A Learning System Integrating Temporal Convolution and Deep Learning for Predictive Modeling of Crude Oil Price

Tong Niu, *Graduate Student Member, IEEE*, Jianzhou Wang, Haiyan Lu, *Senior Member, IEEE*, Wendong Yang, and Pei Du

***Abstract*—Accurately crude oil price prediction remains challenging so far. Despite the abundant research achievements of crude oil price prediction, most of them emphasize the linear and deterministic modeling, which cannot adequately capture the complex nonlinear characteristics and uncertainties involved, thus impeding further developments in the field. In this study, a novel learning system with the aim of obtaining deterministic and probabilistic predictions is presented to model the nonlinear dynamics in crude oil price, composed by the modules of recurrence analysis, outlier detection, data preprocessing, feature selection, predictive modeling based on deep learning, and system evaluation. In particular, the temporal convolution is developed to perform the feature selection, thus improving the generalization of the system. Additionally, the extensions, including the predictive performance test evaluation, convergence investigation, and sensitivity analysis, are carried out. The experimental simulations show that the proposed system can yield the deterministic and probabilistic predictions with higher accuracy and feasibility compared with the benchmarks considered, further indicating its effectiveness.**

***Index Terms*—Deep learning, Oil price, Deterministic and probabilistic prediction, Temporal convolution**

I. Introduction

CRUDE oil, as an important form of non-renewable energy, is usually considered “the blood of modern industries” due to its great socio-economic significance to facilitate national security [1], international economy, and financial sectors being aligned with crude oil markets. However, dramatic fluctuations of crude oil price (COP) caused by geopolitical conflicts [2], economic growth [3], weather situations, and exchange rate [4] will destroy the balance between the supply-side and demand-side in the oil market, further having a significant negative impact on these aspects listed above. In fact, the predictive modeling of COP, with the aim of minimizing prediction bias, is one of the most effective solutions to alleviate the negative impacts above, which can facilitate the decision-making relevant to this field. For example, oil price prediction with a desired accuracy is needed for governments to formulate scientific strategies of oil development, which is also of significance to develop risk-management strategies in response to oil shocks for relevant financial institutions. However, performing the COP prediction has remained challenging due to its nonlinear and chaotic characteristics.

Although some complex factors aforementioned render crude oil price prediction to be a great challenge, the relevant studies being carried out in academia have achieved lots of research results. More specifically, the mainstream relevant to the crude oil price prediction can be divided into two categories, namely, traditional econometric models and artificial intelligence models. Models from the first category can be characterized by statistical distribution theory and linear regression theory. The main econometric models usually contain the error correction model [5], autoregressive integrated moving average model (ARIMA) [6], generalized autoregressive conditional heteroskedasticity (GARCH) [7], and hybrid econometric model (i.e., ARIMA-GARCH) [8].

Although these econometric models above have high computational efficiency, the linear type and limited distribution assumption being difficult to model nonlinear time series, such as the COP, render them difficult to obtain desired predictive performance. Additionally, the further improvement of predictive performance using econometric models faces technological difficulties. For example, they cannot be incrementally learned and can only re-estimate its parameters when the studied dataset is updated.

Currently, the state-of-the-art methods of the crude oil price prediction depend on artificial intelligence models, including artificial neural network (ANN) [9], artificial optimization algorithms [10], and other data-driven methods, such as empirical mode decomposition (EMD) [11]. Besides, some hybrid models [12] based on econometric models and artificial intelligence models above were presented in previous studies to perform the crude oil price prediction. However, most of ANNs suffer from the following drawbacks: 1) over-fitting problem; 2) easy to fall into local optimum. Further, the EMD-type models are theoretically based on an iteration algorithm,

This work was supported by the major project of the national social science of China (Grant No. 17ZDA093) and the China Scholarship Council (No. 201808210268). (Corresponding author: Jianzhou Wang.)
T. Niu is with the Center for Energy, Environment & Economy Research, Zhengzhou University, China (e-mail: niuchenren621@126.com). J. Wang (Corresponding author), W. Yang, and P. Du are with the School of Statistics, Dongbei University of Finance and Economics, China (e-mail: wangjz@dufe.edu.cn, hshwendong@hotmail.com, and renshengdp@126.com). H. Lu is with the School of Computer Science, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia (e-mail: Haiyan.Lu@uts.edu.au).

lacking mathematical foundations, and being sensitive to noise and sampling [13]. More importantly, most of the models based on artificial intelligence models focus mainly on deterministic prediction (i.e., point prediction) and lack of uncertainties analysis and modeling of COP, which cannot effectively provide accurate quantifications of uncertainties surrounding the deterministic prediction.

Deep learning, as an excellent alternative, is a promising direction for the crude oil price prediction due to its excellent ability to model complex nonlinear time series. However, the deep learning has not received enough attention for the probabilistic and deterministic prediction of COP, which may hinder further development in the field. Therefore, a novel learning system based on deep learning and temporal convolution, including the modules of recurrence analysis, outlier detection, data preprocessing, feature selection, deterministic and probabilistic prediction, and system evaluation, is presented in this study to model the nonlinear relationship in COP, with the aim of yielding the desired deterministic and probabilistic predictive results. This study focuses on addressing the following problems:

1) Investigating the inherent recurrent behavior of a dynamic system is challenging due to its complex nonlinearity and chaotic properties. Recurrence analysis, which can reveal its inherent properties from both qualitative and quantitative perspectives, has not received enough attention in the field relevant to COP.

2) Performing outlier detection for nonlinear time series is of significance for improving the generalization and effectiveness of the predictive model. In fact, it is necessary to carry out the outlier detection for COP because of the fact that the uncertainties caused by multiple factors in COP usually bring significant fluctuations of the COP, thus increasing the probability of outliers occurring. However, outlier detection for the COP has not become the focus of current research in the field.

3) Feature selection when performing time series prediction of COP has remained limited because of its complex dynamics. In previous studies, partial autocorrelation function (PACF) was usually utilized to determine appropriate input features. However, the linear computational mechanism of the PACF limits its ability to address complex nonlinear time series.

4) Financial time series usually have long-term memory and complex nonlinearity, which renders its predictive modeling challenging. The ability of traditional econometric models to model the financial time series with nonlinear characteristics is usually restricted due to their linear type.

5) The previous studies of the crude oil price prediction being carried out focused primarily on the deterministic prediction of the COP with the lack of the probabilistic prediction and uncertainty analysis, which largely increases the risk of decision-making in the field.

To facilitate the current research in the field, the following five modules of the proposed learning system are proposed to address the problems above. Further, the physical interpretations of each module are presented as follows.

1) ***Recurrence Analysis Module***: In the module, the embedding dimension and time delay of the COP were calculated using the C-C method. Further, the recurrence plot (RP) and recurrence quantification analysis (RQA) were developed to investigate the complex dynamics in the COP, indicating the presence of outliers from the texture of the RP.

2) ***Outlier Detection Module***: Isolation forest (IF), as an ensemble method, was developed in the learning system to perform outlier detection of the COP. There are two steps when using the IF. In detail, the first step is to establish isolation trees (ITs) based on sub-samples; further, the second step is to calculate the score based on these established ITs, and finally, an outlier can be identified according to its corresponding score under a preset threshold.

3) ***Data Preprocessing Module***: A novel frequency decomposition method (i.e., VMD) was developed in this study to perform the data preprocessing of the COP. In detail, the COP was decomposed into some IMFs. Further, the high-frequency IMF was eliminated to reduce the negative effects of outliers, which is conducive to enhancing the generalization of the prediction model.

4) ***Automatic Feature Selection***: In this study, a temporal convolution model (TCM) is developed to perform automatic feature selection. Further, the output from the TCM will be compressed using a pooling operator, with the aim of reducing the computational complexity of the proposed learning system.

5) ***Prediction Module***: Developing a novel learning system based on deep learning for deterministic and probabilistic prediction of the COP is also a significant focus of this study, which can effectively learn the long-term relationship in the COP, based on its recurrent learning mechanism. Importantly, the properties of the errors yielded by the proposed learning system are investigated based on the theory of probabilistic density estimation in-depth, further producing high-quality prediction intervals of the COP.

The rest of this paper is organized as follows. Section II presents the preliminaries concerning the proposed learning system. The performance evaluation of the learning system is established in Section III. The studied data and its recurrence analysis are investigated in Section IV. Further, the outlier detection of COP is carried out in Section V. Then, two cases are performed to test the performance of the proposed learning system in Section VI. The extensions, including model performance testing, convergence analysis, and sensitivity analysis, are discussed in Section VII. Finally, the conclusions and future scope are put forth in the final section.

II. Proposed Learning System for Crude Oil Price Prediction

In this section, the relevant preliminaries concerning the proposed learning system of crude oil price prediction are presented.

A. Recurrence Analysis

The recurrence analysis can effectively characterize the complex dynamics inherent in a nonlinear time series, such as non-stationarity and periodicity. In this study, the RP and RQA are developed to perform the qualitative and quantitative analysis of the COP, respectively.

1) Recurrence Plot

The RP can be carried out in the phase space of a time series $x=\{x(1),x(2),\cdots,x(n)\}$, based on the two parameters embedding dimension (m) and time delay (τ) calculated by the C-C method

[14]. Further, the corresponding phase space of the time series (**PS**) can be constructed, specified as

$$\mathbf{PS} = \begin{bmatrix} x(1) & \cdots & x(n-(m-1)\tau) \\ x(1+\tau) & \cdots & x(n-(m-2)\tau) \\ x(2+\tau) & \cdots & x(n-(m-3)\tau) \\ \vdots & \ddots & \vdots \\ x(1+(m-1)\tau) & \cdots & x(n-1) \end{bmatrix} \quad (1)$$

where each column in the matrix **PS** represents a point in the phase space.

Further, the matrix **PS** can be translated into an RP according to the following formula.

$$\mathbf{R}_{i,j}(\eta) = \Theta\left(\eta - \left\|x_i - x_j\right\|\right),\ \ i,j = 1,2,\ldots,N \quad (2)$$

where x_i and x_j represent the different points in the phase space. η is the threshold that controls whether the point x_i in the phase space can be considered the recurrence point of x_j, usually set to 0.4–0.5 times the standard deviation of the COP [15]. $\Theta(\cdot)$ denotes the Heaviside function, defined as

$$\Theta(x) = \begin{cases} 1, & \left\|x_i - x_j\right\| < \eta \\ 0, & \left\|x_i - x_j\right\| \geq \eta \end{cases} \quad (3)$$

Notably, if $\Theta(x)$ is equal to 1, a recurrence point will appear on this point (x_i, x_j) in the RP. The identification methods of the RP can refer to [15-16].

2) Recurrence Quantification Analysis

Although the RP can reveal the complex characteristics of a dynamic system, it cannot facilitate the quantitative characterization of the recurrence phenomenon in the RP. Thus, some quantitative indicators used for quantifying the RP were developed in this study, including the recurrence rate (*RR*), determinism (*DET*), entropy (*ENT*), laminarity (*LAM*), and trapping time (*TT*) [15-16].

The *RR* is a crucial metric measuring the percentage of recurrence points in the RP, mainly indicating the recurrence density (i.e., predictability) of a dynamic system. The formula of the RR can be specified as

$$RR = \frac{1}{N^2}\sum_{i,j=1}^{N}\Theta\left(\eta - \left\|x_i - x_j\right\|\right), i \neq j \quad (4)$$

where N denotes the number of recurrence points.

In the RP, the longer the diagonal, the more deterministic the dynamic system is. The determinacy of the dynamic system can be measured by the metric *DET*, defined as the following formula.

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)} \quad (5)$$

where l denotes the length of a diagonal line, $P(l)$ is the corresponding probability, and $P(l)$ can be formulated as $P(l) = \sum_{i,j=1}^{N}\left(1-\mathbf{R}_{i-1,j-1}(\eta)\right)\left(1-\mathbf{R}_{i+l,j+l}(\eta)\right)\prod_{k=0}^{l-1}\mathbf{R}_{i+k,j+k}(\eta)$. Additionally, l_{min} denotes the length of the shortest diagonal, set to 2 in this study.

The metric *ENT* measures the complexity of the RP from the diagonal lines, which can be formulated as

$$ENT = -\sum_{l=l_{min}}^{N} p(l)\cdot In\left(p(l)\right) \quad (6)$$

where $p(l) = P(l)/N_l$ and N_l denotes the number of diagonal lines.

The metric *LAM* reflects the percentage of recurrence points from the vertical lines, which can be defined as follows.

$$LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)} \quad (7)$$

where v denotes the length of vertical line. v_{min} represents minimum length of vertical line, which was set to 2 in this study. Besides, $P(v) = \sum_{i,j=1}^{N}\left(1-\mathbf{R}_{i,j}(\eta)\right)\left(1-\mathbf{R}_{i,j+v}(\eta)\right)\prod_{k=0}^{v-1}\mathbf{R}_{i,j+k}(\eta)$.

Finally, the metric *TT* calculates the average of the vertical lines, which can be presented as

$$TT = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=v_{min}}^{N} P(v)} \quad (8)$$

B. Outlier Detection based on Isolation Forest

Outlier detection is a necessary procedure when performing time series prediction, which facilitates improving predictive performance and reducing the probability of the “over-fitting” problem. The IF model [17] with the advantage that it can effectively avoid the problems of swamping and masking was developed in this study to perform the outlier detection of the COP. In the IF, each data point has a score according to the corresponding average path length, and the score is used for identifying outliers according to the set threshold. The smaller the path length of a data point, the more the corresponding data point will be isolated. The studied IF has some distinguished merits [17], as compared to principal component analysis [18], the six-sigma rule [19], including

- The IF model has high efficiency with linear computational complexity;
- The IF model can effectively address large-scale dataset, with high dimensions;
- The IF model effectively avoids the effects of swamping and masking.

C. Variational Mode Decomposition

Decomposing a complex nonlinear time series into some intrinsic mode functions, with the aim of removing the redundant components inherent in the time series, is of important significance to improve the generalization of predictive models. Variational mode decomposition (VMD) [13] is a novel approach of signal decomposition, widely applied into the fields of the power systems [20], electrocardiogram analysis [21], and intelligent fault diagnosis [22]. The VMD has the following merits [13] compared to other data decomposition methods, such as the family of EMD, including (1) it has the ability to address the data with noise; (2) it effectively avoids the problem of recursive sifting; (3) it has an adaptative mechanism to determine the corresponding filter boundaries. The essential computational mechanism of the

VMD can be specified as follows.

The VMD model can be considered a constrained variational problem:

$$\begin{aligned} &\min_{m_k, c_k} \left\{ \sum_k \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \cdot m_k(t) \right] \cdot e^{-jc_k t} \right\|_2^2 \right\}, \\ &s.t. \ \sum_k m_k = f(t) \end{aligned} \tag{9}$$

where $\delta(\cdot)$ denotes the Dirac distribution. m and c represent the mode set and the corresponding center frequency, respectively. m_k denotes k-th mode decomposed. f represents the original time series to be decomposed.

The constrained variational problem above can be translated into the following unconstrained problem based on the two parameters quadratic penalty η and Lagrangian multipliers λ, which can be presented as

$$\begin{aligned} L(m_k, c_k, \lambda) = \eta \sum_k \left\| \partial_t \left[\left(\left(\delta(t) + \frac{j}{\pi t} \right) \cdot m_k(t) \right) \cdot e^{-jc_k t} \right] \right\|_2^2 \\ + \left\| f - \sum_k m_k \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k m_k \right\rangle \end{aligned} \tag{10}$$

The alternating direction method was utilized to solve (10) by updating m_k, c_k, and λ, as shown in (11), (12), and (13), respectively, until (14) is satisfied. After the iteration is completed, m_k can be obtained.

$$m_k^{n+1}(c) = \frac{\hat{f}(c) - \sum_{i \neq k} \hat{m}_k^{n+1}(c) + \frac{\hat{\lambda}^n(c)}{2}}{1 + 2\alpha (c - c_k)^2} \tag{11}$$

$$c_k^{n+1} = \frac{\int_0^\infty c \left| \hat{m}_k^{n+1}(c) \right|^2 dc}{\int_0^\infty \left| \hat{m}_k^{n+1}(c) \right|^2 dc} \tag{12}$$

$$\hat{\lambda}^{n+1}(c) = \hat{\lambda}^n(c) + \tau \left(\hat{f}(c) - \sum_k \left| \hat{m}_k^{n+1}(c) \right| \right) \tag{13}$$

$$\frac{\left\| \hat{m}_k^{n+1} - \hat{m}_k^n \right\|^2}{\left\| \hat{m}_k^n \right\|^2} < \varepsilon \tag{14}$$

where ε denotes a set tolerance.

D. Deep Convolutional Gated Recurrent Unit Model

In this section, a novel deep convolutional gated recurrent unit (DCGRU) model is presented based on one-dimensional temporal convolution. The preliminaries of the proposed DCGRU are introduced as follows.

1) One-dimension Temporal Convolution (Conv1D)

Considering the input sequence $\boldsymbol{X} = (x_{i,j}) \in \mathbb{R}^{M \times N}$, $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$, and D filters with the length of E and the depth of N in the Conv1D, the weight matrix of each filter can be obtained as $\mathbf{W}^k = (w_{i,j}^k) \in \mathbb{R}^{E \times N}, i = 1, 2, \ldots, E, j = 1, 2, \ldots, N, k = 1, 2, \ldots, D$, and the corresponding bias of each filter can be expressed as $b^k \in \mathbb{R}$. Meanwhile, the filter traverses the length of $\boldsymbol{X}$ by the step s for T positions. Further, the output of the Conv1D layer (i.e., the matrix $\mathbf{H} = (h_{t,k})_{t=1,2,\ldots,T}^{k=1,2,\ldots,D} \in \mathbb{R}^{T \times D}$) can be defined as

$$\begin{aligned} h_{t,k} = f\left(\sum_{i=1}^{N} \sum_{j=1}^{E} w_{i,j}^k \cdot x_{(t-1)\cdot s + t, j} + b^k \right), \\ \forall t \in \{1, 2, \ldots, T\}, \forall k \in \{1, 2, \ldots, D\} \end{aligned} \tag{15}$$

where $f(\cdot)$ is the activation function, set to $ReLU$ function in this study.

After the procedure of one-dimension convolution, performing the pooling operation for the results from the Conv1D layer is a crucial step to improve the efficiency of the predictive model. Further, the results of the pooling operation can be expressed as $\mathbf{H} = (h_{t,j})_{t=1,2,\ldots,T}^{j=1,2,\ldots,N}$, where the definition of $h_{t,j}$ is shown as (16).

$$\begin{aligned} h_{t,j} = max\left(\left\{ x_{(t-1)\cdot s + k, j} \right\} \middle| \forall k \in (1, 2, \ldots, E) \right), \\ \forall t \in \{1, 2, \ldots, T\}, \forall j \in \{1, 2, \ldots, N\} \end{aligned} \tag{16}$$

2) Gated Recurrent Unit (GRU)

The GRU [23] is an improved version of the long short-term memory model (LSTM), which has a more compact structure, implying a higher computational efficiency. The GRU is composed of the update gate (U) and the reset gate (R). In detail, the update gate controls the extent to which the hidden state in the previous moment affects the current moment; further, the reset gate is used to control the extent of forgetting the information in the previous moment. Importantly, the reset gate and update gate are used to learn the short-term and long-term dependency in time series, respectively. The definitions of the update gate and reset gate are presented as

$$\begin{cases} R_t = \sigma(X_t \cdot W_{xR} + H_{t-1} \cdot W_{HR} + b_R) \\ U_t = \sigma(X_t \cdot W_{xU} + H_{t-1} \cdot W_{HU} + b_U) \\ \tilde{H}_t = ReLU(X_t \cdot W_{x\tilde{H}} + R_t \circ H_{t-1} \cdot W_{\tilde{H}\tilde{H}} + b_H) \\ H_t = U_t \circ H_{t-1} + (1 - U_t) \circ \tilde{H}_t \end{cases} \tag{17}$$

where W denotes recurrent weight. Further, H_t and $\tilde{H}_t$ are the hidden state and candidate hidden state, respectively. b is the bias vector. The $ReLU(\cdot)$ represents the activation function.

3) Fully Connected Layer (FCL)

In this study, the information stream from the DCGRU is further addressed in FCL, finally obtaining the final predictive results, as shown in (18).

$$\begin{aligned} O_k = \sum_{j=1}^{l} ReLU\left(\sum_{i=1}^{n} W_{ij} x_i - a_j \right) \cdot W_{jk} - b_k, \\ i = 1, 2, \ldots, n, j = 1, 2, \ldots, l, k = 1, 2, \ldots, m \end{aligned} \tag{18}$$

where W_{ij} and W_{jk} denote the input-hidden weight matrix and hidden-output weight matrix, respectively. Further, the parameters a and b signify the thresholds of hidden layer and output layer, respectively. $\mathbf{X} = \{x_i | i = 1, 2, \ldots, n\}$ is the input matrix. Finally, these parameters n, l, m represent the size of input layer, hidden layer, and output layer, respectively

III. Performance Evaluation for the Learning System

In this section, the predictive performance assessment of the proposed learning system is carried out for the deterministic and probabilistic prediction results.

A. Deterministic Prediction Evaluation

The statistical indicators, including mean absolute error (MAE), mean absolute percent error (MAPE), root mean square error (RMSE), direction accuracy of forecasts (DA), Pearson's correlation coefficient (R), and the index of agreement of forecasts (IA) were employed to test the predictive accuracy of the deterministic prediction yielded by the proposed learning system, which were widely applied in other fields, such as air quality forecasting [24]. Notably, the MAE, MAPE, and RMSE are negative indicators, while the DA, IA, and R are positive indicators.

B. Probabilistic Prediction Evaluation

To measure the performance of the probabilistic prediction results (i.e., prediction intervals), some popular metrics, including prediction interval coverage probability (PICP), prediction interval normalized average width (PINAW) [25-28], coverage width criterion (CWC) [25], Winkler score (Score) [29], and accumulated width deviation (AWD) [30], were used to evaluate the quality of the constructed prediction intervals. Therein, the resolution and reliability of the prediction intervals can be quantified by the metrics PINAW and PICP, respectively; further, the indicators CWC, AWD, and Score can evaluate the sharpness of the prediction intervals.

IV. Studied Data and Recurrence Analysis

In this study, the studied data (i.e., daily and weekly COP), collected from Europe Brent spot price (Dollars per Barrel), was utilized to validate the performance of the proposed learning system. To train the learning system with the aim of learning the nonlinear dynamics in the COP, the daily COP with the length of 7883 (sampling from May 20, 1987 to June 11, 2018) and the weekly COP with the length of 1622 (sampling from May 15, 1987 to June 8, 2018) were divided into the training set and test set. In detail, the daily COP from May 20, 1987 to December 31, 2009 and the weekly COP from May 15, 1987 to December 25, 2009 were considered the training sets. Further, the daily COP covering from January 4, 2010 to June 11, 2018 and weekly COP ranging from January 1, 2009 to June 8, 2018 were considered the test sets.

To investigate the complex nonlinear characteristics of the COP, the RPs and their recurrence density of daily and weekly COP, as shown in Fig. 1, were carried out based on these parameters calculated by the C-C method in Table I. From Fig. 1, the discoveries can be obtained as follows.

TABLE I
Results of C-C Method

Dataset	m	τ
Daily oil price	19	6
Weekly oil price	14	8

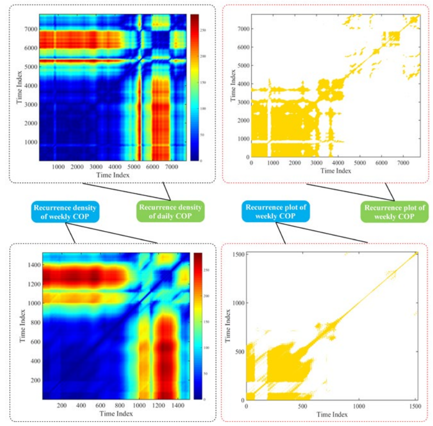


Fig. 1. Recurrence plot of the daily and weekly COP.

1) In the RP of daily COP, there are some vertical navy-blue lines and its clusters, indicating that the presence of laminarity (i.e., weak volatility). Further, some isolated points occur in the RP, which illustrates that the daily COP has chaotic property. Besides, the carmine vertical lines indicate the non-stationarity (i.e., dramatic change) in the daily COP, implying the presence of outliers.

2) From the RP generated by the weekly COP, it can be observed that there are fewer recurrence points in the RP, as compared to the RP of daily COP, thereby indicating that the weekly COP has lower predictability compared to the weekly COP. The interpretation of the carmine vertical lines in the RP of weekly COP is similar to that of the RP of daily COP.

However, the analysis concerning the RPs above is intuitive, lacking quantitative analysis. Therefore, RQA, as shown in Table II, was carried out to obtain the results of quantitative analysis for the RPs. From Table II, it can be observed that the predictability of the daily COP is higher than that of the weekly COP according to these metrics in Table II.

TABLE II
Results of RQA

Dataset	*RR*	*DET*	*ENT*	*LAM*	*TT*
Daily COP	33.8512	98.6962	5.8064	99.1409	98.6380
Weekly COP	18.4311	95.6662	4.9087	96.9953	32.8381

V. Outlier Detection for COP

Effectively detecting the outliers in the studied dataset is an important procedure of data cleaning, which facilitates improving the generalization ability of a prediction model to model nonlinear time series. Based on the analysis in Section IV, there is a high probability that outliers exist in the daily and weekly COP. Therefore, IF model was developed in this section to detect the outliers of the daily and weekly COPs. In detail, the score of each data point of the daily and weekly COPs was presented in Fig. 2; further, the outliers, as shown in Fig. 2, can be identified based on the corresponding score, according to the set threshold 5. In this study, the outliers detected were replaced by their interpolated values using cubic spline interpolation, which is beneficial to improve the explanatory ability and generalization of a prediction model and to ensure high prediction accuracy.

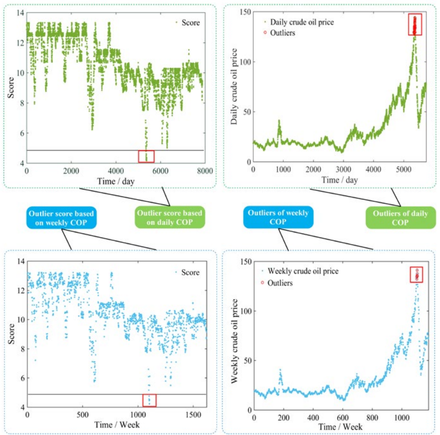


Fig. 2. Results of outlier detection.

VI. Case Study

In this section, the daily and weekly COPs are preprocessed using the VMD model to remove high-frequency noise involved. Further, the preprocessed COPs are utilized for training the proposed learning system, aiming to produce high-quality deterministic and probabilistic predictions.

A. Data Preprocessing

Based on the analytical results from Section V, the COPs contain some outliers, actually reducing the predictive performance. Therefore, the COPs were decomposed into some IMFs in this study, further eliminating the high-frequency IMF (i.e., IMF7), considered as noise component, in the experiments based on the daily and weekly COPs. Additionally, Fig. 3 displays the decomposed IMFs based on the daily COP.

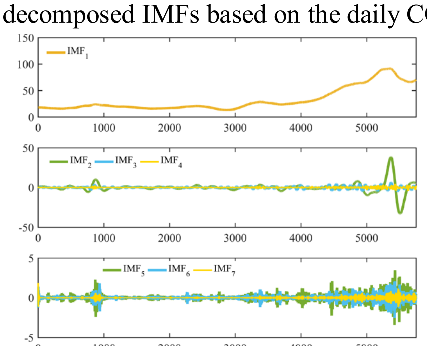


Fig. 3. The results of data decomposition based on VMD.

B. Deterministic Modeling

Deterministic prediction usually produces the single-point forecasts with certain biases, which can be further improved using denoising methods and other strategies. In this section, the daily and weekly COPs were utilized to verify the predictive accuracy and effectiveness of the proposed learning system. The built-in parameters of the learning system were configured, as shown in Table III. Further, the experimental results of the deterministic prediction were quantified using the following metrics MAE, MAPE, RMSE, DA, R, and IA, as shown in Table IV. From the information provided by Table IV, the conclusions can be obtained as follows:

1) The MAE, MAPE, and RMSE are negative indicators. The smaller the values of these indicators, the better the performance of the prediction model. According to the evaluative results reported in Table IV, the proposed learning system can yield more accurate deterministic prediction of the COPs, as compared to the benchmarks, including persistence model (PM), wavelet neural network (WNN), generalized regression neural network (GRNN), Conv1D-GRU, and EMD-Conv1D-GRU.

2) The DA, R, and IA are positive indicators, which means that the larger their values, the better the performance of the predictive model. From the evaluations displayed in Table IV, the proposed system performs best compared to the benchmarks considered.

3) To validate the effectiveness of the VMD, EMD, coupled with Conv1D-GRU, was designed as a benchmark. In accordance with the evaluative results, the proposed learning system has a significant advantage over EMD-Conv1D-GRU, thus demonstrating the superiority of the VMD.

4) Based on the above analysis, the proposed system has the best prediction performance compared to the five benchmarks, which demonstrates that the deep learning model (i.e., Conv1D-GRU), cooperated with the VMD, has great potential and promising application prospects in crude oil price prediction.

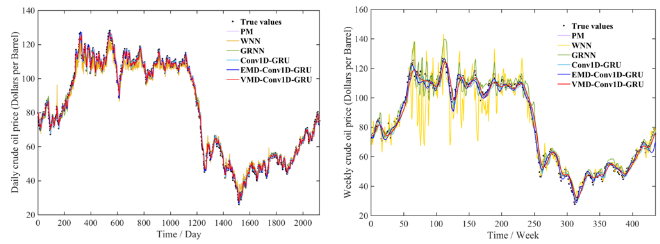
Finally, the result visualizations of the proposed learning system (VMD-Conv1D-GRU) and its benchmarks were carried out in Fig. 4, from which it can be observed that the predictive results yielded from the learning system are closer to the corresponding actual values compared to the benchmarks, which verifies the effectiveness of the proposed learning system further.

TABLE III
Parameters Setting of the Learning System

Model parameters	Default value
Number of the filters in the Conv1D	64
Size of the kernel in the Conv1D	2
Size of the pool in the Conv1D	2
Number of neurons in GRU	600
Number of neurons in the first FCL	600
Number of neurons in the second FCL	200
Size of the input	5
Size of the output	1
The training iteration	300
Batch size	1024
Activation function	*ReLU*
Optimizer	Adam
The moderate bandwidth constraint in VMD	2000
The tolerance ε of VMD	10^{-7}
Number of modes to be decomposed in VMD	7

TABLE IV
Evaluation Results of Deterministic Prediction

Models	MAE	MAPE	RMSE	DA	R	IA
Case I based on daily crude oil price						
PM	2.5976	3.55	3.3583	-	0.9927	0.9963
WNN	2.2970	2.94	3.0994	0.4873	0.9963	0.9966
GRNN	1.6025	2.17	2.0764	0.4892	0.9973	0.9986
Conv1D-GRU	1.0976	1.49	1.4717	0.4969	0.9986	0.9993
EMD-Conv1D-GRU	0.9851	1.39	1.2892	0.6002	0.9987	0.9992
Proposed	0.9691	1.33	1.2798	0.6202	0.9989	0.9995
Case II based on weekly crude oil price						
PM	1.9260	2.64	2.4973	-	0.9960	0.9980
WNN	5.7344	6.33	9.8701	0.5310	0.9378	0.9663
GRNN	4.0050	5.23	5.3200	0.4989	0.9843	0.9909
Conv1D-GRU	3.3122	4.49	4.1156	0.4309	0.9896	0.9943
EMD-Conv1D-GRU	2.2621	3.07	3.0508	0.6574	0.9944	0.9969
Proposed	1.7929	2.40	2.3273	0.6636	0.9968	0.9982

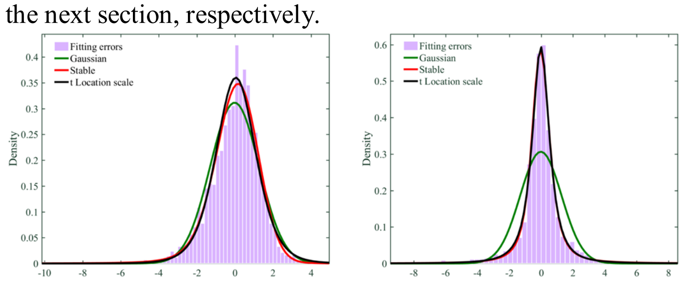


(a) Predictive results of daily crude oil price (b) Predictive results of weekly crude oil price

Fig. 4. Results visualization of predictive results based the daily and weekly COPs.

C. Uncertainty Analysis of Prediction Error

In fact, there are inevitable prediction errors surrounding the deterministic prediction when performing predictive modeling of the COPs, which can be caused by many uncertain factors, such as uncertainties from data measuring, parameter selection of a predictive model, and some macroeconomic factors. Notably, the uncertainties will have a significant negative impact on the risk management and decision-making of the crude oil market. Therefore, quantifying the uncertainties has an important practical significance and necessity for financial management. In the previous studies, the predictive errors were usually assumed to be Gaussian distribution. However, the assumption may lead to some risks for the uncertainty analysis because of lacking sufficient understanding of the statistical characteristics of the predictive error. In this study, to effectively quantify the uncertainties above, the statistical properties of the predictive errors from the proposed learning system were investigated in-depth based on the kernel density estimation. In addition to Gaussian distribution, Stable and t Location-scale (t L-S) distributions, which have gained popularity in previous studies [30] related to the error modeling due to their excellent statistical properties, such as the ability to effectively characterize the peak-to-tail characteristics of prediction errors and estimate the skewed data well, were also developed in this section to model the predictive errors. The corresponding parameters of each distribution above were presented in Table V. Further, the corresponding distribution fitting of each distribution function is illustrated in Fig. 5, from which it can be observed that the fitting performances of the Stable and t L-S distributions are superior to that of the Gaussian distribution. However, there is a phenomenon in Fig. 5 that the performance of the Stable is similar to that of the t L-S. Thus, the comprehensive performances of constructed prediction intervals based on Stable and t L-S were discussed in the next section, respectively.



(a) Error modeling based on daily oil price (b) Error modeling based on weekly oil price

Fig. 5. Distribution fitting of the predictive errors.

TABLE V
RESULTS OF ERROR MODELING

Gaussian		Stable		t L-S	
Case I: Errors based on daily oil price					
Location	-0.0158	α	1.8290	Location	0.0500
Scale	1.2800	β	-0.8301	Scale	1.0669
–	–	Scale	0.8099	n_u	6.6410
–	–	Location	0.1359	–	–
Case II: Errors based on weekly oil price					
Location	-0.0202	α	1.3165	Location	-0.0191
Scale	1.3018	β	0.0600	Scale	0.5906
–	–	Scale	0.4998	*nu*	1.8728
–	–	Location	-0.0267	–	–

Note: α and β denote the first and second shape parameter, respectively. n_u is degree of freedom.

D. Probabilistic Modeling

After the uncertainty analysis of the predictive errors, the prediction intervals of the COPs can be constructed via an optimal distribution (*OP*) based on (19), where L^{α} and U^{α} denote the lower and upper bound of the constructed prediction intervals at the significance level of α, respectively; $\hat{F}$ and E represent the forecast values and fitting errors, respectively, where both $\hat{F}$ and E are produced by the proposed learning system.

$$\left[L^{\alpha}, U^{\alpha}\right] = \left[\hat{F} - OP_{\alpha/2}\sqrt{var(E)}, \hat{F} + OP_{\alpha/2}\sqrt{var(E)}\right] \quad (19)$$

where $var(\cdot)$ represents the function of calculating the variance.

Further, the prediction intervals with different significance levels, including 5%, 10%, and 20%, can be constructed according to (19). Table VI provides the quantitative evaluations of these constructed prediction intervals using a series of statistical metrics, including PINAW, PICP, AWD, Score, and CWC. From Table VI, the findings can be obtained as follows.

1) Theoretically, the constructed prediction intervals are effective if PICP is greater than the corresponding significance level. However, the constructed prediction intervals based on Gaussian at the significance level of 5% for the daily COP are of little value in practical applications, although their PINAWs are smaller compared to that of prediction interval based on other distribution functions, which also applies to Case II based on Gaussian at the significance levels of 5% and 10%. Besides, the prediction intervals based on the Gaussian distribution is valid in the remaining experiments. Therefore, selecting Gaussian distribution to model error distribution needs further consideration in practice.

2) The aim of probabilistic prediction in this study is to construct the prediction intervals with desired resolution (i.e., PINAW) and reliability (i.e., PICP). However, there is a trade-off between resolution and reliability. Therefore, the metrics AWD, Score, and CWC, as comprehensive indicators integrating the resolution and reliability, were applied to assess the overall performance of the proposed learning system and its benchmarks. From the evaluations from Case I in Table VI, the proposed learning system based on the Stable distribution has a significant advantage over the benchmarks according to the metrics AWD, Score, and CWC at different significance levels. However, the comprehensive performance of the constructed prediction intervals based on t L-S distribution is superior to

that of the prediction intervals based on other distributions for most experiments in Case II.

3) Based on the analysis above, it can be concluded that there is currently no optimal distribution of modeling prediction errors, and the performance of different distribution functions need to be analyzed under specific application scenarios. Therefore, the research results obtained in this paper can provide a valuable reference for the uncertainty modeling of the crude oil price in the future.

4) To illustrate the performance of the proposed learning system, Fig. 6 visualizes the constructed prediction intervals generated by the proposed learning system, from which it can be observed that these intervals based on different confidence levels (i.e., 1–α) can cover the true values with high coverage probability, further indicating that it has great application potential. Besides, it is noteworthy that the average width of prediction intervals based on weekly COP is larger than that of the intervals based on daily COP due to its lower resolution and larger uncertainty.

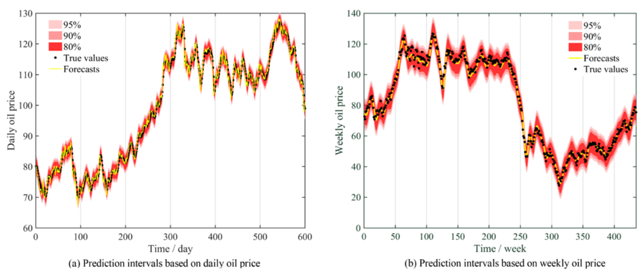


Fig. 6. Prediction intervals of the daily and weekly COPs.

TABLE VI

PERFORMANCE EVALUATION OF PROBABILISTIC PREDICTION

Case I based on daily crude oil price						
α	Distribution	PINAW	PICP	AWD	Score	CWC
5%	t L-S	0.0644	98.12%	0.0026	0.7263	0.0644
	Gaussian	0.0491	94.97%	0.0084	0.7630	1.0657
	Stable	0.0695	98.87%	0.0019	0.6697	0.0625
10%	t L-S	0.0505	95.34%	0.0075	1.1860	0.0505
	Gaussian	0.0412	91.49%	0.0162	1.2877	0.0583
	Stable	0.0583	97.27%	0.0041	1.1143	0.0412
20%	t L-S	0.0370	87.68%	0.0239	1.8720	0.0370
	Gaussian	0.0321	83.07%	0.0385	2.0660	0.0454
	Stable	0.0454	93.93%	0.0113	1.8179	0.0321
Case II based on weekly crude oil price						
5%	t L-S	0.1933	98.62%	0.0011	1.9977	0.1933
	Gaussian	0.1474	93.79%	0.0068	1.8514	1.9758
	Stable	0.2084	99.08%	0.0007	2.1143	0.2084
10%	t L-S	0.1515	94.25%	0.0058	3.3405	0.1515
	Gaussian	0.1237	88.74%	0.0163	3.2403	2.0054
	Stable	0.1749	97.47%	0.0022	3.6126	0.1749
20%	t L-S	0.1110	86.44%	0.0250	5.4857	0.1110
	Gaussian	0.0964	81.38%	0.0404	5.3490	0.0964
	Stable	0.1363	91.95%	0.0102	5.9362	0.1363

VII. EXTENSIONS

In the section, the predictive performance test, model convergence analysis, sensitivity test, and computational complexity analysis are performed to discuss further the effectiveness and feasibility of the proposed learning system

A. Model Performance Test

To further validate the superiority of the proposed learning system, the performance test based on the D-M test [31] was conducted, with the aim of testing the significance of the difference in the prediction errors from the proposed learning system and its benchmarks. The results of the D-M test are presented in Table VII, from which there is a significant difference between the error produced by the proposed learning system and its benchmarks. As a result, the superiority of the learning system can be further confirmed.

TABLE VII

RESULTS OF THE D-M TEST

Case	PM	WNN	GRNN	Conv1D-GRU	EMD-Conv1D-GRU
Case I	-26.14*	-20.69*	-22.33*	-8.82*	-1.99**
Case II	7.73*	-5.99*	-5.63*	-5.47*	-3.62*

B. Convergence Analysis

The model convergence is of significance to investigate the practical feasibility of the proposed learning system. The convergence analysis of Case I-II was performed in this study, as shown in Fig. 7. From Fig. 7, the proposed learning system has excellent convergence in the process of training iteration for the two cases.

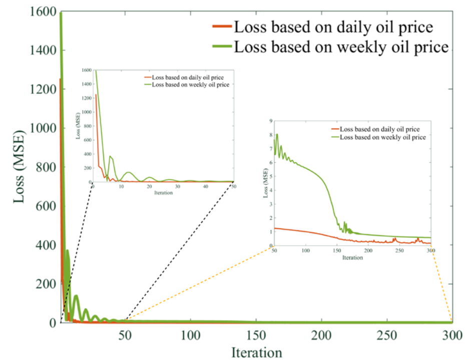


Fig. 7. Convergence visualization of the proposed learning system.

C. Sensitivity Analysis

Investigating the sensitivity of the proposed learning system can reveal its robustness. In this study, the sensitivity analysis based on the perspectives of the iteration and the number of neurons in the GRU was conducted, and the corresponding results are illustrated in Fig. 8. From Fig. 8, the proposed learning system is less sensitive to the training iterations and the size of the neurons after the iteration exceeds 200.

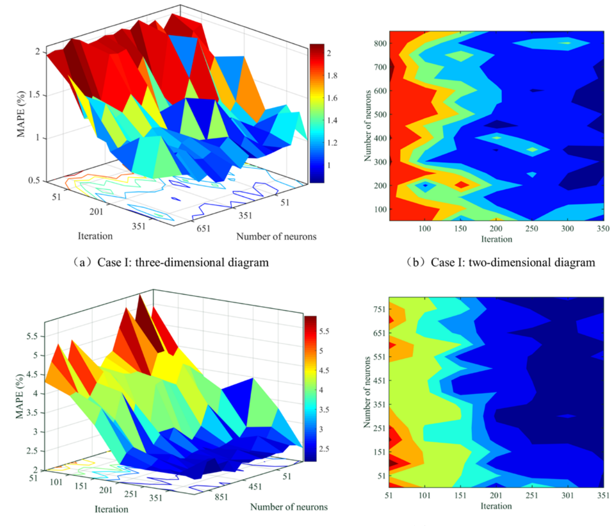


Fig. 8. Sensitivity analysis of the proposed learning system.

D. Computational Complexity Analysis

The computational complexity of the proposed learning system in this study is demonstrated as follows:

1) Complexity of the VMD

The complexity of the VMD depends on the initialization of the center frequency of each mode and the recursive fast Fourier transform [32]. The computational steps mainly include: (1) initialization ($O(2N\log_2(2N))$); (2) update m_k ($O((6A+2M+2D)\cdot K\cdot MI\cdot 2N)$); (3) update c_k ($O((2C+3M+2A)\cdot K\cdot MI\cdot N)$); (4) convergence ($O((4A+2M)\cdot MI\cdot 2N)$); (5) data reconstruction ($O(2N\log_2(2N))$). Therein, *M*, *D*, *C*, and *A* represent multiplication, division, comparison, and addition, respectively. Besides, *N*, *K*, and *MI* denote the length of the studied data to be decomposed, the modes to be decomposed, and maximum iterations. Overall, the computational complexity of the VMD is $O(2N\log_2(2N))$.

2) Complexity of the Conv1D

Considering the sizes of convolution kernel, input channel, and out channel are $M\cdot N$, *IC*, and *OC*, respectively, the number of the filters in the Conv1D is $M\cdot N\cdot IC$, and the filters will be mapped to new channels with a bias. Therefore, the Conv1D needs to take $O((M\cdot N\cdot IC+1)\cdot OC)$ computational efforts.

3) Complexity of the GRU

The GRU has complexity of $O(S(H^2+HK))$ [33], where *S*, *H*, and *K* are the number of items, hidden units, and output units, respectively.

VIII. Conclusion and Future Scope

This study proposes a novel learning system based on the VMD, Conv1D, and GRU-based deep recurrent neural network, aiming to construct deterministic and probabilistic predictions of the daily and weekly COPs. In detail, in order to ensure the generalization of the proposed learning system, the recurrence analysis, outlier detection, and VMD are carried out to cleanse the studied data. In the next, the Conv1D is developed to perform automatic feature selection based on the theory of temporal convolution, and the information from Conv1D is further learned by the GRU structure. The experimental results of deterministic and probabilistic predictions of the COP are evaluated using some indicators, such as MAE, MAPE, Score, and CWC. The corresponding conclusions indicate the superiority of the proposed learning system, as compared to its benchmarks. For example, based on MAPE in the deterministic prediction, the proposed learning system in Case I has the improvement of 62.54%, 54.76%, 38.71%, 10.74%, and 4.32%, respectively, compared to the benchmarks PM, WNN, GRNN, Conv1D-GRU, and EMD-Conv1D-GRU; Besides, the proposed system reflects the improvement of 9.09%, 62.09%, 54.11%, 46.55%, and 21.82% in Case II, respectively, compared to the benchmarks above.

Given the excellent performance of the proposed system, it has great potential to be applied to existing risk management systems of the crude oil market to contribute to better managing the risk of the crude oil market, reducing the transaction costs and improving corresponding transaction efficiency.

However, this study still remains some limitations, including: (1) lack of an analysis on the properties of complex network of the crude oil price; (2) univariate time series prediction; (3) single deep learning model; (4) offline model learning strategy.

To address the limitations above and facilitate the development of this field further, our research directions in the future will focus mainly on the following issues.

● Investigating the properties of recurrence network of the COP based on the theory of complex network;

● Exploring the influencing factors of crude oil prices to improve its predictive accuracy further;

● Developing effective ensemble approaches with time-varying weight for the crude oil price prediction;

● Formulating a feasible prediction model based on online learning.

Acknowledgment

This work was supported by the major project of the national social science of China (Grant No. 17ZDA093) and the China Scholarship Council (No. 201808210268).

References

[1] J. L. Zhang, Y. J. Zhang, and L. Zhang, “A novel hybrid method for crude oil price forecasting,” Energy Econ., vol. 49, pp. 649-659, May. 2015.

[2] M. H. Sahir and A. H. Qureshi, “Specific concerns of Pakistan in the context of energy security issues and geopolitics of the region,” Energy Policy, vol. 29, no. 3, pp. 454-478, Apr. 2007.

[3] L. Tang, Y. Wu, and L. Yu, “A non-iterative decomposition-ensemble learning paradigm using RVFL network for crude oil price forecasting,” Appl. Soft Comput. J., vol. 70, pp. 1097-1108, Sep. 2018.

[4] R. A. Lizardo and A. V. Mollick, “Oil price fluctuations and U.S. dollar exchange rates,” Energy Econ., vol. 32, no. 2, pp. 99-408, Mar. 2010.

[5] A. Lanza, M. Manera, and M. Giovannini, “Modeling and forecasting cointegrated relationships among heavy oil and product prices,” Energy Econ., vol. 27, no. 6, pp. 831-848, Nov. 2005.

[6] Y. Xiang and X. H. Zhuang, “Application of ARIMA Model in Short-Term Prediction of International Crude Oil Price,” Adv. Mater. Res., vol. 798, pp. 979-982, Sep. 2013.

[7] A. Hou and S. Suardi, “A nonparametric GARCH model of crude oil price return volatility,” Energy Econ., vol. 34, no. 2, pp. 618-626, Mar. 2012.

[8] H. Mohammadi and L. Su, “International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models,” Energy Econ., vol. 32, no. 5, pp. 1001-1008, Sep. 2010.

[9] M. A. Boyacioglu and D. Avci, “An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange,” Expert Syst. Appl., vol. 37, no. 12, pp. 7908-7912, Dec. 2010.

[10] S. H. A. Kaboli, J. Selvaraj, and N. A. Rahim, “Long-term electric energy consumption forecasting via artificial cooperative search algorithm,” Energy, , vol. 115, pp. 857-871, Nov. 2016.

[11] K. He, L. Yu, and K. K. Lai, “Crude oil price analysis and forecasting using wavelet decomposed ensemble model,” Energy, vol. 46, no. 1, pp. 564-574, Oct. 2012.

[12] W. Kristjanpoller and M. C. Minutolo, “Forecasting volatility of oil price using an artificial neural network-GARCH model,” Expert Syst. Appl., vol. 65, pp. 233-241, Dec. 2016.

[13] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531-544, Nov. 2014.

[14] H. S. Kim, R. Eykholt, and J. D. Salas, “Nonlinear dynamics, delay times, and embedding windows,” Phys. D Nonlinear Phenom., vol. 127, no. 1-2, pp. 48-60, Mar. 1999.

[15] J. Wang, T. Niu, H. Lu, Z. Guo, W. Yang, and P. Du, “An analysis-forecast system for uncertainty modeling of wind speed: A case study of large-scale wind farms,” Appl. Energy, 2018.

[16] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” Physics Reports., vol. 438, no. 5-6, pp. 237-329, Jan. 2007.

[17] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-Based Anomaly Detection,” ACM Trans. Knowl. Discov. Data, vol. 6, no. 1, Mar. 2012.

[18] D. Wang, “Robust data-driven modeling approach for real-time final product quality prediction in batch process operation,” IEEE Trans. Ind. Informatics, vol. 7, no. 2, pp. 371-377, Jan. 2011.

[19] D. Wang, J. Liu, and R. Srinivasan, “Data-driven soft sensor approach for quality prediction in a refining process,” IEEE Trans. Ind. Informatics, vol. 6, no. 1, pp. 11-17, Jul. 2010.

[20] C. Zhao et al., “Novel Method Based on Variational Mode Decomposition and a Random Discriminative Projection Extreme Learning Machine for Multiple Power Quality Disturbance Recognition,” IEEE Transactions on Industrial Informatics, Sep. 2018.

[21] M. T. Nguyen, B. Van Nguyen, and K. Kim, “Shockable Rhythm Diagnosis for Automated External Defibrillators Using a Modified Variational Mode Decomposition Technique,” IEEE Trans. Ind. Informatics, vol. 13, no. 6, pp. 3037-3046, Aug. 2017.

[22] Y. Wang, Z. Wei, and J. Yang, “Feature Trend Extraction and Adaptive Density Peaks Search for Intelligent Fault Diagnosis of Machines,” IEEE Trans. Ind. Informatics, vol. 15, no. 1, pp. 105-115, Jan. 2019.

[23] K. Cho，B. van Merienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation”, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, pp. 1724～1734, Oct. 2014.

[24] Y. Xu, W. Yang, and J. Wang, “Air quality early-warning system for cities in China,” Atmos. Environ., vol. 148, pp. 239-257, Jan. 2017.

[25] A. Khosravi, S. Nahavandi, and D. Creighton, “Prediction interval construction and optimization for adaptive neurofuzzy inference systems,” IEEE Trans. Fuzzy Syst., vol. 19, no. 5, pp. 983–988, Mar. 2011.

[26] W. Yang, J. Wang, T. Niu, and P. Du, “A novel system for multi-step electricity price forecasting for electricity market management,” Appl. Soft Comput. J., vol. 88, no. 106029, Mar. 2020.

[27] C. Tian and Y. Hao, “Point and interval forecasting for carbon price based on an improved analysis-forecast system,” Appl. Math. Model., vol. 79, pp.126-144, Mar. 2020.

[28] Y. Xu, P. Du, and J. Wang, “Research and application of a hybrid model based on dynamic fuzzy synthetic evaluation for establishing air quality forecasting and early warning system: A case study in China,” Environ. Pollut., vol. 223, pp. 435-448, Apr. 2017.

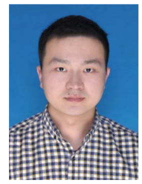
[29] R. L. Winkler, “A decision-theoretic approach to interval estimation,” J. Am. Stat. Assoc., vol. 67, pp. 187–191, Mar. 1972.

[30] J. Wang, T. Niu, H. Lu, W. Yang, and P. Du, “A Novel Framework of Reservoir Computing for Deterministic and Probabilistic Wind Power Forecasting”, IEEE Trans. Sustain. Energy,vol. 11, no. 1, pp. 337-349, Jan. 2020.

[31] F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” J. Bus. Econ. Stat., vol. 13, no. 3, pp. 253-263, Jul. 1995.

[32] Soman K.P., P. Poornachandran, Athira S., and Harikumar K., “Recursive Variational Mode Decomposition Algorithm for Real Time Power Signal Decomposition,” Procedia Technol., vol. 21, pp. 540-546, Nov. 2015.

[33] T. Zhong, Z. Wen, F. Zhou, G. Trajcevski, and K. Zhang, “Session-based Recommendation via Flow-based Deep Generative Networks and Bayesian Inference,” Neurocomputing, vol. 391, pp. 129-141, May 2020.



Tong Niu (GS’19) received the BEc. from Henan University of Economics and Law in 2015, and received M.A.S. and Ph.D. from Dongbei University of Finance and Economics in 2017 and 2020. He has published 20 refereed journal papers. His research interests include computational intelligence, time series forecasting, and machine learning.

Jianzhou Wang received B.S. from Northwest Normal University in 1988, and received MSc and the Ph.D. degree from Lanzhou University in 1998 and 2004, respectively. Currently, he is currently a professor of School of Statistics at Dongbei University of Finance and Economics. He has published over 150 refereed journal papers, and the H-index is 26. His research interests are in the areas of wind energy forecasting, data mining, machine learning, statistics learning, and forecast theory.

Haiyan Lu (SM'15) received the B.Eng. and M.Eng. from the Harbin Institute of Technology, Harbin, China, in 1985 and 1988, respectively, and the Ph.D. degree in engineering from the University of Technology, Sydney, Australia, in 2002. She is currently with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. She has published over 130 refereed journal and conference papers. Her research interests include numerical modeling and simulation of magnetic materials for smart electromagnetic devices, optimization methods, forecasting methods, and recommender systems and their applications in power systems, power markets, and smart grids.

Wendong Yang received the BEc and Ph.D. from the Jilin University of Finance and Economics in 2015 and Dongbei University of Finance and Economics in 2020. Currently, he has published over 20 papers in journals including Applied Energy and Atmospheric Environment. His current research interests include artificial intelligence, big data mining, machine learning, energy forecasting, and air quality early warning.

Pei Du received the B.S. from the Yantai University in 2015 and received the MEc from the Dongbei University of Finance and Economics in 2018. Currently, he is pursuing the Ph.D. degree at Dongbei University of Finance and Economics. He has published over 20 papers in journals including Applied Energy and Applied soft computing. His current research interests include artificial intelligence, optimization algorithms, machine learning, time series forecasting, and air pollution analysis and prediction.