

## Research Article

# A CEEMDAN and XGBOOST-Based Approach to Forecast Crude Oil Prices

Yingrui Zhou,<sup>1</sup> Taiyong Li ,<sup>1,2</sup> Jiayi Shi,<sup>1</sup> and Zijie Qian<sup>1</sup>

<sup>1</sup>School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China

<sup>2</sup>Sichuan Province Key Laboratory of Financial Intelligence and Financial Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China

Correspondence should be addressed to Taiyong Li; [litaiyong@gmail.com](mailto:litaiyong@gmail.com)

Received 13 October 2018; Accepted 13 January 2019; Published 3 February 2019

Guest Editor: Marisol B. Correia

Copyright © 2019 Yingrui Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Crude oil is one of the most important types of energy for the global economy, and hence it is very attractive to understand the movement of crude oil prices. However, the sequences of crude oil prices usually show some characteristics of nonstationarity and nonlinearity, making it very challenging for accurate forecasting crude oil prices. To cope with this issue, in this paper, we propose a novel approach that integrates complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and extreme gradient boosting (XGBOOST), so-called CEEMDAN-XGBOOST, for forecasting crude oil prices. Firstly, we use CEEMDAN to decompose the nonstationary and nonlinear sequences of crude oil prices into several intrinsic mode functions (IMFs) and one residue. Secondly, XGBOOST is used to predict each IMF and the residue individually. Finally, the corresponding prediction results of each IMF and the residue are aggregated as the final forecasting results. To demonstrate the performance of the proposed approach, we conduct extensive experiments on the West Texas Intermediate (WTI) crude oil prices. The experimental results show that the proposed CEEMDAN-XGBOOST outperforms some state-of-the-art models in terms of several evaluation metrics.

## 1. Introduction

As one of the most important types of energy that power the global economy, crude oil has great impacts on every country, every enterprise, and even every person. Therefore, it is a crucial task for the governors, investors, and researchers to forecast the crude oil prices accurately. However, the existing research has shown that crude oil prices are affected by many factors, such as supply and demand, interest rate, exchange rate, speculation activities, international and political events, climate, and so on [1, 2]. Therefore, the movement of crude oil prices is irregular. For example, starting from about 11 USD/barrel in December 1998, the WTI crude oil prices gradually reached the peak of 145.31 USD/barrel in July 2008, and then the prices drastically declined to 30.28 USD/barrel in the next five months because of the subprime mortgage crisis. After that, the prices climbed to more than 113 USD/barrel in April 2011, and, once again, they sharply dropped to about 26 USD/barrel in February 2016. The movement of the crude

oil prices in the last decades has shown that the forecasting task is very challenging, due to the characteristics of high nonlinearity and nonstationarity of crude oil prices.

Many scholars have devoted efforts to trying to forecast crude oil prices accurately. The most widely used approaches to forecasting crude oil prices can be roughly divided into two groups: statistical approaches and artificial intelligence (AI) approaches. Recently, Miao et al. have explored the factors of affecting crude oil prices based on the least absolute shrinkage and selection operator (LASSO) model [1]. Ye et al. proposed an approach integrating ratchet effect for linear prediction of crude oil prices [3]. Morana put forward a semiparametric generalized autoregressive conditional heteroskedasticity (GARCH) model to predict crude oil prices at different lag periods even without the conditional average of historical crude oil prices [4]. Naser found that using the dynamic model averaging (DMA) with empirical evidence is better than linear models such as autoregressive (AR) model and its variants [5]. Gong and Lin proposed several new

heterogeneous autoregressive (HAR) models to forecast the good and bad uncertainties in crude oil prices [6]. Wen et al. also used HAR models with structural breaks to forecast the volatility of crude oil futures [7].

Although the statistical approaches improve the accuracy of forecasting crude oil prices to some extent, the assumption of linearity of crude oil prices cannot be met according to some recent research, and hence it limits the accuracy. Therefore, a variety of AI approaches have been proposed to capture the nonlinearity and nonstationarity of crude oil prices in the last decades [8–11]. Chiroma et al. reviewed the existing research associated with forecasting crude oil prices and found that AI methodologies are attracting unprecedented interest from scholars in the domain of crude oil price forecasting [8]. Wang et al. proposed an AI system framework that integrated artificial neural networks (ANN) and rule-based expert system with text mining to forecast crude oil prices, and it was shown that the proposed approach was significantly effective and practically feasible [9]. Barunik and Malinska used neural networks to forecast the term structure in crude oil futures prices [10]. Most recently, Chen et al. have studied forecasting crude oil prices using deep learning framework and have found that the random walk deep belief networks (RW-DBN) model outperforms the long short term memory (LSTM) and the random walk LSTM (RW-LSTM) models in terms of forecasting accuracy [11]. Other AI-methodologies, such as genetic algorithm [12], compressive sensing [13], least square support vector regression (LSSVR) [14], and cluster support vector machine (ClusterSVM) [15], were also applied to forecasting crude oil prices. Due to the extreme nonlinearity and nonstationarity, it is hard to achieve satisfactory results by forecasting the original time series directly. An ideal approach is to divide the tough task of forecasting original time series into several subtasks, and each of them forecasts a relatively simpler subsequence. And then the results of all subtasks are accumulated as the final result. Based on this idea, a “decomposition and ensemble” framework was proposed and widely applied to the analysis of time series, such as energy forecasting [16, 17], fault diagnosis [18–20], and biosignal analysis [21–23]. This framework consists of three stages. In the first stage, the original time series was decomposed into several components. Typical decomposition methodologies include wavelet decomposition (WD), independent component analysis (ICA) [24], variational mode decomposition (VMD) [25], empirical mode decomposition (EMD) [2, 26] and its extension (ensemble EMD (EEMD)) [27, 28], and complementary EEMD (CEEMD) [29]. In the second stage, some statistical or AI-based methodologies are applied to forecast each decomposed component individually. In theory, any regression methods can be used to forecast the results of each component. In the last stage, the predicted results from all components are aggregated as the final results. Recently, various researchers have devoted efforts to forecasting crude oil prices following the framework of “decomposition and ensemble.” Fan et al. put forward a novel approach that integrates independent components analysis (ICA) and support vector regression (SVR) to forecast crude oil prices, and the experimental results verified the

effectiveness of the proposed approach [24]. Yu et al. used EMD to decompose the sequences of the crude oil prices into several intrinsic mode functions (IMFs) at first and then used a three-layer feed-forward neural network (FNN) model for predicting each IMF. Finally, the authors used an adaptive linear neural network (ALNN) to combine all the results of the IMFS as the final forecasting output [2]. Yu et al. also used EEMD and extended extreme learning machine (EELM) to forecast crude oil prices, following the framework of “decomposition and ensemble.” The empirical results demonstrated the effectiveness and efficiency of the proposed approach [28]. Tang et al. further proposed an improved approach integrating CEEMD and EELM for forecasting crude oil prices, and the experimental results demonstrated that the proposed approach outperformed all the listed state-of-the-art benchmarks [29]. Li et al. used EEMD to decompose raw crude oil prices into several components and then used kernel and nonkernel sparse Bayesian learning (SBL) to forecast each component, respectively [30, 31].

From the perspective of decomposition, although EMD and EEMD are capable of improving the accuracy of forecasting crude oil prices, they still suffer “mode mixing” and introducing new noise in the reconstructed signals, respectively. To overcome these shortcomings, an extension of EEMD, so-called complete EEMD with adaptive noise (CEEMDAN), was proposed by Torres et al. [32]. Later, the authors put forward an improved version of CEEMDAN to obtain decomposed components with less noise and more physical meaning [33]. The CEEMDAN has succeeded in wind speed forecasting [34], electricity load forecasting [35], and fault diagnosis [36–38]. Therefore, CEEMDAN may have the potential to forecast crude oil prices. As pointed out above, any regression methods can be used to forecast each decomposed component. A recently proposed machine learning algorithm, extreme gradient boosting (XGBOOST), can be used for both classification and regression [39]. The existing research has demonstrated the advantages of XGBOOST in forecasting time series [40–42].

With the potential of CEEMDAN in decomposition and XGBOOST in regression, in this paper, we aim at proposing a novel approach that integrates CEEMDAN and XGBOOST, namely, CEEMDAN-XGBOOST, to improve the accuracy of forecasting crude oil prices, following the “decomposition and ensemble” framework. Specifically, we firstly decompose the raw crude oil price series into several components with CEEMDAN. And then, for each component, XGBOOST is applied to building a specific model to forecast the component. Finally, all the predicted results from every component are aggregated as the final forecasting results. The main contributions of this paper are threefold: (1) We propose a novel approach, so-called CEEMDAN-XGBOOST, for forecasting crude oil prices, following the “decomposition and ensemble” framework; (2) extensive experiments are conducted on the publicly-accessed West Texas Intermediate (WTI) to demonstrate the effectiveness of the proposed approach in terms of several evaluation metrics; (3) we further study the impacts of several parameter settings with the proposed approach.

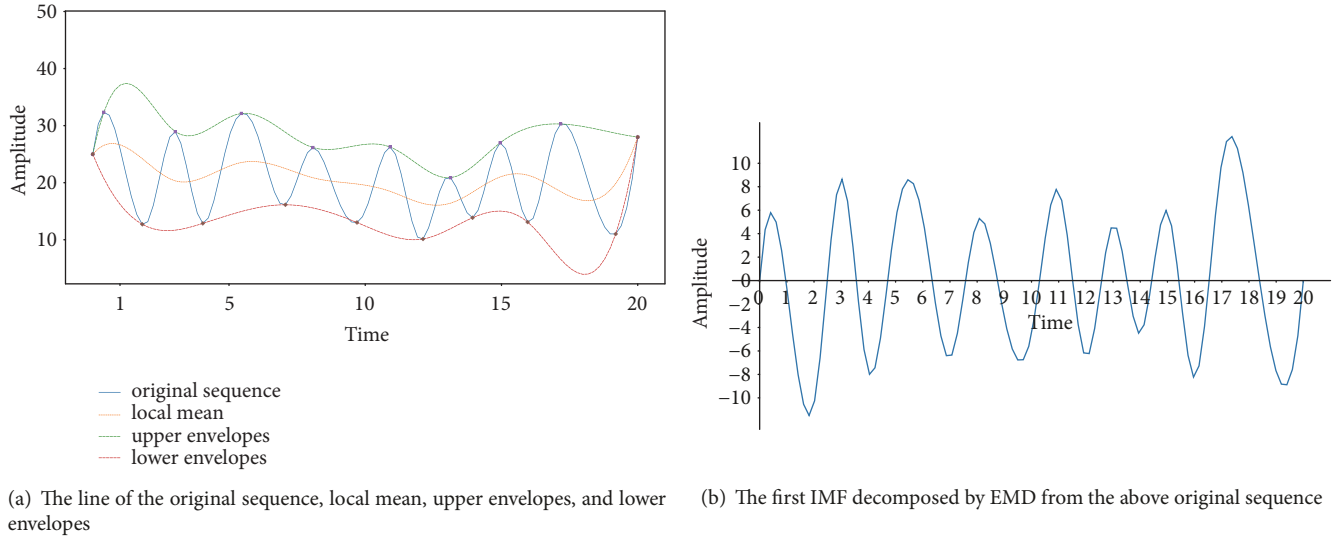


FIGURE 1: An illustration of EMD.

The remainder of this paper is organized as follows. Section 2 describes CEEMDAN and XGBOOST. Section 3 formulates the proposed CEEMDAN-XGBOOST approach in detail. Experimental results are reported and analyzed in Section 4. We also discussed the impacts of parameter settings in this section. Finally, Section 5 concludes this paper.

## 2. Preliminaries

**2.1. EMD, EEMD and CEEMDAN.** EMD was proposed by Huang et al. in 1998, and it has been developed and applied in many disciplines of science and engineering [26]. The key feature of EMD is to decompose a nonlinear, nonstationary sequence into intrinsic mode functions (IMFs) in the spirit of the Fourier series. In contrast to the Fourier series, they are not simply sine or cosine functions, but rather functions that represent the characteristics of the local oscillation frequency of the original data. These IMFs need to satisfy two conditions: (1) the number of local extrema and the number of zero crossing must be equal or differ at most by one and (2) the curve of the “local mean” is defined as zero.

At first, EMD finds out the upper and the lower envelopes which are computed by finding the local extrema of the original sequence. Then, the local maxima (minima) are linked by two cubic spines to construct the upper (lower) envelopes, respectively. The mean of these envelopes is considered as the “local mean.” Meanwhile, the curve of this “local mean” is defined as the first residue, and the difference between original sequence and the “local mean” is defined as the first IMF. An illustration of EMD is shown in Figure 1.

After the first IMF is decomposed by EMD, there is still a residue (the local mean, i.e., the yellow dot line in Figure 1(a)) between the IMF and the original data. Obviously, extrema and high-frequency oscillations also exist in the residue. And EMD decomposes the residue into another IMF and one residue. If the variance of the new residue is not small enough to satisfy the Cauchy criterion, EMD will repeat to

decompose new residue into another IMF and a new residue. Finally, EMD decomposes original sequence into several IMFs and one residue. The difference between the IMF and the residues is defined as

$$r_k[t] = r_{k-1}[t] - IMF_k[t], \quad k = 2, \dots, K, \quad (1)$$

where  $r_k[t]$  is the  $k$ -th residue at the time  $t$  and  $K$  is the total number of IMFs and residues.

Subsequently, Huang et al. thought that EMD could not extract the local features from the mixed features of the original sequence completely. One of the reasons for this is the frequent appearance of the mode mixing. The mode mixing can be defined as the situation that similar pieces of oscillations exist at the same corresponding position in different IMFs, which causes that a single IMF has lost its physical meanings. What is more, if one IMF has this problem, the following IMFs cannot avoid it either. To solve this problem, Wu and Huang extended EMD to a new version, namely, EEMD, that adds white noise to the original time series and performs EMD many times [27]. Given a time series and corresponding noise, the new time series can be expressed as

$$x^i[t] = x[t] + w^i[t], \quad (2)$$

where  $x[t]$  stands for the original data and  $w^i[t]$  is the  $i$ -th white noise ( $i=1,2,\dots,N$ , and  $N$  is the times of performing EMD).

Then, EEMD decomposes every  $x^i[t]$  into  $IMF_k^i[t]$ . In order to get the real  $k$ -th IMF,  $\overline{IMF}_k$ , EEMD calculates the average of the  $IMF_k^i[t]$ . In theory, because the mean of the white noise is zero, the effect of the white noise would be eliminated by computing the average of  $IMF_k^i[t]$ , as shown in

$$\overline{IMF}_k = \frac{1}{N} \sum_{i=1}^N IMF_k^i[t]. \quad (3)$$

However, Torres et al. found that, due to the limited number of  $x^i[t]$  in empirical research, EEMD could not completely eliminate the influence of white noise in the end. For this situation, Torres et al. put forward a new decomposition technology, CEEMDAN, on the basis of EEMD [32].

CEEMDAN decomposes the original sequence into the first IMF and residue, which is the same as EMD. Then, CEEMDAN gets the second IMF and residue, as shown in

$$\overline{IMF}_2 = \frac{1}{N} \sum_{i=1}^N E_1(r_1[t] + \varepsilon_1 E_1(w^i[t])), \quad (4)$$

$$r_2[t] = r_1[t] - \overline{IMF}_2, \quad (5)$$

where  $E_1(\cdot)$  stands for the first IMF decomposed from the sequence and  $\varepsilon_i$  is used to set the signal-to-noise ratio (SNR) at each stage.

In the same way, the  $k$ -th IMF and residue can be calculated as

$$\overline{IMF}_k = \frac{1}{N} \sum_{i=1}^N E_1(r_{k-1}[t] + \varepsilon_{k-1} E_{k-1}(w^i[t])), \quad (6)$$

$$r_k[t] = r_{k-1}[t] - \overline{IMF}_k, \quad (7)$$

Finally, CEEMDAN gets several IMFs and computes the residue, as shown in

$$R[t] = x[t] - \sum_{k=1}^K \overline{IMF}_k. \quad (8)$$

The sequences decomposed by EMD, EEMD, and CEEMDAN satisfy (8). Although CEEMDAN can solve the problems that EEMD leaves, it still has two limitations: (1) the residual noise that the models contain and (2) the existence of spurious modes. Aiming at dealing with these issues, Torres et al. proposed a new algorithm to improve CEEMDAN [33].

Compared with the original CEEMDAN, the improved version obtains the residues by calculating the local means. For example, in order to get the first residue shown in (9), it would compute the local means of  $N$  realizations  $x^i[t] = x[t] + \varepsilon_0 E_1(w^i[t]) (i=1, 2, \dots, N)$ .

$$r_1[t] = \frac{1}{N} \sum_{i=1}^N M(x^i[t]), \quad (9)$$

where  $M(\cdot)$  is the local mean of the sequence.

Then, it can get the first IMF shown in

$$IMF_1 = x[t] - r_1[t]. \quad (10)$$

For the  $k$ -th residue and IMF, they can be computed as (11) and (12), respectively:

$$r_k[t] = \frac{1}{N} \sum_{i=1}^N M(r_{k-1}[t] + \varepsilon_{k-1} E_k(w^i[t])), \quad (11)$$

$$IMF_k = r_{k-1}[t] - r_k[t]. \quad (12)$$

The authors have demonstrated that the improved CEEMDAN outperformed the original CEEMDAN in signal decomposition [33]. In what follows, we will refer to the improved version of CEEMDAN as CEEMDAN, unless otherwise stated. With CEEMDAN, the original sequence can be decomposed into several IMFs and one residue, that is, the tough task of forecasting the raw time series, can be divided into forecasting several simpler subtasks.

2.2. XGBOOST. Boosting is the ensemble method that can combine several weak learners into a strong learner as

$$\widehat{y}_i = \theta(x_i) = \sum_{k=1}^K f_k(x_i), \quad (13)$$

where  $f_k(\cdot)$  is a weak learner and  $K$  is the number of weak learners.

When it comes to the tree boosting, its learners are decision trees which can be used for both regression and classification.

To a certain degree, XGBOOST is considered as tree boost, and its core is the Newton boosting instead of Gradient Boosting, which finds the optimal parameters by minimizing the loss function ( $\theta$ ), as shown in

$$L(\theta) = \sum_{i=1}^n l(\widehat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k), \quad (14)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \alpha \|\omega\|^2, \quad (15)$$

where  $\Omega(f_k)$  is the complexity of the  $k$ -th tree model,  $n$  is the sample size,  $T$  is the number of leaf nodes of the decision trees,  $\omega$  is the weight of the leaf nodes,  $\gamma$  controls the extent of complexity penalty for tree structure on  $T$ , and  $\alpha$  controls the degree of the regularization of  $f_k$ .

Since it is difficult for the tree ensemble model to minimize loss function in (14) and (15) with traditional methods in Euclidean space, the model uses the additive manner [43]. It adds  $f_t$  that improves the model and forms the new loss function as

$$L^t = \sum_{i=1}^n l(y_i, \widehat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t), \quad (16)$$

where  $\widehat{y}_i^{(t)}$  is the prediction of the  $i$ -th instance at the  $t$ -th iteration and  $f_t$  is the weaker learner at the  $t$ -th iteration.

Then, Newton boosting performs a Taylor second-order expansion on the loss function  $l(y_i, \widehat{y}_i^{(t-1)} + f_t(x_i))$  to obtain  $g_i f_t(x_i) + (1/2) h_i f_t^2(x_i)$ , because the second-order approximation helps to minimize the loss function conveniently and quickly [43]. The equations of  $g_i$ ,  $h_i$  and the new loss function are defined, respectively, as

$$g_i = \frac{\partial l(y_i, \widehat{y}_i^{(t-1)})}{\widehat{y}_i^{(t-1)}}, \quad (17)$$

$$h_i = \frac{\partial^2 l(y_i, \widehat{y}_i^{(t-1)})}{\widehat{y}_i^{(t-1)}}, \quad (18)$$

$$L_t = \sum_{i=1}^n \left[ l(y_i, \widehat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t). \quad (19)$$

Assume that the sample set  $I_j$  in the leaf node  $j$  is defined as  $I_j = \{i \mid q(x_i) = j\}$ , where  $q(x_i)$  represents the tree structure from the root to the leaf node  $j$  in the decision tree, (19) can be transformed into the following formula, as show in

$$\begin{aligned} \widetilde{L}^{(t)} &= \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2} \alpha \sum_{j=1}^T \omega_j^2 \\ &= \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + a \right) \omega_j^2 \right] + \gamma T. \end{aligned} \quad (20)$$

The formula for estimating the weight of each leaf in the decision tree is formulated as

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + a}, \quad (21)$$

According to (21), as for the tree structure  $q$ , the loss function at the leaf node  $j$  can be changed as

$$\widetilde{L}^{(t)}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{\left( \sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + a} + \gamma T, \quad (22)$$

Therefore, the equation of the information gain after branching can be defined as

*Gain*

$$\begin{aligned} &= \frac{1}{2} \left[ \frac{\left( \sum_{i \in I_{jL}} g_i \right)^2}{\sum_{i \in I_{jL}} h_i + a} + \frac{\left( \sum_{i \in I_{jR}} g_i \right)^2}{\sum_{i \in I_{jR}} h_i + a} - \frac{\left( \sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + a} \right] \\ &\quad - \gamma, \end{aligned} \quad (23)$$

where  $I_{jL}$  and  $I_{jR}$  are the sample sets of the left and right leaf node, respectively, after splitting the leaf node  $j$ .

XGBOOST branches each leaf node and constructs basic learners by the criterion of maximizing the information gain.

With the help of Newton boosting, the XGBOOST can deal with missing values by adaptively learning. To a certain extent, XGBOOST is based on the multiple additive regression tree (MART), but it can get better tree structure by learning with Newton boosting. In addition, XGBOOST can also subsample among columns, which reduces the relevance of each weak learner [39].

### 3. The Proposed CEEMDAN-XGBOOST Approach

From the existing literature, we can see that CEEMDAN has advantages in time series decomposition, while XGBOOST

does well in regression. Therefore, in this paper, we integrated these two methods and proposed a novel approach, so-called CEEMDAN-XGBOOST, for forecasting crude oil prices. The proposed CEEMDAN-XGBOOST includes three stages: decomposition, individual forecasting, and ensemble. In the first stage, CEEMDAN is used to decompose the raw series of crude oil prices into  $k+1$  components, including  $k$  IMFs and one residue. Among the components, some show high-frequency characteristics while the others show low-frequency ones of the raw series. In the second stage, for each component, a forecasting model is built using XGBOOST, and then the built model is applied to forecast each component and then get an individual result. Finally, all the results from the components are aggregated as the final result. Although there exist a lot of methods to aggregate the forecasting results from components, in the proposed approach, we use the simplest way, i.e., addition, to summarize the results of all components. The flowchart of the CEEMDAN-XGBOOST is shown in Figure 2.

From Figure 2, it can be seen that the proposed CEEMDAN-XGBOOST based on the framework of “decomposition and ensemble” is also a typical strategy of “divide and conquer”; that is, the tough task of forecasting crude oil prices from the raw series is divided into several subtasks of forecasting from simpler components. Since the raw series is extremely nonlinear and nonstationary while each decomposed component has a relatively simple form for forecasting, the CEEMDAN-XGBOOST has the ability to achieve higher accuracy of forecasting crude oil prices. In short, the advantages of the proposed CEEMDAN-XGBOOST are threefold: (1) the challenging task of forecasting crude oil prices is decomposed into several relatively simple subtasks; (2) for forecasting each component, XGBOOST can build models with different parameters according to the characteristics of the component; and (3) a simple operation, addition, is used to aggregate the results from subtasks as the final result.

## 4. Experiments and Analysis

*4.1. Data Description.* To demonstrate the performance of the proposed CEEMDAN-XGBOOST, we use the crude oil prices from the West Texas Intermediate (WTI) as experimental data (the data can be downloaded from <https://www.eia.gov/dnav/pet/hist/RWTCD.htm>). We use the daily closing prices covering the period from January 2, 1986, to March 19, 2018, with 8123 observations in total for empirical studies. Among the observations, the first 6498 ones from January 2, 1986, to September 21, 2011, accounting for 80% of the total observations, are used as training samples, while the remaining 20% ones are for testing. The original crude oil prices are shown in Figure 3.

We perform multi-step-ahead forecasting in this paper. For a given time series  $x_t$  ( $t = 1, 2, \dots, T$ ), the  $m$ -step-ahead forecasting for  $x_{t+m}$  can be formulated as

$$\widehat{x}_{t+m} = f(x_{t-(l-1)}, x_{t-(l-2)}, \dots, x_{t-1}, x_t), \quad (24)$$

where  $\widehat{x}_{t+m}$  is the  $m$ -step-ahead predicted result at time  $t$ ,  $f$  is the forecasting model,  $x_i$  is the true value at time  $i$ , and  $l$  is the lag order.

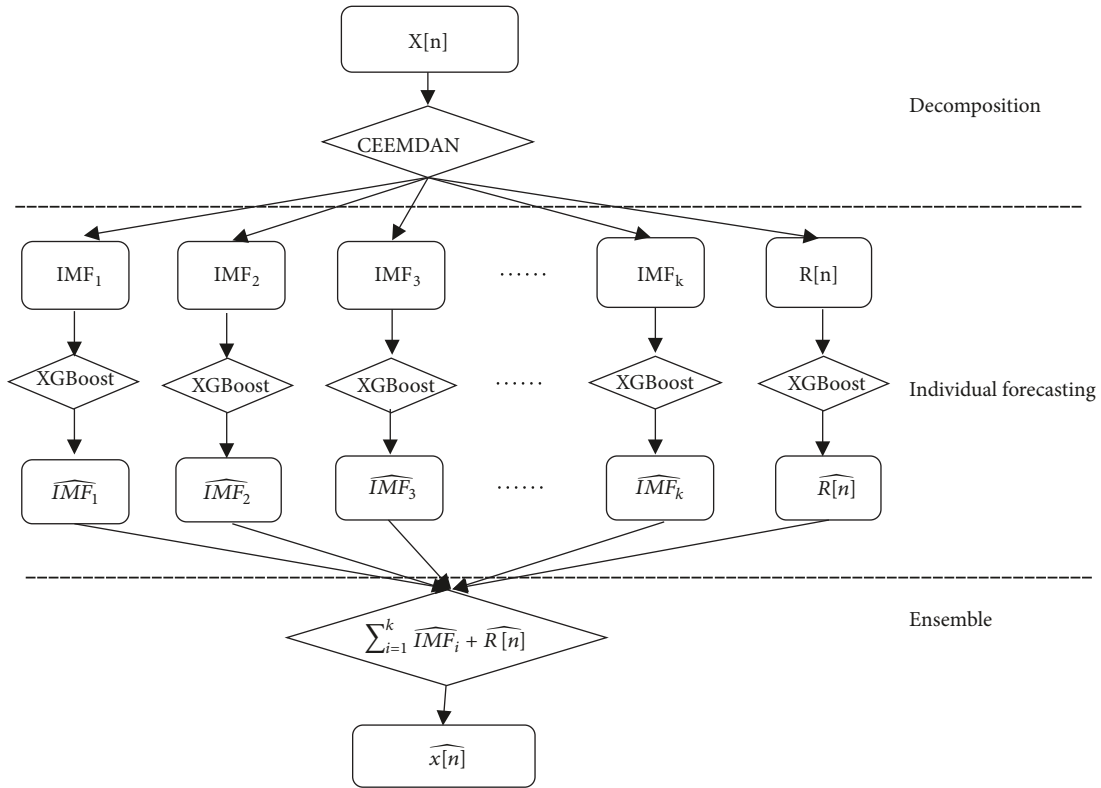


FIGURE 2: The flowchart of the proposed CEEMDAN-XGBOOST.

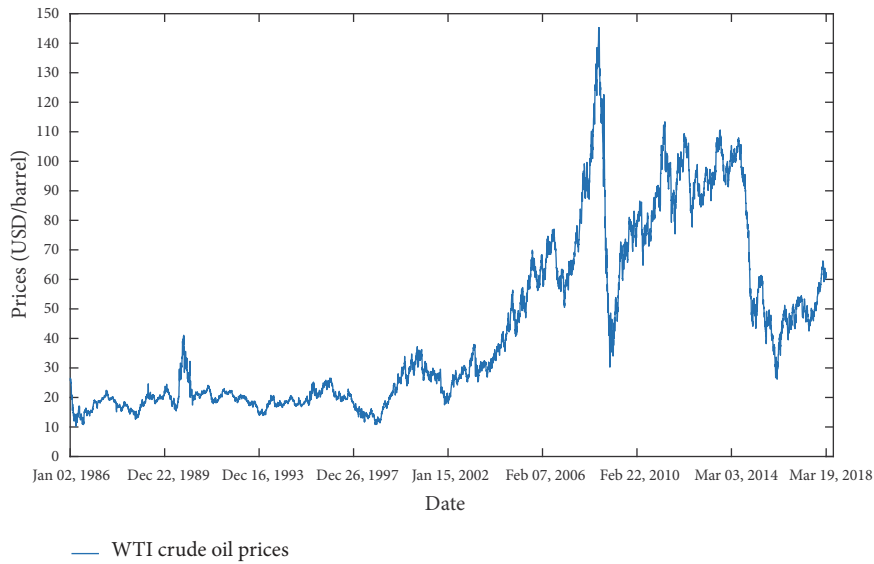


FIGURE 3: The original crude oil prices of WTI.

For SVR and FNN, we normalize each decomposed component before building the model to forecast the component individually. In detail, the normalization process can be defined as

$$x'_t = \frac{x_t - \mu}{\sigma}, \quad (25)$$

where  $x'_t$  is the normalized series of crude oil prices series,  $x_t$  is the data before normalization,  $\mu$  is the mean of  $x_t$ , and  $\sigma$  is the standard deviation of  $x_t$ . Meanwhile, since normalization is not necessary for XGBOOST and ARIMA, for models with these two algorithms, we build forecasting models from each of the decomposed components directly.

**4.2. Evaluation Criteria.** When we evaluate the accuracy of the models, we focus on not only the numerical accuracy but also the accuracy of forecasting direction. Therefore, we select the root-mean-square error (RMSE) and the mean absolute error (MAE) to evaluate the numerical accuracy of the models. Besides, we use directional statistic (Dstat) as the criterion for evaluating the accuracy of forecasting direction. The RMSE, MAE, and Dstat are defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_t - \widehat{y}_t)^2}, \quad (26)$$

$$\text{MAE} = \frac{1}{N} \left( \sum_{i=1}^N |y_t - \widehat{y}_t| \right), \quad (27)$$

$$\text{Dstat} = \frac{\sum_{i=2}^N d_i}{N-1}, \quad (28)$$

$$d_i = \begin{cases} 1, & (y_t - y_{t-1})(\widehat{y}_t - \widehat{y}_{t-1}) > 0 \\ 0, & (y_t - y_{t-1})(\widehat{y}_t - \widehat{y}_{t-1}) < 0, \end{cases}$$

where  $y_t$  is the actual crude oil prices at the time  $t$ ,  $\widehat{y}_t$  is the prediction, and  $N$  is the size of the test set.

In addition, we take the Wilcoxon signed rank test (WSRT) for proving the significant differences between the forecasts of the selected models [44]. The WSRT is a nonparametric statistical hypothesis test that can be used to evaluate whether the population mean ranks of two predictions from different models on the same sample differ. Meanwhile, it is a paired difference test which can be used as an alternative to the paired Student t-test. The null hypothesis of the WSRT is whether the median of the loss differential series  $d(t) = g(e_a(t)) - g(e_b(t))$  is equal to zero or not, where  $e_a(t)$  and  $e_b(t)$  are the error series of model  $a$  and model  $b$  respectively, and  $g(\cdot)$  is a loss function. If the  $p$  value of pairs of models is below 0.05, the test rejects the null hypothesis (there is a significant difference between the forecasts of this pair of models) under the confidence level of 95%. In this way, we can prove that there is a significant difference between the optimal model and the others.

However, the criteria defined above are global. If some singular points exist, the optimal model chosen by these criteria may not be the best one. Thus, we make the model confidence set (MCS) [31, 45] in order to choose the optimal model convincingly.

In order to calculate the  $p$ -value of the statistics accurately, the MCS performs bootstrap on the prediction series, which can soften the impact of the singular points. For the  $j$ -th model, suppose that the size of a bootstrapped sample is  $T$ , and the  $t$ -th bootstrapped sample has the loss functions defined as

$$L_{j,t} = \frac{1}{T} \sum_{t=h+1}^{h+T} |y_t - \widehat{y}_t|, \quad (29)$$

Suppose that a set  $M_0 = \{m_i, i = 1, 2, 3, \dots, n\}$  that contains  $n$  models, for any two models  $j$  and  $k$ , the relative values of the loss between these two models can be defined as

$$d_{j,k,t} = L_{j,t} - L_{k,t}, \quad (30)$$

According to the above definitions, the set of superior models can be defined as

$$M^* \equiv \{m_j \in M_0 : E(d_{j,k,t}) \leq 0, \forall m_k \in M_0\}, \quad (31)$$

where  $E(\cdot)$  represents the average value.

The MCS repeatedly performs the significant test in  $M_0$ . At each time, the worst prediction model in the set is eliminated. In the test, the hypothesis is the null hypothesis of equal predictive ability (EPA), defined as

$$H_0 : E(d_{j,k,t}) = 0, \quad \forall m_j, m_k \in M \subset M_0 \quad (32)$$

The MCS mainly depends on the equivalence test and elimination criteria. The specific process is as follows.

*Step 1.* Assuming  $M = M_0$ , at the level of significance  $\alpha$ , use the equivalence test to test the null hypothesis  $H_{0,M}$ .

*Step 2.* If it accepts the null hypothesis and then it defines  $M_{1-\alpha}^* = M$ , otherwise it eliminates the model which rejects the null hypothesis from  $M$  according to the elimination criteria. The elimination will not stop until there are not any models that reject the null hypothesis in the set  $M$ . In the end, the models in  $M_{1-\alpha}^*$  are thought as surviving models.

Meanwhile, the MCS has two kinds of statistics that can be defined as

$$T_R = \max_{j,k \in M} |t_{j,k}|, \quad (33)$$

$$T_{SQ} = \max_{j,k \in M} t_{j,k}^2, \quad (34)$$

$$t_{j,k} = \frac{\overline{d_{j,k}}}{\sqrt{\text{var}(\overline{d_{j,k}})}}, \quad (35)$$

$$\overline{d_{j,k}} = \frac{1}{T} \sum_{t=h+1}^{h+T} d_{j,k,t}, \quad (36)$$

where  $T_R$  and  $T_{SQ}$  stand for the range statistics and the semiquadratic statistic, respectively, and both statistics are based on the  $t$ -statistics as shown in (35)-(36). These two statistics ( $T_R$  and  $T_{SQ}$ ) are mainly to remove the model whose  $p$ -value is less than the significance level  $\alpha$ . When the  $p$ -value is greater than the significance level  $\alpha$ , the models can survive. The larger the  $p$ -value, the more accurate the prediction of the model. When the  $p$ -value is equal to 1, it indicates that the model is the optimal forecasting model.

**4.3. Parameter Settings.** To test the performance of XGBOOST and CEEMDAN-XGBOOST, we conduct two groups of experiments: single models that forecast crude

TABLE 1: The ranges of the parameters for XGBOOST by grid search.

Parameter	Description	range
Booster	Booster to use.	{'gblinear', 'gbtree'}
N_estimators	Number of boosted trees.	{100,200,300,400,500}
Max_depth	Maximum tree depth for base learners.	{3,4,5,6,7,8}
Min_child_weight	Maximum delta step we allow each tree's weight estimation to be.	{1,2,3,4,5,6}
Gamma	Minimum loss reduction required to make a further partition on a leaf node of the tree.	{0.01, 0.05,0.1,0.2,0.3 }
Subsample	Subsample ratio of the training instance.	{0.6,0.7,0.8,0.9,1}
Colsample	Subsample ratio of columns when constructing each tree.	{0.6,0.7,0.8,0.9,1}
Reg_alpha	L1 regularization term on weights	{0.01,0.05,0.1}
Reg_lambda	L2 regularization term on weights	{0.01,0.05,0.1}
Learning_rate	Boosting learning rate	{0.01,0.05,0.07,0.1,0.2}

oil prices with original sequence, and ensemble models that forecast crude oil prices based on the “decomposition and ensemble” framework.

For single models, we compare XGBOOST with one statistical model, ARIMA, and two widely used AI-models, SVR and FNN. Since the existing research has shown that EEMD significantly outperforms EMD in forecasting crude oil prices [24, 31], in the experiments, we only compare CEEMDAN with EEMD. Therefore, we compare the proposed CEEMDAN-XGBOOST with EEMD-SVR, EEMD-FNN, EEMD-XGBOOST, CEEMDAN-SVR, and CEEMDAN-FNN.

For ARIMA, we use the Akaike information criterion (AIC) [46] to select the parameters ( $p-d-q$ ). For SVR, we use RBF as kernel function and use grid search to optimize  $C$  and  $\gamma$  in the ranges of  $2^{\{0,1,2,3,4,5,6,7,8\}}$  and  $2^{\{-9,-8,-7,-6,-5,-4,-3,-2,-1,0\}}$ , respectively. We use one hidden layer with 20 nodes for FNN. We use a grid search to optimize the parameters for XGBOOST; the search ranges of the optimized parameters are shown in Table 1.

We set 0.02 and 0.05 as the standard deviation of the added white noise and set 250 and 500 as the number of realizations of EEMD and CEEMDAN, respectively. The decomposition results of the original crude oil prices by EEMD and CEEMDAN are shown in Figures 4 and 5, respectively.

It can be seen from Figure 4 that, among the components decomposed by EEMD, the first six IMFs show characteristics of high frequency while the remaining six components show characteristics of low frequency. However, regarding the components by CEEMDAN, the first seven ones show clear high-frequency and the last four show low-frequency, as shown in Figure 5.

The experiments were conducted with Python 2.7 and MATLAB 8.6 on a 64-bit Windows 7 with 3.4 GHz I7 CPU and 32 GB memory. Specifically, we run FNN and MCS with MATLAB, and, for the remaining work, we use Python. Regarding XGBoost, we used a widely used Python package (<https://xgboost.readthedocs.io/en/latest/python/>) in the experiments.

TABLE 2: The RMSE, MAE, and Dstat by single models with horizon = 1, 3, and 6.

Horizon	Model	RMSE	MAE	Dstat
1	XGBOOST	<b>1.2640</b>	<b>0.9481</b>	0.4827
	SVR	1.2899	0.9651	0.4826
	FNN	1.3439	0.9994	0.4837
	ARIMA	1.2692	0.9520	<b>0.4883</b>
3	XGBOOST	<b>2.0963</b>	<b>1.6159</b>	0.4839
	SVR	2.2444	1.7258	<b>0.5080</b>
	FNN	2.1503	1.6512	0.4837
	ARIMA	2.1056	1.6177	0.4901
6	XGBOOST	<b>2.9269</b>	2.2945	0.5158
	SVR	3.1048	2.4308	<b>0.5183</b>
	FNN	3.0803	2.4008	0.5028
	ARIMA	2.9320	<b>2.2912</b>	0.5151

**4.4. Experimental Results.** In this subsection, we use a fixed value 6 as the lag order, and we forecast crude oil prices with 1-step-ahead, 3-step-ahead, and 6-step-ahead forecasting; that is to say, the horizons for these three forecasting tasks are 1, 3, and 6, respectively.

**4.4.1. Experimental Results of Single Models.** For single models, we compare XGBOOST with state-of-the-art SVR, FNN, and ARIMA, and the results are shown in Table 2.

It can be seen from Table 2 that XGBOOST outperforms other models in terms of RMSE and MAE with horizons 1 and 3. For horizon 6, XGBOOST achieves the best RMSE and the second best MAE, which is slightly worse than that by ARIMA. For horizon 1, FNN achieves the worst results among the four models; however, for horizons 3 and 6, SVR achieves the worst results. Regarding Dstat, none of the models can always outperform others, and the best result of Dstat is achieved by SVR with horizon 6. It can be found that the RMSE and MAE values gradually increase with the increase of horizon. However, the Dstat values do not show such discipline. All the values of Dstat are around 0.5, i.e., from



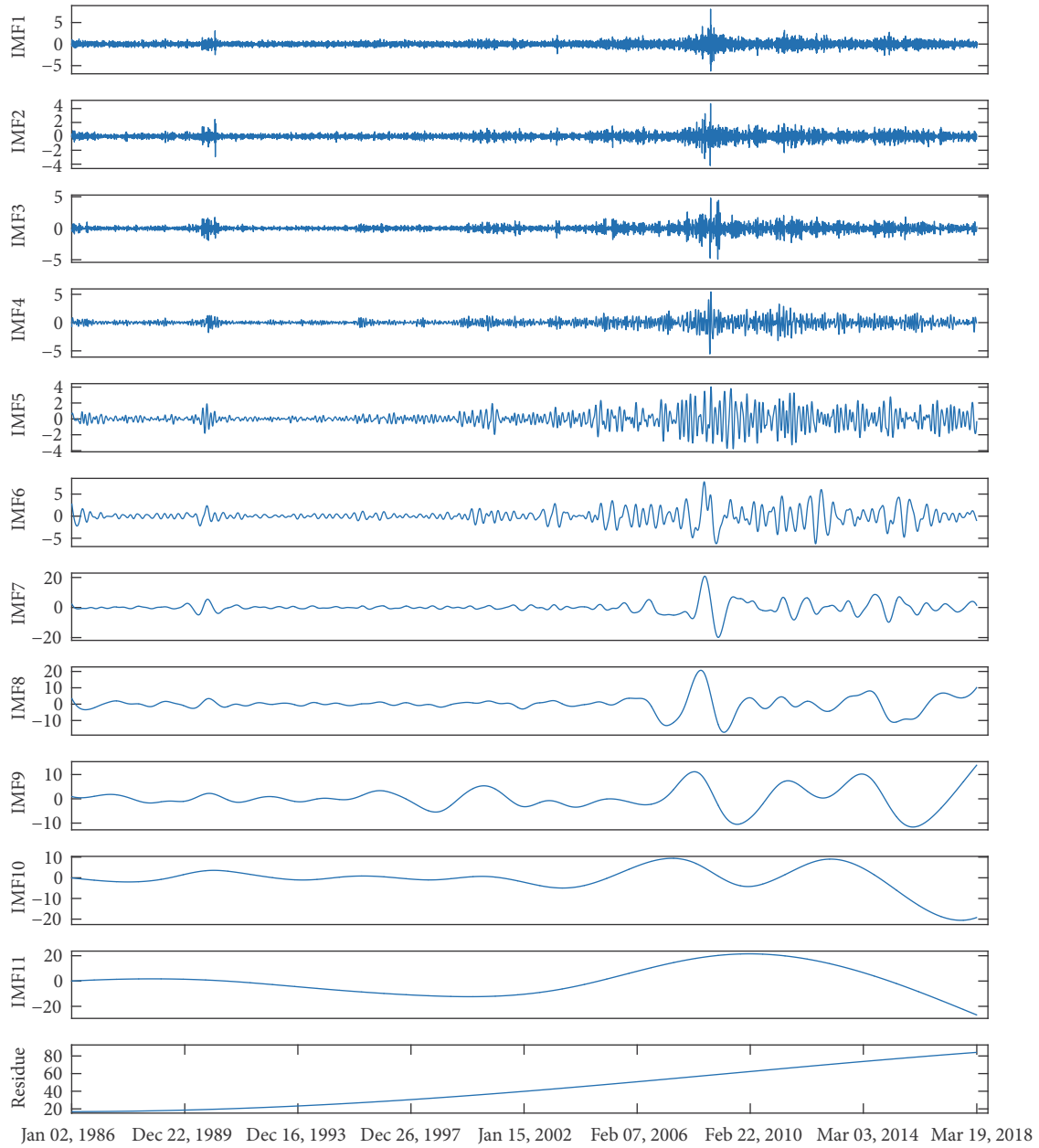


FIGURE 4: The IMFs and residue of WTI crude oil prices by EEMD.

0.4826 to 0.5183, which are very similar to the results of random guesses, showing that it is very difficult to accurately forecast the direction of the movement of crude oil prices with raw crude oil prices directly.

To further verify the advantages of XGBOOST over other models, we report the results by WSRT and MCS, as shown in Tables 3 and 4, respectively. As for WSRT, the  $p$ -value between XGBOOST and other models except ARIMA is below 0.05, which means that there is a significant difference among the forecasting results of XGBOOST, SVR, and FNN in population mean ranks. Besides, the results of MCS show that the  $p$ -value of  $T_R$  and  $T_{SQ}$  of XGBOOST is always equal to 1.000 and prove that XGBOOST is the optimal model among

TABLE 3: Wilcoxon signed rank test between XGBOOST, SVR, FNN, and ARIMA.

	XGBOOST	SVR	FNN	ARIMA
XGBOOST	1	4.0378e-06	2.2539e-35	5.7146e-01
SVR	4.0378e-06	1	4.6786e-33	0.7006
FNN	2.2539e-35	4.6786e-33	1	6.9095e-02
ARIMA	5.7146e-01	0.7006	6.9095e-02	1

all the models in terms of global errors and most local errors of different samples obtained by bootstrap methods in MCS. According to MCS, the  $p$ -values of  $T_R$  and  $T_{SQ}$  of SVR on the horizon of 3 are greater than 0.2, so SVR becomes the

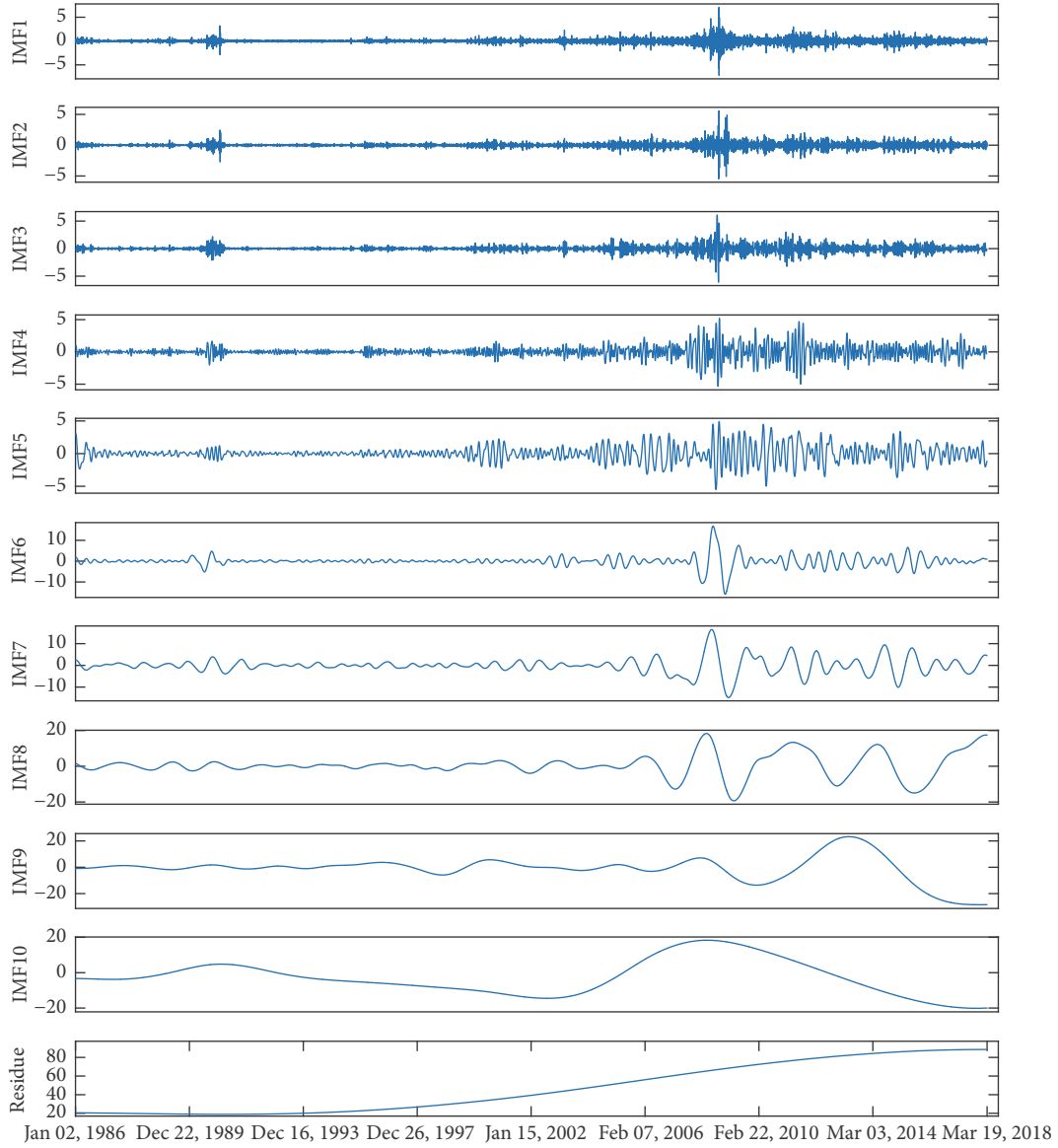


FIGURE 5: The IMFs and residue of WTI crude oil prices by CEEMDAN.

TABLE 4: MCS between XGBOOST, SVR, FNN, and ARIMA.

	HORIZON=1		HORIZON=3		HORIZON=6	
	$T_R$	$T_{SQ}$	$T_R$	$T_{SQ}$	$T_R$	$T_{SQ}$
XGBOOST	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SVR	0.0004	0.0004	0.4132	0.4132	0.0200	0.0200
FNN	0.0002	0.0002	0.0248	0.0538	0.0016	0.0022
ARIMA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

survival and the second best model on this horizon. When it comes to ARIMA, ARIMA almost performs as good as XGBOOST in terms of evaluation criteria of global errors but it does not pass the MCS. It indicates that ARIMA does not perform better than other models in most local errors of different samples.

*4.4.2. Experimental Results of Ensemble Models.* With EEMD or CEEMDAN, the results of forecasting the crude oil prices by XGBOOST, SVR, and FNN with horizons 1, 3, and 6, are shown in Table 5.

It can be seen from Table 5 that the RMSE and MAE values of the CEEMDAN-XGBOOST are the lowest ones

TABLE 5: The RMSE, MAE, and Dstat by the models with EEMD or CEEMDAN.

Horizon	Model	RMSE	MAE	Dstat
1	CEEMDAN-XGBOOST	<b>0.4151</b>	<b>0.3023</b>	0.8783
	EEMD-XGBOOST	0.9941	0.7685	0.7109
	CEEMDAN-SVR	0.8477	0.7594	<b>0.9054</b>
	EEMD-SVR	1.1796	0.9879	0.8727
	CEEMDAN-FNN	1.2574	1.0118	0.7597
	EEMD-FNN	2.6835	1.9932	0.7361
3	CEEMDAN-XGBOOST	<b>0.8373</b>	<b>0.6187</b>	0.6914
	EEMD-XGBOOST	1.4007	1.0876	0.6320
	CEEMDAN-SVR	1.2399	1.0156	<b>0.7092</b>
	EEMD-SVR	1.2366	1.0275	<b>0.7092</b>
	CEEMDAN-FNN	1.2520	0.9662	0.7061
	EEMD-FNN	1.2046	0.8637	0.6959
6	CEEMDAN-XGBOOST	<b>1.2882</b>	<b>0.9831</b>	0.6196
	EEMD-XGBOOST	1.7719	1.3765	0.6165
	CEEMDAN-SVR	1.3453	1.0296	<b>0.6683</b>
	EEMD-SVR	1.3730	1.1170	0.6485
	CEEMDAN-FNN	1.8024	1.3647	0.6422
	EEMD-FNN	2.7786	2.0495	0.6337

among those by all methods with each horizon. For example, with horizon 1, the values of RMSE and MAE are 0.4151 and 0.3023, which are far less than the second values of RMSE and MAE, i.e., 0.8477 and 0.7594, respectively. With the horizon increases, the corresponding values of each model increase, in terms of RMSE and MAE. However, the CEEMDAN-XGBOOST still achieves the lowest values of RMSE and MAE with each horizon. Regarding the values of Dstat, all the values are far greater than those by random guesses, showing that the “decomposition and ensemble” framework is effective for directional forecasting. Specifically, the values of Dstat are in the range between 0.6165 and 0.9054. The best Dstat values in all horizons are achieved by CEEMDAN-SVR or EEMD-SVR, showing that, among the forecasters, SVR is the best one for directional forecasting, although corresponding values of RMSE and MAE are not the best. As for the decomposition methods, when the forecasters are fixed, CEEMDAN outperforms EEMD in 8, 8, and 8 out of 9 cases in terms of RMSE, MAE, and Dstat, respectively, showing the advantages of CEEMDAN over EEMD. Regarding the forecasters, when combined with CEEMDAN, XGBOOST is always superior to other forecasters in terms of RMSE and MAE. However, when combined with EEMD, XGBOOST outperforms SVR and FNN with horizon 1, and FNN with horizon 6 in terms of RMSE and MAE. With horizons 1 and 6, FNN achieves the worst results of RMSE and MAE. The results also show that good values of RMSE usually are associated with good values of MAE. However, good values of RMSE or MAE do not always mean good Dstat directly.

For the ensemble models, we also took a Wilcoxon signed rank test and an MCS test based on the errors of pairs of models. We set 0.2 as the level of significance in MCS, and 0.05 as the level of significance in WSRT. The results are shown in Tables 6 and 7.

From these two tables, it can be seen that, regarding the results of WSRT, the  $p$ -value between CEEMDAN-XGBOOST and any other models except EEMD-FNN are below 0.05, demonstrating that there is a significant difference on the population mean ranks between CEEMDAN-XGBOOST and any other models except EEMD-FNN. What is more, the MCS shows that the  $p$ -value of  $T_R$  and  $T_{SQ}$  of CEEMDAN-XGBOOST is always equal to 1.000 and demonstrates that CEEMDAN-XGBOOST is the optimal model among all models in terms of global errors and local errors. Meanwhile, the  $p$ -values of  $T_R$  and  $T_{SQ}$  of EEMD-FNN are greater than other models except CEEMDAN-XGBOOST and become the second best model with horizons 3 and 6 in MCS. Meanwhile, with the horizon 6, the CEEMDAN-SVR is also the second best model. Besides, the  $p$ -values of  $T_R$  and  $T_{SQ}$  of EEMD-SVR and CEEMDAN-SVR are up to 0.2 and they become the surviving models with horizon 6 in MCS.

From the results of single models and ensemble models, we can draw the following conclusions: (1) single models usually cannot achieve satisfactory results, due to the non-linearity and nonstationarity of raw crude oil prices. As a single forecaster, XGBOOST can achieve slightly better results than some state-of-the-art algorithms; (2) ensemble models can significantly improve the forecasting accuracy in terms of several evaluation criteria, following the “decomposition and ensemble” framework; (3) as a decomposition method, CEEMDAN outperforms EEMD in most cases; (4) the extensive experiments demonstrate that the proposed CEEMDAN-XGBOOST is promising for forecasting crude oil prices.

**4.5. Discussion.** In this subsection, we will study the impacts of several parameters related to the proposed CEEMDAN-XGBOOST.

#### 4.5.1. The Impact of the Number of Realizations in CEEMDAN.

In (2), it is shown that there are  $N$  realizations  $x^i[t]$  in CEEMDAN. We explore how the number of realizations in CEEMDAN can influence on the results of forecasting crude oil prices by CEEMDAN-XGBOOST with horizon 1 and lag 6. And we set the number of realizations in CEEMDAN in the range of {10,25,50,75,100,250,500,750,1000}. The results are shown in Figure 6.

It can be seen from Figure 6 that, for RMSE and MAE, the bigger the number of realizations is, the more accurate results the CEEMDAN-XGBOOST can achieve. When the number of realizations is less than or equal to 500, the values of both RMSE and MAE decrease with increasing of the number of realizations. However, when the number is greater than 500, these two values are increasing slightly. Regarding Dstat, when the number increases from 10 to 25, the value of Dstat increases rapidly, and then it increases slowly with the number increasing from 25 to 500. After that, Dstat decreases slightly. It is shown that the value of Dstat reaches the top values with the number of realizations 500. Therefore, 500 is the best value for the number of realizations in terms of RMSE, MAE, and Dstat.

TABLE 6: Wilcoxon signed rank test between XGBOOST, SVR, and FNN with EEMD or CEEMDAN.

	CEEMDAN-XGBOOST	EEMD-XGBOOST	CEEMDAN-SVR	EEMD-SVR	CEEMDAN-FNN	EEMD-FNN
CEEMDAN-XGBOOST	1	3.5544e-05	1.8847e-50	0.0028	1.6039e-187	0.0726
EEMD-XGBOOST	3.5544e-05	1	4.5857e-07	0.3604	8.2912e-82	0.0556
CEEMDAN-SVR	1.8847e-50	4.5857e-07	1	4.9296e-09	5.7753e-155	8.6135e-09
EEMD-SVR	0.0028	0.3604	4.9296e-09	1	2.5385e-129	0.0007
CEEMDAN-FNN	1.6039e-187	8.2912e-82	5.7753e-155	2.5385e-129	1	8.1427e-196
EEMD-FNN	0.0726	0.0556	8.6135e-09	0.0007	8.1427e-196	1

TABLE 7: MCS between XGBOOST, SVR, and FNN with EEMD or CEEMDAN.

	HORIZON=1		HORIZON=3		HORIZON=6	
	$T_R$	$T_{SQ}$	$T_R$	$T_{SQ}$	$T_R$	$T_{SQ}$
CEEMDAN-XGBOOST	1	1	1	1	1	1
EEMD-XGBOOST	0	0	0	0	0	0.0030
CEEMDAN-SVR	0	0.0002	0.0124	0.0162	0.8268	0.8092
EEMD-SVR	0	0	0.0008	0.004	0.7872	0.7926
CEEMDAN-FNN	0	0	0.0338	0.0532	0.2924	0.3866
EEMD-FNN	0	0.0002	0.4040	0.4040	0.8268	0.8092

4.5.2. *The Impact of the Lag Orders.* In this section, we explore how the number of lag orders impacts the prediction accuracy of CEEMDAN-XGBOOST on the horizon of 1. In this experiment, we set the number of lag orders from 1 to 10, and the results are shown in Figure 7.

According to the empirical results shown in Figure 7, it can be seen that as the lag order increases from 1 to 2, the values of RMSE and MAE decrease sharply while that of Dstat increases drastically. After that, the values of RMSE or MAE remain almost unchanged (or increase very slightly) with the increasing of lag orders. However, for Dstat, the value increases sharply from 1 to 2 and then decreases from 2 to 3. After the lag order increases from 3 to 5, the Dstat stays almost stationary. Overall, when the value of lag order is up to 5, it reaches a good tradeoff among the values of RMSE, MAE, and Dstat.

4.5.3. *The Impact of the Noise Strength in CEEMDAN.* Apart from the number of realizations, the noise strength in CEEMDAN, which stands for the standard deviation of the white noise in CEEMDAN, also affects the performance of CEEMDAN-XGBOOST. Thus, we set the noise strength in the set of {0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07} to explore how the noise strength in CEEMDAN affects the prediction accuracy of CEEMDAN-XGBOOST on a fixed horizon 1 and a fixed lag 6.

As shown in Figure 8, when the noise strength in CEEMDAN is equal to 0.05, the values of RMSE, MAE and Dstat achieve the best results simultaneously. When the noise strength is less than or equal to 0.05 except 0.03, the values of RMSE, MAE and Dstat become better and better with the increase of the noise strength. However, when the strength is greater than 0.05, the values of RMSE, MAE and Dstat

become worse and worse. The figure indicates that the noise strength has a great impact on forecasting results, and an ideal range for it is about 0.04-0.06.

## 5. Conclusions

In this paper, we propose a novel model, namely, CEEMDAN-XGBOOST, to forecast crude oil prices. At first, CEEMDAN-XGBOOST decomposes the sequence of crude oil prices into several IMFs and a residue with CEEMDAN. Then, it forecasts the IMFs and the residue with XGBOOST individually. Finally, CEEMDAN-XGBOOST computes the sum of the prediction of the IMFs and the residue as the final forecasting results. The experimental results show that the proposed CEEMDAN-XGBOOST significantly outperforms the compared methods in terms of RMSE and MAE. Although the performance of the CEEMDAN-XGBOOST on forecasting the direction of crude oil prices is not the best, the MCS shows that the CEEMDAN-XGBOOST is still the optimal model. Meanwhile, it is proved that the number of the realizations, lag, and the noise strength for CEEMDAN are the vital factors which have great impacts on the performance of the CEEMDAN-XGBOOST.

In the future, we will study the performance of the CEEMDAN-XGBOOST on forecasting crude oil prices with different periods. We will also apply the proposed approach for forecasting other energy time series, such as wind speed, electricity load, and carbon emissions prices.

## Data Availability

The data used to support the findings of this study are included within the article.

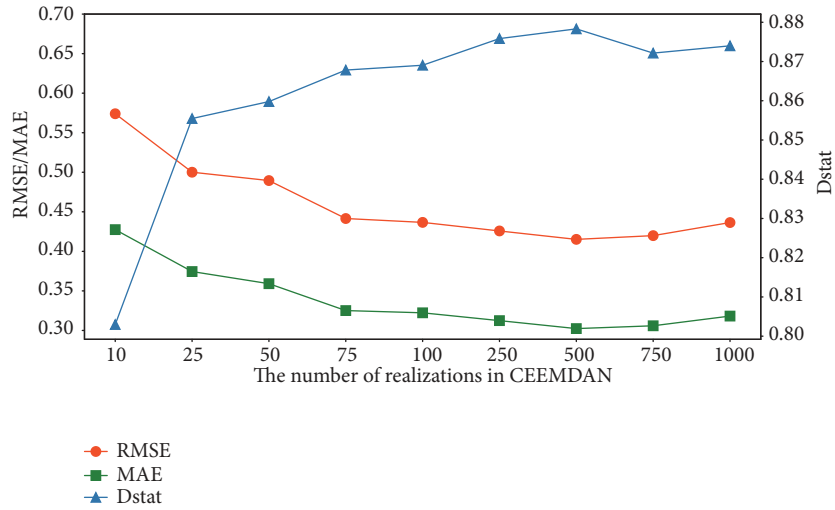


FIGURE 6: The impact of the number of IMFs with CEEMDAN-XGBOOST.

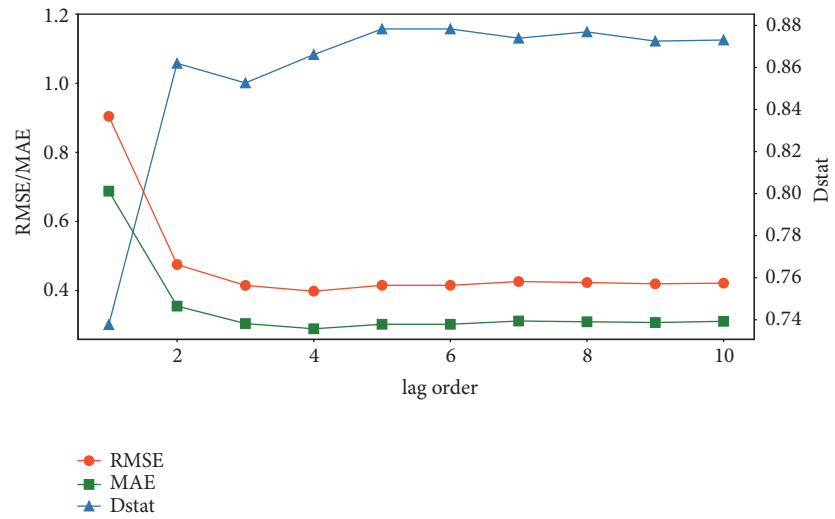


FIGURE 7: The impact of the number of lag orders with CEEMDAN-XGBOOST.

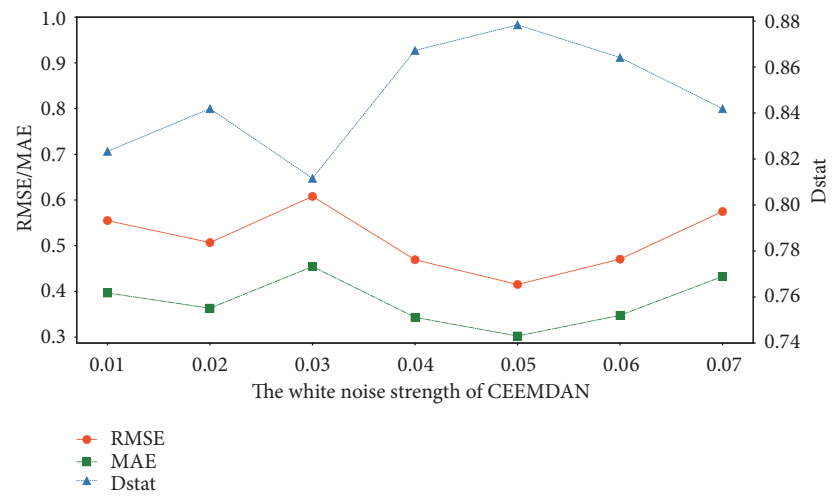


FIGURE 8: The impact of the value of the white noise of CEEMDAN with CEEMDAN-XGBOOST.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (Grants no. JBK1902029, no. JBK1802073, and no. JBK170505), the Natural Science Foundation of China (Grant no. 71473201), and the Scientific Research Fund of Sichuan Provincial Education Department (Grant no. 17ZB0433).

## References

- [1] H. Miao, S. Ramchander, T. Wang, and D. Yang, "Influential factors in crude oil price forecasting," *Energy Economics*, vol. 68, pp. 77–88, 2017.
- [2] L. Yu, S. Wang, and K. K. Lai, "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm," *Energy Economics*, vol. 30, no. 5, pp. 2623–2635, 2008.
- [3] M. Ye, J. Zyren, C. J. Blumberg, and J. Shore, "A short-run crude oil price forecast model with ratchet effect," *Atlantic Economic Journal*, vol. 37, no. 1, pp. 37–50, 2009.
- [4] C. Morana, "A semiparametric approach to short-term oil price forecasting," *Energy Economics*, vol. 23, no. 3, pp. 325–338, 2001.
- [5] H. Naser, "Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (DMA) approach," *Energy Economics*, vol. 56, pp. 75–87, 2016.
- [6] X. Gong and B. Lin, "Forecasting the good and bad uncertainties of crude oil prices using a HAR framework," *Energy Economics*, vol. 67, pp. 315–327, 2017.
- [7] F. Wen, X. Gong, and S. Cai, "Forecasting the volatility of crude oil futures using HAR-type models with structural breaks," *Energy Economics*, vol. 59, pp. 400–413, 2016.
- [8] H. Chiroma, S. Abdul-Kareem, A. Shukri Mohd Noor et al., "A Review on Artificial Intelligence Methodologies for the Forecasting of Crude Oil Price," *Intelligent Automation and Soft Computing*, vol. 22, no. 3, pp. 449–462, 2016.
- [9] S. Wang, L. Yu, and K. K. Lai, "A novel hybrid AI system framework for crude oil price forecasting," in *Data Mining and Knowledge Management*, Y. Shi, W. Xu, and Z. Chen, Eds., vol. 3327 of *Lecture Notes in Computer Science*, pp. 233–242, Springer, Berlin, Germany, 2004.
- [10] J. Baruník and B. Malinská, "Forecasting the term structure of crude oil futures prices with neural networks," *Applied Energy*, vol. 164, pp. 366–379, 2016.
- [11] Y. Chen, K. He, and G. K. F. Tso, "Forecasting crude oil prices—a deep learning based model," *Procedia Computer Science*, vol. 122, pp. 300–307, 2017.
- [12] R. Tehrani and F. Khodayar, "A hybrid optimized artificial intelligent model to forecast crude oil using genetic algorithm," *African Journal of Business Management*, vol. 5, no. 34, pp. 13130–13135, 2011.
- [13] L. Yu, Y. Zhao, and L. Tang, "A compressed sensing based AI learning paradigm for crude oil price forecasting," *Energy Economics*, vol. 46, no. C, pp. 236–245, 2014.
- [14] L. Yu, H. Xu, and L. Tang, "LSSVR ensemble learning with uncertain parameters for crude oil price forecasting," *Applied Soft Computing*, vol. 56, pp. 692–701, 2017.
- [15] Y.-L. Qi and W.-J. Zhang, "The improved SVM method for forecasting the fluctuation of international crude oil price," in *Proceedings of the 2009 International Conference on Electronic Commerce and Business Intelligence (ECBI '09)*, pp. 269–271, IEEE, China, June 2009.
- [16] M. S. AL-Musaylh, R. C. Deo, Y. Li, and J. F. Adamowski, "Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting," *Applied Energy*, vol. 217, pp. 422–439, 2018.
- [17] L. Huang and J. Wang, "Forecasting energy fluctuation model by wavelet decomposition and stochastic recurrent wavelet neural network," *Neurocomputing*, vol. 309, pp. 70–82, 2018.
- [18] H. Zhao, M. Sun, W. Deng, and X. Yang, "A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing," *Entropy*, vol. 19, no. 1, Article ID 14, 2017.
- [19] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A Novel Fault Diagnosis Method Based on Integrating Empirical Wavelet Transform and Fuzzy Entropy for Motor Bearing," *IEEE Access*, vol. 6, pp. 35042–35056, 2018.
- [20] Q. Fu, B. Jing, P. He, S. Si, and Y. Wang, "Fault Feature Selection and Diagnosis of Rolling Bearings Based on EEMD and Optimized Elman\_AdaBoost Algorithm," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5024–5034, 2018.
- [21] T. Li and M. Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, p. 285, 2016.
- [22] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Computers in Biology and Medicine*, vol. 38, no. 1, pp. 1–13, 2008.
- [23] J. Lee, D. D. McManus, S. Merchant, and K. H. Chon, "Automatic motion and noise artifact detection in holter ECG data using empirical mode decomposition and statistical approaches," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1499–1506, 2012.
- [24] L. Fan, S. Pan, Z. Li, and H. Li, "An ICA-based support vector regression scheme for forecasting crude oil prices," *Technological Forecasting & Social Change*, vol. 112, pp. 245–253, 2016.
- [25] E. Jianwei, Y. Bao, and J. Ye, "Crude oil price analysis and forecasting based on variational mode decomposition and independent component analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 484, pp. 412–427, 2017.
- [26] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, 1998.
- [27] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [28] L. Yu, W. Dai, L. Tang, and J. Wu, "A hybrid grid-GA-based LSSVR learning paradigm for crude oil price forecasting," *Neural Computing and Applications*, vol. 27, no. 8, pp. 2193–2215, 2016.
- [29] L. Tang, W. Dai, L. Yu, and S. Wang, "A novel CEEMD-based eelm ensemble learning paradigm for crude oil price forecasting," *International Journal of Information Technology & Decision Making*, vol. 14, no. 1, pp. 141–169, 2015.

- [30] T. Li, M. Zhou, C. Guo et al., "Forecasting crude oil price using EEMD and RVM with adaptive PSO-based kernels," *Energies*, vol. 9, no. 12, p. 1014, 2016.
- [31] T. Li, Z. Hu, Y. Jia, J. Wu, and Y. Zhou, "Forecasting Crude Oil Prices Using Ensemble Empirical Mode Decomposition and Sparse Bayesian Learning," *Energies*, vol. 11, no. 7, 2018.
- [32] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 4144–4147, IEEE, Prague, Czech Republic, May 2011.
- [33] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: a suitable tool for biomedical signal processing," *Biomedical Signal Processing and Control*, vol. 14, no. 1, pp. 19–29, 2014.
- [34] T. Peng, J. Zhou, C. Zhang, and Y. Zheng, "Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine," *Energy Conversion and Management*, vol. 153, pp. 589–602, 2017.
- [35] S. Dai, D. Niu, and Y. Li, "Daily peak load forecasting based on complete ensemble empirical mode decomposition with adaptive noise and support vector machine optimized by modified grey Wolf optimization algorithm," *Energies*, vol. 11, no. 1, 2018.
- [36] Y. Lv, R. Yuan, T. Wang, H. Li, and G. Song, "Health Degradation Monitoring and Early Fault Diagnosis of a Rolling Bearing Based on CEEMDAN and Improved MMSE," *Materials*, vol. 11, no. 6, p. 1009, 2018.
- [37] R. Abdelkader, A. Kaddour, A. Bendiabdellah, and Z. Derouiche, "Rolling Bearing Fault Diagnosis Based on an Improved Denoising Method Using the Complete Ensemble Empirical Mode Decomposition and the Optimized Thresholding Operation," *IEEE Sensors Journal*, vol. 18, no. 17, pp. 7166–7172, 2018.
- [38] Y. Lei, Z. Liu, J. Ouazri, and J. Lin, "A fault diagnosis method of rolling element bearings based on CEEMDAN," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 231, no. 10, pp. 1804–1815, 2017.
- [39] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2016*, pp. 785–794, ACM, New York, NY, USA, August 2016.
- [40] B. Zhai and J. Chen, "Development of a stacked ensemble model for forecasting and analyzing daily average PM2.5 concentrations in Beijing, China," *Science of the Total Environment*, vol. 635, pp. 644–658, 2018.
- [41] S. P. Chatzis, V. Siakoulis, A. Petropoulos, E. Stavroulakis, and N. Vlachogiannakis, "Forecasting stock market crisis events using deep and statistical machine learning techniques," *Expert Systems with Applications*, vol. 112, pp. 353–371, 2018.
- [42] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 591–608, 2017.
- [43] J. Friedman, T. Hastie, and R. Tibshirani, "Special Invited Paper. Additive Logistic Regression: A Statistical View of Boosting: Rejoinde," *The Annals of Statistics*, vol. 28, no. 2, pp. 400–407, 2000.
- [44] J. D. Gibbons and S. Chakraborti, "Nonparametric statistical inference," in *International Encyclopedia of Statistical Science*, M. Lovric, Ed., pp. 977–979, Springer, Berlin, Germany, 2011.
- [45] P. R. Hansen, A. Lunde, and J. M. Nason, "The model confidence set," *Econometrica*, vol. 79, no. 2, pp. 453–497, 2011.
- [46] H. Liu, H. Q. Tian, and Y. F. Li, "Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction," *Applied Energy*, vol. 98, no. 1, pp. 415–424, 2012.

