to (2):

$$\begin{aligned} H &= \left[\, h(x_1) \quad h(x_2) \quad \ldots \quad h(x_N) \,\right]^T \\ &= \begin{bmatrix} \xi(w_1 x_1 + b_{11}) & \ldots & \xi(w_1 x_n + b_{1n}) \\ \ldots & \ldots & \ldots \\ \xi(w_m x_{1m} + b_{m1}) & \ldots & \xi(w_m x_n + b_{mn}) \end{bmatrix}_{n*m} \end{aligned} \tag{2}$$

where ξ is the activation function of the enhancement layer, w and b are the weight and bias of the enhancement layer, respectively, and $h(x)$ is the row vector of matrix H.

The output matrix H of the enhancement layer is determined after randomly generating the enhancement layer neuron parameter (w_i, b_i) and given training samples according to any continuous sampling distribution probability, which is transformed to solve the minimum norm and least square solution of formula (1) β, as shown in (3):

$$\beta = [X|H]^{+}T \tag{3}$$

where $[X|H]^{+}$ is the Moore–Penrose generalised inverse of matrix $[X|H]$, and β is the weight of output layers [23]–[26].

The standard breadth-learning algorithm is based on the principle of empirical risk minimisation, and its training process is prone to overfitting. Additionally, when there are many outliers in the training sample, the output matrix H of the enhancement layer is uncertain, which weakens the generalisation ability and robustness of the model. The regularisation theory can be used to resolve these issues [27], [28]. The mathematical regularised BL (regularised BL, RBL) model is shown in (4):

$$\begin{aligned} &\min_{\beta \in R^{L*m}} \frac{1}{2}\|\beta\|_p^{\sigma_1} + \frac{C}{2}\|\beta\|_q^{\sigma_2} \\ &s.t. \ \left\{ h(x_i)\beta = t^T - \zeta_i^T \right\} \quad i = 1, 2, \ldots, N \end{aligned} \tag{4}$$

where $\sigma_1 > 0$, $\sigma_2 > 0$, $\|\bullet\|_p$ is the L_p norm of a vector or matrix. Therefore, the output of regularised BL can be obtained using (5):

$$f(x) = [X|H]^T \left(\frac{I}{C} + [X|H][X|H]^T\right)^+ T \tag{5}$$

where $C \geq 0$ is the regularisation or penalty parameter used to balance the experience risk and model complexity, and I is the number of training samples [22], [29].

B. ORTHOGONAL TRIGONOMETRIC DECOMPOSITION

QR decomposition achieves excellent performance for solving linear problems, with a simple and effective calculation process. During the solving process, QR decomposition maintains its efficiency, regardless of the size of the Hussain matrix [30]. The algorithm can be described as follows: Let $A = CD$ be the maximum rank decomposition of A, and use the Gram–Schmidt standard orthogonality for the r linearly independent column vectors of C Method, as shown in (6):

$$C = (\mu_1 \quad \mu_2 \quad \dots \quad \mu_r) \tag{6}$$

where μ_r represents the column vector of matrix A.

$$(\mu_1 \quad \mu_2 \quad \dots \quad \mu_r) = (\overset{0}{\alpha_1} \quad \overset{0}{\alpha_2} \quad \dots \quad \overset{0}{\alpha_r}) \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1r} \\ 0 & k_{22} & \dots & k_{2r} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & k_{rr} \end{bmatrix} \tag{7}$$

where $\overset{0}{\alpha_1}\ \overset{0}{\alpha_2}\ \dots\ \overset{0}{\alpha_r}$ is a pairwise orthogonal unit vector and $k_{11}, \dots, k_{rr}$ is greater than zero; thus,

$$C = Q \cdot K \tag{8}$$

where Q is defined as shown in (9):

$$Q = (\overset{0}{\alpha_1} \quad \overset{0}{\alpha_2} \quad \dots \quad \overset{0}{\alpha_r}) \quad s.t.\{Q^H Q = E\} \tag{9}$$

where E is the identity matrix. When $KD = R$, the orthogonal triangular decomposition of matrix A can be defined using (10):

$$A = Q \cdot R \tag{10}$$

where Q is an orthogonal matrix and R is an upper triangular matrix [31], [32].

C. COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITION ADAPTIVE NOISE

EMD is a new method for processing non-stationary signals proposed by Huang *et al.* in 1998 [33]. EMD continuously extracts the various scale components of the original signal from a high to low frequency, and finally obtains a residual component with a frequency that is close to zero. EEMD is an improved method based on EMD, which mainly adds Gaussian white noise to the original signal to solve the modal aliasing problem of the EMD method [34]–[37]. In practice, the white noise of EEMD will not be completely cancelled; therefore, CEEMDAN was proposed. CEEMDAN adaptively adds white noise according to the signal, and obtains the Intrinsic Mode Function (IMF) component by calculating the unique residual signal; this ensures that the reconstructed signal is almost completely the same as the original signal [17], [38]–[40]. The digital signal $x(\bar{t})$ with noise can be described as follows:

$$x(\bar{t}) = x(t) + \sigma n(t), \quad t = 1, 2, \dots, M \tag{11}$$

where $x(t)$ is the ideal noiseless signal, $n(t)$ is the noise obeying $N(0,1)$, and σ is the variance of the noise.

CEEMDAN is an optimisation algorithm based on EMD, and the specific steps are as follows:

1) Gaussian white noise $x(t)$ with different amplitudes is added to signal $n^k(t)$ to generate several new signals.

$$x^k(t) = x(t) + \sigma_k n^k(t) \tag{12}$$

2) The EMD method is used to decompose $x^k(t)$ and obtain the first IMF, and the average value is then calculated.

$$\overline{IMF_1(n)} = \frac{1}{K}\sum_{k=1}^{K} IMF_1^k(t) \tag{13}$$

3) The residual signal component R_n of CEEMDAN is calculated:

$$R_n = \begin{cases} x(t) - IMF_1(n), n = 1 \\ R_{n-1} - IMF_n(n), n > 1 \end{cases} \tag{14}$$

4) Denoted as E_j, i.e. the j-th component after EMD decomposition of the signal, the second IMF is acquired as follows:

$$IMF_{(2)}(n) = \frac{1}{\mathrm{K}}\sum_{k=1}^{K} E_1\left\{R_1(t) + \sigma_1 E_1(n^k(t))\right\} \tag{15}$$

5) The L-th residual component is calculated by analogy:

$$R_L(t) = R_{L-1}(t) = \overline{IMF_L(t)} \tag{16}$$

6) The L+1 IMF is calculated as follows:

$$IMF_{(L+1)}(n) = \frac{1}{\mathrm{K}}\sum_{k=1}^{K} E_L\left\{R_L(t) + \sigma_L E_L(n^k(t))\right\} \tag{17}$$

7) Finally, the original signal is represented as,

$$x(t) = \sum\nolimits_{m=1}^{M} \overline{IMF}_m(t) + R(t) \tag{18}$$

III. CONSTRUCTION OF THE INTELLIGENT DROUGHT PREDICTION MODEL

A. QR-RBL PREDICTION MODEL

This study proposes a new enhanced intelligent broad learning prediction model (QR-BL model) that uses the QR orthogonal triangular decomposition method to decompose the broad learning weight matrix, which replaces the traditional singular value decomposition (SVD) method, thereby optimising the output board learning layer. The QR-BL model has a higher efficiency than that of the traditional BL model.

To ensure that the model had a high generalisation ability, the regularisation was introduced to the QR-BL model, and a regular BL model was constructed based on orthogonal triangle decomposition (QR-RBL model). The basic implementation is as follows:

Let $M+N=L$, then N input neurons and M enhancement layer neurons are defined as (19):

$$Q_L R_L = A_L. \tag{19}$$

where

$$A_L = \begin{bmatrix} A_1 & A_2 & \dots & A_l \end{bmatrix},$$
$$Q_L = \begin{bmatrix} q_1 & q_2 & \dots & q_l \end{bmatrix},$$

and

$$R_L = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1l} \\ 0 & r_{22} & \dots & r_{2l} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & r_{ll} \end{bmatrix}.$$

After expanding (19), we can get (20) as follows:

$$\begin{cases} a_1 = q_1 r_{11} \\ a_2 = q_1 r_{12} + q_2 r_{22} \\ \dots \\ a_l = q_1 r_{1l} + q_2 r_{2l} + \dots + q_l r_{ll} \end{cases} \tag{20}$$

From this, (21) and (22) can be derived:

$$a_1^T a_1 = r_{11} q_1^T q_1 r_{11} = r_{11}^2$$
$$s.t. \begin{cases} r_{11} = \sqrt{a_1^T a_1} \\ q_1 = \dfrac{h_1}{r_{11}}, \end{cases} \quad \text{and} \tag{21}$$
$$a_l = q_l r_{ll} = a_l - \sum_{i=1}^{L-1} q_i r_{il}$$
$$s.t. \begin{cases} r_{ll} = \sqrt{a_l^T a_l} \\ q_l = \dfrac{h_l}{r_{ll}}. \end{cases} \tag{22}$$

According to the related properties of the inversion of the block matrix, we can infer the following:

$$\begin{bmatrix} B & D \\ 0 & C \end{bmatrix}^{-1} = \begin{bmatrix} B^{-1} & -B^{-1} D C^{-1} \\ 0 & C^{-1} \end{bmatrix} \tag{23}$$

where B and C are reversible, hence

$$R_{l+1}^{-1} = \begin{bmatrix} R_l & r_{l+1} \\ 0 & r_{l+1,l+1} \end{bmatrix}^{-1} = \begin{bmatrix} R_l^{-1} & -R_l^{-1} r_{l+1} r_{l+1,l+1}^{-1} \\ 0 & r_{l+1,l+1}^{-1} \end{bmatrix} \tag{24}$$

Finally, the regularisation board learning can be defined as follows:

$$f(x) = A^T \left(\frac{I}{C} + AA^T \right)^+ T = R_{l+1}^{-1} Q_{l+1}^T T$$
$$= \begin{bmatrix} R_l^{-1} & -R_l^{-1} r_{l+1} r_{l+1,l+1}^{-1} \\ 0 & r_{l+1,l+1}^{-1} \end{bmatrix} \begin{bmatrix} Q_l^T \\ q_{l+1}^T \end{bmatrix} T = \begin{bmatrix} f_l - R_l^{-1} r_{l+1} f_{l+1}^{\mathrm{T}} \\ f_{l+1}^T \end{bmatrix}. \tag{25}$$

Therefore, L consecutive neural nodes are updated, and the calculation cost of QR-RBL is $O(L^2Nm)$ [41].

B. IMPROVED CEEMDAN BASED ON EXTREMUM EXTENSION

The end effect refers to the phenomenon where the third-order spline function is used to obtain the upper and lower envelope of the signal during the EMD decomposition process, the end point is divergent owing to the lack of extreme point constraints, and the decomposition result is distorted. To improve the boundary effect, we can consider the known data information to extend or estimate the data at both ends of the original signal. Pearson's correlation coefficient was used as a waveform matching function to delay the original signal boundary in this study. The specific steps are as follows:

As shown in Figure 2, for a given meteorological sequence $x(t)$, the two leading extreme points on the left end are recorded as M_0 and N_0 as the maximum and minimum values, respectively, and the waveform from the starting point on the left end to N_0 is defined as W_0 while the wavelength is defined as I.

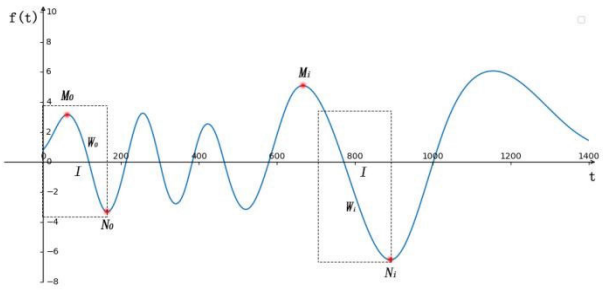


FIGURE 2. Extension of extreme values of meteorological series.

The algorithm can be defined as follows: Let the minimum value set of $x(t)$ be $E_{\min}$, and the maximum value set be $E_{\max}$. Using each minimum point in $E_{\min}$, except N_0, as a reference point, define the i-th minimum point as N_i, and define the adjacent sub-signal with N_i as the end point and wavelength I as waveform W_i. Using the Pearson correlation coefficient M_p calculate the matching degree of waveform W_i and W_0:

$$M_p(W_i, W_0) = \frac{\sum_{i=1}^{N} \left(W_i - \bar{W}_i\right)\left(W_0 - \bar{W}_0\right)}{\sqrt{\sum_{i=1}^{N} \left(W_i - \bar{W}_i\right)^2 \sum_{i=1}^{n} \left(W_0 - \bar{W}_0\right)^2}} \tag{26}$$

where $\overline{W_i}$ and $\overline{W_0}$ represent the mean value of W_i and W_0, respectively. Calculate the distance scale D between the

sequences W_i and W_0.

$$D = \frac{\sum_{i=1}^{N} |W_i - W_0|}{N} \tag{27}$$

If $M_p(W_i, W_0) < \alpha D(\alpha$ is a constant), take the sub-signal containing two extreme points at the left end of W_i as the continuation signal of the sequence $x(t)$. If $M_p(W_i, W_0) \geq \alpha D$, the average of the two minimum points N_0 and N_1 are taken at the leftmost end as the minimum point $P_{\min}$ of the extension, and $t(N_{\min}) = t(N_1) - t(N_0)$, where t represents the time position of the point. The average of the two maximum points M_0 and M_1 at the leftmost end is taken as the extended maximum value $P_{\max}$, and $t(M_{\max}) = t(M_1) - t(M_0)$.

C. CONSTRUCTION OF PARALLEL CEEMDAN-QR-BL METHOD

The solution of the width learning output matrix was redefined with orthogonal triangular decomposition, the CEEMDAN end effect with extreme extension was solved, the improved CEEMDAN with the improved width learning model was integrated, and this was combined with the parallel thinking to build a 'decomposition-synthesis' strategy based on the parallel CEEMDAN-QR-BL model; its basic realisation is as follows:

1) First, the two ends of the original signal were extended by the extreme value extension method, and then the extended signal with CEEMDAN was decomposed to obtain N different proof modal components to complete the transition from the unsteady-state timing signal to the steady-state timing signal.

2) Autocorrelation analysis of IMFs and $x(t)$ was carried out, and the wavelet high parameter threshold processing was applied to the selected high-frequency IMF components, and wavelet low parameter threshold processing to the low-frequency IMF components. Finally, the original signal was reconstructed from the filtered IMF components to obtain the denoised result. The wavelet threshold function is defined as follows:

$$\hat{W}_{j,k} = \begin{cases} sgn\left(W_{j,k}\right) \\ \left(|W_{j,k}| - a\dfrac{\lambda}{\sqrt[N]{|W_{j,k}|^N - \lambda^N}} \right), & |W_{j,k}| \geq \lambda \\ 0, & |W_{j,k}| < \lambda \end{cases} \tag{28}$$

In this formula, N is the number of wavelet decomposition layers, sgn is the symbolic function, and λ is the threshold. There are many ways to select the threshold; this study used heuristic threshold estimation rules.

3) Each signal component (IMF) was used as the input of the QR-BL model, and parallel prediction was performed on the IMF to establish a QR-BL parallel computing model.

4) Sequence recombination of the predicted future signal was performed to obtain the future predicted value of the original signal. The model principle is shown in Figure 3.

D. DROUGHT PREDICTION METHOD BASED ON PARALLEL CEEMDAN-QR-BL

This study used intelligent calculation methods to predict the future meteorological status, and then classified future droughts based on the existing drought evaluation indicators, which have important practical guiding significance for the regional drought prevention and mitigation work. This study used the parallel CEEMDAN-QR-BL model to predict the future monthly average rainfall and monthly average temperature. The De Martonne aridity index was used as the drought grade classification standard to establish the drought risk assessment model. The De Martonne index is defined in (29).

$$I_{\text{dm}} = \frac{12R}{T + 10} \tag{29}$$

where R is the monthly precipitation and T is the monthly average temperature. Index values below 30 indicate the occurrence of droughts. Index values between 10 and 30 indicate a moderate drought, while values below 10 indicate a severe drought. The drought model was established as follows:

1) Some cities in Henan Province are taken as experimental objects. Their monthly urban meteorological data are obtained, and illegal data, such as null and non-numerical data, are filtered through a data cleaning technology.

2) The filtered data are decomposed by CEEMDAN into several modal components (IMF).

3) Based on the wavelet threshold, using the original signal as a benchmark, the high-frequency IMF components are subjected to wavelet high parameter threshold processing, and the low frequency IMF components are subjected to wavelet low parameter threshold processing.

4) The test and training data are imported into the intelligent model in parallel (proportion of the test set was 0.35), and predictions are made for the next two, four, and six months.

5) The predicted components are reorganised to generate the output sequence.

6) The forecast data are used as the input of the De Martonne index to classify droughts. The detailed data processing flow is shown in Figure 4.

E. MODEL EVALUATION

The Ens Nash efficiency coefficient was used to evaluate the credibility and stability of the prediction model, which ranges from negative infinity to one. Ens values being close to one indicates that the studied model is of good quality and has high credibility. Ens values being close to zero indicates that the simulation result is close to the average of the observed value; that is, the overall result is credible, but the process simulation error is large. Ens values being well below zero indicates that the model is not credible. The relative error (RE) and mean absolute error (MAE) were also determined. The real-time error and overall error were evaluated separately. The definition of each indicator is as

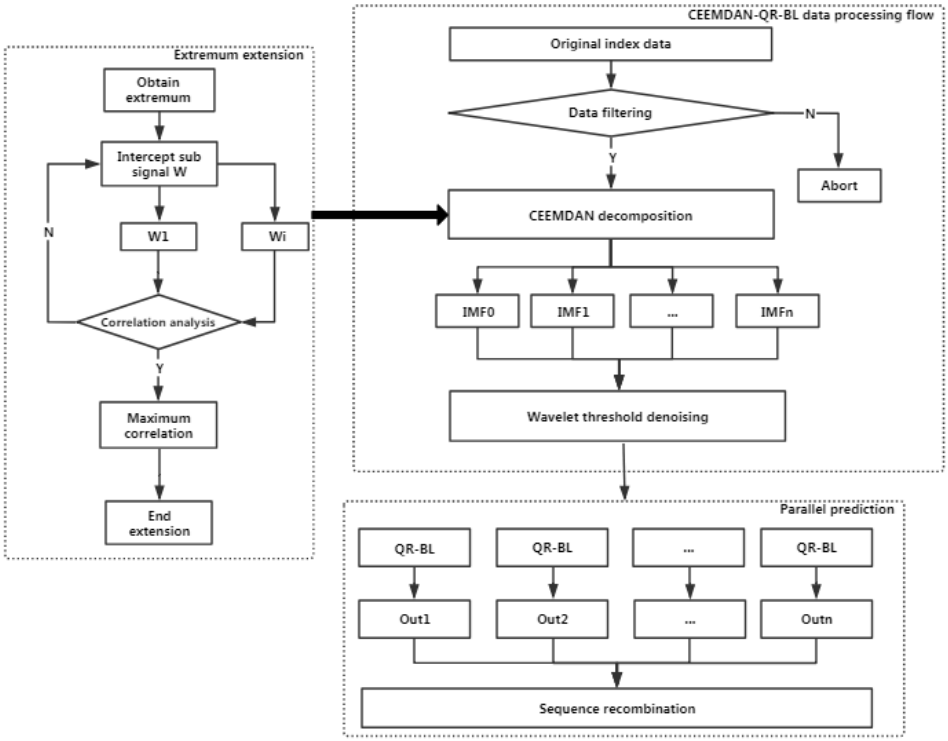


FIGURE 3. Structure flowchart of parallel CEEMDAN-QR-BL.

follows:

$$E_{ns} = 1 - \left[\frac{\sum_{i=1}^{n} \left(Q_o - Q_f\right)^2}{\sum_{i=1}^{n} \left(Q_o - \bar{Q}_o\right)^2} \right] \tag{30}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| Q_f - Q_o \right| \tag{31}$$

$$RE = \left(\frac{Q_o - Q_f}{Q_o} \right) * 100 \tag{32}$$

where Q_o, Q_f, $\overline{Q_o}$, $\overline{Q_f}$ are the predicted value of the observed value, the average observed value, and the average predicted value, respectively. In addition, an energy-based evaluation index θ is used to evaluate the end effect, where $\theta > 0$ and $\theta = 0$ indicate that the end effect has no influence, and the larger the value of θ, the greater the end effect. The definition of the indicator is as follows:

$$\theta = \frac{\left| \sqrt{\sum_{i=1}^{n} RMS_i^2} - RMS_{orginal} \right|}{RMS_{orginal}} \tag{33}$$

where $RMS_{orginal}$ represents the effective value of the original signal, RMS_i represents the effective value of the i-th IMF, and n represents the total number of IMFs. RMS is defined as $RMS = \sqrt{\frac{\sum_{i=1}^{m} s^2(i)}{m}}$ and $s(i)$ represents the original signal sequence. m represents the number of samples.

IV. RESULTS AND ANALYSIS

Kaifeng City, Henan Province, was selected as the research object of this study. Precipitation is unevenly spatiotemporally distributed in Kaifeng City, which often triggers drought or flood disasters and impacts local agriculture. The meteorological data of Kaifeng from 1951 to 2014 were selected as the research object and used to compare the CEEMDAN-QR-BL, EMD-SVR, BL, and QR-BL models. Figure 5 shows the improvement effect of the extreme value extension method for the end effect in the meteorological data.

Figure 5 shows part of the sample points at the end of the meteorological data. The black, blue, and red curves in Figures 5(a)–(f) represent the known endpoint extreme values; the endpoint extreme value is unknown and the extreme value extension is extended and the endpoint is also unknown. Table 2 numerically compares the evaluation index θ obtained with and without extreme value extension. In the decomposition process of CEEMDAN, the upper and lower envelopes diverge at both ends of the data sequence, and this divergence will gradually move inward as the operation proceeds. By comparing the IMF with the unknown endpoint extreme value, it was found that the extreme value extension method introduced in this article has a good inhibitory effect on the CEEMDAN endpoint effect.

The black, blue, green, red, and dark-red curves in Figures 6(a)–(f) represent the observed values and values predicted by the CEEMDAN-QR-BL, EMD-SVR, BL, and QR-BL methods. Figure 6 indicates that a stable average temperature signal influenced the algorithm's prediction

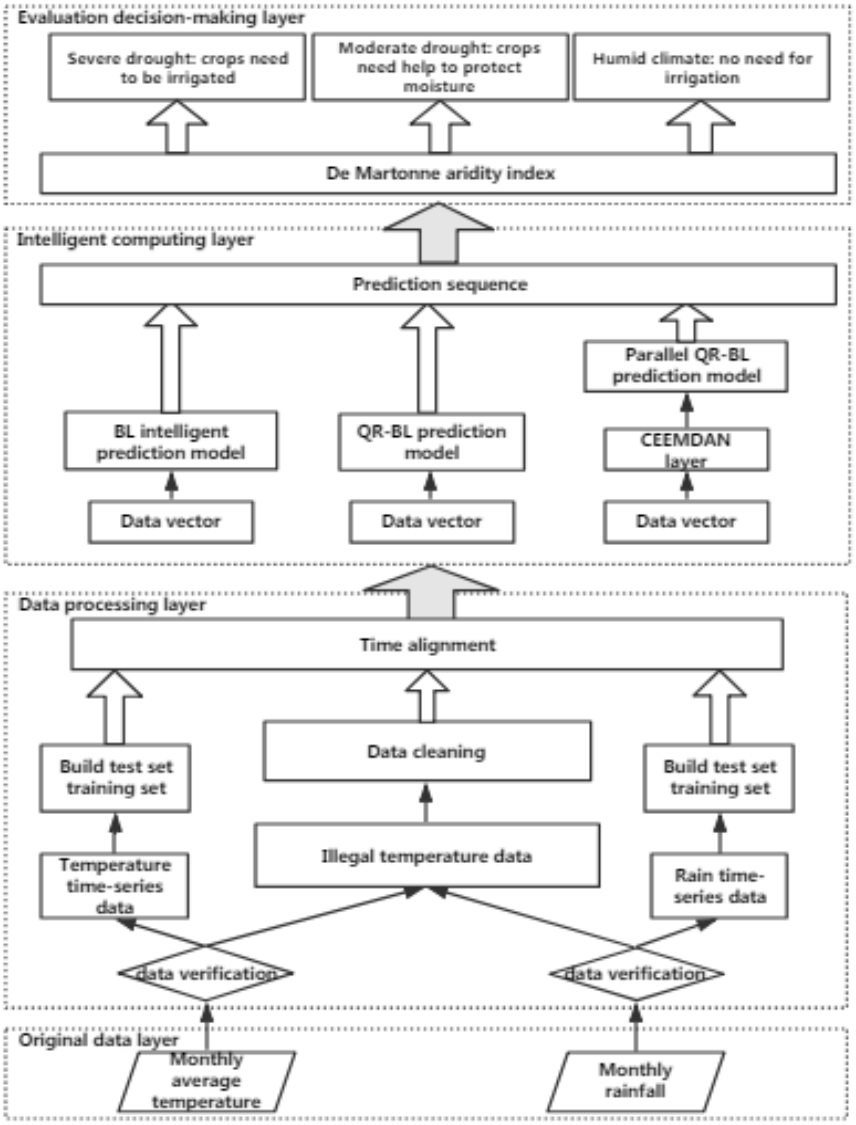


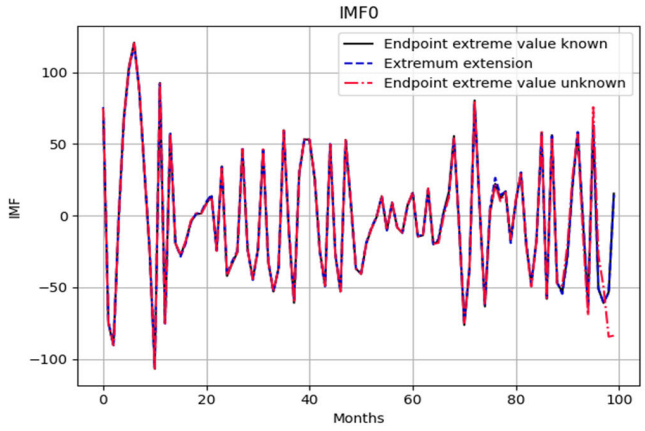
FIGURE 4. Flowchart of drought model based on CEEMDAN-QR-BL.

TABLE 1. Comparison of numerical results of end effect evaluation index.

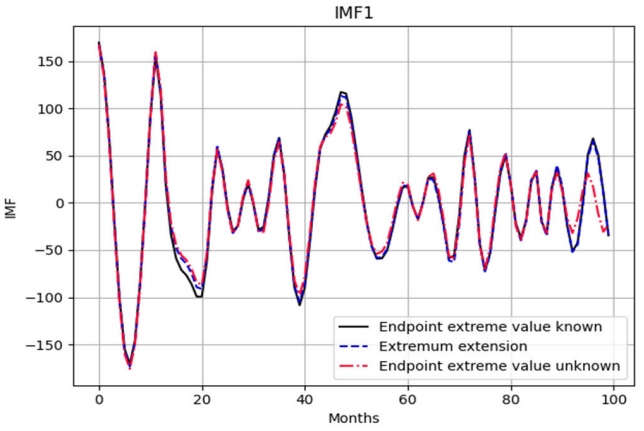
Data processing scenario	θ
Extremum extension	1.4506
Endpoint extreme value unknown	1.7564

performance. CEEMDAN-QR-BL performed better than the BL and EMD-SVR algorithm in low-latency future rainfall prediction. As the number of forecast months increases, the advantages of CEEMDAN-QR-BL over BL and EMD-SVR remained. Overall, the mixed model is better than the single model, and CEEMDAN-QR-BL performs best.

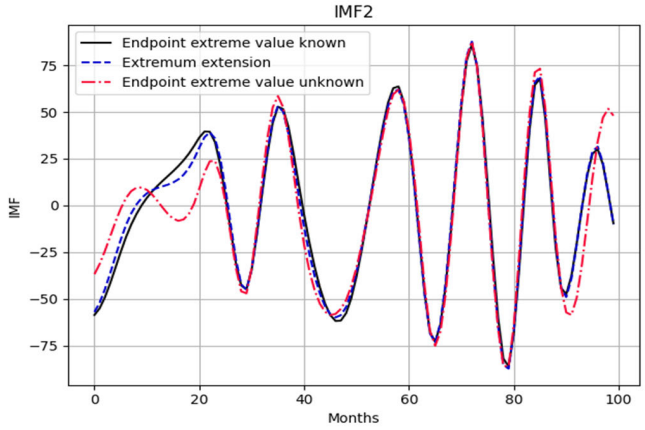
In Figure 7, the blue, green, red, and dark-red curves represent the real-time error and Ens performance of the CEEMDAN-QR-BL, EMD-SVR, BL, and QR-BL methods, respectively. The figure shows that CEEMDAN-QR-BL performed better than the BL, EMD-SVR, and QR-BL models when predicting weather conditions for the next two, four, and six months. The overall CEEMDAN-QR-BL had levels of 0.5–0.7, which are slightly higher than the 0.5–0.6 level of BL and EMD-SVR, and considerably higher than the -0.5 to 0.1 level of BL and QR-BL. Additionally, the real-time



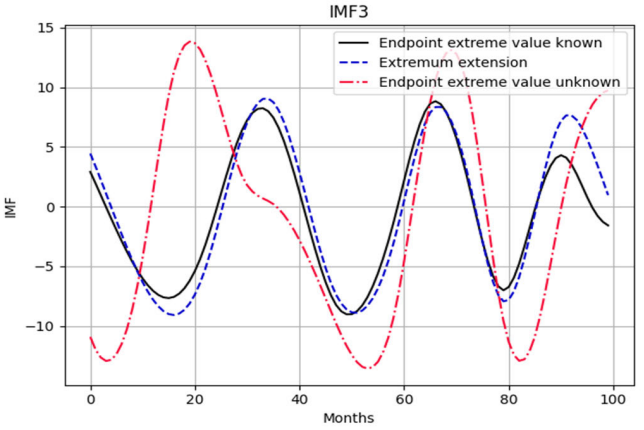
(a) Performance of extreme value extension in IMF0



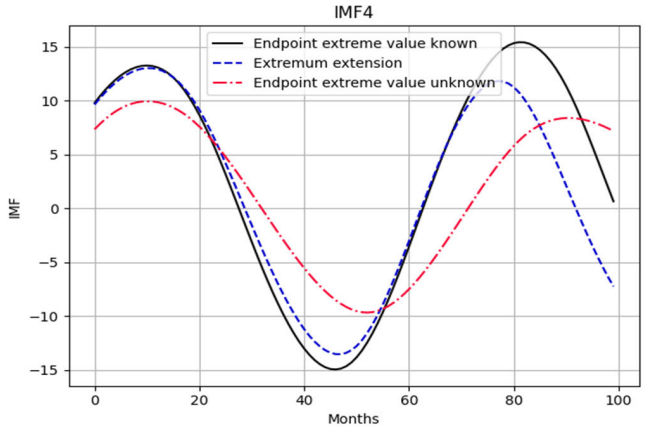
(b) Performance of extreme value extension in IMF1



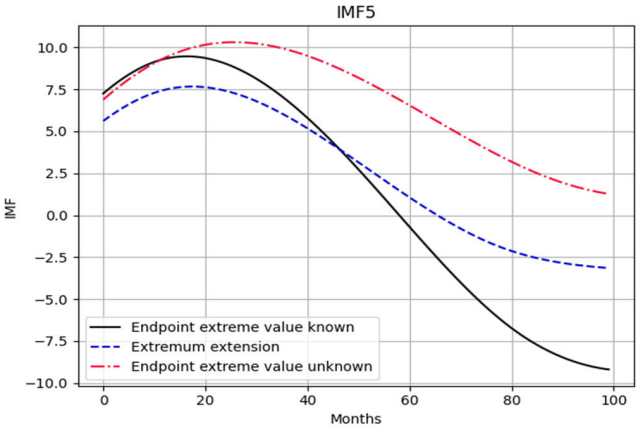
(c) Performance of extreme value extension in IMF2



(d) Performance of extreme value extension in IMF3



(e) Performance of extreme value extension in IMF4

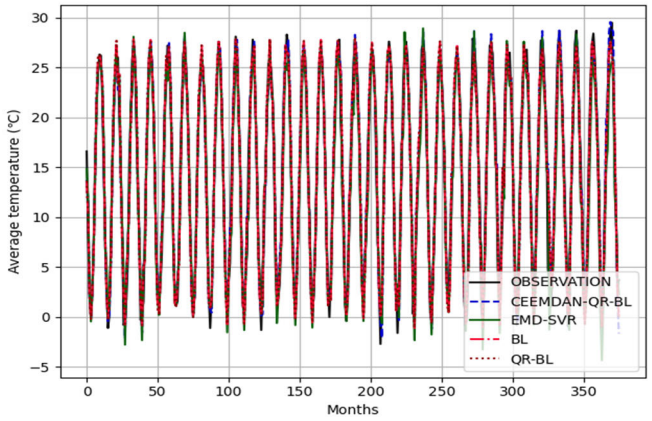


(f) Performance of extreme value extension in IMF5

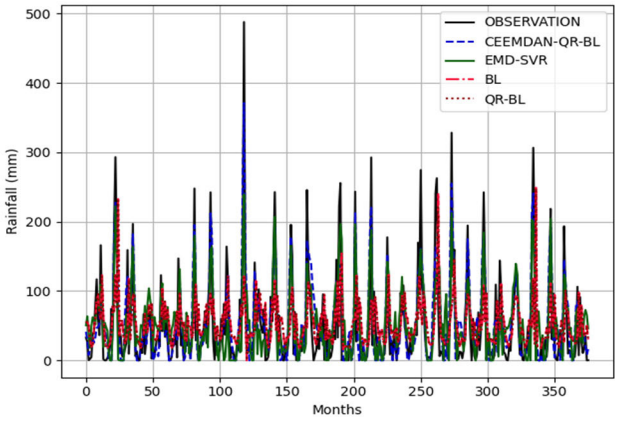
FIGURE 5. Performance of extreme value extension in different IMFs.

errors of CEEMDAN-QR-BL were lower than those of BL, EMD-SVR, and QR-BL, excluding individual points. Overall, CEEMDAN-QR-BL exhibited a higher stability and accuracy than those of the BL, EMD-SVR, and QR-BL models.

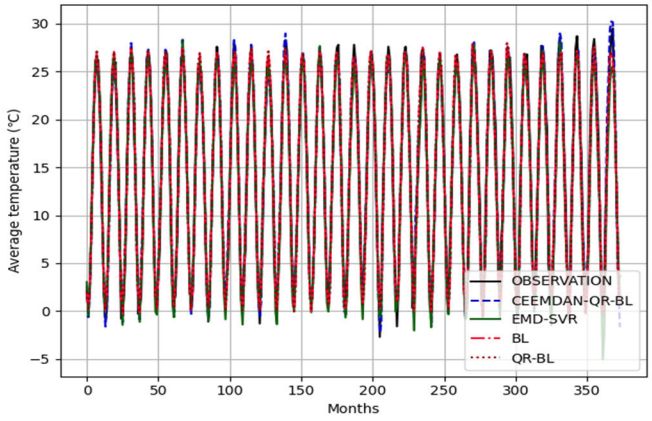
Figure 8 and Table 2 show the gap between the Ens and MAE of the BL, EMD-SVR, and CEEMDAN-QR-BL models. Table 3 shows the gap between the runtime of BL, QR-BL, Serial CEEMDAN-QR-BL, and Parallel



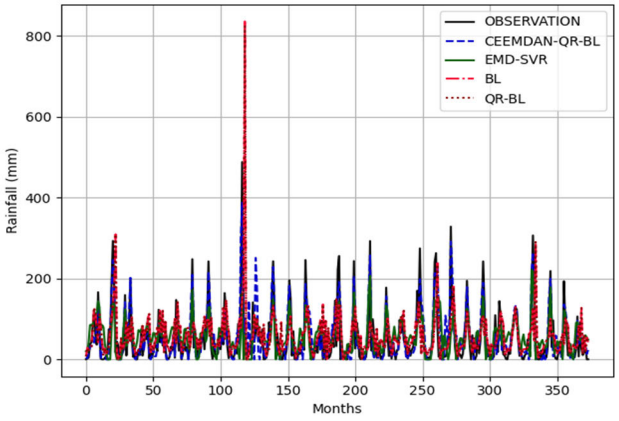
(a) Forecasted average temperature for the next two months



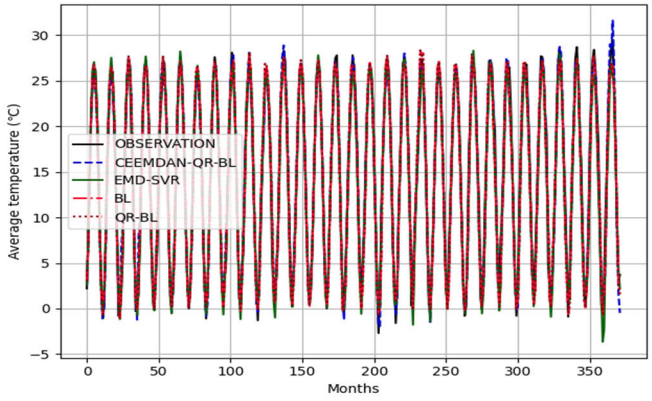
(b) Forecasted rainfall for the next two months



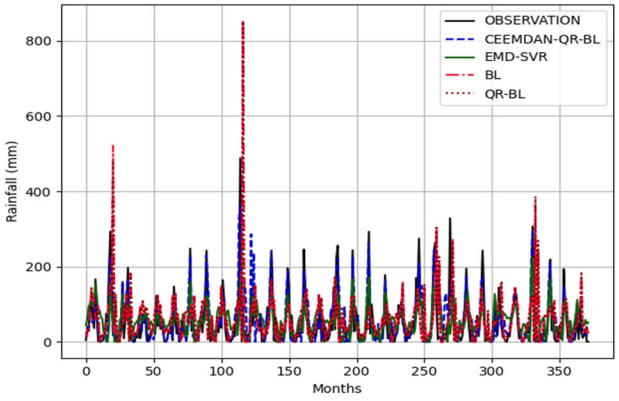
(c) Forecasted average temperature for the next four months



(d) Forecasted rainfall for the next four months



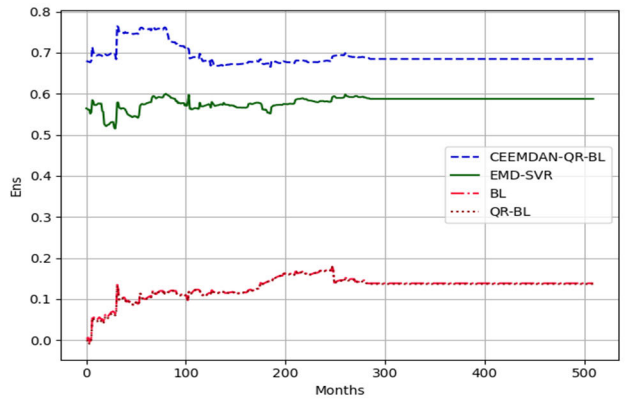
(e) Forecasted average temperature for the next six months



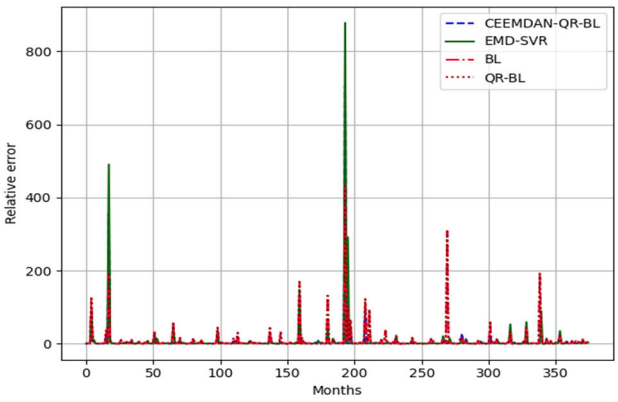
(f) Forecasted rainfall for the next six months

FIGURE 6. Forecast curves of the drought indices for different periods by the intelligent computing model.

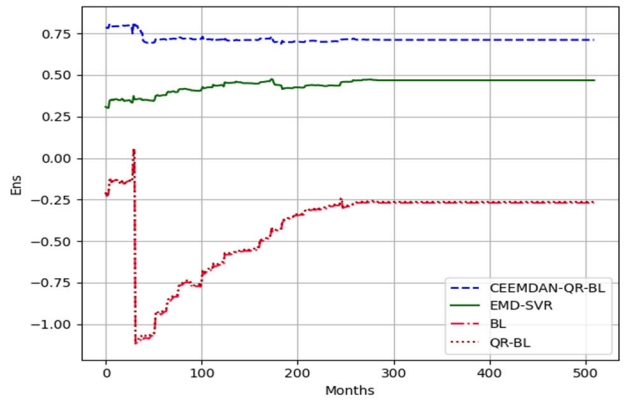
CEEMDAN-QR-BL models. The number of prediction days greatly influenced the model error; however, the error level of the CEEMDAN hybrid model was lower than that of the BL and EMD-SVR models. The runtime of the algorithm can indicate the efficiency of the algorithm. When compared with only BL, the prediction efficiency of QR-BL improved by 62.29%. The original BL model had a calculation prediction time of 1.22 s, while that of parallel QR-BL was 0.98 s.



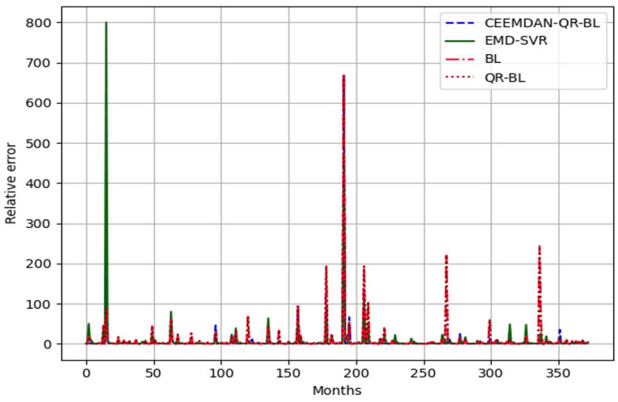
(a) Ens of the forecast for the next two months



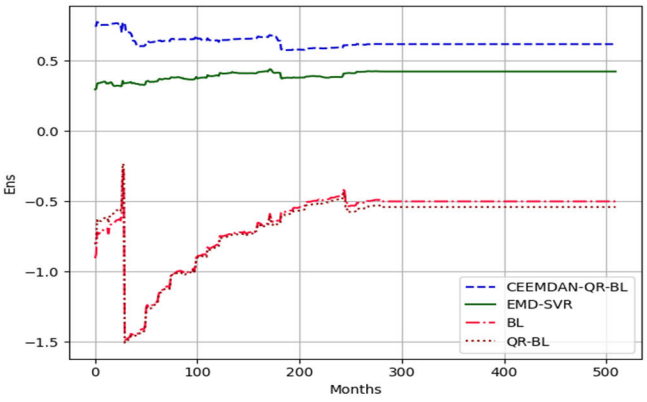
(b) Error in the forecast for the next two months



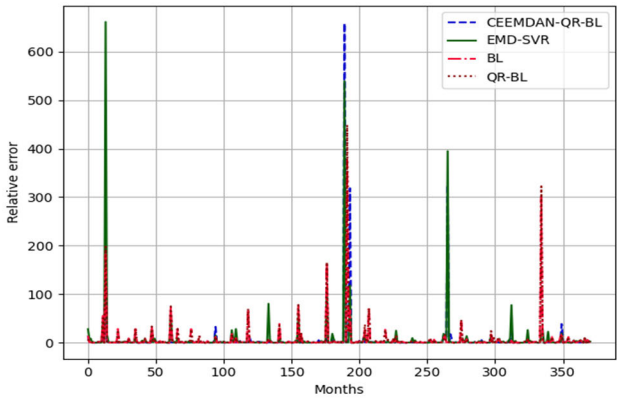
(c) Ens of the forecast for the next four months



(d) Error in the forecast for the next four months



(e) Ens of the forecast for the next six months



(f) Error in the forecast error for the next six months

FIGURE 7. Real-time indicator curve of the intelligent calculation model under different forecast times.

The above evaluation indicates that the CEEMDAN-QR-BL model exhibited a better accuracy and efficiency than those of the BL model.

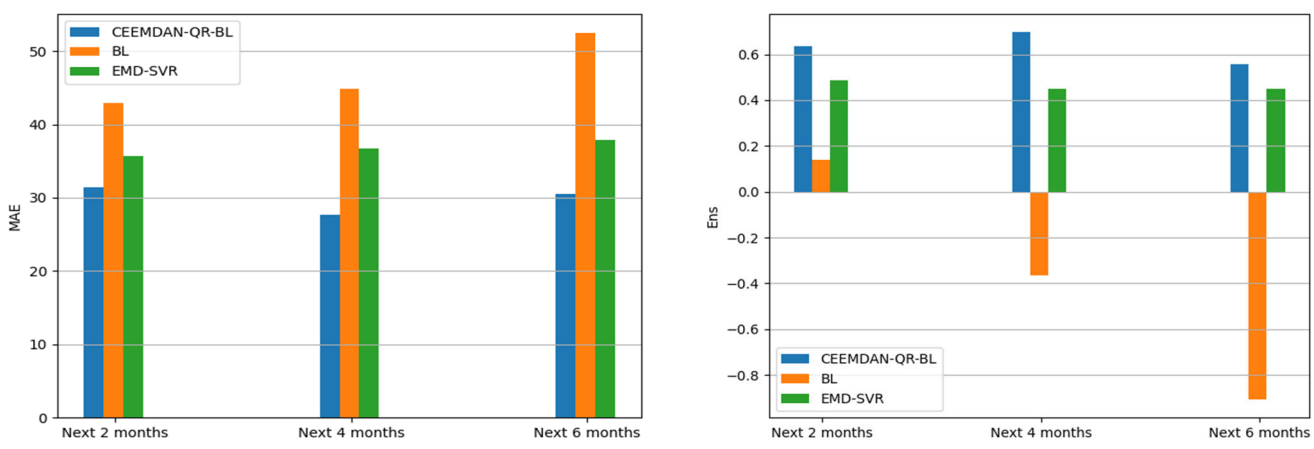
Figure 9 compares the future predictions and calculated risk levels of the different models. The blue, orange, green, and red bars represent the actual situation, CEEMDAN-QR-BL, BL, and EMD-SVR fitting results, respectively. The results indicate that the frequency of severe and moderate droughts in the city has remained the same, but it is much lower than the frequency of non-drought events. The CEEMDAN-QR-BL model exhibited higher accuracy and stability than those of the BL and EMD-SVR algorithms.

TABLE 2. Comparison of the numerical results of various evaluation indicators.

Foresight period	Method	Ens	MAE
Next two months	CEEMDAN-QR-BL	0.634964	31.413031
	BL	0.138337	42.876178
	EMD-SVR	0.488432	35.625005
Next four months	CEEMDAN-QR-BL	0.695893	27.665982
	BL	-0.366530	44.898675
	EMD-SVR	0.448242	36.734932
Next six months	CEEMDAN-QR-BL	0.557349	30.482894
	BL	-0.905956	52.432619
	EMD-SVR	0.450908	37.922336

TABLE 3. Comparison of computing efficiency of different models.

Model	Data pre-processing time (s)	Calculation and prediction time (s)
BL	0	1.22
QR-BL	0	0.98
Serial CEEMDAN-QR-BL	28.12	8.73
Parallel CEEMDAN-QR-BL	28.12	0.017

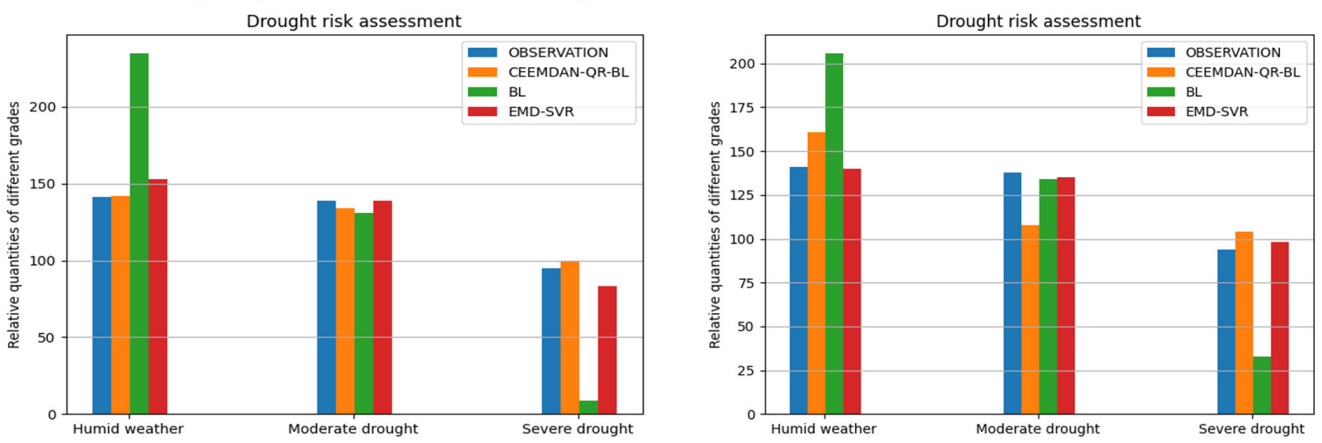


(a) MAE evaluation chart

(b) Ens evaluation chart

FIGURE 8. Evaluation of the performance of the intelligent computing model.

Figure 10 reflects the proportions of drought levels in the city from January to December 2009–2014, and indicates that few droughts occurred during the summer, whereas severe droughts occurred in the autumn and winter and moderate droughts frequently occurred in the spring.



(a) Comparison of the prediction levels of different algorithms for the next two months

(b) Comparison of the prediction levels of different algorithms for the next four months

FIGURE 9. Comparison of the risk levels of the intelligent calculation models at different forecast times.

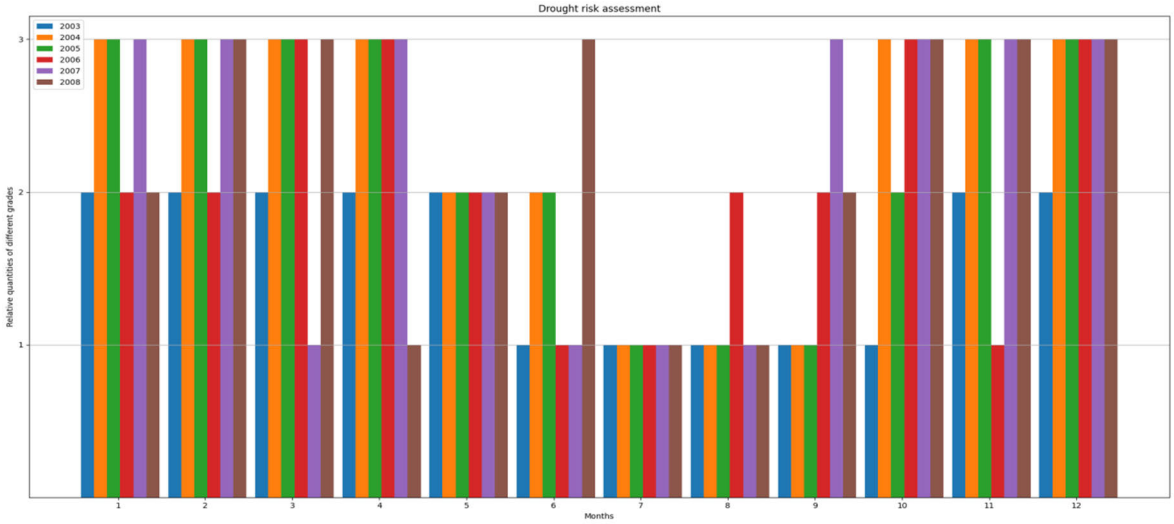


FIGURE 10. Changes in the drought levels in the city from 2009 to 2014.

V. CONCLUSION

Effective drought control can contribute to resolving a country's livelihood problems. In this study, weather time series data were used to predict droughts in a city and drought risk was determined by data analysis, mining, and other related technologies. The CEEMDAN-QR-BL model was found to be superior to the original BL model, as indicated by its higher accuracy (MAE), efficiency, and stability. However, both the introduced decomposition and extremum extension processes caused the model efficiency to decrease. Although the introduced parallelism can solve the problem to a certain extent, it still cannot fundamentally solve the consumption caused by CEEMDAN decomposition and endpoint extension.

REFERENCES

[1] Z. Hao, V. P. Singh, and Y. Xia, "Seasonal drought prediction: Advances, challenges, and future prospects," *Rev. Geophys.*, vol. 56, no. 1, pp. 108–141, Mar. 2018.

[2] Z. Hao, F. Hao, and V. P. Singh, "A general framework for multivariate multi-index drought prediction based on multivariate ensemble streamflow prediction (MESP)," *J. Hydrol.*, vol. 539, pp. 1–10, Aug. 2016.

[3] C. Gao, H. Chen, S. Sun, V. Ongoma, W. Hua, H. Ma, B. Xu, and Y. Li, "A potential predictor of multi-season droughts in Southwest China: Soil moisture and its memory," *Natural Hazards*, vol. 91, no. 2, pp. 553–566, Mar. 2018.

[4] Z. Li, T. Chen, Q. Wu, G. Xia, and D. Chi, "Application of penalized linear regression and ensemble methods for drought forecasting in Northeast China," *Meteorol. Atmos. Phys.*, vol. 132, no. 1, pp. 113–130, Feb. 2020.

[5] H. Park, K. Kim, and D. K. Lee, "Prediction of severe drought area based on random forest: Using satellite image and topography data," *Water*, vol. 11, no. 4, p. 705, Apr. 2019.

[6] L. Xu, N. Chen, and X. Zhang, "A comparison of large-scale climate signals and the North American multi-model ensemble (NMME) for drought prediction in China," *J. Hydrol.*, vol. 557, pp. 378–390, Feb. 2018.

[7] S. Poornima and M. Pushpalatha, "Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network," *Soft Comput.*, vol. 23, no. 18, pp. 8399–8412, Sep. 2019.

[8] A. Kaur and S. K. Sood, "Cloud-fog based framework for drought prediction and forecasting using artificial neural network and genetic algorithm," *J. Exp. Theor. Artif. Intell.*, vol. 32, no. 2, pp. 273–289, Mar. 2020.

[9] N. Nabipour, M. Dehghani, A. Mosavi, and S. Shamshirband, "Short-term hydrological drought forecasting based on different nature-inspired optimization algorithms hybridized with artificial neural networks," *IEEE Access*, vol. 8, pp. 15210–15222, 2020.
[10] K. F. Fung, Y. F. Huang, and C. H. Koo, "Coupling fuzzy–SVR and boosting–SVR models with wavelet decomposition for meteorological drought prediction," *Environ. Earth Sci.*, vol. 78, no. 24, p. 693, Dec. 2019.
[11] V. Bari, B. Cairo, B. De Maria, D. Tonon, G. Rossato, L. Faes, and A. Porta, "An empirical mode decomposition approach to assess the strength of heart period-systolic arterial pressure variability interactions," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 2573–2576.
[12] L. Wang and Y. Shao, "Fault feature extraction of rotating machinery using a reweighted complete ensemble empirical mode decomposition with adaptive noise and demodulation analysis," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106545.
[13] Y. Yu, D. Chi, T. Chen, M. Sun, and C. Luan, "Drought forecast application of BP prediction models based on EMD in Ling river basin," *J. Shenyang Agricult. Univ.*, vol. 45, no. 1, pp. 68–72, 2014.
[14] N. Agana and A. Homaifar, "EMD-based predictive deep belief network for time series prediction: An application to drought forecasting," *Hydrology*, vol. 5, no. 1, p. 18, Feb. 2018.
[15] E. Meng, S. Huang, Q. Huang, W. Fang, L. Wu, and L. Wang, "A robust method for non-stationary streamflow prediction based on improved EMD-SVM model," *J. Hydrol.*, vol. 568, pp. 462–478, Jan. 2019.
[16] A. Guo and Y. Wang, "A hybrid model for ET (evapotranspiration) forecasting based on EEMD for ET-based water resources management," in *Proc. AGUFM*, 2015, pp. H33D–1642.
[17] T. Liang, G. Xie, S. Fan, and Z. Meng, "A combined model based on CEEMDAN, permutation entropy, gated recurrent unit network, and an improved bat algorithm for wind speed forecasting," *IEEE Access*, vol. 8, pp. 165612–165630, 2020.
[18] S. Feng and C. L. P. Chen, "Fuzzy broad learning system: A novel neuro-fuzzy model for regression and classification," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 414–424, Feb. 2020.
[19] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.
[20] Z. Liu and C. L. P. Chen, "Broad learning system: Structural extensions on single-layer and multi-layer neural networks," in *Proc. Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Dec. 2017, pp. 136–141.
[21] M. Han, S. Feng, C. L. P. Chen, M. Xu, and T. Qiu, "Structured manifold broad learning system: A manifold perspective for large-scale chaotic time series analysis and prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 9, pp. 1809–1821, Sep. 2019.
[22] J. Jin, Z. Liu, and C. L. P. Chen, "Discriminative graph regularized broad learning system for image recognition," *Sci. China Inf. Sci.*, vol. 61, no. 11, Nov. 2018, Art. no. 112209.
[23] H. Zhao, J. Zheng, J. Xu, and W. Deng, "Fault diagnosis method based on principal component analysis and broad learning system," *IEEE Access*, vol. 7, pp. 99263–99272, 2019.
[24] Y. Kong, X. Wang, Y. Cheng, and C. Chen, "Hyperspectral imagery classification based on semi-supervised broad learning system," *Remote Sens.*, vol. 10, no. 5, p. 685, Apr. 2018.
[25] M. Xu, M. Han, C. L. P. Chen, and T. Qiu, "Recurrent broad learning systems for time series prediction," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1405–1417, Apr. 2020.
[26] S. Sui, C. L. P. Chen, S. Tong, and S. Feng, "Finite-time adaptive quantized control of stochastic nonlinear systems with input quantization: A broad learning system based identification method," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8555–8565, Oct. 2020.
[27] J. Lin, Z. Liu, C. L. P. Chen, and Y. Zhang, "Quaternion broad learning system: A novel multi-dimensional filter for estimation and elimination tremor in teleoperation," *Neurocomputing*, vol. 380, pp. 78–86, Mar. 2020.
[28] H. Lei, Y. Wen, Z. You, A. Elazab, E.-L. Tan, Y. Zhao, and B. Lei, "Protein–protein interactions prediction via multimodal deep polynomial network and regularized extreme learning machine," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1290–1303, May 2019.
[29] X. Liu, T. Qiu, C. Chen, H. Ning, and N. Chen, "An incremental broad learning approach for semi-supervised classification," in *Proc. IEEE Int. Conf. Dependable, Autonomic Secure Comput., Int. Conf Pervasive Intell. Comput., Int. Conf Cloud Big Data Comput., Int. Conf Cyber Sci. Technol. Congr. (DASC/PiCom/CBDCom/CyberSciTech)*, Aug. 2019, pp. 250–254.
[30] J. N. Franklin, *Matrix Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1968.
[31] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, vol. 14. Mathematics of Computation, 1992.
[32] R. A. Brualdi, and H. J. Ryser, *Combinatorial Matrix Theory*, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[33] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
[34] P. O. Junior, A. K. Tiwari, H. Padhan, and I. Alagidede, "Analysis of EEMD-based quantile-in-quantile approach on spot-futures prices of energy and precious metals in India," *Resour. Policy*, vol. 68, Oct. 2020, Art. no. 101731.
[35] Y. Li, S. Gao, S. Zhang, H. He, P. Xian, and C. Yuan, "The baseline wander correction based on the improved ensemble empirical mode decomposition (EEMD) algorithm for grounded electrical source airborne transient electromagnetic signals," *Geosci. Instrum., Methods Data Syst.*, vol. 9, no. 2, pp. 443–450, Nov. 2020.
[36] Y.-S. Kao, K. Nawata, and C.-Y. Huang, "Predicting primary energy consumption using hybrid ARIMA and GA-SVR based on EEMD decomposition," *Mathematics*, vol. 8, no. 10, p. 1722, Oct. 2020.
[37] P. E. Suryani, "Perbandingan metode eemd dan emd untuk mereduksi noise pada sinyal seismik," *Jurnal Ilmiah Teknosains*, vol. 6, no. 1, pp. 20–24, 2020.
[38] T. Li, Z. Qian, and T. He, "Short-term load forecasting with improved CEEMDAN and GWO-based multiple kernel ELM," *Complexity*, vol. 2020, pp. 1–20, Feb. 2020.
[39] H. M. Nazir, I. Hussain, M. Faisal, A. M. Shoukry, M. A. W. Sharkawy, F. F. Al-Deek, and M. Ismail, "Dependence structure analysis of multisite river inflow data using vine copula-CEEMDAN based hybrid model," *PeerJ*, vol. 8, Nov. 2020, Art. no. e10285.
[40] M. F. Bari and S. A. Fattah, "Epileptic seizure detection in EEG signals using normalized IMFs in CEEMDAN domain and quadratic discriminant classifier," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101833.
[41] Y.-P. Zhao, Z.-Q. Li, P.-P. Xi, D. Liang, L. Sun, and T.-H. Chen, "Gram–Schmidt process based incremental extreme learning machine," *Neurocomputing*, vol. 241, pp. 1–17, Jun. 2017.

YANG LIU was born in Kaifeng, Henan, China, in 1979. She received the B.S. degree in computer application technology from Shanxi University, Taiyuan, in 2002, the M.S. degree in computer application technology from the University of Electronic Science and Technology of China, Chengdu, in 2005, and the Ph.D. degree in information and signal processing from the Institute of Optoelectronics Technology, Chinese Academy of Sciences, Chengdu, in 2012.

From 2002 to 2005, she has worked with the Tenth Research Institute of China Electronics Technology Group and became a Software Engineer. Since 2012, she has been working with the North China University of Water Resources and Hydropower. In 2015, she became an Associate Professor and a Postgraduate Tutor. Her research interests include intelligent water conservancy, machine learning, and mathematical sequence analysis. She has published more than 30 articles, including more than 20 articles included by SCI and EI, and presided over more than 20 scientific research projects.

LIHU WANG received the B.S. degree in computer science from Xuchang University, in 2019. He is currently pursuing the M.S. degree in computer science with the North China University of Water Resources and Electric Power. His research interests include water conservancy big data, intelligent cloud computing, machine learning, and deep learning frontier knowledge and applications. His awards and honors include the North China University of Water Resources and Electric Power Scholarship and the Internet Plus Innovation and Entrepreneurship Competition Contest Provincial Third Prize.