



Research article

Crude oil prices and volatility prediction by a hybrid model based on kernel extreme learning machine

Hongli Niu* and Yazhi Zhao

School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

* **Correspondence:** Email: niuhongli@ustb.edu.cn.

Abstract: In view of the important position of crude oil in the national economy and its contribution to various economic sectors, crude oil price and volatility prediction have become an increasingly hot issue that is concerned by practitioners and researchers. In this paper, a new hybrid forecasting model based on variational mode decomposition (VMD) and kernel extreme learning machine (KELM) is proposed to forecast the daily prices and 7-day volatility of Brent and WTI crude oil. The KELM has the advantage of less time consuming and lower parameter-sensitivity, thus showing fine prediction ability. The effectiveness of VMD-KELM model is verified by a comparative study with other hybrid models and their single models. Except various commonly used evaluation criteria, a recently-developed multi-scale composite complexity synchronization (MCCS) statistic is also utilized to evaluate the synchrony degree between the predictive and the actual values. The empirical results verify that 1) KELM model holds better performance than ELM and BP in crude oil and volatility forecasting; 2) VMD-based model outperforms the EEMD-based model; 3) The developed VMD-KELM model exhibits great superiority compared with other popular models not only for crude oil price, but also for volatility prediction.

Keywords: crude oil prediction; variational mode decomposition; kernel extreme learning machine; hybrid model; volatility

1. Introduction

As the benchmark of oil market, crude oil has a strong impact on the global economic growth, social stability and national security [1]. In the last two decades, the prediction of crude oil either for

prices or volatility has attracted extensive attention of scholars. This is because the accurate prediction of crude oil price is beneficial to perfect the plans of corresponding production, marketing and investing, to regulate market risks and to enhance future's gainings of the oil-related industries [2], and the oil price volatility is the core of asset pricing, asset allocation and risk management. But in practice, the crude oil prediction is always a great challenging task [3]. One reason is that numerous information factors usually affect the crude oil prices, including fundamental supply-demand relationship [4], external uncertainties factors [5] and unexpected event impact such as epidemic disease. For instance, affected by the coronavirus disease 2019 (COVID-19) pandemic, crude oil prices have exhibited tremendous downturn on April 20 and even reached a historic negative value. The market was observed with great uncertainty and volatility. These factors expand the uncertainty of the prediction results and lowering the prediction accuracy. Therefore, scholars are always seeking a better and more effective forecasting method. In this backdrop, this paper is devoted to propose an effective crude oil prediction model, which can better extract the real information in crude oil prices and volatility so as to achieve accurate forecasting.

In literatures, kinds of prediction algorithms have been proposed, which can be mainly classified into three groups, namely econometrics approaches, artificial intelligence (AI) and hybrid models. Crude oil prices have the characters of highly nonlinear, irregular and complex, the econometric models cannot effectively extract these features. The artificial intelligence algorithms have become popular in dealing with nonlinear and non-stationary time series, like artificial neural networks (ANNs) [6], support vector regression (SVR) [7], least squares support vector regression (LSSVR) [8] and various other deep learning models [2]. However, these AI-technologies often suffer from the disadvantages of long running time and parameter sensitivity [9,10]. For example, ANNs use iterative learning process such as gradient descent method to adjust parameters, which requires a lot of time. Besides it usually trapped into a local optimal solution and the fixed hidden neurons also effect the result. SVR and LSSVR apply iterative learning algorithm, like grid search approach or trial and error technique, to compute the parameters of regularization and kernel, which also face the time-consuming and parameter sensitivity problems.

In recent years, the ideas of randomization and some non-iterative algorithms have been proposed to overcome limitations of AI-models and display excellent performance in speediness and prediction accuracy [9–11], which possess the features of random fixed parameters, random mapping characteristics and unnecessarily to set stop condition, learning rate training times and other parameters during training procedure [12]. Among them, Huang et al. introduced a novel machine learning algorithm known as Extreme Learning Machine (ELM), which randomly chooses input weights between the input layer and the hidden layer, leading to less consuming of time [9]. Meanwhile, weights between the hidden layer and output layer are computed through inversion of matrix and computation, involving lower complexity of computation. But the randomly selected weights will lead to the output changes of different trial runs, so that the system becomes not robust. Later an improved ELM model called Kernel based Extreme Learning Machine (KELM) was developed [13,14] in which the hidden layer feature map is defined by the kernel matrix. After introducing the kernel function into ELM, the stability of forecasting is greatly improved. It has been seen extensive application in many fields with higher performance, easier implementation and faster training speed [15–19].

Since the single prediction models are limited, more and more hybrid models are utilized combining various single algorithms for predicting prices of crude oil, particularly following the

decomposition-ensemble learning paradigm. Some typical decomposition techniques are wavelet decomposition [20], empirical mode decomposition (EMD) and their developed approaches [9]. But EMD-based methods have generally been proved to have some shortcomings, such as, boundary effects, noise sensitivity, mode overlap and lacking of accurate mathematical basis. These may have a negative impact on the precision of decomposition, resulting in distortion of results. Different from EMD, variational mode decomposition (VMD) is a completely non-recursive model which can decompose the original data into multiple components with a specific bandwidth in the spectral domain [21]. Compared with existing decomposition algorithm, such as EEMD and EMD, VMD is more sensitive to noise and sampling. The superiority of VMD method has been indicated in VMD-based decomposition-ensemble models for crude oil prices in some few works [6,22–24].

Briefly, the main contributions of this work can be briefly described in three aspects. Firstly, we follow the “Decomposition-Ensemble” framework and develop a hybrid model VMD-KELM which shows excellence applicability in forecasting the international crude oil prices. The VMD algorithm which can effectively extract intrinsic features and smoother the nonlinear and complex characteristics of crude oil data, while the KELM prediction model is capable in overcoming the time-consuming and parameter sensitivity problem of iterative process. Compared with other hybrid VMD-based and ensemble empirical mode decomposition based (EEMD-based) models as well as single models, the VMD-KELM model demonstrates powerful predictive capabilities of crude oil time series. To the best of our knowledge, this VMD-KELM model has not yet been used for crude oil data. Secondly, this paper also focuses on the prediction of crude oil volatility. The existing works concentrate mostly on the crude oil price forecasting but relatively rare on volatility by the non-econometrics model. This paper validates that the proposed model VMD-KELM has the superior performance in volatility prediction. Thirdly, a recently developed multi-scale composite complexity synchronization (MCCS) statistic [25] from complexity theory is utilized to evaluate the mode, which offers a new perspective to show the forecasting performance. Overall, this study, on one hand, complements the existing decomposition-ensemble learning paradigm in terms of precision of crude oil prediction. On the other hand, it fills in the method literature of crude oil volatility forecasting (using decomposition technology plus promising randomized algorithms).

The remaining of this paper is organized as follows: the main algorithms and performance evaluation measures are given in Section 3. Section 4 depicts the dataset. In Section 5, the prediction effects of VMD-based KELM model for crude oil prices and volatility are analyzed empirically, meanwhile the comparison results with the EEMD-based hybrid models and single models are demonstrated. Section 6 concludes.

2. Literature review

In crude oil price prediction, scholars have proposed a large amount of algorithms. In general, these algorithms can be divided into three categories which are econometrics approaches, artificial intelligence (AI) and hybrid models which integrate two or more single models in any of the above type. In the first type of traditional economic models, autoregressive integrated moving average (ARIMA), random walk (RW), vector auto regression (VAR), generalized autoregressive conditional heteroskedasticity (GARCH) and error correction models (ECM) are comprehensively utilized in forecasting the crude oil price [26–28] as well as volatility [29–32]. For example, Kanjilal and Ghosh [26] used ECM to explore the fluctuation of crude oil prices [26]. Xiang and Zhuang [28]

performed a prediction of Brent crude oil price by ARIMA model and suggested that ARIMA (1,1,1) model can be used as short-term prediction of international crude oil price. Marchese [29] et al. compared the prediction ability of short-memory multivariate GARCH models and long-memory multivariate models, showing the superiority of long-memory multivariate models in predicting crude oil data. Wang and Wu examined the prediction effectiveness of univariate and multivariate GARCH-class models with energy market volatility and found that univariate models allowing for asymmetric effects have higher prediction accuracy than other models [31]. Klein and Walther shown that the mixture memory GARCH (MMGARCH) outperform other predicting models (GARCH, EGARCH, and APARCH, etc.) in predicting volatility and value at risk [32]. Traditional economic models require the processed data to be linear, and this assumption is very difficult to realize. Therefore, they cannot predict the nonlinear and non-stationary time series well.

In order to avoid the shortcomings of economic models, some nonlinear and emerging artificial intelligence algorithms have been with popularity in crude oil price forecasting. The mainstream artificial intelligence algorithms adopted widely include artificial neural networks (ANNs) [6,20,25,33–41], support vector machine (SVM) [7,43], least squares support vector regression (LSSVR) [8,44]. For instance, Lahmiri applied the generalized regression neural network (GRNN) to forecast day ahead energy prices and shown that GRNN is a promising tool for prediction of energy prices [6]. Azadeh et al proposed a flexible algorithm based on artificial neural network (ANN) to predict long-term oil prices [34]. Chiroma et al applied genetic algorithm and neural network (GANN) to forecast WTI crude oil prices and shown that GANN outperform ten BP models in prediction accuracy and computational efficiency [36]. Yu et al. utilized a LSSVR ensemble learning paradigm with uncertain parameters to forecast WTI price and the empirical results verified the prediction effectiveness of the proposed model [44]. Wu et al. added crude oil news as input data and used ANN to predict crude oil prices and made a good progress [41]. They applied the convolutional neural network to extract text features from online crude oil news to show the explanatory power of text features for crude oil price prediction [42]. These AI algorithms often hold the disadvantages of long running time and parameter sensitivity [9,10]. More stable models are waiting to be found.

In recent years, the hybrid models are becoming more and more popular. Following the “Decomposition-Ensemble” framework, there are models based on some typical decomposition techniques, namely, wavelet decomposition [20,25,45], empirical mode decomposition (EMD) and their developed approaches [9,46–49]. For example, Jammazi et al. implemented a HTW-MPNN model combining multilayer back propagation neural network and Harr A trous wavelet decomposition to achieve prediction of crude oil prices and shown that HTW-MBPNN performs better than the traditional BPNN [20]. Tang et al tested the prediction effects of crude oil prices by employing several randomized algorithms like extreme learning machine (ELM), random vector functional link network (RVFLN) and random kitchen sinks (RKS) combined with EEMD method [11]. Wu et al improved the ensemble empirical mode decomposition (EEMD) model, proposed a novel EEMD-LSTM model to predict the crude oil spot price of West Texas Intermediate (WTI) [48]. Recently, a novel decomposition technique originated in signal processing, named variational mode decomposition (VMD), has been adopted as an effective decomposition approach. Lahmiri [6] employed the VMD and neural network for day-ahead energy prices forecasting, the results shown superiority of VMD in decomposition. Bisoi et al predicted the crude oil prices based on VMD and the robust random vector functional link network (RVFLN) [23]. Li et al. proposed VMD-AI models for crude oil price forecasting, and compared the results of VMD-AI, EEMD-AI and AI models [22].

The empirical results implied that the hybrid VMD models are superior to hybrid EEMD models and single models.

3. Methodologies

3.1. Variational mode decomposition (VMD)

Variational mode decomposition (VMD) is a non-recursive and adaptive data decomposition technique with multi-resolution [21]. It aims to disintegrate an input data X into several discrete subseries called intrinsic mode function (IMF) $X_k (k = 1, 2, \dots, K)$, where each IMF has limited bandwidth in spectral domain and needs to be mostly compact around a center pulse ω_k identified along with the decomposition. The bandwidth of every mode will be computed in steps as: firstly for a mode $X_k(t)$, an associated analytical signal is calculated through the method of Hilbert transform,

$$\left(\delta(t) + \frac{j}{\pi t}\right) * X_k(t). \quad (1)$$

$*$ denotes the convolution and δ is Dirac distribution. The frequency spectrum is then transferred to its respective central frequency estimated,

$$\left[\left(\delta(t) + \frac{j}{\pi t}\right)\right] e^{-j\omega_k t} \quad (2)$$

Finally, H1 Gaussian smoothness of the demodulated signal is utilized to compute the mode bandwidth, which is squared L2 norm of the gradient. After the bandwidth estimation, the resulting constrained variational problem is expressed as

$$\min_{\{X_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \|\partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * X_k(t) \right] e^{-j\omega_k t} \|_2^2 \right\} \quad (3)$$

Such that

$$\sum_{k=1}^K X_k = X \quad (4)$$

where X denotes the decomposed original data, K is the number of modes, $\{X_k\}$ is the set of IMFs $\{X_1, X_2, \dots, X_K\}$ and ω_k is the central pulsation set i.e., $\{\omega_1, \omega_2, \dots, \omega_K\}$.

By combining the quadratic penalty term with Lagrange multipliers, the constrained problem could be converted into an unconstrained problem. The discussion is as follows:

$$\begin{aligned} L(\{X_k\}, \{\omega_k\}, \lambda) := & \alpha \sum_{k=1}^K \|\partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * X_k(t) \right] e^{-j\omega_k t} \|_2^2 \\ & + \|X(t) - \sum_{k=1}^K X_k(t)\|_2^2 + \langle \lambda(t), X(t) - \sum_{k=1}^K X_k(t) \rangle \end{aligned} \quad (5)$$

where $\lambda(t)$ is Lagrangian multiplier and α represents the balance parameter of the data fidelity constraint. In order to deal with this problem, the alternate direction multiplier method (ADMM) is employed to solve the saddle point of the augmented Lagrangian. It is believed that bi-directional update of X_k and ω_k is helpful to the analysis process of VMD, and the solutions of X_k and ω_k is expressed as follows:

$$\hat{X}_k^{n+1} = \frac{\hat{X}(\omega) - \sum_{i \neq k} \hat{X}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (6)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{X}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{X}_k(\omega)|^2 d\omega} \quad (7)$$

where $\hat{X}(\omega)$, $\hat{X}_i(\omega)$, $\hat{\lambda}(\omega)$, and $\hat{X}_k^{n+1}(\omega)$ denote the Fourier transforms of $X(t)$, $X_i(t)$, $\lambda(t)$ and $X_k^{n+1}(t)$ respectively and n refers to the total iterations number.

3.2. Kernel-based extreme learning machine (KELM)

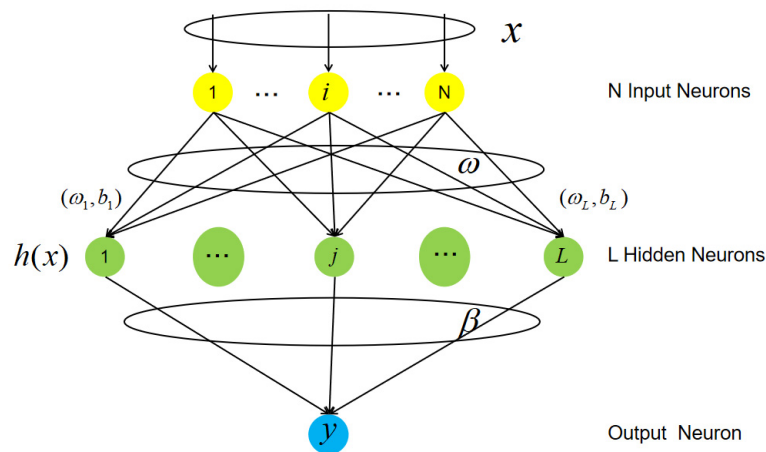


Figure 1. Architecture of ELM model.

Extreme learning machine (ELM) [9] is an improved learning algorithm of single hidden layer feed forward neural network (SLFN). Figure 1 shows its architecture. ELM randomly selects weights between the input layer and the hidden layer without iterative learning, and determines the weights between the hidden layer and output layer by matrix inversion. The following represents the output function of ELM with L hidden node:

$$f(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta \quad (8)$$

where $\beta = [\beta_1, \beta_2, \dots, \beta_L]$ is the vector of the output weight that connects the hidden nodes to the output nodes. $h(x)$ is ELM feature mapping function that maps the data from the N -dimensional input space to the feature space H of L -dimensional hidden layer. $H = \{h_{ij}\} (i = 1, 2, 3, \dots, N, j = 1, 2, \dots, L)$ denotes randomized matrix in the hidden layer of neural network. The output weights β is determined by the least square (LS) approach:

$$\hat{\beta} = H^+ Y \quad (9)$$

where the norm of $\hat{\beta}$ is the minimum and unique among all the LS solutions of the linear system $H\beta = Y$ (Eq. (8)). $Y = [y_1^T, y_2^T, \dots, y_N^T]^T$ indicates the target matrix. H^+ represents the Moore-Penrose generalized inverse [9] of output matrix H for the hidden layer, which is given as

$H = H^T(HH^T)^{-1}$. To get more stable generalization, a regularization parameter C is usually added to HH^T diagonal, and the output weight β is computed by:

$$\beta = H^T \left(\frac{1}{C} + HH^T \right)^{-1} Y \quad (10)$$

Nevertheless, ELM might face the problems of time-consuming and poor stability caused by randomly parameters assignment. A kernel extreme learning machine (KELM) was developed [13] combining the kernel function theory with ELM. The random feature mapping in ELM is replaced by the kernel mapping, and the kernel matrix based on Mercer Theorem is presented by

$$K_{ELM} = HH^T, \text{ and } K_{ELM}(x_i, x_j) = h(x_i)h(x_j). \quad (11)$$

Hence, the output function can be written as:

$$f(x) = h(x)\beta = h(x)H^T \left(K_{ELM} + \frac{1}{C} \right)^{-1} Y = [(x, x_1), (x, x_2), \dots, (x, x_N)] \left(K_{ELM} + \frac{1}{C} \right)^{-1} Y \quad (12)$$

The five kernel functions that meet the Mercer condition include: Sigmoid kernel, Polynomial kernel, Radial basis function kernel and Wavelet kernel etc. This paper selects Radial basis function (RBF) kernel as it can realize the nonlinear mapping and improve the generalization capabilities of ELM [13,14], which is given:

$$K(x_i, x_j) = \exp\{-\|x_i - x_j\|^2 / (2\sigma^2)\} \quad (13)$$

The optimal regularization factor C and kernel width σ are evaluated in trial and error.

3.3. VMD-KELM hybrid model framework

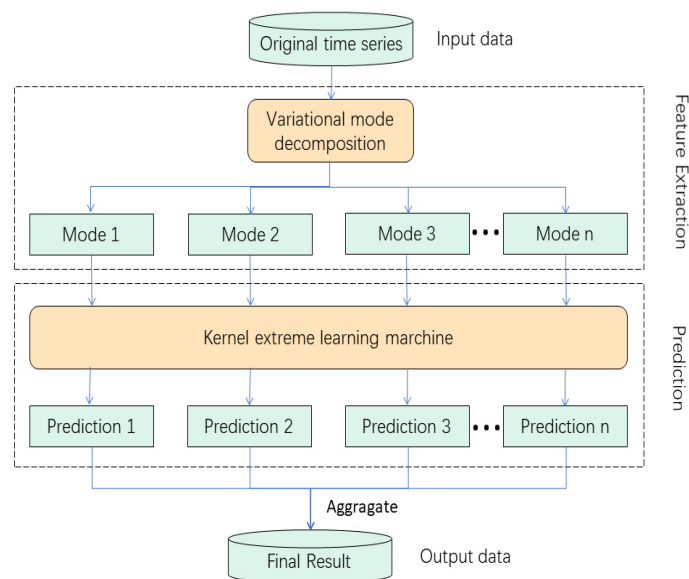


Figure 2. General framework of the proposed model VMD-KELM.

VMD-KELM follows a typical decomposition-ensemble training paradigm, which consists of three major processes, that is, data decomposition through the VMD technique, individual forecasting by the KELM algorithm and results integration through linear aggregation. Figure 2 displays the schematic depiction of implement procedures for VMD-KELM model. Specifically, it can be achieved briefly in the following steps:

- 1) Data decomposing. The historical data series $X(t), (t = 1, 2, \dots, T)$ is separated by VMD technique into an ensemble of modes $\{IMF_k | k = 1, 2, \dots, K\}$, each of which will be a new time series that KELM is prepared to forecast separately.
- 2) Individual forecasting. The KELM is introduced to predict all the extracted IMF series. For each series $IMF_k(t)$, it is split into training and testing set. The exact KELM model is constructed based on the training data, which is further employed to predict the testing dataset. Through the KELM learning process, the prediction output $\widehat{IMF}_k(t)$ is obtained.
- 3) Results ensemble. All the individual forecasted outputs are added linearly to form the final integrated prediction results $\widehat{IMF} = \sum_{k=1}^K \widehat{IMF}_k$.

To illustrate the process more clearly, the pseudo-code of model is described as follows:

Algorithm

//the meaning of letters is in the Note.

// The input nodes are N .

//Data decomposition

- 1) Using VMD to decompose the original data series $\mathbf{X}(t)$, the decomposition process is:

$$\{IMF_k(t) | k = 1, 2, \dots, K; t = 1, 2, \dots, T_{train} + T_{test}\} = \mathbf{X}(t).$$

//Individual forecasting

- 2) For each $\{IMF_k(t) | k = 1, 2, \dots, K; t = 1, 2, \dots, T_{train}\}$ is as input to train the model.

//Forecasting the crude oil price from day $T_{train+1}$ to $T_{train+test}$.

- 3) count $\leftarrow 1$. // prediction counter, which is a temporary variable.

Repeat

- 4) count \leftarrow count + 1.

- 5) Using the well-trained model to predict the $T_{train+count}$ th day's value of crude oil prices.

This can be written as:

$$\mathbf{Output} = \text{Model}(\{IMF_k(t) | k = 1, 2, \dots, K; t = \{T_{train-N+1}, T_{train-N+2}, \dots, T_{train}\}\}).$$

Until count == T_{test}

Note: T_{train} is the length of training set of $X(t)$, T_{test} is the length of testing set of $X(t)$. Model() is the well-trained model, $IMF_k(t)$ is the modes decomposed by VMD, K is the number of decomposed mode, Output is the prediction value of the well-trained model.

3.4. Performance evaluation

3.4.1. Commonly-used metrics

This work adopts seven commonly-used criteria to examine the robustness and superiority of the model from different aspects. Table 1 sums them, in which the Mean absolute error (MAE), Mean absolute percent error (MAPE), Root mean square error (RMSE), Theil inequality coefficient (TIC)

and correlation coefficient R are selected to measure the level accuracy and D_{stat} is used to measure the directional accuracy. The better performance corresponds to smaller MAE, MAPE, RMSE and TIC, larger R and D_{stat} . The closer P_{MAE} , P_{MAPE} and P_{RMSE} to 0, the smaller the difference is between the two models. In this paper, VMD-KELM model is taken as a benchmark method.

Table 1. Commonly-used performance evaluation metrics.

$MAE = \frac{1}{T} \sum_{t=1}^T X_t - \hat{X}_t $	$D_{stat} = \frac{1}{T} \sum_{t=1}^T a_t$
$MAPE = 100 \times \frac{1}{T} \sum_{t=1}^T \left \frac{X_t - \hat{X}_t}{X_t} \right $	$a_t = 1 \text{ if } (\hat{X}_{t+1} - X_t)(X_{t+1} - X_t) \geq 0, \text{ otherwise } a_t = 0.$
$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (X_t - \hat{X}_t)^2}$	$P_{MAE} = \left \frac{MAE_2 - MAE_1}{MAE_2} \right $
$TIC = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^T (X_t - \hat{X}_t)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^T X_t^2 + \frac{1}{T} \sum_{t=1}^T \hat{X}_t^2}}$	$P_{MAPE} = \left \frac{MAPE_2 - MAPE_1}{MAPE_2} \right $
$R = \frac{\sum_{t=1}^T (X_t - \bar{X}_t)^2}{\sqrt{\sum_{t=1}^T (X_t - \bar{X}_t)^2 \sum_{t=1}^T (\hat{X}_t - \bar{\hat{X}}_t)^2}}$	$P_{RMSE} = \left \frac{RMSE_2 - RMSE_1}{RMSE_2} \right $

Note: T is the data length, X_t is the real data, and \hat{X}_t is the predicted data. MAE_1 , $MAPE_1$, $RMSE_1$ denote the criteria of the VMD-KELM.

3.4.2. Diebold-Mariano (DM) test

To illustrate the superiority of the proposed model from statistical perspective, we apply DM test to evaluate the predicting performance of VMD-KELM against other models [47]. The DM test investigates the null hypothesis of forecast accuracy equality against the alternative of different forecasting capabilities between the target model A and its benchmark model B. The DM statistic is written as:

$$S = \frac{\bar{g}}{(\hat{V}_{\bar{g}}/N)^{1/2}} \quad (14)$$

where $\bar{g} = 1/N \sum_{t=1}^N [(x_t - \hat{x}_{A,t})^2 - (x_t - \hat{x}_{B,t})^2]$, $\hat{V}_{\bar{g}} = \gamma_0 + 2 \sum_{l=1}^{\infty} \gamma_l$, ($\gamma_l = cov(g_t, g_{t-l})$). $\hat{x}_{A,t}$ and $\hat{x}_{B,t}$ denote the predicted value of x_t by the forecasting model A, B respectively.

3.4.3. Multi-scale composite complexity synchronization

Multi-scale composite complexity synchronization (MCCS) is a recently-used new method for measuring the synchronization of two data, which can be used to evaluate the synchronization degree between the prediction results and the original time series [24]. MCCS algorithm combines the theory of sample entropy (SampEn) and complexity-invariant distance (CID), which can be briefly

described in the following steps:

1) Given $X_t = \{X_1, X_2, \dots, X_T\}$ and $\hat{X}_t = \{\hat{X}_1, \hat{X}_2, \dots, \hat{X}_T\}$ are the real data and predicted data with length T respectively, the generalized complex-invariant distance (GCID) between them is calculated in the following process:

$$\Delta X = \{X_2 - X_1, X_3 - X_2, \dots, X_T - X_{T-1}\} \quad (15)$$

$$\Delta \hat{X} = \{\hat{X}_2 - \hat{X}_1, \hat{X}_3 - \hat{X}_2, \dots, \hat{X}_T - \hat{X}_{T-1}\} \quad (16)$$

$$\|X\|^q = \sum_{t=1}^T |X_t|^q, \quad \|\hat{X}\|^q = \sum_{t=1}^T |\hat{X}_t|^q \quad (17)$$

$$|X - \hat{X}|^q = \{|X_1 - \hat{X}_1|^q, |X_2 - \hat{X}_2|^q, \dots, |X_T - \hat{X}_T|^q\} \quad (18)$$

Then, GCID between X_t and \hat{X}_t is calculated employing Minkowski distance instead of Euclidean distance, that is,

$$GCID = (X, \hat{X}, q) = |X - \hat{X}|^q \times \frac{MAX\{\|\Delta X\|^q, \|\Delta \hat{X}\|^q\}}{MIN\{\|\Delta X\|^q, \|\Delta \hat{X}\|^q\}} \quad (19)$$

where q is set to 2 here, which denotes the power exponent.

2) Calculate the composite complexity synchronization (CCS) between X_t and \hat{X}_t combining SampEn and GCID. Firstly, the SampEn is computed for X_t (the same for \hat{X}_t),

$$SampEn(X, m, \epsilon) = -\log \left[\frac{c^{m+1}(\epsilon)}{c^m(\epsilon)} \right] \quad (20)$$

where m is the space-dimension set as 2 and ϵ is the tolerance that equals to k ($0.1 \leq k \leq 0.25$) times the standard deviation of the data. Lastly, CCS is measured as:

$$CCS = (X, \hat{X}, q) = [SampEn(|X - \hat{X}|^q, m, \epsilon) \times GCID(X, \hat{X}, q)]^{\frac{1}{q}} \quad (21)$$

3) Compute the MCCS values. MCCS approach considers the multiple time scales of CSS. Firstly for X_t and \hat{X}_t , the coarse-grained sequences with scale factor τ , $X(\tau) = \{X_1^\tau, X_2^\tau, \dots, X_{\lfloor \frac{T}{\tau} \rfloor}^\tau\}$

and $\hat{X}(\tau) = \{\hat{X}_1^\tau, \hat{X}_2^\tau, \dots, \hat{X}_{\lfloor \frac{T}{\tau} \rfloor}^\tau\}$ can be obtained respectively:

$$X_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} X_i, \quad \hat{X}_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} \hat{X}_i, \quad 1 \leq j \leq \left\lfloor \frac{T}{\tau} \right\rfloor \quad (22)$$

where $\left\lfloor \frac{T}{\tau} \right\rfloor$ is the number of coarse-grained sequences that are separated from the original sequence for any τ . Then, MCCS between actual and predicted data X_t and \hat{X}_t is

$$MCCS = (X, \hat{X}, q, \tau) = CCS(X^{(\tau)}, \hat{X}^\tau, q) \quad (23)$$

The smaller values of MCCS are, the high the synchronization of two time series is.

4. Data description and preparation

The daily closing prices and volatility of the two typical energy, the Brent crude oil spot and WTI crude oil spot are selected as the prediction samples. The original prices dataset are comprised of 2000 daily observations for Brent oil from Oct 07, 2013 to Aug 16, 2021, and 2000 daily observations of WTI oil from Aug 28, 2013 to Aug 16, 2021, which are gathered from the energy information administration (EIA). The realized volatility RV_t at time t for the daily prices is calculated by:

$$RV_t = \sqrt{\frac{1}{\rho_t} \sum_{t=1}^{\rho_t} (r_t - \bar{r}_t)^2} \quad (24)$$

where ρ_t is the number of days remaining after time t , r_t is the logarithmic returns of daily prices, defined as $r_t = \log P_t - \log P_{t-1}$ ($t = 1, 2, \dots, T$) and \bar{r}_t is the average mean of r_t . The realized volatility is the value obtained by observing how much the crude oil price has changed during ρ_t days, which is considered as the historical volatility. We take 7-day volatility as the prediction target. Figure 3 shows the evolution dynamics of daily closing prices and 7-day volatility. Further the dataset of daily prices and volatility are partitioned into the training set that accounts for first 80% of the samples of 1600 data points (1594 data points for volatility) and the testing set that accounts for the last 20% of the samples with 400 data points (399 points for volatility) respectively. That is, the training set of Brent crude oil prices lasts from Oct 07, 2013 to Jan 15, 2020, and testing set lasts from Jan 16, 2020 to Aug 16, 2021. For WTI oil, the prices dataset is trained from November 29, 2011 to Jan 10, 2020 and tested from Jan 11, 2020 to December Aug 16, 2021.

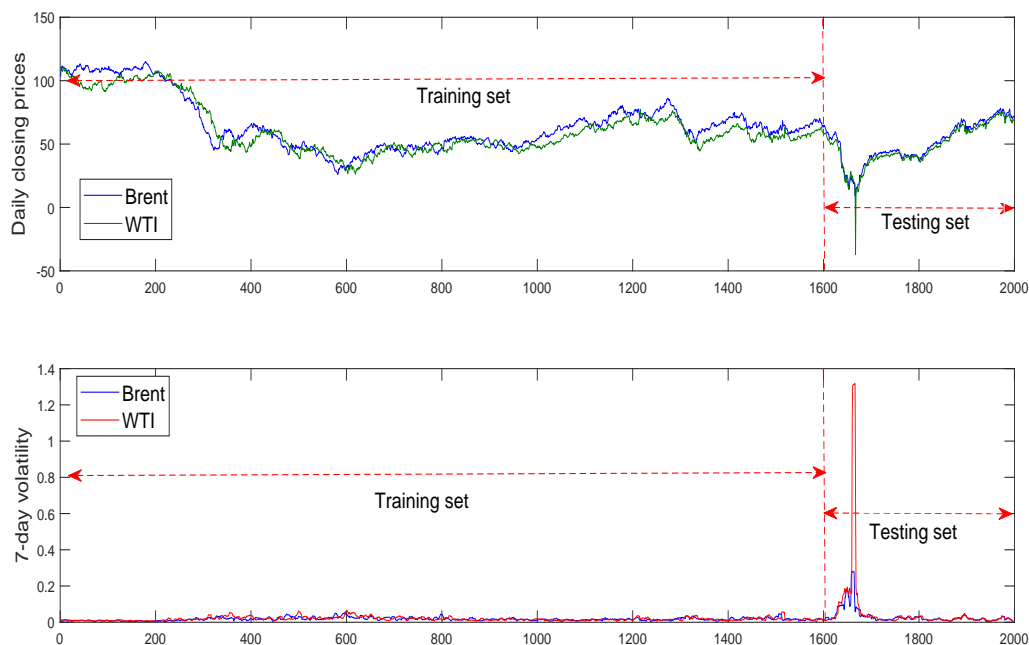


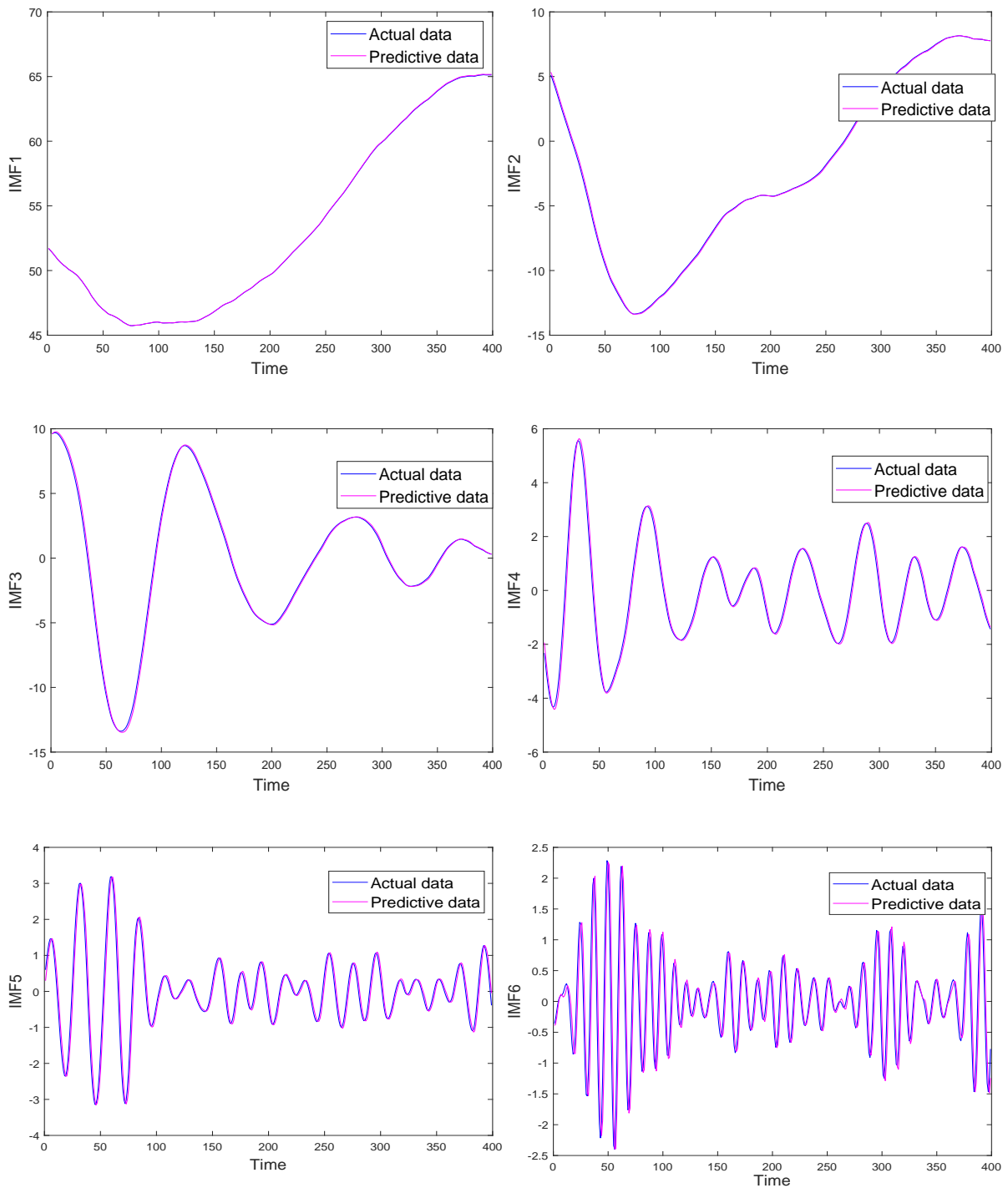
Figure 3. Daily closing prices and 7-day volatility of Brent and WTI crude oil.

5. Empirical results and discussions

5.1. Forecasting prices by different models

In this subsection, we employ the proposed VMD-KELM hybrid model to make a prediction of crude oil prices. The performance analysis of VMD-KELM is conducted by comparing it with other three type approaches, including single models (KELM, ELM and back propagation neural network (BPNN)), VMD-based models (VMD-ELM and VMD-BPNN) and EEMD-based models (EEMD-KELM, EEMD-ELM and EEMD-BPNN). According to the “decomposition and ensemble” learning paradigm, firstly the prices are decomposed by VMD technique. For comparison, the number of decomposed components K by VMD is set to be the same as that obtained by the EEMD technique [50], which can adaptively decompose the original series data without any pre-set parameters. Both Brent and WTI oil, have 11 decomposed IMFs representing different local oscillations embodied in the price series. Later, the corresponding KELM prediction model is constructed for each composed IMF subseries. In parameters setting of the KELM model, a historical lag of order 5 is taken for the energy price series, which means there are five input nodes. It is determined by autocorrelation and partial correlation analysis. For KELM, when the input node is limited, the number of hidden layer nodes will be the same as the input nodes number under the effect of kernel function, which is also 5. So we also set hidden layer nodes as 5 for ELM and BPNN as an experimental comparison. That is, the $5 \times 5 \times 1$ (where the input and output parameters are numeric) is set for all the prediction models. Suitable parameters of the regularization C and kernel width σ in KELM are selected basically based on trials and errors approach. C is searched within the range $[10, 100]$ with the interval 10 and it finds the value of 100 for all the models. The kernel width σ is searched ranging from $[0, 1]$ with the interval 0.1. Suitable σ of 0.1, 0.3 and 0.1 is found for KELM, EEMD-KELM and VMD-KELM respectively for Brent oil prices, while that of 1, 0.9 and 0.2 for WTI oil prices.

Figure 4 demonstrates comparisons between the predicted results and every original subcomponent on the testing set for Brent crude oil prices by VMD-KELM. Roughly, the prediction results of the subsequences with lower frequencies are closer to the true values than those with higher frequencies, which shows that VMD-KELM algorithm seems to have a higher prediction accuracy for low-frequency information. Table 2 further lists the MAPE values of VMD-KELM predicting all the IMF series of Brent and WTI oil, in which the results of other hybrid models are also listed for comparison. It can be seen clearly that the smoothest IMF1 with the lowest frequency characteristics has the lowest MAPE value in almost all sequences, indicating that the hybrid model has a better prediction effect on low frequency data than high frequency data. Taking Brent crude oil as an example, VMD-KELM, VMD-ELM and VMD-BPNN models have the maximum MAPE value for IMF7 prediction, while IMF11 produces the maximum MAPE value in EEMD-based hybrid models.



continued on next page

Figure 4. Predictive results of decomposed subseries on testing set for Brent crude oil prices by VMD-KELM.

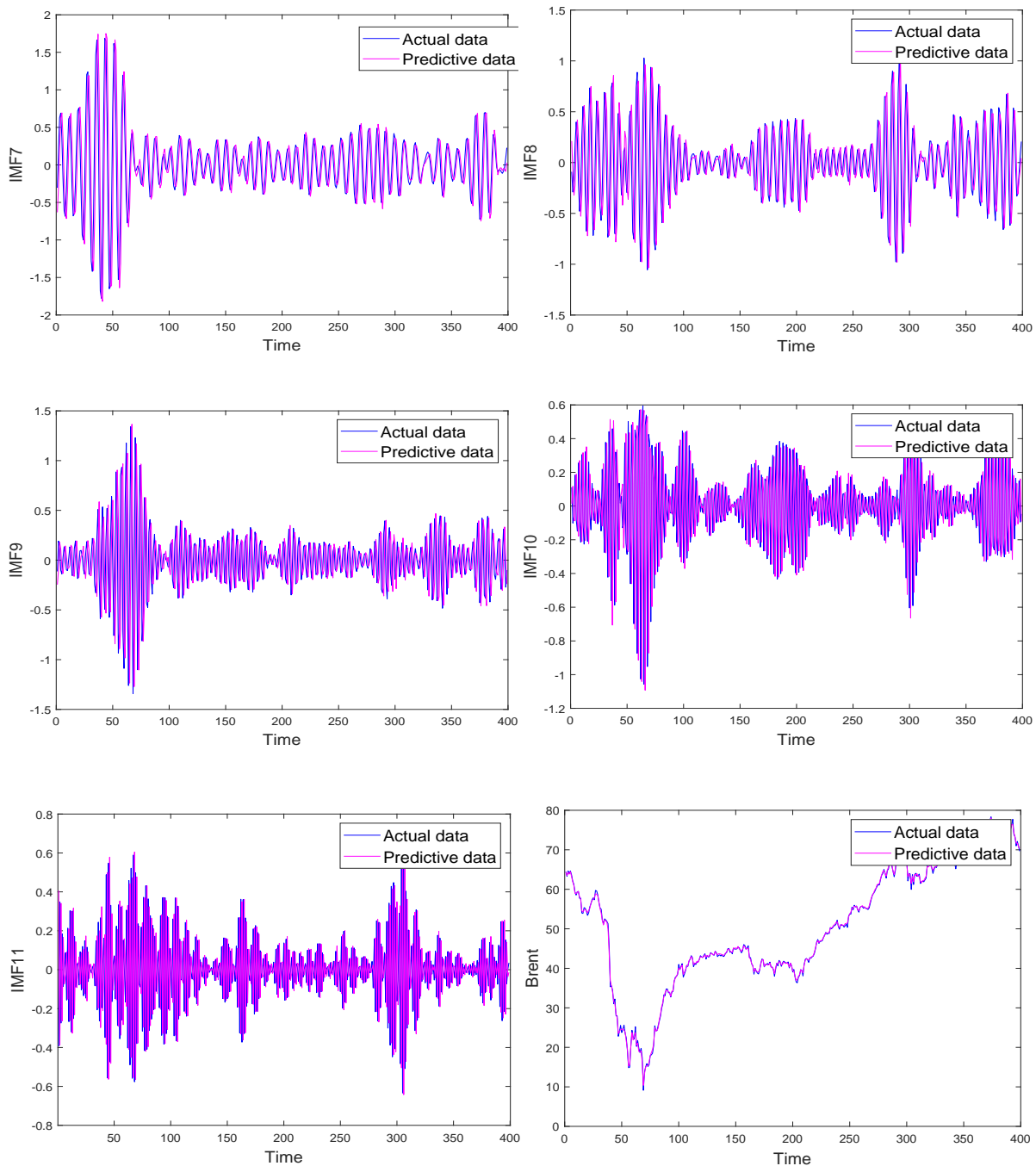


Figure 4. Predictive results of decomposed subseries on testing set for Brent crude oil prices by VMD-KELM.

Figure 5 shows the predictive results about testing data of Brent and WTI oil price for different models. Due to COVID-19, crude oil prices fluctuated sharply around August 20, 2020. The curves are all gathered together, indicating that the predicted curve of each model is highly close to the real price curve. According to enlarged view, we can see that the VMD-KELM has a good performance of the large fluctuation. The prediction value of VMD-KELM is much closer to the true value. So, we can boldly guess that target model has the highest accuracy for crude oil price prediction.

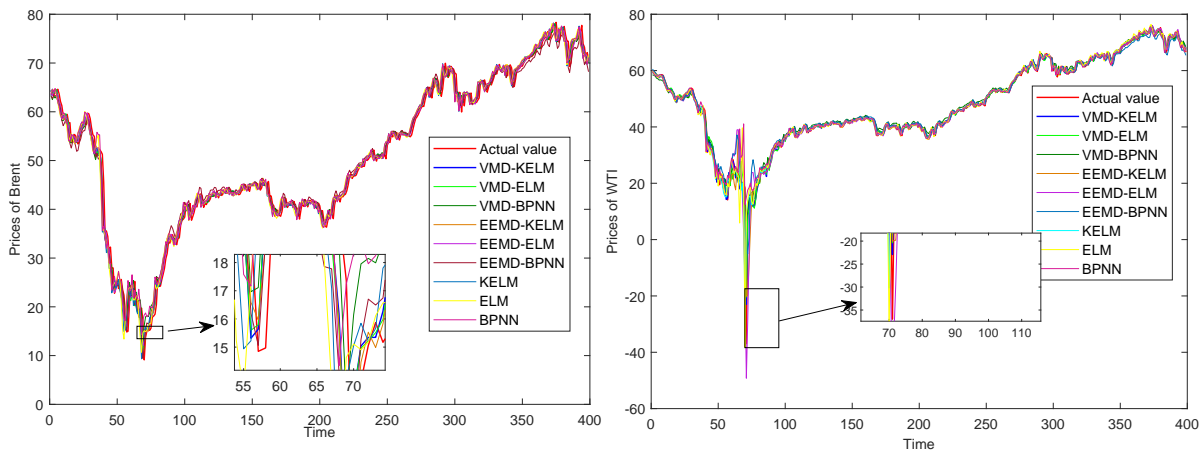


Figure 5. Predictive values of Brent and WTI crude oil prices for different models.

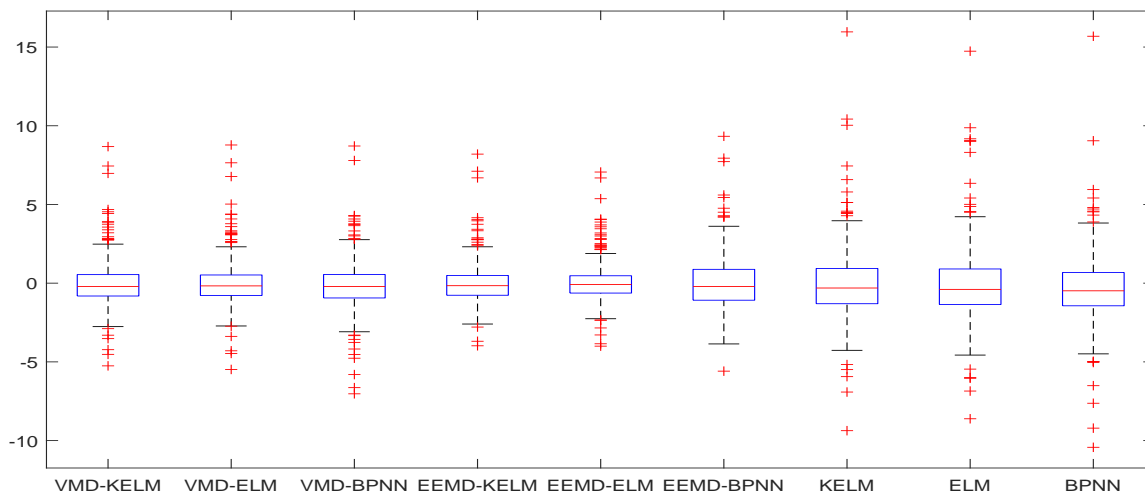


Figure 6. Relative errors and boxplots of predictive results of Brent for different models.

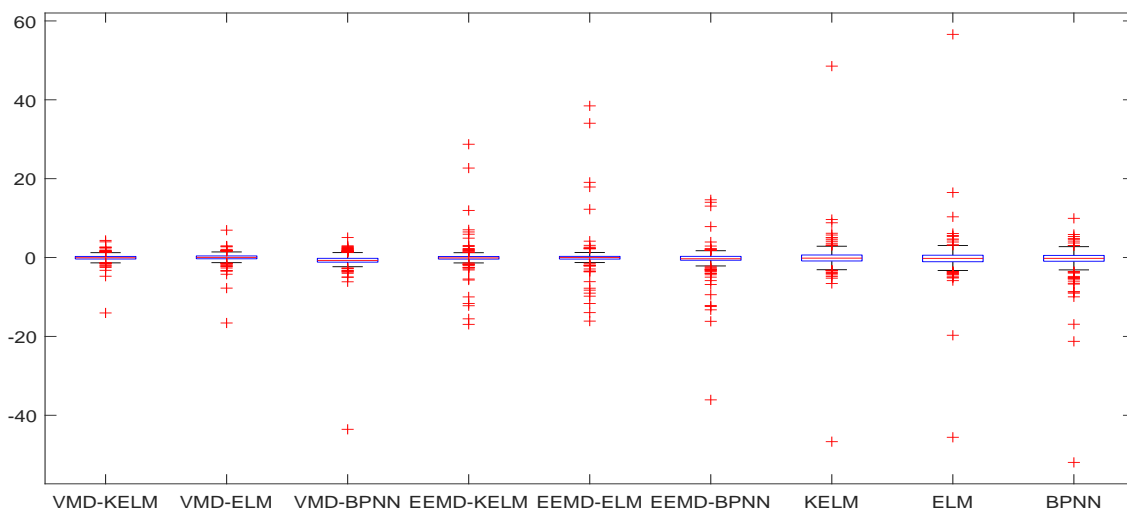


Figure 7. Relative errors and boxplots of predictive results of WTI for different models.

Table 2. The MAPE(%) values of the decomposed subseries for oil prices.

Model	Brent										
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10	IMF11
VMD-KELM	0.0343	7.293	25.59	186.0	122.9	210.4	2510	287.2	381.9	606.3	288.1
VMD-ELM	0.0207	7.930	28.02	179.9	120.0	229.1	2041	286.9	400.1	624.8	289.9
VMD-BPNN	0.641	10.22	28.33	203.4	125.9	220.9	2373	293.9	391.9	595.6	286.1
EEMD-KELM	0.0329	0.0231	0.0193	1.638	4.558	12.55	28.72	83.57	351.2	173.7	628.1
EEMD-ELM	0.0214	0.2185	0.0546	1.323	6.327	13.47	41.30	86.78	335.5	218.0	384.9
EEMD-BPNN	3.6422	0.0453	0.9992	15.47	12.33	13.21	26.47	80.87	341.2	193.6	784.4

Model	WTI										
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10	IMF11
VMD-KELM	0.0311	12.62	32.26	127.4	428.1	394.2	620.1	364.8	430.2	475.8	212.0
VMD-ELM	0.1838	11.74	32.9	134.1	411.1	419.9	601.2	367.7	377.3	465.1	216.7
VMD-BPNN	0.8669	14.55	40.14	114.9	395.9	421.4	580.1	374.2	350.1	473.4	232.1
EEMD-KELM	0.0097	0.0245	0.0221	1.305	9.034	16.23	66.03	78.39	122.1	256.3	269.9
EEMD-ELM	0.132	0.0581	0.0454	4.989	11.02	16.33	73.78	118.6	121.7	440.4	366.8
EEMD-BPNN	0.0625	0.281	2.095	5.926	8.808	17.41	80.52	100.3	163.4	397.4	441.5

Note: IMF1-IMF11 of EEMD, which originally corresponds to the high-frequency to low-frequency, is arranged in reverse and here represents the low-frequency to high-frequency to keep consistent with results of VMD.

The boxplots of the relative forecasting errors for each model are shown in Figures 6 and 7. The prediction errors of single model are much larger than those of hybrid model. The median of the VMD-KELM model is closest to 0, and the absolute values of the upper and lower quartiles are the smallest. The smallest values also appear for the absolute values of the upper and lower limits, the number of outliers as well. All above indicate that the relative error of the target model is relatively smaller and more concentrated, illustrating the closest predicted values of VMD-KELM model to the true values. As for WTI crude oil, similar results are exhibited and it is evidently observed that single models have larger predictive errors than the hybrid models.

In order to quantitatively measure the predictive effect of each model, various evaluation metrics are calculated in Table 3, and the bar graphs of them are exhibited in Figure 8. In the figure, the heights of MAE, MAPE, RMSE, TIC bars for the single models are almost twice than those for the hybrid models, while the heights of R and D_{stat} bars are quite lower for the single models. It indicates that the hybrid models based on EEMD algorithm and VMD algorithm have better performances than single models. Specifically, taking three KELM-related models for WTI crude oil as an example (in Table 3), MAE, MAPE, RMSE and TIC value of VMD-KELM model is reduced about 61.1, 63.5, 72.4 and 72.3% respectively, compared to the results of single KELM model, while the value of R and D_{stat} is increased about 2.7 and 11.8% respectively. Compared to KELM model, EEMD-KELM model reduces MAE value about 36.3%, MAPE value about 7.5%, RMSE value about 29.4% and TIC value about 29.5%, but increases R about 1.5% and D_{stat} about 9.2%. These illustrate that the prediction accuracy of each model is really improved by decomposing the original time series. Therefore, it is meaningful to choose the decomposition-ensemble prediction strategy in this paper. Besides, the evaluation criteria of VMD-based hybrid model are smaller than those of EEMD-based hybrid model. Taking Brent oil as an example, MAPE value of VMD-KELM,

VMD-ELM and VMD-BPNN model is respectively reduced about 28.5, 27.6 and 20.1% compared to that of the corresponding EEMD-KELM, EEMD-ELM and EEMD-BPNN model. It manifests that VMD algorithm can better eliminate the noise of the original data than EEMD algorithm, resulting in a better prediction accuracy. Moreover, KELM model performs best among three single models, illustrating that KELM model which is chosen as the basic target model is appropriate. Among the three VMD models, the first four evaluation criteria of the target model hold the smallest values, while the latter two are slightly higher, suggesting the best prediction of VMD-KELM in forecasting crude oil price.

Table 3. Evaluation results of prediction of crude oil prices for different models.

Model	Brent					
	MAE	MAPE (%)	RMSE	TIC	R	D_{stat}
VMD-KELM	0.3225	0.8056	0.4353	0.0137	0.9957	0.9047
EEMD-KELM	0.4446	1.1509	0.6089	0.0121	0.9967	0.8571
VMD-ELM	0.3606	0.9035	0.4975	0.0135	0.9958	0.9047
EEMD-ELM	0.4812	1.2579	0.6873	0.0116	0.9996	0.8596
VMD-BPNN	0.5339	1.618	0.7834	0.0151	0.9746	0.8647
EEMD-BPNN	0.8045	1.8545	0.9815	0.0159	0.97	0.7293
KELM	1.0783	2.7863	1.5855	0.0213	0.9896	0.7719
ELM	1.1907	3.1036	1.7152	0.022	0.9889	0.7494
BPNN	1.1401	3.1711	1.6508	0.021	0.9899	0.7644
Model	WTI					
	MAE	MAPE (%)	RMSE	TIC	R	D_{stat}
VMD-KELM	0.5225	1.4808	1.0326	0.0102	0.9978	0.8496
EEMD-KELM	0.8564	3.7555	2.6411	0.026	0.9856	0.8296
VMD-ELM	0.5986	1.7360	1.2502	0.0123	0.9968	0.8320
EEMD-ELM	0.9765	4.5677	3.4017	0.0334	0.976	0.8571
VMD-BPNN	1.1306	3.1945	2.5345	0.0248	0.9868	0.7619
EEMD-BPNN	1.1313	4.2848	2.8686	0.0282	0.983	0.7544
KELM	1.3442	4.0613	3.7429	0.0369	0.9709	0.7594
ELM	1.4958	4.6922	4.1951	0.0413	0.9633	0.7469
BPNN	1.4177	4.6027	3.4843	0.0343	0.9343	0.7419

The P_{MAE} , P_{MAPE} and P_{RMSE} values of each model can be intuitively Table 4, where VMD-KELM is taken as the benchmark. These metrics quantitatively show how much higher is the error of each model than that of VMD-KELM. Taking MAPE of Brent crude oil as an instance, VMD-KELM improves the prediction accuracy of VMD-ELM, VMD-BPNN, EEMD-KELM, EEMD-ELM, EEMD-BPNN and single models by 10.83, 50.21, 30.00, 35.95, 56.56% and more than 60% respectively. Besides, VMD-ELM show the smallest P_{MAE} , P_{MAPE} and P_{RMSE} than other models, implying that the forecasting performance of VMD-ELM is only second to VMD-KELM in Brent price prediction, meanwhile VMD-KELM is only second to VMD-KELM in WTI price prediction.

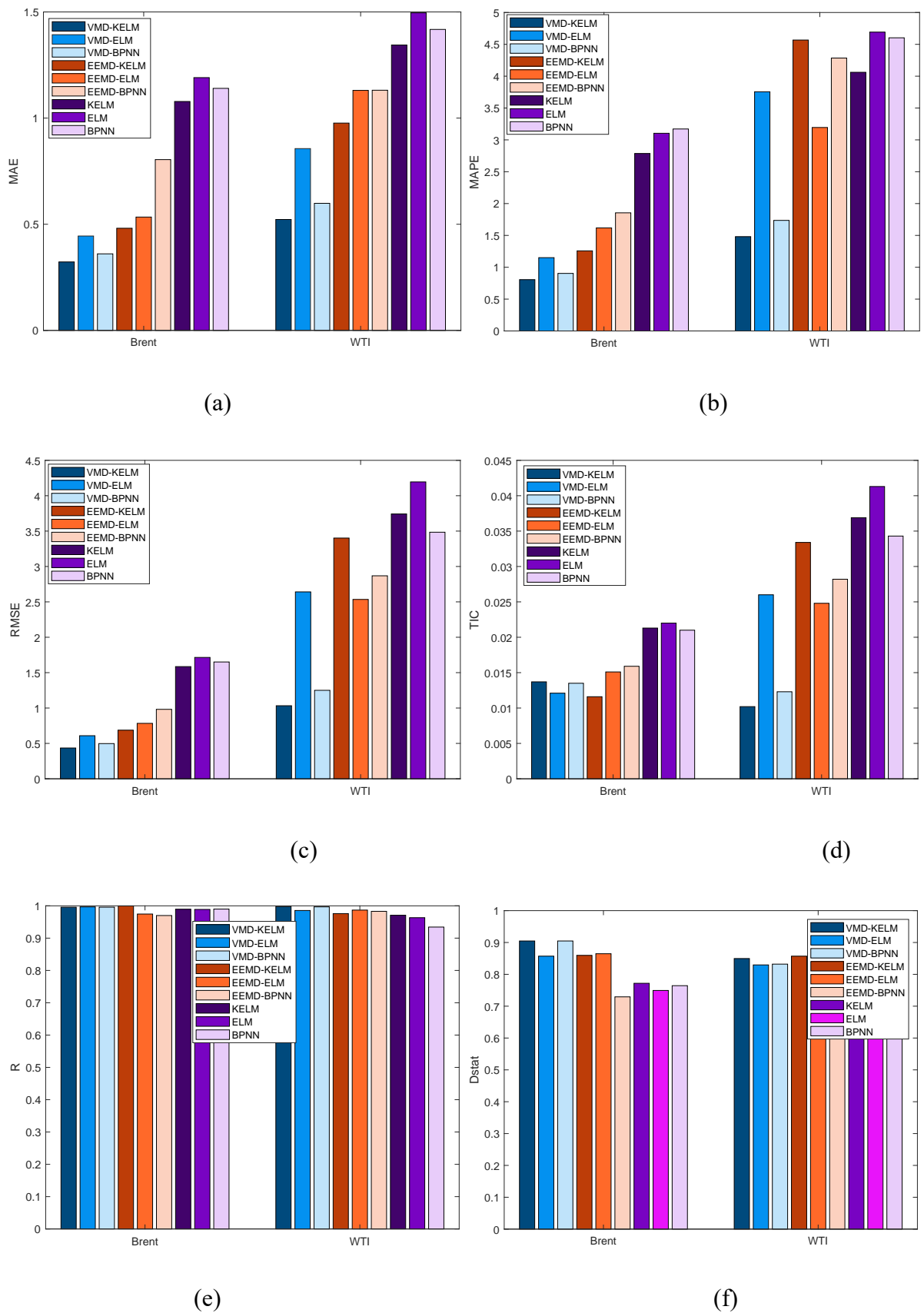


Figure 8. Evaluation criteria of crude oil price prediction for different models.

Table 4. P_{MAE} , P_{MAPE} , P_{RMSE} for different models (VMD-KELM as benchmark).

Model	Brent			WTI		
	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}
EEMD-KELM	0.2746	0.3000	0.2850	0.3899	0.6057	0.6090
VMD-ELM	0.1057	0.1083	0.1250	0.1271	0.1470	0.1741
EEMD-ELM	0.3298	0.3595	0.3666	0.4649	0.6758	0.6964
VMD-BPNN	0.3960	0.5021	0.4443	0.5379	0.5365	0.5926
EEMD-BPNN	0.5991	0.5656	0.5564	0.5381	0.6544	0.6400
KELM	0.7009	0.7108	0.7254	0.6113	0.6354	0.7241
ELM	0.7292	0.7404	0.7462	0.6507	0.6844	0.7539
BPNN	0.7171	0.7459	0.7363	0.6314	0.6783	0.7036

The DM test is further performed to verify the superiority of the proposed model VMD-KELM against the compared models from a statistical point of view. The results are listed in Table 5. For Brent data, the VMD-KELM as target model, the p-values (except for VMD-ELM) are much smaller than the significance level of 1%, demonstrating that the VMD-KELM model has statistically better prediction effectiveness under the confidence level 99%. Compared with VMD-ELM, the VMD-KELM is statistically significant better at confidence level of 90%. Besides, as for the WTI data, the confidence level reaches 95%. The results statistically confirm the superiority of VMD-KELM.

Table 5. DM test results of different models against VMD-KELM.

	EEMD-KELM	VMD-ELM	EEMD-ELM	VMD-BPNN	EEMD-BPNN	KELM	ELM	BPNN
VMD-KELM	5.2443	1.4851	4.8520	-2.6545	-5.4205	-4.6961	-5.4806	-4.6585
Brent	(0.000)	(0.069)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
VMD-KELM	-2.2787	-1.7631	-2.1400	-2.2571	-2.3538	-1.6846	-1.7724	-1.7260
WTI	(0.011)	(0.039)	(0.016)	(0.012)	(0.009)	(0.046)	(0.038)	(0.042)

Note: The p-values of DM-test are given in the brackets.

Furthermore, the MCCS is applied as an accuracy evaluation method to describe the synchronization degree between the predictive and the actual data series. Figure 9 plots the MCCS between the forecasting values and the actual prices for different predicting models at different scales. The corresponding results at scale τ of 1–5 are shown in Tables 6 and 7. With the scale increasing, the MCCS value for each index shows a downward trend. Clearly, the MCCS values of the VMD-KELM model are the smaller ones in majority of scales for both Brent and WTI, which means that the forecasting results of the proposed VMD-KELM model have the largest synchrony with the actual crude oil prices. It indicates that the VMD-KELM model is propitious for enhancing the accuracy in crude oil price forecasting. For the Brent crude oil, the top three lines in the plot imply the relatively departure of the forecasting results of the ELM, BPNN and KELM models from actual ones. The bottom three lines are VMD-KELM, EEMD-ELM and EEMD-KELM, implying the relatively closer of predictive value to the actual one. These illustrate the superiority of hybrid models. The similar behavior of MCCS is shown for WTI crude oil.

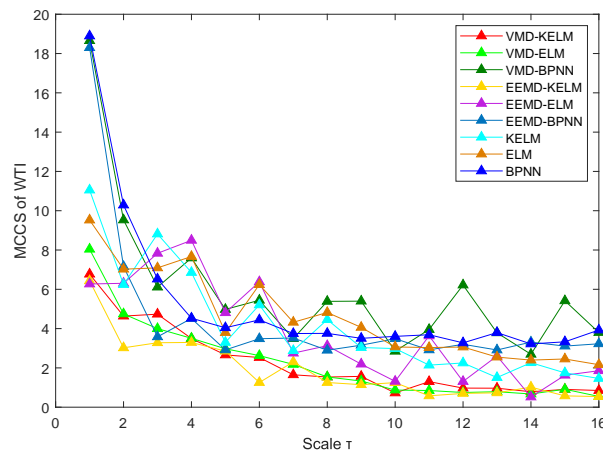


Figure 9. MCCS of Brent and WTI crude oil prices for different models

Table 6. MCCS value of different predictive models for Brent crude oil price.

Scale	VMD			EEMD			Single		
	KELM	ELM	BPNN	KELM	ELM	BPNN	KELM	ELM	BPNN
$\tau = 1$	24.05	23.48	25.90	21.02	23.15	27.10	26.00	31.28	25.13
$\tau = 2$	10.96	10.20	12.88	9.473	10.66	14.45	18.29	21.55	18.34
$\tau = 3$	9.160	8.735	9.507	8.048	7.260	10.53	15.64	17.67	16.26
$\tau = 4$	6.608	6.34	8.714	6.329	5.632	10.11	10.03	10.81	11.42
$\tau = 5$	4.202	4.188	5.919	4.399	3.414	7.103	7.931	9.016	9.236

Table 7. MCCS value of different predictive models for WTI crude oil price.

Scale	VMD			EEMD			Single		
	KELM	ELM	BPNN	KELM	ELM	BPNN	KELM	ELM	BPNN
$\tau = 1$	6.771	8.040	18.65	6.433	6.270	18.30	11.05	9.525	18.89
$\tau = 2$	4.636	4.747	9.530	3.022	6.306	7.158	6.243	7.024	10.29
$\tau = 3$	4.734	3.999	6.111	3.280	7.833	3.581	8.815	7.090	6.523
$\tau = 4$	3.494	3.506	7.594	3.295	8.494	4.516	6.857	7.669	4.528
$\tau = 5$	2.659	2.974	4.976	2.853	4.808	2.916	3.298	3.799	4.043

5.2. Forecasting volatility by different models

In this subsection we will investigate the prediction effects of the referred nine models on the 7-volatility of crude oil. From the prediction result of volatility, it can be understood whether the proposed model has scalability. Figure 10 depicts the 7-day realized volatility and predictive volatility of energy prices on testing set for different models. It is seen that the realized volatility has many fluctuations making prediction more difficult relative to the price. All the forecasting curves in the graph are very compact and highly close to the curves of the actual volatility. Through the enlarged view of the peak, it is observable that VMD-KELM fits the sharp volatility caused by the COVID-19 very well. The prediction value of VMD-KELM is closest to the actual value. Besides,

the ELM model exceeds the true value of the peak, while BP is much lower than the true value of the peak. We can guess that the target model of VMD-KELM has a better prediction effect in prediction of volatility series when the fluctuation is large.

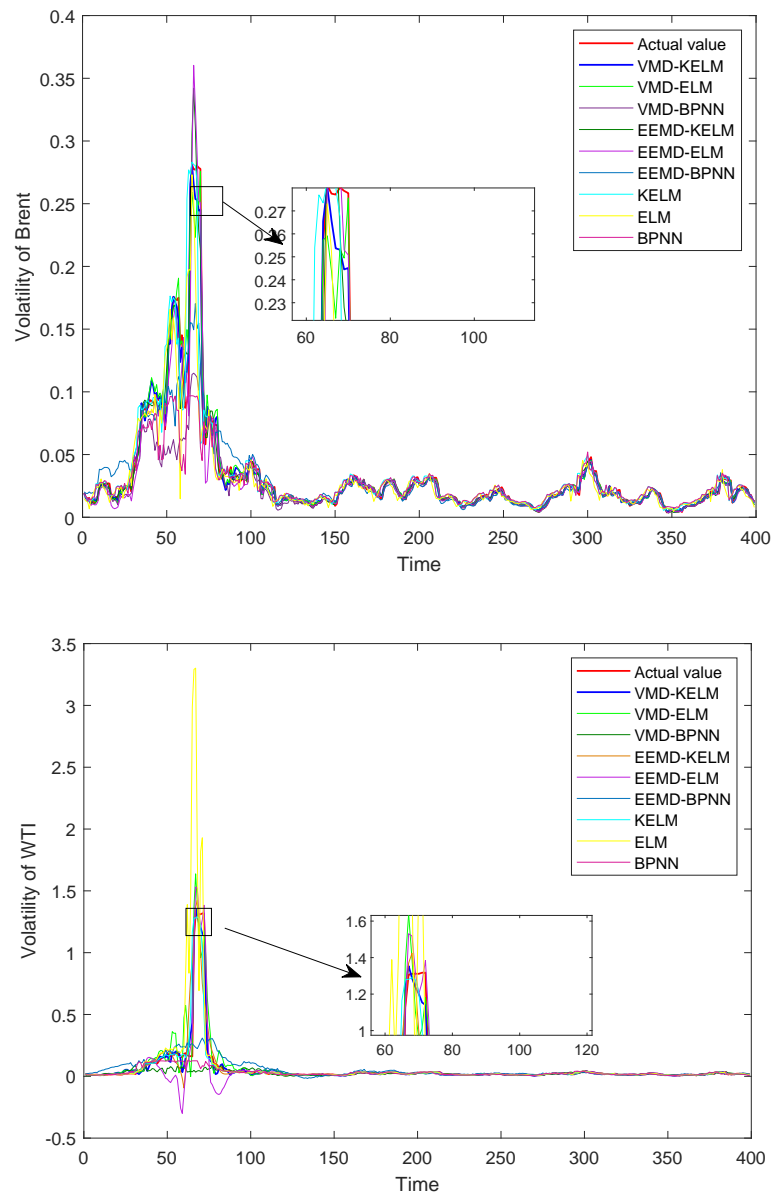


Figure 10. Predictive value of volatility of Brent and WTI crude oil for different models.

Figure 11 shows bar graphs of the performance evaluation metrics for each model intuitively. It is significantly obvious that the first four criteria of single models are quite higher than those of hybrid models, while the latter two are lower. It implies that overall prediction effect of all hybrid models is superior to that of single models. Table 8 lists the specific values of the evaluation measures of each model. Taking WTI volatility for an example, KELM model holds the largest R and D_{stat} while ELM model has the largest MAE, MAPE, RMSE. Besides, the error of a single model is significantly greater than that of its corresponding hybrid EEMD model and hybrid VMD model. Specifically, MAE, RMSE and TIC of VMD-KELM model are respectively 0.0092, 0.0268

and 0.0752, decreased by 24.5, 64.2, 64.4% compared to KELM model. The difference of MAPE value is small. In directional accuracy, D_{stat} value is 0.7043 that is similar to the KELM model (0.7168). It illustrates that prediction accuracy of each model for volatility is really improved when the volatility time series is decomposed by the VMD decomposition algorithm. Through the observation, it can be seen that a single KELM model has a good predictive effect on extreme data. Therefore, the decomposition-ensemble prediction strategy is a promising strategy for predicting volatility. The error of the VMD-based hybrid model is smaller than that of the EEMD based hybrid model. Taking Brent volatility for instance, the MAPE value of the VMD-KELM model (7.4215) is reduced about 13.1% compared to the EEMD-KELM model (8.5444), the MAPE value of the VMD-ELM model (7.9708) is reduced about 12.8% compared to the EEMD-ELM model (9.1458), and the VMD-BPNN model's MAPE value (13.1988) is reduced about 32.7% compared to the EEMD-BPNN model (19.6297). It shows that the VMD algorithm can better extract the features of the original time series, bringing about a higher prediction result. Besides, the errors of VMD-KELM are the smallest. Among the three VMD models, VMD-KELM has the smallest level accuracy measure of MAPE and the largest direction accuracy measure of D_{stat} . For WTI crude oil, the conclusions are similar. Therefore, we can conclude that VMD-KELM has a superiority in prediction of crude oil volatility.

Table 8. Evaluation criteria values of different models for prediction of 7-day volatility.

Model	Brent					
	MAE	MAPE (%)	RMSE	TIC	R	D_{stat}
VMD-KELM	0.0023	7.4215	0.0048	0.0453	0.9937	0.7419
VMD-ELM	0.0035	8.5444	0.0094	0.0875	0.9758	0.7318
VMD-BPNN	0.0027	7.9708	0.0067	0.0619	0.988	0.7243
EEMD-KELM	0.0039	9.1458	0.01	0.0934	0.9728	0.6917
EEMD-ELM	0.0086	13.1988	0.0279	0.3249	0.7618	0.7168
EEMD-BPNN	0.008	19.6297	0.0207	0.2156	0.8765	0.604
KELM	0.007	22.7126	0.0205	0.19	0.8789	0.6917
ELM	0.0088	25.0281	0.0224	0.2238	0.8538	0.6366
BPNN	0.0102	25.2103	0.0292	0.3319	0.7362	0.6541
Model	WTI					
	MAE	MAPE (%)	RMSE	TIC	R	D_{stat}
VMD-KELM	0.0092	18.0679	0.0268	0.0752	0.9878	0.7043
VMD-ELM	0.0179	20.5352	0.078	0.2197	0.8919	0.6591
VMD-BPNN	0.0192	24.0717	0.0597	0.1599	0.9381	0.7043
EEMD-KELM	0.0212	25.4667	0.0657	0.1757	0.9245	0.6391
EEMD-ELM	0.0385	33.0761	0.1705	0.8008	0.5812	0.609
EEMD-BPNN	0.0403	81.5429	0.1426	0.5329	0.5622	0.5614
KELM	0.0122	15.1719	0.0749	0.2115	0.9007	0.7168
ELM	0.0396	23.5341	0.2367	0.4466	0.4513	0.6917
BPNN	0.0307	26.3427	0.1611	0.7159	0.3555	0.6366

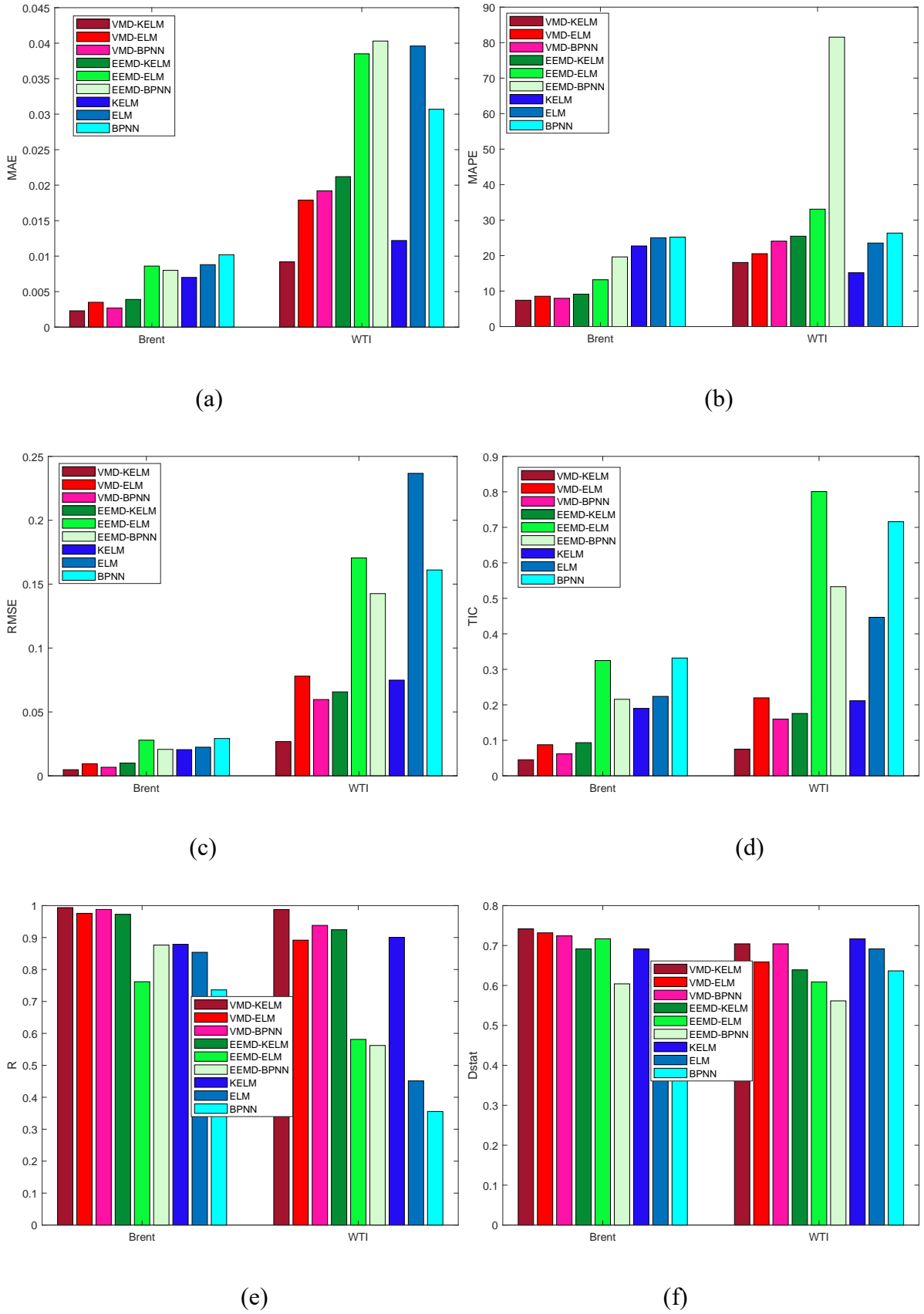


Figure 11. Comparisons of volatility forecasting for different models.

6. Conclusions

The prediction of crude oil price and volatility is of great significant for guidance for investors and government decision-making. To improve their accuracy of prediction, this paper aims to introduce an effective decomposition-ensemble model, VMD-KELM, to predict both crude oil price and volatility. In the model, the VMD method is used to decompose the original series into several subseries with different frequencies, and then each subseries is predicted by the kernel extreme learning machine. Finally, the prediction result is obtained by summarizing the results for the subsequences. Brent, WTI daily prices and their 7-day volatility series are applied to evaluate the forecasting accuracy. The following aspects of conclusions are obtained: 1) The VMD-KELM model which combines the advantage of variational mode decomposition and kernel extreme learning machine, possesses more robust decomposition ability and strong prediction ability. 2) Compared to the ELM and BPNN algorithm, KELM model shows a better prediction performance with relatively low values of MAE, MAPE, RMSE, TIC and higher values of R and D_{stat} . 3) The superiority of the decomposition-ensemble strategy is demonstrated by the fact that the prediction accuracies of the hybrid models are greatly higher than those of the single models. Besides, the VMD-based model has a better prediction effect than its EEMD-based model, especially for forecasting the high-frequency time series. 4) The proposed VMD-KELM model shows a more powerful ability than other models in improving the precision of forecasting crude oil volatility. However, this work only considers the impact of historical price or volatility data on the crude oil forecasting. There are many other potential factors (including structured and non-structured text data etc.) that affect crude oil price and volatility. We think they are worth for further being analyzed and considered and hope it will be instructive for further research.

Acknowledgment

The work was partially supported by the Fundamental Research Funds for the Central Universities (No. FRF-BR-20-04B).

Conflict of interests

The authors declare there is no conflict of interests.

References

1. G. Wu, Y. Z. Zhang, Does China factor matter? An econometric analysis of international crude oil prices, *Energy Policy*, **72** (2014), 78–86.
2. Y. Zhao, J. P. Li, L. Yu, A deep learning ensemble approach for crude oil price forecasting, *Energy Econ.*, **66** (2017), 9–16.
3. L. Yu, Y. Zhao, L. Tang, Ensemble forecasting for complex time series using sparse, representation and neural networks, *J. Forecast.*, **36** (2017).
4. M. Monge, L. A. Gil-Alana, P. D. G. Fernando, Crude oil price behaviour before and after military conflicts and geopolitical events, *Energy*, **120** (2017), 79–91.
5. D. W. Jones, P. N. Leiby, I. K. Paik, Oil price shocks and the macroeconomy: what has been learned since 1996, *Energy J.*, **25** (2004), 1–32.

6. S. Lahmiri, Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices, *IEEE Syst. J.*, **11** (2017), 1907–1910.
7. S. R. Li, Y. L. Ge, Crude oil price prediction based on a dynamic correcting support vector regression machine, *Abstr. Appl. Anal.*, **2013** (2013), 666–686.
8. L. Yu, W. Dai, L. Tang, J. Wu, A hybrid grid-GA-based LSSVR learning paradigm for crude oil price forecasting, *Neural Comput. Appl.*, **27** (2016), 2193–2215.
9. G. B. Huang, Q. Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feed forward neural networks, in *2004 IEEE international joint conference on neural networks*, **2** (2004), 985–990.
10. G. B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, *IEEE Trans. Syst., Man, Cybern.*, **42** (2012), 513–529.
11. L. Tang, Y. Wu, L. Yu, A randomized-algorithm-based decomposition-ensemble learning methodology for energy price forecasting, *Energy*, **157** (2018), 526–538.
12. L. Zhang, P. Suganthan, A survey of randomized algorithms for training neural networks, *Inf. Sci.*, **364** (2016), 146–155.
13. G. B. Huang, An insight into extreme learning machines: random neurons, random features and kernels, *Cognit. Comput.*, **6** (2014), 376–390.
14. S. Ding, Y. Zhang, X. Xu, L. Bao, A novel extreme learning machine based on hybrid kernel function, *J. Comput. Phys.*, **8** (2013), 2110–2117.
15. M. Pal, A. E. Maxwell, T. A. Warner, Kernel-based extreme learning machine for remote sensing image classification, *Remote Sens. Lett.*, **4** (2013), 853–862.
16. C. Chen, W. Li, H. Su, K. Liu, Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine, *Remote Sens.*, **6** (2014), 5795–5814.
17. W. Y. Deng, Q. H. Zheng, Z. M. Wang, Cross-person activity recognition using reduced kernel extreme learning machine, *Neural Networks*, **53** (2014), 1–7.
18. S. Shamshirband, K. Mohammadi, H. L. Chen, G. N. Samy, D. Petković, C. Ma, Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: A case study for Iran, *J. Atmos. Sol.-Terr. Phys.*, **134** (2015), 109–117.
19. I. Majumder, P. K. Dash, R. Bisoi, Variational mode decomposition based low rank robust kernel extreme learning machine for solar irradiation forecasting, *Energy Convers. Manage.*, **171** (2018), 787–780.
20. R. Jammazi, C. Aloui, Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling, *Energy Econ.*, **34** (2012), 828–841.
21. K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Process.*, **62** (2014), 531–544.
22. J.C. Li, S.W. Zhu, Q.Q. Wu, Monthly crude oil spot price forecasting using variational mode decomposition, *Energy Econ.*, **83** (2019), 240–253.
23. R. Bisoi, P. K. Dash, S. P. Mishra, Modes decomposition method in fusion with robust random vector functional link network for crude oil price forecasting, *Appl. Soft Comput. J.*, **80** (2019), 475–493.
24. H. Niu, K. Xu, W. Wang, A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network, *Appl. Intell.*, **50** (2020), 4296–4309.
25. L. Huang, J. Wang, Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network, *Energy*, **151** (2018), 875–888.

26. K. Kanjilal, S. Ghosh, Dynamics of crude oil and gold price post 2008 global financial crisis-New evidence from threshold vector error-correction model, *Resour. Policy*, **52** (2017), 358–365.
27. L. Lin, Y. Jiang, H. Xiao, Z. Zhou, Crude oil price forecasting based on a novel hybrid long memory GARCH-M and wavelet analysis model, *Phys. A*, **543** (2020), 123532.
28. Y. Xiang, X. H. Zhuang, Application of ARIMA model in short-term prediction of international crude oil price, *Adv. Mater. Res.*, **798–799** (2013), 979–982.
29. M. Marchese, L. Kyriakou, M. Tamvakis, F. D. Iorio, Forecasting crude oil and refined products volatilities and correlations: New evidence from fractionally integrated multivariate GARCH models, *Energy Econ.*, **88** (2020), 104757.
30. Y. Wei, Y. Wang, D. Huang, Forecasting crude oil market volatility: Further evidence using GARCH-class models, *Energy Econ.*, **32** (2010), 1477–1484.
31. Y. Wang, C. Wu, Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models? *Energy Econ.*, **34** (2012), 2167–2181.
32. T. Klein, T. Walther, Oil price volatility forecast with mixture memory GARCH, *Energy Econ.*, **58** (2016), 46–58.
33. H. Hu, L. Wang, S. X. Lv, Forecasting energy consumption and wind power generation using deep echo state network, *Renewable Energy*, **154** (2020), 598–613.
34. A. Azadeh, M. Moghaddam, M. Khakzad, V. Ebrahimipour, A flexible neural network fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting, *Comput. Ind. Eng.*, **62** (2012), 421–430.
35. H. Niu, J. Wang, Financial time series prediction by a random data-time effective RBF neural network, *Soft Comput.*, **18** (2014), 497–508.
36. H. Chiroma, S. Abdulkareem, T. Herawan, Evolutionary neural network model for West Texas Intermediate crude oil price prediction, *Appl. Energy*, **142** (2015), 266–273.
37. J. Barunk, B. Malinsk, Forecasting the term structure of crude oil futures prices with neural networks, *Appl. Energy*, **164** (2016), 366–379.
38. M. G. Wang, L. F. Zhao, R. J. Du, C. Wang, L. Chen, L. X. Tian, et al., A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms, *Appl. Energy*, **220** (2016), 480–495.
39. J. Wang, J. Wang, Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations, *Energy*, **102** (2016), 365–374.
40. M. M. Mostafa, A. A. El-Masry, Oil price forecasting using gene expression programming and artificial neural networks, *Econ. Modell.*, **54** (2016), 40–53.
41. B. Wu, L. Wang, S. Wang, Y. R. Zeng. Forecasting the U.S. oil markets based on social media information during the COVID-19 pandemic, *Energy*, **226** (2021), 120403.
42. B. Wu, L. Wang, S. X. Lv, Y. R. Zeng. Effective crude oil price forecasting using new text-based and big-data-driven model, *Measurement*, **168** (2021), 108468.
43. J. L. Zhang, Y. J. Zhang, L. Zhang, A novel hybrid method for crude oil price forecasting, *Energy Econ.*, **49** (2015), 649–659.
44. L. Yu, H. Xu, L. Tang, LSSVR ensemble learning with uncertain parameters for crude oil price forecasting, *Appl. Soft Comput.*, **56** (2017), 692–701.
45. L. Peng, L. wang, D. Xia, Q. Gao, Effective energy consumption forecasting using empirical wavelet transform and long short-term memory, *Energy*, **238** (2022), 121756.
46. L. Tang, W. Dai, L. Yu, S. Wang, A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting, *Int. J. Inf. Technol. Decis. Making*, **14** (2015), 141–169.

47. L. Yu, Z. S. Wang, L. Tang, A decomposition-ensemble model with data characteristic driven reconstruction for crude oil price forecasting, *Appl. Energy*, **156** (2015), 251–267.
48. Y. X. Wu, Q. B. Wu, J. Q. Zhu, Improved EEMD-based crude oil price forecasting using LSTM networks, *Phys. A*, **516** (2019), 114–124.
49. H. Abdollahi, S. B. Ebrahimi, A new hybrid model for forecasting Brent crude oil price, *Energy*, **200** (2020), 117520.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)