



Forecasting carbon futures price: a hybrid method incorporating fuzzy entropy and extreme learning machine

Peng Chen¹ · Andrew Vivian²  · Cheng Ye³

Accepted: 29 October 2021 / Published online: 30 December 2021
© The Author(s) 2021

Abstract

In this paper, we propose a novel hybrid model that extends prior work involving ensemble empirical mode decomposition (EEMD) by using fuzzy entropy and extreme learning machine (ELM) methods. We demonstrate this 3-stage model by applying it to forecast carbon futures prices which are characterized by chaos and complexity. First, we employ the EEMD method to decompose carbon futures prices into a couple of intrinsic mode functions (IMFs) and one residue. Second, the fuzzy entropy and K-means clustering methods are used to reconstruct the IMFs and the residue to obtain three reconstructed components, specifically a high frequency series, a low frequency series, and a trend series. Third, the ARMA model is implemented for the stationary high and low frequency series, while the extreme learning machine (ELM) model is utilized for the non-stationary trend series. Finally, all the component forecasts are aggregated to form final forecasts of the carbon price for each model. The empirical results show that the proposed reconstruction algorithm can bring more than 40% improvement in prediction accuracy compared to the traditional fine-to-coarse reconstruction algorithm under the same forecasting framework. The hybrid forecasting model proposed in this paper also well captures the direction of the price changes, with strong and robust forecasting ability, which is significantly better than the single forecasting models and the other hybrid forecasting models.

Keywords Carbon futures price · EEMD · Fuzzy entropy · K-means clustering method · ARMA · Extreme learning machine

We thank three anonymous referees for their constructive and valuable comments that help us to improve the paper. Any remaining errors are our own responsibility. Peng Chen is also grateful for the grant from the Humanities and Social Science Foundation of the Ministry of Education of China (No. 20YJC790012) and the National Natural Science Foundation of China (No. 72071095).

✉ Andrew Vivian
A.J.Vivian@lboro.ac.uk

¹ Department of Finance and Institute of Finance, School of Economics, Jinan University, Guangzhou 510632, China

² School of Business and Economics, Loughborough University, Leicestershire LE11 3TU, UK

³ Department of Finance, School of Economics, Jinan University, Guangzhou 510632, China

1 Introduction

The increasing global attention on environmental issues such as climate change in recent decades, has raised interest in understanding the topics related to it. Carbon markets are one important mechanism to tackle climate change. A major step was the establishment of the European Union Emissions Trading System (EU ETS) which has become the biggest and most important carbon exchange market in the world; as such it attracts the attention of global organizations to govern carbon emissions as well as international investors as an opportunity for the investment in this market. However, as an emerging policy-based market, the carbon market is determined by internal market mechanism (such as the quantity of allowances issued) as well as external factors (such as macroeconomic factors; Zhang and Wei, 2010). The determinants together lead to strong fluctuations in the carbon market, which is characterized by chaos and high volatility with nonlinear, non-stationary phenomena evident in its prices. Further understanding the patterns of price fluctuations and predicting movements more accurately is of great importance to practitioners, policymakers and academics. For policymakers' it will be useful to gauge when the supply of carbon permits needs adjusting to avoid prices moving to extreme levels.¹ For investors it is critical to know how the price is likely to move so that they can exploit this in their trading decisions. Moreover, from an academic perspective, our paper contributes to the literature by demonstrating that a new hybrid method that features fuzzy entropy and extreme learning machine is effective in forecasting carbon futures price.²

Over the last decade, a growing body of empirical studies on the forecasting of carbon prices has emerged (Byun and Cho, 2013; Fan et al., 2015; Zhu et al., 2018; Song et al., 2019). In earlier studies, traditional statistical and econometric models, such as linear regression, ARMA models and VAR models have been widely applied to carbon prices forecasting. The models achieve a desirable prediction performance when the prices series is linear or near-linear. However, the dynamics of carbon futures prices series is characterized by the features of nonlinearity and irregularity (Arouri et al., 2012). Due to the limitations of traditional econometric models on the nonlinear pattern of price series, recent scholars turn to some emerging nonlinear artificial intelligence (AI) models or statistical learning, such as artificial neural network (ANN), support vector machines (SVM) and least squares support vector machines (LSSVM), to forecast carbon prices. Many experiments have demonstrated the superiority in carbon price prediction of AI-based models when compared to traditional statistical models (e.g., Tsai and Kuo, 2013; Fan et al., 2015; Jiang and Wu, 2015). Although the AI-based models have strong prediction ability, they have their own drawbacks in that they are very sensitive to parameters selection, which may often lead to long training time or no convergence. To remedy these shortcomings, the extreme learning machine (ELM) proposed by Huang et al. (2006) has recently been applied to time series prediction (Sun and Zhang, 2017).

Moreover, some hybrid models have been proposed to forecast carbon futures prices and achieve a better performance. For example, Zhu and Wei (2013) develop a novel hybrid method integrating the ARIMA with LSSVM model to forecasting carbon prices. Zhang et al. (2017) propose a hybrid model that combines signal processing technology, econometric

¹ Extremely low prices or excessively high prices will be problematic. Very low prices provide very little incentive for companies to reduce carbon emissions. On the other hand very high prices can place much higher costs on firms which could lead to financial distress for them.

² Plausibly, greater predictability of carbon prices should help avoid large persistent deviations from "efficient prices" (e.g. bubbles) since market participants would notice large deviations quickly and trade accordingly to rectify this.

models and neural network for carbon prices forecasting. Huang et al. (2021) combine econometric and neural network methods to propose a novel decomposition-ensemble paradigm VMD-GARCH/LSTM-LSTM model for carbon price forecasting, providing evidence of superiority in forecasting accuracy.

In addition, empirical mode decomposition (EMD) method proposed by Huang et al. (1997), which can not only automatically determine the total decomposed number of the original signal, but also can extract linear stationary components from nonlinear and non-stationary original time series, has also been widely used to forecast time series (Guo et al., 2012; Jiang et al., 2020; Wang, et al., 2016; Zhang et al., 2008), including carbon price (Zhu, 2012; Zhu et al., 2017, 2018, 2019); these features help explain why it has been (recently) preferred to wavelet decomposition. Prior work has used EMD along with other methods such as the AI-based models and contributed by demonstrating it is effective for forecasting (Zhu et al., 2017).

Motivated by hybrid methods and the “decomposition-and-ensemble” principle, the paper proposes a new hybrid method for forecasting carbon futures prices. This method, incorporates the ensemble EMD, and introduces i) Fuzzy entropy and ii) extreme learning machine (ELM) into the model. This leads to a novel hybrid multiscale nonlinear ensemble ELM approach to forecasting carbon futures prices. The advantages of the novel hybrid approach to forecasting are as follows. First, compared with other popular forecasting methods, this novel hybrid framework provides more accurate and efficient prediction of the nonstationary and nonlinear carbon futures prices. Second, this paper is the first to use the fuzzy entropy analysis and K-means clustering method to reconstruct the IMFs and the residue decomposed by the EEMD; the approach adopted in this paper is therefore truly novel. Compared to traditional fine-to-coarse reconstruction algorithm that ignores the complexity of the decomposed IMFs as well as directly using the residue as the trend component, the proposed reconstruction algorithm (using the Fuzzy Entropy and K-means clustering methods) under the decomposition-ensemble framework is more efficient and accurate to reconstruct the three components, (i.e., i) high frequency, ii) low frequency and iii) trend). Third, the proposed hybrid model makes full use of the advantages of econometric model as well as artificial intelligence model in that the ARMA and the ELM forecasting model are established for the stationary and non-stationary components, respectively. Compared to the traditional methods (such as Back-Propagation method) in dealing with the non-stationary trend series, the Extreme Learning Machine (ELM) works efficiently without iteration. In addition to working efficiently, the proposed method is easily processed and requires less sample for the training set than many alternatives. Finally, this paper first proposes a time-varying cross-validation method to select the optimal parameters for the ELM model. This is due to the fact that there is short-term memory in the carbon futures prices since the series is not a random walk (Feng et al., 2011).

The remainder of the paper is organized as follows: Sect. 2 gives a brief review of the related literature. Section 3 describes the proposed novel hybrid forecasting model and outlines the main integrated modules, including EEMD, fuzzy entropy, extreme learning machine (ELM). Section 4 outlines the measures of forecast errors and describes the data; Sect. 5 provides empirical analysis; Sect. 6 presents forecast results; Sect. 7 concludes.

2 Literature review

Many different aspects of the carbon market have been examined by the rapidly developing literature. Montagnoli and De Vries (2010) provide early evidence of trading thickness and

efficiency of the European carbon market during phase I (2005–2007) and the beginning of phase II (2008–2009). They find that during phase I the market is characterized by inefficiency but during phase II it becomes increasingly efficient. Feng et al. (2012) provide early evidence on tail risk in the European carbon market using extreme value theory to effectively estimate value at risk; their main results indicate that downside risk is high in the market and it is of greater magnitude than upside risk. Relatedly, Jiao et al. (2018) emphasise the important role that macroeconomic fundamentals can play in enhancing value at risk estimates. Rootzén and Johnsson (2016) take a different perspective and looks at the impact of the carbon price on car production (especially steel cost) finding that even a relatively high carbon price (100€/tCO₂) has a very modest impact on the cost of a car. In contrast, Zhang et al. (2016) project carbon emissions in China under various scenarios finding that carbon emissions can be reduced at modest cost by 2050. Recently, Tan et al. (2020) examine how connected the European carbon market is to energy and financial markets; they find that carbon is connected more closely to energy markets than to equity and bond markets.

Earlier studies in the literature have widely employed traditional statistical and econometric models for the forecasting of carbon prices and volatility. For example, Byun and Cho (2013) employ the GARCH family of models to forecast carbon futures volatility and find that GJR-GARCH is more effective than TGARCH and GARCH in prediction. When using the AR-GARCH and regime-switching models for stochastic modeling the price dynamics of CO₂ emission allowances, Benz and Trück (2009) find that these models adequately capture characteristics like skewness, excess kurtosis and in particular different phases of volatility behavior in the returns. Sanin et al. (2015) incorporate a time-varying jump probability to the hybrid ARMAX-GARCH model and find that this model obtains a better performance of carbon volatility prediction than the standard ARMAX-GARCH framework does. Although GARCH and Markov switching multifractal (MSM) models might capture different facets of the volatility process in carbon prices, Segnon et al. (2017) find that MSMs in most cases encompass GARCH and FIGARCH when comparing the volatility forecasting between different models. Chevallier (2011) proposes a nonparametric modeling approach to the carbon prices and volatility showing that it reduces the prediction error for conditional mean by almost 15% compared to linear AR models. The above econometric models can produce high prediction accuracy when the data used in the studies is linear or stationary. In fact, carbon price is a time series with nonlinear, non-stationary, chaotic and multifractal characteristics (Feng et al., 2011; Zhu et al., 2017), so implementing traditional econometric models may no longer be the most suitable approach for forecasting carbon prices and may lead to a larger forecasting error.

Due to the limitations of econometric models, some scholars investigate the usefulness of nonlinear artificial intelligence (AI) models or statistical learning for carbon price forecasting, such as artificial neural network (ANN), support vector machines (SVM), and least squares support vector machines (LSSVM), among other. Among the ANN models, the multi-layer perceptron (MLP) model is the most widely used in time series prediction (de Oliveira et al., 2013). For example, when using the multi-layer perceptron (MLP) model to predict the carbon prices, Fan et al. (2015) find the model possesses good prediction performance in both level and directional measurement. Although the ANN method has a strong prediction power, it highly depends on the network structure (topology, connections, neurons number) and their operational parameters (learning rate, momentum, etc.). Thus, there exist some drawbacks in the models, including long training time to get an ANN network, no convergence of the optimization algorithm, and easily falling into the local optimal solution. However, the support vector machine model (SVM) proposed by Vapnik (1995) can avoid the issue of falling into the local optimal solution and obtain the globally optimal solution, which has

been widely used in time series prediction. For example, Zhu and Wei (2013) use the least square support vector machines (LSSVM) method to predict carbon prices, and show that the prediction performance is better than ARIMA and ANN do. When proposing a hybrid ARMA and SVM models to predict the carbon price, Zhu and Wei (2013) find that the hybrid forecasting model can improve the prediction accuracy.

Although the SVM model can obtain a better forecasting power, it spends a lot of time on training process and needs to determine the hyper-parameters. To cope with the above limitations, Huang et al. (2006) propose the extreme learning machine (ELM) based on a single hidden layer feedforward neural network, in which the algorithm randomly generates the connection weights between the input layer and the hidden layer, and the threshold of the neurons in the hidden layer. This method not only succeeds in overcoming the drawback of long training time of the ANN and SVM models, but also can avoid falling into the local optimal solution and thus quickly find the optimal solution. Therefore, the ELM has been applied to time series prediction in various fields recently including rainfall (Taormina and Chau, 2015) and CO₂ emissions (Sun, Wang and Zhang, 2017).

In fact, carbon futures price is a chaotic and highly volatile time series featuring non-linearity, non-stationarity as well as mixed frequency characteristics. However, traditional econometric models such as the ARMA model, GARCH, and VAR methods are primarily suitable for the prediction of stationary time series, so their predictive power is very limited for non-stationary carbon price series. Further, for the single statistical learning models, such as ANN, SVM and ELM model, although they are very suitable for modelling nonlinear time series, their prediction power is also limited because the carbon price is characterized by components with different frequencies.

To remedy the above limitations, hybrid methods have been proposed for forecasting carbon prices as well as other time series in the literature (e.g., Liu et al., 2012; Liu and Shi, 2–13; Zhu and Wei, 2013; Li et al., 2017; Zhang et al., 2017; Huang et al., 2021). For example, Zhu and Wei (2013) use the ARMA-LSSVM model to predict the carbon prices and find that these combined forecasting models can achieve a better forecasting performance than a single forecasting model. However, there are some characteristics of mixed frequency (high and low) components in carbon price series, so that the prediction power of the above-combined forecasting model may be incomplete, imperfect or limited. Therefore, some forecasting methods based on decomposition-reconstruction principle are proposed for time series prediction, such as wavelet decomposition (Meng et al., 2016; Yu et al., 2015). However, the wavelet decomposition method requires one to determine the number of decomposed layers and wavelet functions in advance. In contrast, the empirical mode decomposition (i.e., EMD) method proposed by Huang et al. (1997), can not only automatically determine the total decomposed number of the original signal, but can also extract linear stationary components from nonlinear and non-stationary original time series. Consequently, this method has been widely used to forecast time series (Guo et al., 2012; Wang, et al., 2016; Zhang et al., 2008), including carbon price prediction (Zhu, 2012; Zhu et al., 2017, 2018) and carbon volatility (Tang et al., 2017). For example, Zhu et al. (2018) use a novel multiscale nonlinear ensemble leaning paradigm incorporating EMD and LSSVM with kernel function prototype for carbon price forecasting and demonstrate that the proposed model can obtain higher level and directional predictions and higher robustness. However, there exists the mode mixing problem in the EMD method. Further, an ensemble empirical mode decomposition (EEMD), proposed by Wu and Huang (2009) for solving the mode mixing problem, has been used in time series prediction (Ren et al., 2015; Wang et al., 2015; Zhang et al., 2017). For example, Zhu et al. (2017) use an EEMD-based LSSVM to predict

the carbon prices, and demonstrate that the proposed model achieves high accuracy both in level and directional predictions.

3 Methodology

3.1 Ensemble empirical mode decomposition (EMD)

The method of empirical mode decomposition (i.e., EMD) proposed by Huang et al. (1997) has been widely used in the literature to decompose a nonlinear and non-stationary time series into a set of intrinsic mode functions (IMFs) and a residue.³ However, one major limitation is the mode mixing (i.e., the occurrence of disparate scales across different IMFs) in the EMD decomposition. To address the problem of mode mixing, Wu and Huang (2009) combine EMD with noise-assisted analysis method to propose the ensemble EMD (EEMD) method. The basic principle of the EEMD algorithm is that it defines the true IMF components as the mean of an ensemble of trials; each trial consists of the signal plus a Gaussian white noise term. The EEMD method is defined as follows:

Step 1: Setting the total number of white noises added is M , and the standard deviation of white noise is generally 0.1 ~ 0.4 times of the standard deviation of the original signal. And add the Gaussian white noise $n_i(t)$ with zero mean and constant standard deviation in the original signal $x(t)$.

$$x_i(t) = x(t) + n_i(t) \quad (1)$$

Step 2: Decompose $x(t)$ into several IMF components and a residue by the EMD method, and the $imf_{ij}(t)$ is the j th IMF component after adding the i th Gaussian white noise.

Step 3: Repeat steps (2), (3) M times, and then calculate the average value of all the corresponding IMF components to eliminate the effect of the white noise added and obtain the final IMF component:

$$imf_j(t) = \frac{1}{M} \sum_{i=1}^M imf_{ij}(t) \quad (2)$$

where imf_j is the j -th component decomposed by the EEMD method.

3.2 Fuzzy entropy

Entropy is a property of the thermodynamic system that can measure the disorder in the dynamic states of time series. Due to the advantage that entropy can be visualized, different entropy methods have been proposed to analyze the complexity characteristics of time series (Chen et al., 2009), including for quantifying the efficiency of financial markets (Ortiz-Cruz et al., 2012). In particular, entropy can quantitatively estimate the complexity of hidden patterns in data, but in fact, the boundaries between patterns are ambiguous and difficult to determine the relationship between patterns, thus the fuzzy entropy method based on fuzzy theory was proposed to calculate the fuzzy similarity between different hidden patterns by using membership function. Therefore, the similarity of two time series can be measured by the fuzzy entropy method. For a given time series $x(t)$, $t = 1, 2, \dots, T$, where T is the length of $x(t)$, and then the fuzzy entropy can be calculated as follows:

³ The method of empirical mode decomposition (i.e., EMD) can be found in Appendix A.1.

Step 1: For a given time series $x(t)$, the pattern dimension is set to m , and then the m -dimension vector is defined as:

$$X_i^m = [x(i), x(i + 1), \dots, x(i + m - 1)], 1 \leq i \leq T - m + 1 \tag{3}$$

Step 2: Compute the distance between vector $X(i)$ and $X(j)$:

$$d_{ij}^m = d[X_i^m, X_j^m] = \max_{k=0,1,\dots,m-1} \{|[x(i + k) - x_0(i)] - [x(j + k) - x_0(j)]|\}$$

$$i, j = 1, 2, \dots, T - m, i \neq j \tag{4}$$

Step 3: Given parameter n and tolerance parameter r , then the similarity D_{ij}^m between X_i^m and X_j^m is estimated by the fuzzy membership function $\mu(d_{ij}^m, n, r)$, i.e.,

$$D_{ij}^m = \mu(d_{ij}^m, n, r) = \exp(-(d_{ij}^m / r)^n) \tag{5}$$

Step 4: Compute the following function after calculating the similarity D_{ij}^m :

$$\varphi^m(n, r) = \frac{1}{T - m + 1} \sum_{i=1}^{T-m+1} \left(\frac{1}{T - m} \sum_{j=1, i \neq j}^{T-m+1} D_{ij}^m \right) \tag{6}$$

Similarly, for X_i^{m+1} :

$$\varphi^{m+1}(n, r) = \frac{1}{T - m} \sum_{i=1}^{T-m} \left(\frac{1}{T - m - 1} \sum_{j=1, i \neq j}^{T-m} D_{ij}^{m+1} \right) \tag{7}$$

Step 5: The final fuzzy entropy of time series $x(t)$ can be calculated as follows:

$$FuzzyEn(m, n, r, T) = \ln \varphi^m(n, r) - \ln \varphi^{m+1}(n, r) \tag{8}$$

In the above Eq. (8), m is the pattern dimension, r is tolerance parameter, and T is the length of series. Following the method in Chen et al. (2007), the embedded dimension m usually takes 2 or 3. And r represents the width of the boundary of the fuzzy function, and will lose a lot of statistical information when r is set too large. Therefore, r is generally $0.1 \sim 0.25 \sigma_{SD}$, where σ_{SD} is the standard deviation of the original series, and n take 2 or 3.

3.3 Extreme learning machine (ELM)

The extreme learning machine (ELM), proposed by Huang et al. (2006), is based on single-hidden layer feedforward neural networks. The algorithm randomly generates the connection weight matrix between input and hidden layers and the threshold of neurons in the hidden layer, and there is no need to adjust the connection weight and the threshold during the training process of network. The optimal solution can be achieved as long as the number of neurons in the hidden layer is determined. Compared with the traditional neural network model based on gradient descent method, the advantage of the ELM method is that it does not require iteration to obtain the optimal solution, which leads to faster learning speed and better generalization performance.

For given n different samples (x_i, t_i) , where $x_i \in R^m$, $t_i \in R^q$ are m, q dimension vectors, respectively. $x_i = (x_{1i}, x_{2i}, \dots, x_{mi})^T$, $t_i = (t_{1i}, t_{2i}, \dots, t_{qi})^T$, and the total number of neurons of the hidden layer neurons is L . The topology of ELM network is shown in Fig. 1.

Define the connection weight matrix between the input and the hidden layers as W , the threshold of neurons in the hidden layer as b , and the connection weight between the hidden and the output layers as β .

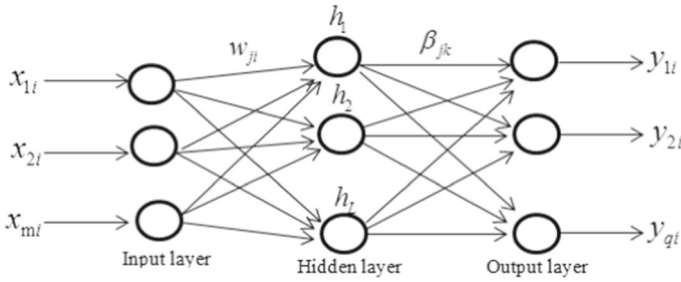


Fig. 1 The network of extreme learning machine (ELM)

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ w_{L1} & w_{L2} & \cdots & w_{Lm} \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_L \end{bmatrix}, \beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2q} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{L1} & \beta_{L2} & \cdots & \beta_{Lq} \end{bmatrix} \tag{9}$$

where w_{ji} is the connection weight between the i th neuron of the input layer and the j th neuron of the hidden layer; b_j represents the threshold of j th neuron in the hidden layer; and β_{jk} represents the connection weight between the j th neuron of the hidden layer and the k th neuron of the output layer.

Then the input matrix X and the output layer matrix Y are defined as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}; Y = \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{q1} \\ y_{12} & y_{22} & \cdots & y_{q2} \\ \vdots & \vdots & \cdots & \vdots \\ y_{1n} & y_{2n} & \cdots & y_{qn} \end{bmatrix} \tag{10}$$

The activated function of the neurons in the hidden layer is set as $G(x)$, thus the output matrix of the hidden layer is calculated by

$$H = \begin{bmatrix} G(w_1x_1 + b_1) & G(w_2x_1 + b_2) & \cdots & G(w_Lx_1 + b_L) \\ G(w_1x_2 + b_1) & G(w_2x_2 + b_2) & \cdots & G(w_Lx_2 + b_L) \\ \vdots & \vdots & \cdots & \vdots \\ G(w_1x_n + b_1) & G(w_2x_n + b_2) & \cdots & G(w_Lx_n + b_L) \end{bmatrix} \tag{11}$$

In Eq. (11), $w_i = (w_{i1}, w_{i2}, \dots, w_{im})$ represents the connection weight between the hidden layer and the input layer. Finally, the output matrix is calculated as:

$$Y = H\beta \tag{12}$$

When the activated function $G(x)$ is infinitely differentiable, not all the parameters of the single layer feedforward neural network (SLFN) need to be adjusted. Here, W and b can be selected randomly before the process of training, and remains unchanged during the process of training. Therefore, the training time of the ELM algorithm is shorter than that of the BP algorithm. The connection weight β between the hidden and the output layers can be obtained by solving the least squares of the following equation:

$$\min_{\beta} \|H\beta - T'\| \tag{13}$$

Then, the solution is $\hat{\beta} = H^+T'$, where H^+ is Moore–Penrose generalized inverse of the output matrix in the hidden layer.

3.4 The proposed EEMD-FuzzyEn–ARMA-ELM hybrid forecasting model

Due to the non-stationary and nonlinear properties of carbon futures prices, a single ARMA model cannot fit the non-stationary carbon price series well. The hybrid model under the decomposition-ensemble framework is proposed to separate the original carbon prices into linear and stationary component, and nonlinear and nonstationary trend. In order to take advantage of the ARMA model's prediction power on stationary time series as well as the ELM model on nonstationary time series, the proposed hybrid model incorporating the ARMA and ELM model, namely, EEMD-FuzzyEn–ARMA-ELM, are established for the linear stationary components and non-stationary components, respectively. Thus, the proposed hybrid model makes full use of the advantages of econometric model as well as artificial intelligence model for forecasting.

The overall framework of the hybrid forecasting model is shown in Fig. 2. The main steps are as follows:

Step 1 : Use EEMD method to decompose carbon price into a series of intrinsic mode functions (IMFs) and a residue.

Step 2: Calculate the fuzzy entropy values of each IMF and the residue, which are used as the clustering feature variables. Further, the k-means clustering method is used to classify the decomposed IMFs and the residue into different clusters. Following the literature of the reconstruction of decomposed IMFs, the components are generally reconstructed into three different sequences (i.e., three clusters), including the high frequency (*High*), the low frequency (*Low*) and the trend series (*Trend*). As a result, three different series are reconstructed from the original carbon price.

Step 3: The stability of these reconstructed components, including the high frequency, low frequency and trend series, is examined via the unit root test. Then the ARMA and ELM models are established for the stationary high and low frequency series, and non-stationary trend series respectively.

Step 4: The predicted results of the high-frequency (*High*), low-frequency (*Low*) and trend series (*Trend*) based on ARMA model and ELM model are denoted as *High_pred*, *Low_pred*, and *Trend_pred*, respectively, and then the final prediction of the original time series based on the hybrid forecasting model is obtained by aggregating *High_pred*, *Low_pred*, and *Trend_pred*, i.e.,

$$pred_value = High_pred + Low_pred + Trend_pred \quad (14)$$

In sum, this study employs the EEMD method to decompose the nonlinear and non-stationary carbon futures prices, which effectively separate the stationary and nonstationary components. In order to reduce the training time and the number of elements requiring training for the forecasting model, fuzzy entropy and K-means clustering methods are used to reconstruct the IMFs and the residue. We obtain three different frequency series, including the high frequency series representing the carbon market fluctuation component, the low-frequency series representing the major event-driven component, and the trend series

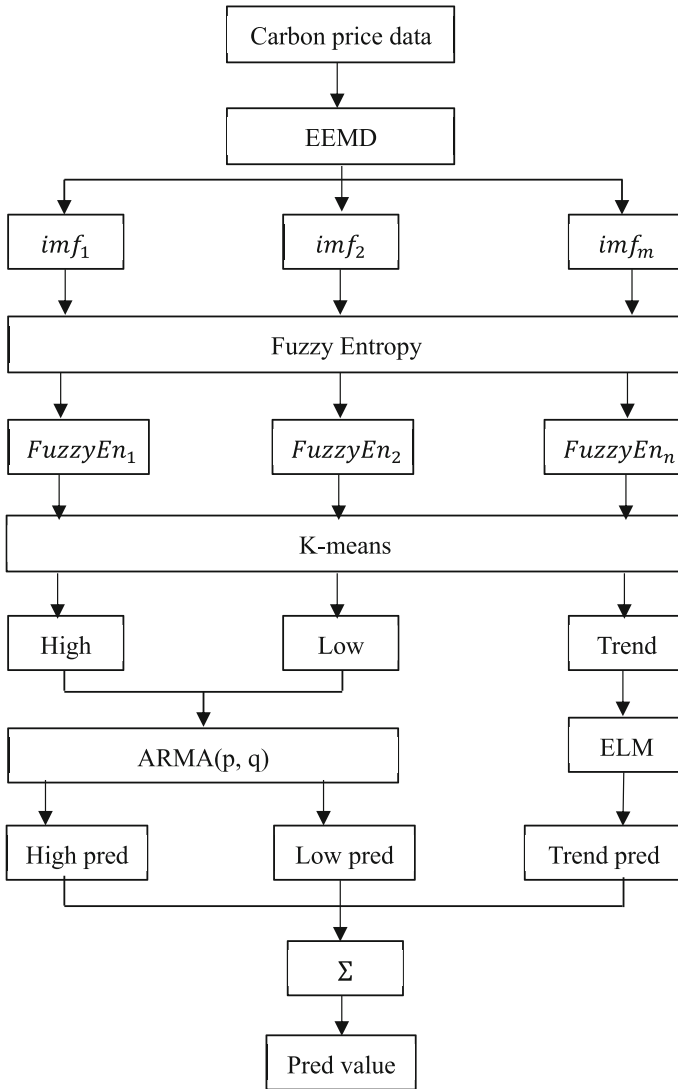


Fig. 2 The overall framework of the proposed hybrid model

representing the tendency of the carbon prices respectively. Further, we take advantages of the prediction power of the ARMA model as well as the ELM model, in which the ARMA model is established for the stationary high and low frequency series, while the extreme learning machine (ELM) model is established for the non-stationary trend series. Finally, the forecasting results of all the components based on the forecasting models are aggregated as the final forecasting of the carbon prices.

4 Data and evaluation criteria

4.1 Evaluation criteria

In order to evaluate the predication performance of the models, we use four different measures of forecast error, including root mean square error (RMSE), mean absolute percentage error (MAPE), correlation coefficient (R) and direction statistic (DS), as the evaluation criteria.⁴

Moreover, we implement the adjusted Diebold-Mariano test (i.e., DM, there) proposed by Harvey et al. (1997) to evaluate the predictive accuracy between different forecast models. When forecast errors follow heavy-tailed distributions, the test framework in the seminal paper by Diebold and Mariano (1995) is found to be quite seriously over-sized, which may generate some misleading finding. The adjusted Diebold-Mariano test enables us to alleviate the problem through modifying their test statistic. The null hypothesis of the adjusted Diebold-Mariano test is that the benchmark model A is not inferior to any of the alternative model B, i.e., the two models have the same of expected prediction accuracy. Thus, it can be written as follows:

$$H_0 : E[g(e_t^A) - g(e_t^B)] = 0 \quad (15)$$

where e_t^A and e_t^B denotes the forecast errors of model A and model B, and loss function g represents the equality of forecast mean squared errors. Thus, the DM statistics can be defined as:

$$S_1 = \left[\frac{n+1 - 2h + n^{-1}h(h-1)}{n} \right] \frac{\bar{d}}{\sqrt{V(\bar{d})}} \quad (16)$$

Here

$$\bar{d} = n^{-1} \sum_{t=1}^n dt, d_t = g(e_t^A) - g(e_t^B), t = 1, \dots, n \quad (17)$$

and the variance of \bar{d} is:

$$V(\bar{d}) = n^{-1} [\gamma_0 + 2 \sum_{k=1}^{h-1} \gamma_k] \quad (18)$$

Where $\gamma_k = n^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})$; γ_0 is the variance of d_t ; h refers to the h-steps ahead forecasts; and n is the sample size in the testing set.

In addition, we further adopt the approach of model confidence set (MCS) proposed by Hansen et al. (2011) to identify the potential models with superior forecasting ability without a benchmark model to be specified. A MCS is a subset of models that are viewed as the best forecasting models with a certain level of confidence.⁵

Consider a set, M_0 , containing a number of models ($i = 1, \dots, m_0$). The models are evaluated in terms of a loss function over the sample period,⁶ and the loss associated with model i in period t as $L_{i,t}$. Thus, we define the relative performance of different models,

⁴ The methods of evaluation criteria of prediction performance can be found in Appendix A.3.

⁵ The detailed method of model confidence set (MCS) can refer to the paper of Hansen et al. (2011).

⁶ In the paper, we use four different loss functions (i.e., *MSE*, *MAE*, *MAPE*, and *QLIKE*) as the forecasting error benchmark. The details of loss functions can be found in Appendix A3.

$d_{ij,t} \equiv L_{i,t} - L_{j,t}$ for all $i, j \in M_0$. Then the set of superior models can be defined by:

$$M^* \equiv \{i \in M_0 : E(d_{ij,t}) \leq 0 \text{ for all } j \in M_0\}. \quad (19)$$

The MCS is completed via a sequence of significance tests to trim the set of candidate models, M_0 . At each step, the null hypothesis is conducted as follows:

$$H_{0,M} : E(d_{ij,t}) = 0 \text{ for all } i, j \in M \subset M_0 \quad (20)$$

The test aims for the full set of candidate models, $M = M_0$, and if H_0 is rejected, the worst performing model is eliminated from M . The trimming procedure is repeated until the first non-rejection occurs, and the set of surviving models is the model confidence set (MCS), M^* .⁷ We employ the range statistic and the semi-quadratic statistic for the test as follows:

$$T_R = \max_{i,j \in M} \frac{|\bar{d}_{ij}|}{\sqrt{\text{var}(d_{ij})}}, T_{SQ} = \sum_{i < j} \frac{(\bar{d}_{ij})^2}{\text{var}(\bar{d}_{ij})} \quad (21)$$

In the paper, we consider the confidence (significance) level of 90% (10%) and compute the MCS p-values based on the range statistics using the circular block bootstrap. If the MCS p-value is less than 0.1, it indicates that the model should be deleted from the set of M_0 , namely, the forecasting ability of the model is weaker than that of other models.

4.2 Data

The European Climate Exchange (ECX) is the carbon trading market with the largest trading volume under the EU ETS, accounting for 80% of the total trading volume in the EU. We collect daily prices of the carbon futures contracts which mature in December 2016 and December 2017, which are denoted as the Dec16 and the Dec17, respectively. For the two futures contracts, the daily data from January 2, 2012 to December 30, 2016 for the Dec16 and January 2, 2013 to 13 October 2017 for the Dec17, without including public holidays, respectively, is obtained from the website of ECX (<http://www.theice.com>). It results in a total of 1302 and 1241 observations for the two samples, respectively.⁸

For prediction modeling, the sample is divided into the in-sample training set and the out-of-sample testing set, where the training set is used to estimate and optimize the proposed model, and the testing set is used to evaluate the prediction performance of the established model. The detailed division of the two contracts are reported in Table 1. We provide robustness exercises for other forecast sample lengths in Sect. 5. The closing prices of the Dec16 and Dec17 futures contracts are shown in Fig. 15. As shown in the figure, the futures prices illustrate a pattern of time-varying characteristics (e.g., time-varying volatility) and non-normality (e.g., large price movements); in the next section we test the series for nonlinearity as well as non-stationarity.

4.3 Nonstationarity and nonlinearity tests of carbon prices

In this section, we employ the augmented Dicky–Fuller (ADF) and Brock–Decher–Scheikman (BDS) tests to examine the nonstationarity and nonlinearity of Carbon futures prices. In

⁷ The detailed algorithm for constructing the model confidence set (MCS) can refer to the paper of Hansen et al. (2011).

⁸ This approach to forecasting carbon price of futures contract is also adopted in important prior work such as Zhu and Wei (2021).

Table 1 Samples of carbon futures prices

Contracts	Sample	Size	Sample period
Dec16	Sample set	1302	2-Jan-2012 ~ 30-Dec-2016
	Training set	1202	2-Jan-2012 ~ 12-Aug-2016
	Testing set	100	12-Aug-2016 ~ 30-Dec-2016
Dec17	Sample set	1241	2-Jan-2013 ~ 13-Oct-2017
	Training set	1141	2-Jan-2013 ~ 26-May-2017
	Testing set	100	29-May-2017 ~ 13-Oct-2017

Table 2 ADF test result

		Test critical values:			
		1% level	5% level	10% level	
Contract	t-Statistic	-3.436	-2.864	-2.568	Prob.*
Dec16	-2.288				0.176
Dec17	-2.400				0.142

Table 3 BDS test result

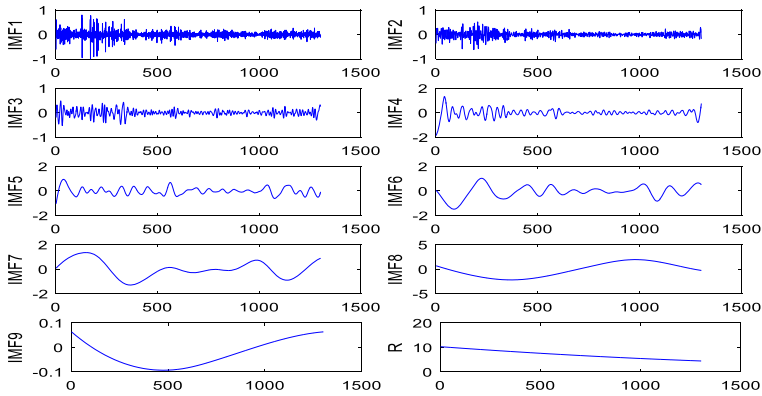
Contracts	m-Dimensional space							
	2		3		4		5	
	t-Stat	Prob	t-Stat	Prob	t-Stat	Prob	t-Stat	Prob
Dec16	0.032	0.000	0.059	0.000	0.081	0.000	0.093	0.000
Dec17	0.031	0.000	0.060	0.005	0.085	0.000	0.098	0.000

the BDS test, the embedding dimension is usually set to 2–5, and the dimensional distance is set to 0.7 times of the variance of data. The results of the two tests are reported in Tables 2 and 3. As shown in the tables, the ADF test indicates that carbon futures prices are nonstationary at all the levels of significance. Further, the BDS test indicates that the prices are nonlinear at the significance level of 1%. These findings confirm prior results that carbon futures are non-linear, non-stationary series and thus alternative approaches to traditional econometrics are worth investigating.

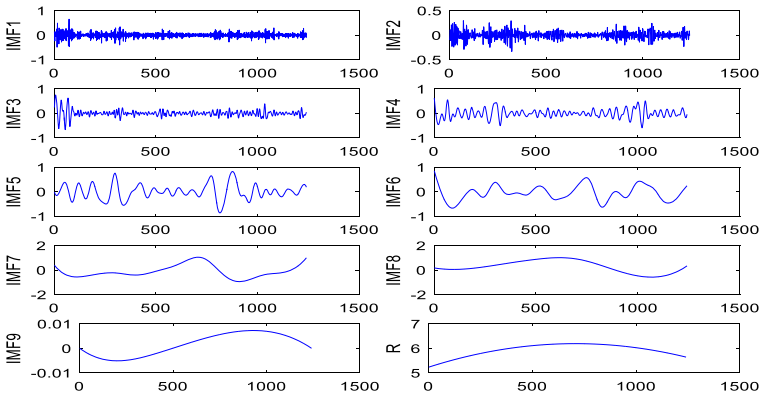
5 Empirical results

5.1 EEMD decomposition

We implement the proposed hybrid model outlined in Sect. 3.4 step by step. First, we apply EEMD method to decompose the carbon prices of the Dec16 and the Dec17. Here the standard deviation of white noise of EEMD decomposition is set to 0.2 times the standard deviation of the carbon futures price, and the number of white noises added is 100. The results of decomposition are reported in Fig. 3.



(a) The EEMD decomposition of carbon futures prices for Dec 16



(b) The EEMD decomposition of carbon futures prices for Dec 17

Fig. 3 The EEMD decomposition of carbon futures prices

As shown in Fig. 3, the carbon prices of Dec 16 and Dec 17 are decomposed into nine IMF functions and one residue (R). The IMFs are arranged by their frequencies, from the highest to the lowest. Since the residue (R) represents the overall trend of the original series, it can be seen from Fig. 3 that the Dec 16 contract shows a downward trend during the sample period, while the tendency of Dec 17 rises at first and then falls during the entire training period.

5.2 The identification of different frequency components (fuzzy entropy and K-means cluster methods)

After obtaining the decomposed components, we use fuzzy entropy as well as the K-means cluster methods to reconstruct the IMFs and the residue. The fuzzy entropy method is employed to calculate the complexity of the IMFs and the residue, where the embedded dimension $m = 2$, tolerance parameter $r = 0.2\sigma_{SD}$, and $n = 2$. The fuzzy entropy values of each IMF and the residue of the Dec 16 and Dec 17 are presented in Fig. 4. As shown in the figure, the fuzzy entropy values of the IMF1 and IMF2 are bigger than 1, while the fuzzy

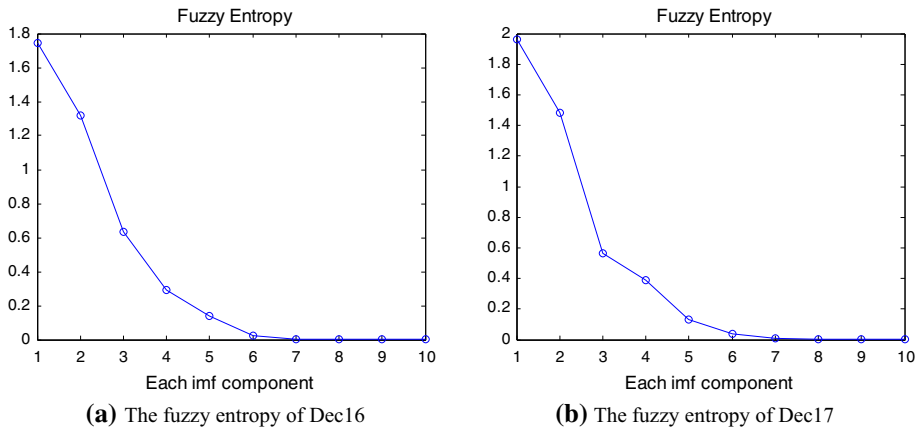


Fig. 4 The fuzzy entropy of the IMFs and the residue

entropy values of the IMF3-IMF9 and the residue are less than 1 for the Dec16. This indicates that the IMF1 and IMF2 have the highest complexity while all other IMFs and the residue have the lower complexity.⁹ In particular, starting from the fuzzy entropy of IMF4, it gradually declines to about 0, indicating the decreasing degree of the complexity. All these show that the high frequency components with the characteristics of higher uncertainty have higher complexity than the low frequency components. Similarly, we find very similar complexity of the different IMFs (residue) of the carbon futures prices for the Dec17.

Since the number of the IMFs and the residue decomposed by the EEMD is large, it takes a lot of observations to adequately train the model as well as long training time for the final modeling if the prediction is directly constructed based on the IMFs. Thus, it is necessary to reconstruct the IMFs and the residue before prediction. In general, high frequency components have the highest fuzzy entropy while low frequency components have minimal fuzzy entropy; the fuzzy entropy values decline monotonically as you move from the first IMF to the second IMF and so on. Since the IMFs with similar frequencies generally have approximate fuzzy entropy values, the fuzzy entropy is a good measure of the similarity for the decomposed IMFs and the residue, which can be used as a clustering factor to reconstruct the decomposed components by the K-means clustering method.

Then, the K-means clustering method is applied to identify different categories for the components of the carbon futures prices for the Dec16 and Dec17 contracts. For the Dec16 contract, the IMF1 and IMF2 are clustered into high-frequency components, IMF3 is clustered into low-frequency components and other components (including the residue) are clustered into trend, respectively. Similarly, for the Dec17 contract, IMF1-IMF2 are clustered into high-frequency components, IMF3-IMF4 are low-frequency components, and other components (including the residue) belong to trend, respectively. Further, the IMFs in each category are aggregated to obtain the high frequency, low frequency and trend series, which reflect the short-term fluctuations (randomness), the strong periodic fluctuations (such as the impact of major events), and the tendency pattern of carbon futures markets, respectively.

The reconstructed high frequency, low frequency and trend series are shown in Fig. 5. For comparisons, the trend series and the real price series are shown in the same figure. As

⁹ The lower entropy value indicates higher relevant feature with respect to the classification and contributes with more information while the higher entropy value indicates less contribution to the classification.

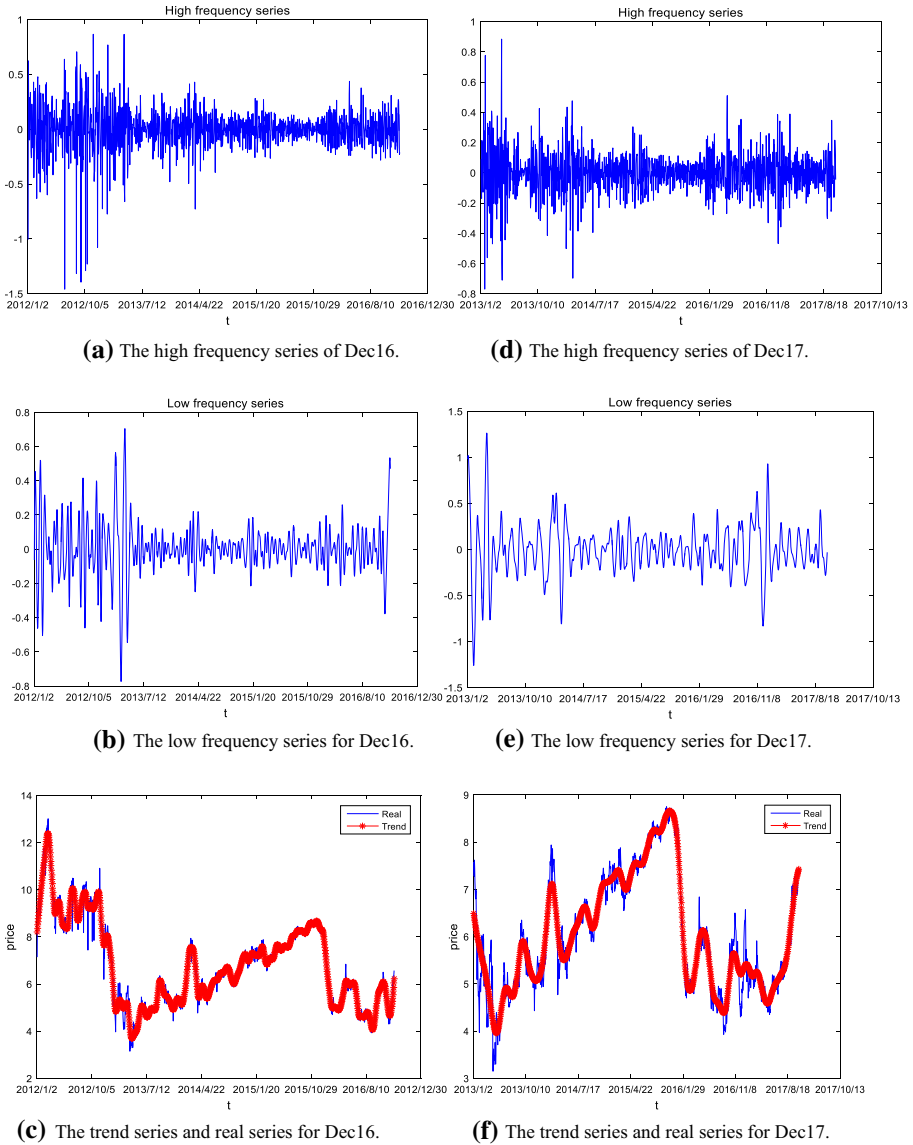


Fig. 5 The reconstructed high frequency, low frequency and trend series

shown in Fig. 5, the trend series accurately depicts the overall trend of the closed prices of carbon futures without losing too much information.

5.3 The identification and modeling of the ARMA model for high and low frequencies series

Under the decomposition-ensemble framework, we first need to further test the stationarity of the three reconstructed series. The results of the Augmented Dickey-Fuller (ADF) test are

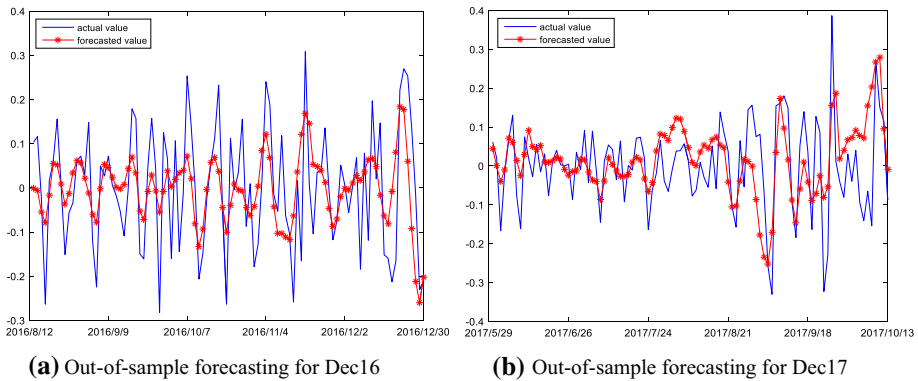


Fig. 6 Out-of-sample forecasting result of high frequency series

reported in Table 13 in “Appendix”. As shown in the table, the high and low frequency series of the two contracts are stationary, while the trend series are non-stationary. The high and low frequency series of both contracts are stationary, as anticipated, which supports the use of an ARMA model. However, the trend series of both contracts are non-stationary, also as anticipated, which supports the use of the extreme learning machine (ELM) model to analyze this component.

Using the above reconstructed stationary series, including the high and low frequencies series, to perform the experiments, we employ the ARMA(p, q) model for the prediction. The structure and parameters of the ARMA model are identified by the autocorrelation and partial correlation analysis. Therefore, the optimal ARMA models for the high and low frequency series are identified by the AIC criteria.¹⁰

5.3.1 Out-of-sample prediction of high frequency series

Finally, the final forecasting results of the high frequency series of the Dec16 and Dec17 are reported in Fig. 6. As shown in the figure, the ARMA model established in this paper can accurately capture the magnitude and direction of the high frequency series changes for the Dec16 and Dec17, and it has strong prediction power and is very suitable for the prediction of high frequency series.

5.3.2 Out-of-sample prediction of low frequency series

The results of the prediction of low frequency series of Dec16 and Dec17 are shown in Fig. 7. As shown in the figure, the magnitude and direction of changes are almost completely captured by the ARMA model established in this paper. Thus, the ARMA model has remarkable prediction power on the low frequency series.

5.4 The ELM modeling of trend series

From the results of stability test in the above section, we find that the trend component is non-stationary. Thus, this section employs the ELM method for modeling the trend series

¹⁰ The detailed results of the identification of ARMA model can be found in Appendix C1.

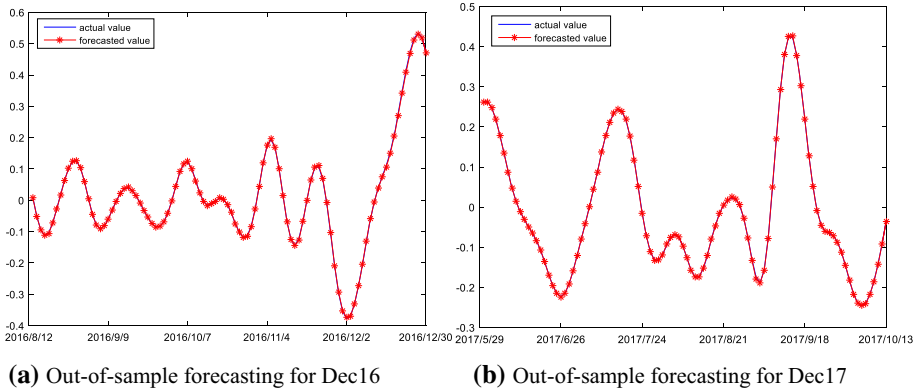


Fig. 7 Out-of-sample forecasting result of low frequency series

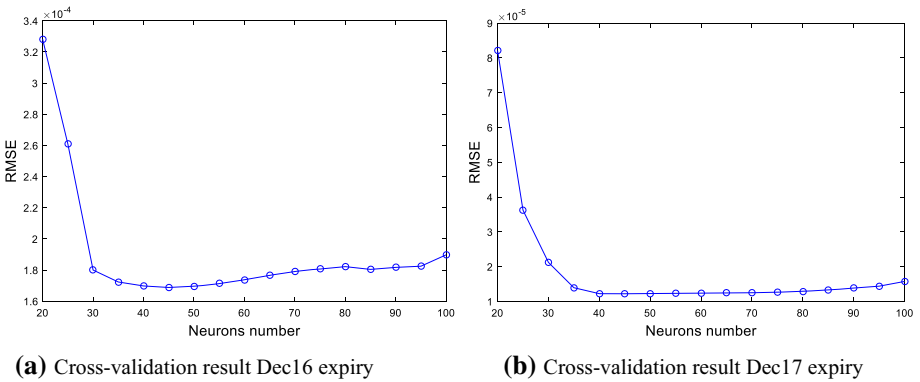


Fig. 8 The RMSE of the cross-validation and grid search methods

because of its strong prediction power in other contexts for time-series which are nonlinear and non-stationary. Before training the ELM model, it is necessary to identify the number and the activation function of neurons in the hidden layer. Therefore, we divide the training sample into a training set and a test set, and then use time-varying cross-validation and grid search methods to identify the optimal number and the optimal activation function of neurons in the hidden layer.¹¹

The results of the time-varying cross-validation and grid search methods are reported in Fig. 8. For the trend series of the Dec16 and the Dec17, the root mean square error (RMSE) of the cross-validation monotonically decreases when the number of neurons in the hidden layer is in Liu and Shi (2013); Zhu & Wei, (2013), while the RMSE basically remains the same or increases slightly as the number of neurons increases during the interval of [45,100]. It indicates that the RMSE of the cross-validation method is smallest when the number of neurons in the hidden layer is around 45. Therefore, the number of neurons in the ELM model is set to 45 for the trend series of the two contracts.

¹¹ The detailed method of time-varying cross-validation can be found in Appendix A4.

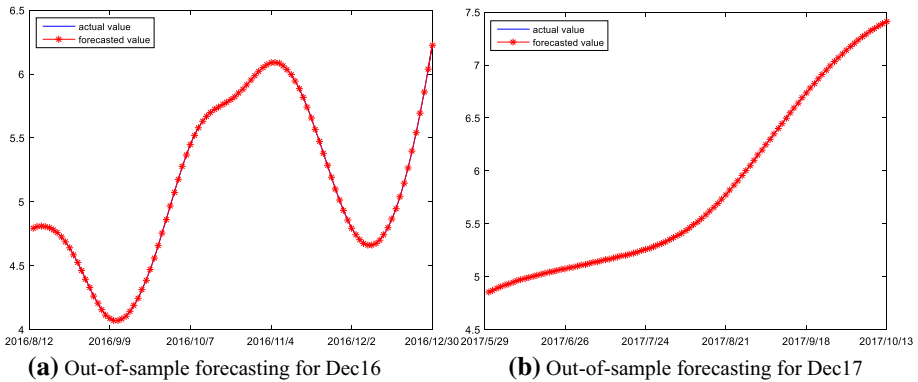


Fig. 9 Out-of-sample forecasting result of trend frequency series

The forecasting results of the Dec16 and the Dec17 are shown in Fig. 9; the figure shows the curves of out-of-sample forecast are almost coincident with their real curves. This indicates that the change magnitudes and movement directions of the trend series are remarkably well captured by the ELM model. This is mainly attributed to the following: on the one hand, the original prices are decomposed into high-frequency, low-frequency and trend components via using the EEMD method, which may solve the problem of mixed frequency components in the prices; on the other hand, when using the reconstruction algorithm proposed in this paper, the decomposed components of carbon prices is reconstructed into three different frequency series according to the complexity of the components. This reconstruction algorithm is advantageous in that it makes the structure of the high/low frequency and trend series simpler, which facilitates the ELM modeling. Thus, all this indicates that the ELM model is highly suitable for the prediction of trend series component.

5.5 Forecasting results of the proposed hybrid model

5.5.1 Comparison between the proposed hybrid model and other models

We aggregate the forecasting results of the three different frequency series (including high frequency, low frequency and trend series) for the Dec16 and the Dec17 and generate the ensemble forecasting result based on the proposed EEMD-FuzzyEn-ARMA-ELM hybrid model. For comparisons, we employ the single ARMA and ELM models, Random walk with drift item, multiscale ensemble prediction models, including the EEMD-FuzzyEn-ARMA and EEMD-FuzzyEn-ELM models, to predict carbon prices as well. The evaluation of out-of-sample forecasting results for the two carbon futures contracts is reported in Table 4.

Table 4, in general, shows that the prediction performance of the proposed EEMD-FuzzyEn-ARMA-ELM hybrid model is very good relative to that of the other models considered for both the Dec16 and the Dec17. In term of the forecast accuracy, documented by the MAPE and RMSE, all the multiscale ensemble prediction models perform much better than the single prediction models (e.g. ARMA and RW). This indicates that the hybrid forecasting models based on the decomposition-reconstruction principle have strong prediction power and substantially improves the prediction accuracy compared to the single models. Furthermore, among the ensemble prediction models, the prediction accuracy of the

Table 4 Comparison of out-of-sample forecasting for different models

Contract	Error	ARMA	ELM	RW	EEMD-FuzzyEn-ARMA	EEMD-FuzzyEn-ELM	EEMD-FuzzyEn-ARMA -ELM
Dec16	MAPE (%)	3.108	3.113	3.030	1.997	1.776	1.661
	RMSE	0.1995	0.1941	0.1890	0.1318	0.1156	0.1054
	R	0.9254	0.9295	0.9323	0.9674	0.9755	0.9798
	DS	0.5960	0.5857	0.5758	0.8182	0.6869	0.8182
Dec17	MAPE (%)	1.735	1.724	1.700	1.232	1.174	1.220
	RMSE	0.1497	0.1496	0.1491	0.1004	0.0929	0.0979
	R	0.9675	0.9681	0.9677	0.9857	0.9872	0.9863
	DS	0.5152	0.5354	0.5455	0.7172	0.6869	0.7172

Here RW refers to Random walk with drift item. In the bold cells, the number in each cell is the statistics of measuring forecast error for the proposed hybrid models in the paper

hybrid models incorporating ELM is superior over that of the hybrid model incorporating ARMA; thus the novel incorporation of ELM into the model is very beneficial. This is mainly attributed to the advantage of ELM in forecasting non-linear and non-stationary series, while the ARMA model is useful for traditional linear, stationary series it is not designed to capture non-linearities or non-stationarities.

The correlation coefficient (R) for the hybrid models is higher than the single ARMA model, the ELM model or random walk model with drift term. The higher correlation coefficient represents the higher degree of association between the predicted and actual values, indicating the hybrid models greatly improve the forecasting accuracy. In terms of the direction prediction, the direction statistics DS , we find that the proposed hybrid models achieve the highest rates of directional accuracy. The hybrid models perform much better than the simple models and rates of 70–80% are impressive given we are looking at noisy daily data. In addition, the ARMA model outperforms the ELM model in capturing the direction prediction of carbon price changes, either in the single model or in the ensemble models.

Following Harvey et al. (1997), we use the adjusted Diebold-Mariano (DM) test to assess the differences between the performance of two forecasts in the paper. The results of the comparison of the forecasting performance for the two contracts are reported in Table 5. We can draw the following conclusions. First, the prediction accuracy of the ensemble models (EEMD-FuzzyEn-ARMA, EEMD-FuzzyEn-ELM and EEMD-FuzzyEn-ARMA-ELM) is significantly better than that of the single models (ARMA, ELM, RW). Second, the proposed hybrid EEMD-FuzzyEn-ARMA-ELM model, in general, demonstrates the superior performance of carbon price prediction than other proposed ensemble models and all the single models. However, for the Dec17 contract statistically there is not a difference between the prediction performance of EEMD-FuzzyEn-ARMA-ELM and EEMD-FuzzyEn-ELM; nevertheless, this underlines that models that feature EEMD, Fuzzy Entropy and Extreme Learning Machine (as both these do) are highly suitable for this application.

In addition, we employ the method of model confidence set (i.e., MCS) proposed by Hansen et al. (2011) to examine the potential models with superior forecasting ability. To obtain the statistics of TR and TSQ and their corresponding MCS p-values, we consider the confidence level of 90% (i.e., significance level $\alpha = 0.10$) for the MCS as suggested by Hansen et al. (2011). The block length is set to 3, and the number of bootstrap samples is 10,000. Table 6 reports the empirical results of the MCS test of the forecasting performance for forecasting models. As shown in the table, under different loss functions (i.e., MSE , MAE , $MAPE$, and $QLIKE$), only hybrid models ever survive in the model confidence set. In particular, the proposed hybrid method of EEMD-FuzzyEn-ARMA-ELM consistently enters the MCS for the Dec 16 and Dec 17.¹² This indicates that our proposed hybrid model generates forecasts that are statistically better than it, which is generally consistent the DM test findings (Table 5).

Figure 10 demonstrates the out-of-sample prediction results of carbon futures prices for the Dec16 and Dec17 using the proposed forecasting model. As shown in the figure, the proposed hybrid model has a strong power in the forecasting of carbon future prices, since it can not only accurately measure the magnitude of carbon futures price changes, but also capture the future direction information in price changes.

¹² Our proposed hybrid method of EEMD-FuzzyEn-ARMA-ELM also enters the MCS for the Dec 20, which can be found in Table 12 in Sect. 6 of Robustness test.

Table 5 The results of the adjusted DM test for different forecasting models

	Alternative model	Benchmark model					
		ARMA	ELM	RW	EEMD-FuzzyEn-ARMA	EEMD-FuzzyEn-ELM	EEMD-FuzzyEn-ARMA-ELM
Dec16	ELM	-0.8825 (0.1898)					
	Rw	-1.972** (0.1357)	-1.106 (0.1357)				
	EEMD-FuzzyEn-ARMA	-3.983*** (0.0000)	-3.628*** (0.0000)	-3.647*** (0.0000)			
	EEMD-FuzzyEn-ELM	-4.794*** (0.0000)	-4.710*** (0.0000)	-4.800*** (0.0000)	-2.087** (0.0197)		
	EEMD-FuzzyEn-ARMA-ELM	-4.993*** (0.0000)	-4.898 (0.0000)	-5.008*** (0.0000)	-3.336*** (0.0000)	-2.728*** (0.0038)	
Dec17	ELM	-0.0468 (0.4814)					
	Rw	-0.3181 (0.3756)	-0.1359 (0.4461)				
	EEMD-FuzzyEn-ARMA	-2.106** (0.0189)	-2.1916** (0.015)	-2.103** (0.0190)			
	EEMD-FuzzyEn-ELM	-2.508*** (0.0069)	-2.6239*** (0.0050)	-2.512*** (0.0068)	-1.080 (0.1414)		
	EEMD-FuzzyEn-ARMA-ELM	-2.209** (0.0148)	-2.301** (0.0118)	-2.208** (0.0148)	-2.646*** (0.0047)	0.7711 (0.2212)	

Test of Diebold-Mariano (DM) examines the null hypothesis that there is no difference of the forecasting performance between the benchmark model and the alternative model. *** indicates $p < .01$; ** indicates $p < .05$; and * indicates $p < .10$. The value in parentheses is P value. RW refers to the Random walk with drift item

Table 6 The MCS p-value of different models for forecasting carbon futures prices

Contract	Model	MSE		MAE		MAPE		QLIKE	
		T _R	T _{SQ}	T _R	T _{SQ}	T _R	T _{SQ}	T _R	T _{SQ}
Dec16	M1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	M2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	M3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	M4	0.0122	0.0039	0.0200	0.0120	0.0214	0.0117	0.0065	0.0021
	M5	0.0122	0.0052	0.0669	0.0669	0.0532	0.0532	0.0065	0.0037
	M6	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>
Dec17	M1	0.0442	0.0290	0.0035	0.0060	0.0008	0.0045	0.0228	0.0145
	M2	0.0438	0.0226	0.0028	0.0053	0.0008	0.0037	0.0206	0.0126
	M3	0.0442	0.0270	0.0040	0.0088	0.0017	0.0062	0.0228	0.0137
	M4	<u>0.1392</u>	<u>0.2210</u>	<u>0.6397</u>	<u>0.5781</u>	<u>0.7353</u>	<u>0.6479</u>	<u>0.2715</u>	<u>0.2893</u>
	M5	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>0.4606</u>
	M6	<u>0.4925</u>	<u>0.4925</u>	<u>0.6397</u>	<u>0.5781</u>	<u>0.7353</u>	<u>0.6479</u>	<u>0.4606</u>	<u>1.0000</u>

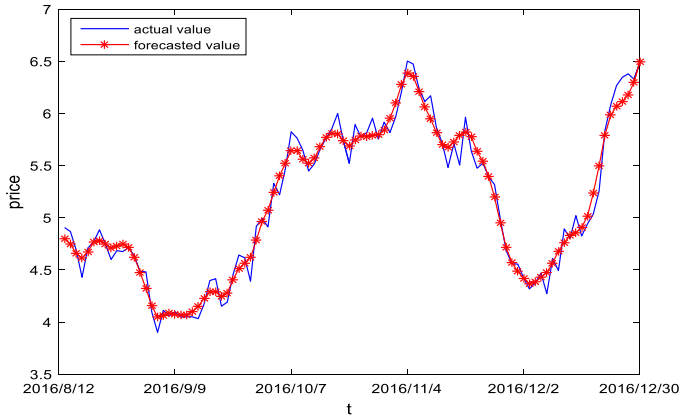
The six forecasting models (referred to as M1-M6) are ARMA, ELM, RW, EEMD-FuzzyEn-ARMA, EEMD-FuzzyEn-ELM, EEMD-FuzzyEn-ARMA-ELM, respectively. The numbers in the cells are the p-values corresponding to the MCS test using 10,000 Bootstrap simulations. The p-value is marked in bold and underlined when it is greater than 0.10, indicating the survival model during the MCS test achieves a better predictive ability

5.5.2 Comparison of prediction performance of the proposed models between different reconstruction algorithms (FuzzyEn-K-mean and fine-to-coarse)

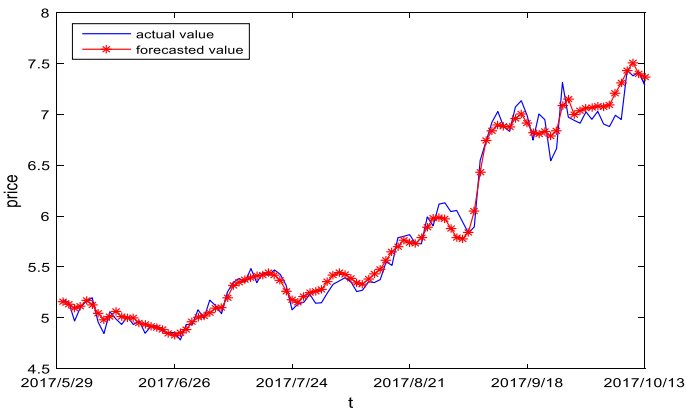
In order to examine the prediction power of the proposed reconstruction algorithms, we further compare the differences of prediction performance for the hybrid models under different reconstruction algorithms, between the proposed reconstruction algorithm and the classical fine-to-coarse reconstruction algorithm.

Table 7 reports the out-of-sample prediction result of the proposed hybrid models using different reconstruction algorithms, including the proposed FuzzyEn-K-mean and traditional fine-to-coarse algorithms. As shown in the table, compared to the fine-to-coarse algorithm, the proposed Fuzzy Entropy and K-mean clustering method is superior. The results show that with the same conditions, the performance of the hybrid models incorporating the proposed algorithm (including EEMD-FuzzyEn-ARMA, EEMD-FuzzyEn-ELM, and EEMD-FuzzyEn-ARMA) is much better than that of the models incorporating fine-to-coarse algorithm. For the Dec16, this reduces the horizontal prediction error index (MAPE) by 38.72%, 45.70% and 49.03%, respectively, and reduces the RMSE error index by 34.49%, 44.64% and 47.61%, respectively. Further, in term of the direction statistics DS index, the hybrid model under the proposed reconstruction algorithm can achieve a higher prediction accuracy by 49.99%, 28.30% and 49.99%, respectively. The prediction results for the Dec17 is very similar to the case of the Dec16, which confirms the robustness of the strong prediction performance of the hybrid model under proposed algorithm in our paper.

Further, we report the direction prediction results of the high-frequency series under two different reconstruction algorithms in Fig. 11. As shown in the figure (as well as in Table 5),



(a) Actual prices and forecasted prices for Dec16



(b) Actual price and forecasted prices for Dec17.

Fig. 10 Actual price and forecasting prices for Dec16 and Dec17

we find that the hybrid model under the proposed reconstruction algorithm (using the fuzzy entropy and K-means clustering methods) outperforms the model under the traditional fine-to-coarse algorithm in capturing the prediction of direction movements in the carbon prices of both the Dec16 and Dec17. For example, the proposed hybrid EEMD-FuzzyEn-ARMA-ELM model accurately predicts the 75 direction changes of high-frequency series among 100 predicted observations for the Dec16, while EEMD-FTC-ARMA-ELM only accurately predicts 51 direction changes among the 100 observations. Similarly, the hybrid model under the proposed reconstruction algorithm demonstrates a higher power of direction prediction than the model under the traditional fine-to-coarse algorithm for the Dec17.

In order to compare the forecasting performance of the hybrid models with the proposed FuzzyEn-K-mean to the traditional fine-to-coarse algorithms, we use the adjusted Diebold-Mariano test (Harvey et al., 1997) to assess their predictive performance. The results of tests of equal predictive performance between two reconstruction algorithms for the Dec16 and Dec17 are reported in Table 8. We can draw the following conclusions. Overall, with the same conditions, the prediction accuracy of the ensemble models based on the proposed

Table 7 Comparison of prediction performance for the hybrid models under different reconstruction algorithms

Contract	Error	EEMD-FTC-ARMA	EEMD-FTC-ELM	EEMD-FTC-ARMA-ELM	EEMD-FuzzyEn-ARMA	EEMD-FuzzyEn-ELM	EEMD-FuzzyEn-ARMA-ELM
Dec16	MAPE (%)	3.259	3.271	3.259	1.997 (38.72%)	1.776 (45.70%)	1.661 (49.03%)
	RMSE	0.2012	0.2088	0.2012	0.1318 (34.49%)	0.1156 (44.64%)	0.1054 (47.61%)
	R	0.9305	0.9213	0.9305	0.9674 (3.97%)	0.9755 (4.84%)	0.9798 (5.30%)
	DS	0.5455	0.5354	0.5455	0.8182 (49.99%)	0.6869 (28.30%)	0.8182 (49.99%)
Dec17	MAPE (%)	1.662	1.627	1.662	1.232 (25.87%)	1.174 (27.84%)	1.220 (26.59%)
	RMSE	0.1412	0.1391	0.1412	0.1004 (28.90%)	0.0929 (33.21%)	0.0979 (30.67%)
	R	0.9704	0.9713	0.9704	0.9857 (1.48%)	0.9872 (1.64%)	0.9863 (1.64%)
	DS	0.5051	0.5354	0.5051	0.7172 (41.99%)	0.6869 (28.30%)	0.7172 (41.99%)

Here FTC refers to the fine-to-course reconstruction algorithm. **In the bold cells**, the first number in each cell is the statistics of measuring forecast error for the proposed hybrid models in the paper, and the value in parentheses is the percent of the increase in prediction accuracy using the same model under different reconstruction algorithms between proposed reconstruction and fine-to-course

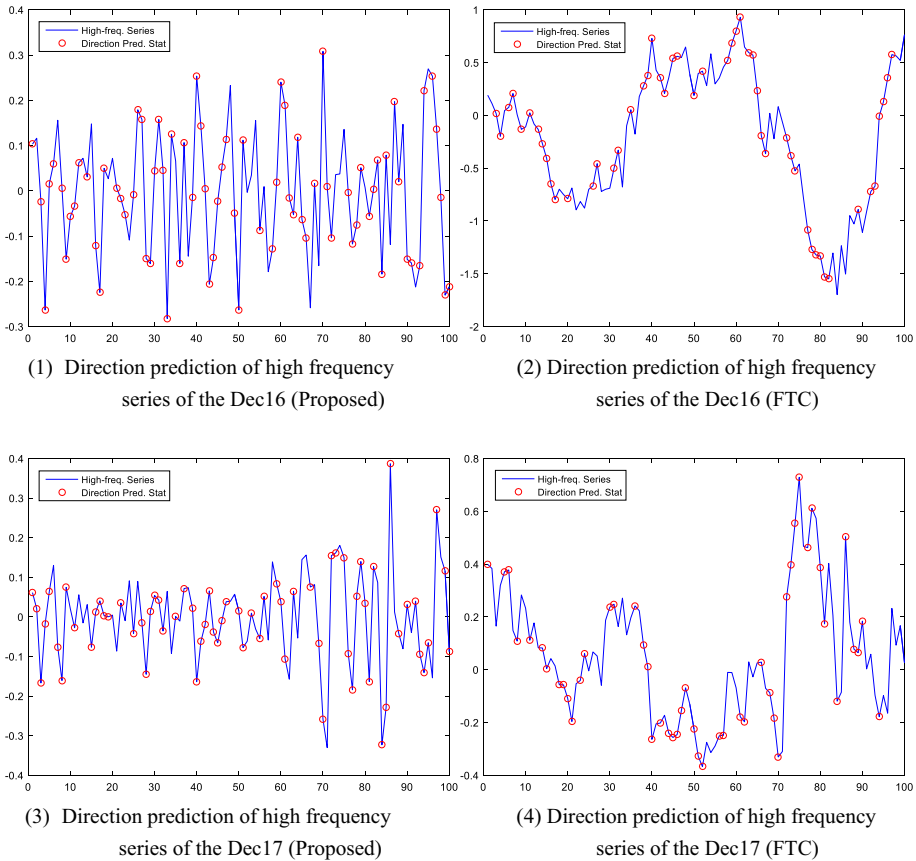


Fig. 11 The direction prediction of high frequency series under different reconstruction algorithms Notes: “Proposed” refers to the proposed reconstruction algorithm, using the Fuzzy Entropy and K-means clustering methods in the paper; “FTC” represents the traditional fine-to-coarse reconstruction algorithm

reconstruction algorithm (using the Fuzzy entropy and K-means clustering methods), including EEMD-FuzzyEn-ARMA, EEMD-FuzzyEn-ELM and EEMD-FuzzyEn-ARMA-ELM, is significantly better than that of the ensemble models based on the traditional fine-to-coarse algorithm.

In sum, our evidence clearly supports the proposed hybrid model which demonstrates superior performance in carbon price prediction than other ensemble models and all the single models. Compared to the fine-to-coarse algorithm, the proposed Fuzzy Entropy and K-mean clustering methods has the advantage that they can easily analyze the complexity of the IMFs and the residue and thus reconstruct the three different frequency series more efficiently and accurately for prediction. Further, we benefit from the advantages of the prediction power of the ARMA model for the stationary high and low frequency series and the extreme learning machine (ELM) method for the non-stationary trend series. The forecasts of all three components (high frequency, low frequency and trend) based on these forecasting models are aggregated to form the final forecast of the carbon price. Therefore, the new hybrid

Table 8 The results of the adjusted DM test for the hybrid models under different reconstruction algorithms

Contract	Alternative model	Benchmark model		
		EEMD-FTC-ARMA	EEMD-FTC-ELM	EEMD-FTC-ARMA-ELM
Dec 16	EEMD-FuzzyEn-ARMA	-4.136*** (0.000)	-3.878*** (0.000)	-4.137*** (0.000)
	EEMD-FuzzyEn-ELM	-5.386*** (0.000)	-4.883*** (0.000)	-5.387*** (0.000)
	EEMD-FuzzyEn-ARMA-ELM	-5.707*** (0.000)	-5.147*** (0.000)	-5.707*** (0.000)
Dec 17	EEMD-FuzzyEn-ARMA	-1.980** (0.025)	-1.994** (0.025)	-1.979** (0.025)
	EEMD-FuzzyEn-ELM	-2.458*** (0.008)	-2.495*** (0.007)	-2.458*** (0.008)
	EEMD-FuzzyEn-ARMA-ELM	-2.102** (0.019)	-2.123** (0.018)	-2.208** (0.015)

Test of Diebold-Mariano (DM) examines the null hypothesis that there is no difference of the forecasting performance between the benchmark model and the alternative model. “FuzzyEn” refer to the proposed reconstruction algorithm of Fuzzy entropy and K-means clustering methods; “FTC” represents the traditional fine-to-coarse reconstruction algorithm. *** indicates $p < .01$; ** indicates $p < .05$; and * indicates $p < .10$. The value in parentheses is P value

forecasting model under the decomposition-reconstruction principle proposed in this paper significantly improves the prediction accuracy of carbon futures prices.

6 Further analysis

6.1 Robustness tests

In this section, for robustness testing of the prediction ability of the proposed hybrid models, we proceed a couple of out-of-sample predictions using alternative testing samples for the Dec16, including the size of testing samples with 200, 300, 400 and 500 observations, respectively. This means that we reserve the samples of 200, 300, 400, and 500 as the testing sets, while the other observations of the Dec16 are used to train the proposed models. We follow the same steps used in the above section to examine the prediction performance for different models.

Table 9 reports the comparison of the forecasting results of these four different periods. We find that the results of robustness tests of the prediction performance of the hybrid models are consistent with the previous findings. First, the results show that the hybrid models outperform the single models in the prediction of the price level. For example, when using the testing set of 200 observations, compared with the single ARMA model, the ELM model and the random walk model with drift term, the proposed hybrid models increase the prediction accuracy by more than 44%, on average than these single models. Similar, we find the very same results of prediction when using the testing set of 300, 400 or 500 observations. Second, in terms of the direction prediction of price movements, the direction statistics DS is more than 70% for the proposed hybrid models, while only about 50% for the single models. In particular, the hybrid models incorporating the ARMA model, including EEMD-FuzzyEn-ARMA-ELM and EEMD-FuzzyEn-ARMA models, achieve a high level of prediction accuracy in the direction movements at about 77% and 76% for the testing sample of 200 observations of the Dec16.

Figure 12 presents the forecasting results of different models using different testing sample periods. As shown in the figure, in terms of the level prediction, the indicators of the MAPE, the RMSE and the correlation coefficients (R) of the hybrid forecasting model (i.e., EEMD-FuzzyEn-ARMA-ELM) proposed in this paper are significantly better than those of all other forecasting models under different periods. Second, the proposed hybrid models outperform the single models in the direction prediction of price movements. Although the EEMD-FuzzyEn-ARMA-ELM model is not always superior to other hybrid models in terms of price direction prediction, its forecast performance still ranks the second in all forecasting models. The high DS statistics of the proposed hybrid model, which exceeds 70% in all five forecasting periods, demonstrates high ability to capture the direction of price movements within different forecasting periods.

6.2 The COVID-19 pandemic period

The COVID-19 pandemic has a substantial impact on financial markets, including the carbon futures market as well as the economies of many countries during 2020. How reliable is the forecasting performance of the hybrid proposed model during this period on pandemic and great uncertainty? We therefore follow the same framework to examine the extreme data

Table 9 The forecasting results for the Dec 16 using different testing samples

Size	Error	ARMA	ELM	RW	EEMD-FuzzyEn-ARMA	EEMD-FuzzyEn-ELM	EEMD-FuzzyEn-ARMA-ELM
200 Obs	MAPE(%)	2.740	2.754	2.688	1.783	1.568	1.510
	RMSE	0.1902	0.1903	0.1817	0.1209	0.1082	0.1029
	R	0.9246	0.9243	0.9314	0.9698	0.9757	0.9781
	Ds	0.5226	0.5075	0.5176	0.7739	0.7337	0.7588
300 Obs	MAPE(%)	2.459	2.525	2.458	1.425	1.405	1.324
	RMSE	0.1813	0.1857	0.1764	0.1037	0.1025	0.0969
	R	0.9811	0.9801	0.9823	0.9938	0.9939	0.9945
	Ds	0.5017	0.4883	0.5084	0.7291	0.7592	0.7358
400 Obs	MAPE(%)	2.068	2.070	2.054	1.265	1.210	1.139
	RMSE	0.1645	0.1614	0.1590	0.0983	0.0939	0.0884
	R	0.9882	0.9887	0.9892	0.9958	0.9962	0.9966
	Ds	0.4987	0.4937	0.4987	0.7118	0.6842	0.7118
500 Obs	MAPE(%)	1.951	1.958	1.928	1.250	1.178	1.082
	RMSE	0.1611	0.1596	0.1564	0.0999	0.0947	0.0866
	R	0.9870	0.9872	0.9878	0.9950	0.9955	0.9962
	Ds	0.4850	0.4810	0.4990	0.7114	0.6814	0.7094

Here RW refers to Random walk with drift item. In the bold cells, the number in each cell is the statistics of measuring forecast error for the proposed hybrid models in the paper.

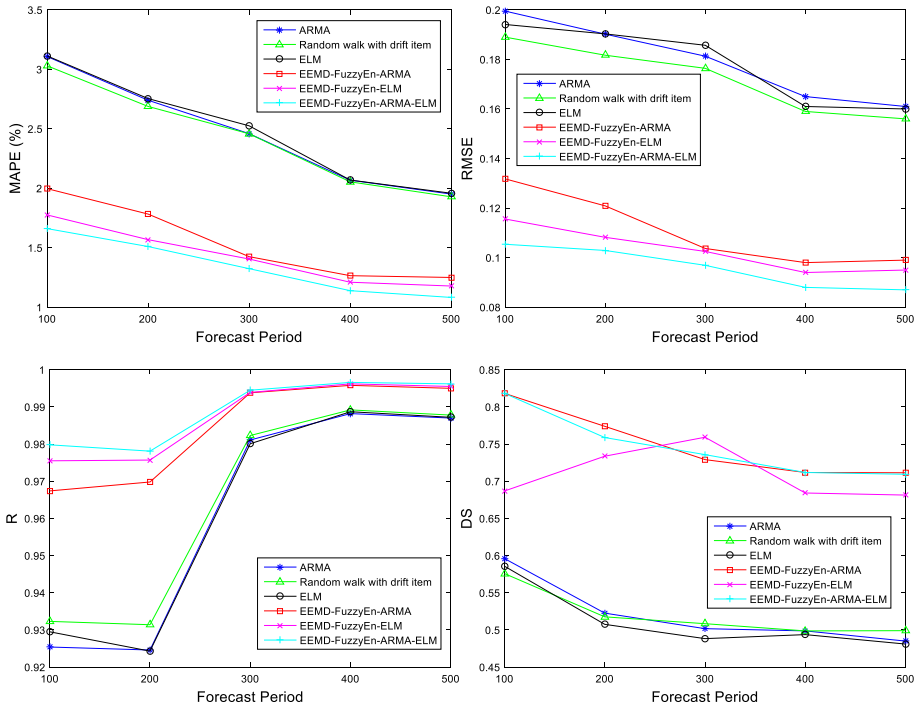


Fig. 12 The forecasting results of different models across different testing sample sizes

during the pandemic period.¹³ Figure 13 presents the out-of-sample prediction result of carbon futures prices for the Dec20 using the proposed hybrid forecasting model. As shown in the figure, the result demonstrates that our proposed hybrid model has a strong power in the forecasting of carbon futures prices during the pandemic period, capturing the magnitude of carbon futures price changes as well as the direction information in price dynamics well. In fact, this should not be surprising since the hybrid model is well suited to capture the effects of high uncertainty witnessed during the COVID pandemic period.

The result of out-of-sample forecasting performance for the carbon futures contracts of the Dec20 are reported in Table 10. As shown in the table, in general, it indicates that the prediction performance of the proposed EEMD-FuzzyEn-ARMA-ELM hybrid model is very good relative to that of other models. In term of the forecast accuracy, documented by the MAPE and RMSE, all the multiscale ensemble prediction models perform much better than the single prediction models (e.g. ARMA, ELM, and RW).

This further indicates that the hybrid forecasting models based on the decomposition-reconstruction principle have strong prediction power and substantially improves the prediction accuracy even during the pandemic period. Among the ensemble prediction models, the prediction accuracy of the hybrid models incorporating both ARMA and ELM is superior over that of the hybrid model incorporating only ARMA or ELM. Furthermore, in

¹³ We collect the daily prices of carbon futures contracts maturing in December 2020 (i.e., the Dec20), covering the period from January 1, 2016 to December 31, 2020. It results in a sample of total 1305 observations, which is divided into two sets, including the training set (i.e., 1044 observations) and the testing set (i.e., 261 observations), respectively.

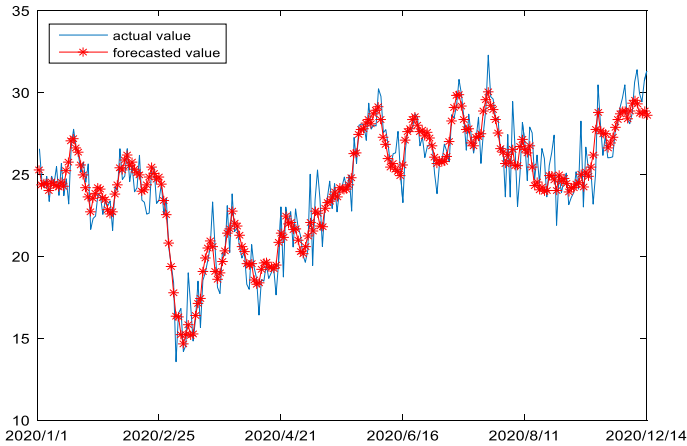


Fig. 13 Actual price and forecasting prices for Dec20

terms of the correlation coefficient (R), the higher coefficient indicates that our proposed hybrid model outperforms other models in forecasting accuracy. Lastly, based on the direction prediction, the direction statistics (DS), we find that the proposed hybrid model achieve the highest rates of directional accuracy. In addition, even considering the sample of the pandemic period, the hybrid models generally perform much better than the simple models. We further use the adjusted Diebold-Mariano (DM) test to compare the forecasting performance across different models for the Dec20 contract. As shown in Table 11, it indicates the prediction accuracy of the ensemble models incorporating the Extreme Learning Machine (i.e., ELM) is significantly better than that of the single models. Second, the proposed hybrid EEMD-FuzzyEn-ARMA-ELM model, demonstrates superior forecasting performance for carbon price than to all other proposed models.

Finally, we employ the MCS test to examine the potential group of models with superior forecasting ability for the Dec20. Table 12 reports the empirical results of the MCS test of the forecasting performance for different models. As shown in the table, under different loss functions, the proposed hybrid method of EEMD-FuzzyEn-ARMA-ELM is consistently the only model in the MCS. This is broadly consistent with the empirical findings of the MCS test for the Dec16 and Dec17. This further emphasizes the importance of incorporating Fuzzy Entropy and Extreme Learning Machine when forecasting the carbon price.

In summary, the empirical evidence provided indicates that the forecasting results obtained by the hybrid prediction model based on EEMD decomposition and reconstruction are significantly better than the single prediction models across different testing sample periods as well as the COVID-19 pandemic period, which indicates the proposed framework proposed in this paper is reasonable and robust. The hybrid prediction model using the decomposition-reconstruction algorithm consistently achieves more accurate prediction results both in terms of predicting the magnitude of carbon futures price change as well as the direction carbon futures price movements.

Table 10 Comparison of out-of-sample forecasting performance for the Dec20

Contract	Error	ARMA	ELM	RW	EEMD-FuzzyEn-ARMA	EEMD-FuzzyEn-ELM	EEMD- FuzzyEn-ARMA -ELM
Dec20	MAPE (%)	5.1043	5.4555	5.0823	4.8577	4.1012	3.9311
	RMSE	1.5330	1.6031	1.5315	1.4666	1.2239	1.1776
	R	0.8220	0.8053	0.8223	0.8381	0.8869	0.8952
	DS	0.5060	0.4940	0.5060	0.7390	0.6707	0.7390

Here RW refers to Random walk with drift item. In the bold cells, the number in each cell is the statistics of measuring forecast error for the proposed hybrid models in the paper

Table 11 Results of the adjusted DM test of different forecasting models for Dec20

Alternative model	Benchmark model					
	ARMA	ELM	RW	EEMD-FuzzyEn-ARMA	EEMD-FuzzyEn-ELM	EEMD-FuzzyEn-ARMA-ELM
Dec20 ELM	2.0828 (0.019)					
RW	-0.2625 (0.397)	-2.218** (0.014)				
EEMD-FuzzyEn-ARMA	-0.899 (0.185)	-1.800** (0.0365)	-0.877 (0.1906)			
EEMD-FuzzyEn-ELM	-4.702*** (0.0000)	-5.017*** (0.0000)	-4.664*** (0.0000)	-4.0687*** (0.0000)		
EEMD-FuzzyEn-ARMA-ELM	-4.920*** (0.0000)	-5.202*** (0.0000)	-4.873*** (0.0000)	-5.1563*** (0.0000)	-1.936** (0.0270)	

Test of Diebold-Mariano (DM) examines the null hypothesis that there is no difference of the forecasting performance between the benchmark model and the alternative model. *** indicates $p < .01$; ** indicates $p < .05$; and * indicates $p < .10$. The value in parentheses is P value. RW refers to the Random walk with drift item

Table 12 MCS p -value of forecasting models for Dec20carbon futures prices

Contract	Model	MSE		MAE		MAPE		QLIKE	
		T _R	T _{SQ}	T _R	T _{SQ}	T _R	T _{SQ}	T _R	T _{SQ}
Dec20	M1	0.0003	0.0000	0.0000	0.0001	0.0006	0.0000	0.0002	0.0001
	M2	0.0003	0.0000	0.0001	0.0001	0.0006	0.0000	0.0002	0.0001
	M3	0.0003	0.0001	0.0001	0.0001	0.0006	0.0000	0.0002	0.0001
	M4	0.0011	0.0014	0.0009	0.0010	0.0035	0.0025	0.0051	0.0026
	M5	0.0458	0.0458	0.0563	0.0563	0.0402	0.0402	0.0251	0.0251
	M6	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>	<u>1.0000</u>

Notes The six forecasting models (referred to as M1-M6) are ARMA, ELM, RW, EEMD-FuzzyEn-ARMA, EEMD-FuzzyEn-ELM, EEMD-FuzzyEn-ARMA-ELM, respectively. The numbers in the cells are the p -values corresponding to the MCS test using 10,000 Bootstrap simulations. The p -value is marked in bold and underlined when it is greater than 0.10, indicating the survival model during the MCS test achieves a better predictive ability

7 Conclusions

In this paper, we propose a novel hybrid model that builds on ensemble empirical mode decomposition (EEMD) models by incorporating fuzzy entropy (*FuzzyEn*) and extreme learning machine (ELM) methods. We successfully apply this method to forecast EU carbon futures prices that are characterized by complexity and chaos.

First, we find strong forecasting performance of the EEMD model that includes *FuzzyEn* and ELM both in terms of predictive accuracy and in terms of sign prediction. This model also performs substantially better than simpler models such as ARMA and random walk; for predictive accuracy it can reduce forecast errors by about 40% relative to these models. Second, we undertake time-varying cross-validation tests whether the ELM model is suitable for the prediction of trend series; these tests further support the hybrid model proposed. Third, we undertake several robustness tests that suggest the main results are qualitatively very similar to changes in the length of the forecast sample. Finally, the hybrid model developed in this paper has the potential to be used in a wide range of future applications especially to series that are non-stationary and that feature complex and chaotic properties. In particular, there is scope for models featuring fuzzy entropy and extreme learning machine to be applied in many (other) areas of Management Science, Economics and Finance.

The paper has important practical and policy implications. First, the hybrid model proposed in the paper is highly advantageous in asset price prediction, which is of great significance for financial investors. The proposed model enables us to predict asset prices, as illustrated for carbon futures prices, more accurately, thus encouraging investors' participation in asset markets and helping improve trading performance. Second, it is helpful for governments when determining the supply of carbon permits to avoid extreme levels of prices thereby facilitating the market to encourage reductions in carbon emissions. In particular the supply of carbon permits can be adjusted in response to the price levels and trend provided by the model. In addition, an interesting extension of our work would be to investigate whether forecast accuracy can be further improved by including the impact of climate change in the proposed hybrid model.

Acknowledgements We thank three anonymous referees for their constructive and valuable comments that help us to improve the paper. Any remaining errors are our own responsibility. Peng Chen is also grateful for the grant from the Humanities and Social Science Foundation of the Ministry of Education of China (No. 20YJC790012) and the National Natural Science Foundation of China (No. 72071095).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A

Empirical mode decomposition (EMD)

The Empirical mode decomposition method, first proposed by Huang et al. (1998), can be used to decompose a given series into a series of intrinsic mode functions (IMFs) and a residue. The IMFs must meet the following conditions: (1) the difference between the number of extreme points and the number of zero crossings in the whole series is less than or equal to 1; (2) the mean of the upper and lower envelopes is 0 at any time. The specific decomposition steps are as follows:

Step 1: Find out all extreme points of time series $x(t)$, and obtain local maxima x_{\max} and local minima x_{\min} . Then use the cubic spline interpolation method to obtain its upper and lower envelopes (i.e., $e_{\max}(t)$ and $e_{\min}(t)$). Finally, calculate the mean envelope $m(t)$ as follows: $m(t) = \frac{e_{\min}(t) + e_{\max}(t)}{2}$.

Step 2: Compute the difference between $x(t)$ and $m(t)$ to obtain $h(t)$: $h(t) = x(t) - m(t)$;

Step 3: Check the properties of $h(t)$ whether meets the conditions of IMFs. If it meets, then $h(t)$ is defined as a new IMF; otherwise, let $x(t) = h(t)$, return to step 1 and repeat the above steps until the conditions of IMFs are met;

Step 4: let the residue $r(t) = x(t) - imf(t)$;

Step 5: Take $r(t)$ as a new series to be decomposed, repeat the above steps until the n -th order IMF component $imf_n(t)$ or its residue $r_n(t)$ is either less than the default value or become a monotonic function;

Finally, $x(t)$ is decomposed into the sum of all IMF components and a residue: $x(t) = \sum_{i=1}^n imf_i(t) + r_n(t)$, where n is the number of IMFs, and $r_n(t)$ is the final residue.

ARIMA

The ARIMA model consists of auto-regressive (AR) model and moving average (MA) model. For a given time series y_t , the AR (p) model can be defined as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \quad (\text{A1})$$

And the MA (q) model is defined as:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \tag{A2}$$

where the random term ε_t is independent white noise.

So the autoregressive integrated moving average (ARIMA (p, d, q)) model is to combine the AR(p) and MA (q) models, where d is the order of the difference, which indicates that the series y_t is a stationary series after the d order difference. Then the ARIMA (p, d, q) model can be defined as follows:

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t \tag{A3}$$

where L is the lag operator, that is $Ly_t = y_{t-1}$, $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ and $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$ are p -order, and q -order lag polynomials, respectively. $\phi_i, \theta_i, i = 1, 2, \dots, p$ are the coefficients of AR and MA models respectively, where $(1 - L)^d$ represents d -order differential operation.

Additional evaluation criteria and loss functions

Here we present different measures to evaluate the forecasting performance of the models, including root mean square error (RMSE), mean absolute percentage error (MAPE), correlation coefficient (R) and direction statistic (DS), as the evaluation criteria. Let y_i and \hat{y}_i ($i = 1, 2, \dots, N$) refer to the actual and predicted values, where N is the sample size of the testing set. Root mean square error (RMSE) and mean absolute percentage error (MAPE) are calculated as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \tag{A4}$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{A5}$$

Correlation coefficient (R) and direction statistic (DS) are defined as follows:

$$R = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \text{mean}(\bar{y}))(\hat{y}_i - \text{mean}(\hat{y}))}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \text{mean}(\bar{y}))^2} * \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - \text{mean}(\hat{y}))^2}} \tag{A6}$$

The correlation coefficient R measures the degree of linear similarity between the actual and predicted values. The larger value of R means the higher degree of collinearity between the actual and predicted values, indicating a better forecasting performance of the model. The direction statistic (DS) indicator can be used to evaluate the forecasting performance of the model from the perspective of price change directions (ups or downs). It is calculated as:

$$DS = \frac{1}{N} \sum_{i=1}^N w_i \tag{A7}$$

where w_i is as follows;

$$w_i = \begin{cases} 1, & \text{if } (y_{i+1} - y_i)(\hat{y}_{i+1} - \hat{y}_i) \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{A8}$$

Similarly, the larger DS indicator indicates much more accurate forecasting of the direction for carbon price changes.

Moreover, when conducting the test of model confidence set (i.e., MCS) of Hansen et al. (2011) to identify the potential superior forecasting models, we use four different loss functions (i.e., *MSE*, *MAE*, *MAPE*, and *QLIKE*) as the forecasting error benchmark to compare the forecasting performance. Among them, *RMSE* and *MAPE* are already presented as Eqs. A4 and A5, respectively; similarly, *MAE* and *QLIKE* can be expressed as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{A9}$$

$$QLIKE = \frac{1}{N} \sum_{i=1}^N (\ln \hat{y}_i + y_i / \hat{y}_i) \tag{A10}$$

where N is the number of forecasting data points.

Time-varying cross-validation method

See Fig. 14.

In the traditional K -fold cross-validation method, the training sample is divided into K equal parts and then the prediction errors for all K cases are obtained. Next, the mean of K prediction errors is calculated and finally the parameter set with the smallest error is selected. However, since the traditional cross-validation method ignores the long-term memory characteristics commonly associated with carbon futures; thus this major drawback means it is not well suited to our application. Therefore, we propose a time-varying cross-validation method, and the framework of the method is shown in Fig. 14.

In Fig. 14, TS is the reserved sample of the model, IN is the first round sample used for model training, $S_i = IN + i - 1$ is the i -th round sample ($i = 1, 2, \dots, TS$) used for model training, and h_{max} is the sample size of validation set. Therefore, an observation from the reserved sample is added to the training set in order when the cross-validation is carried out. The number of observations of training set increases in a rolling manner, and eventually the $TS + 1$ training models are obtained. In contrast to the traditional cross-validation method ignoring the serial correlation of time series, the proposed method considers the characteristics of the serial correlation instead of splitting the training sample equally. Here the sample size of the reserved sample TS and the validation set h_{max} are set to 20 and 100 (the sample size is

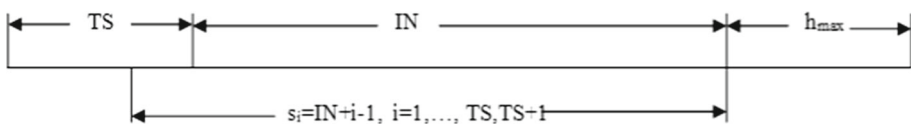


Fig. 14 Time-varying cross-validation method

equal to the sample size of testing set) respectively, and the range of the number of neurons in the hidden layer searched by the grid search method is set as [20, 25... 100].

Appendix B

See Tables 13 and 14 and Fig. 15 .

Table 13 The results of ADF test

		Test critical values				
		1% level	5% level	10% level		
Contract	Reconstructed Series	t-Statistic	-3.436	-2.864	-2.568	Prob.*
Dec16	High	-20.82				0.0000
	Low	-12.56				0.0000
	Trend	-2.230				0.1958
Dec17	High	-18.98				0.0000
	Low	-9.074				0.0000
	Trend	-1.9416				0.8769

Table 13 presents the result of the stationarity test (ADF) of the three reconstructed series (high frequency, low frequency and trend series) of the Dec16 and the Dec17 contracts. The result indicates that the high and low frequency series of the two contracts are stationary, while the trend series of the contracts are non-stationary

Table 14 The identification of high/low frequency components (using fine-to-coarse reconstruction algorithms)

Contract	Stat	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆	IMF ₇	IMF ₈	IMF ₉
Dec 16	mean	0.002	0.003	0.005	-0.002	-0.009	-0.097	-0.058	-1.472	-1.832
	t	0.420	0.618	0.748	-0.200	-0.543	-4.306	-1.932	-2.927	-3.606
	P-value	0.675	0.537	0.455	0.841	0.587	0.000	0.054	0.003	0.000
Dec 17	mean	0.001	0.001	0.004	-0.001	-0.001	-0.048	-0.176	0.193	0.194
	t	0.241	0.251	0.669	-0.070	-0.063	-2.785	-7.422	5.939	5.972
	P-value	0.809	0.802	0.504	0.944	0.950	0.005	0.000	0.000	0.000

Table 14 presents the identification of high–frequency and low-frequency components using the fine-to-coarse algorithm. The mean of the IMF component C_i ($i = 1, 2, \dots, 9$) for the Dec 16 and Dec 17 is examined via t-test. As shown, starting from the C_6 of both the Dec 16 and Dec 17, the mean value of the IMF series is not equal to be 0 at the significance level of 1%. Therefore, we use the fine-to-coarse algorithm to reconstruct $IMF1 \sim IMF5$ to the high frequency series, and reconstruct $IMF6 \sim IMF9$ to the low frequency series

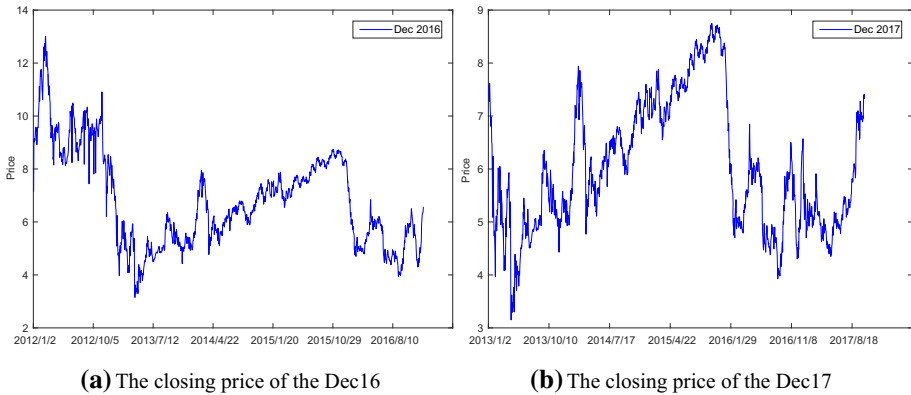


Fig. 15 The closing price of the Dec16 and Dec17

Appendix C

Result of identification of ARMA model

We identify the optimal ARMA models for the high frequency series of the Dec16 and the Dec17 by the AIC criteria. The result shows that the optimal ARMA model is identified as the form of ARMA (4, 4) for the high frequency series of the Dec16. Similarly, the structure of ARMA model for the high frequency series of the Dec17 is identified as ARMA (5, 4). Further, the optimal ARMA model for the high frequency series of the training sample of the Dec16 is:

$$\begin{aligned}
 High_t = & -0.211High_{t-1} + 0.782High_{t-2} - 0.203High_{t-3} - 0.340High_{t-4} \\
 & + 0.106High_{t-5} + \varepsilon_t + 1.243\varepsilon_{t-2} - 0.544\varepsilon_{t-4}
 \end{aligned} \tag{C1}$$

And the optimal ARMA model for high frequency series of the Dec17 is:

$$\begin{aligned}
 High_t = & 0.334High_{t-1} + 0.442High_{t-3} - 0.659High_{t-4} + 0.297High_{t-5} \\
 & + \varepsilon_t + 0.552\varepsilon_{t-1} + 0.545\varepsilon_{t-2} + 0.649\varepsilon_{t-3} - 0.858\varepsilon_{t-4}
 \end{aligned} \tag{C2}$$

Similarly, the optimal ARMA model for the low frequency series of the Dec16 is:

$$\begin{aligned}
 Low_t = & 3.123Low_{t-1} - 4.023Low_{t-2} + 2.510Low_{t-3} - 0.651Low_{t-4} \\
 & + \varepsilon_t - 0.747\varepsilon_{t-1} - 0.495\varepsilon_{t-2} - 0.409\varepsilon_{t-3} - 0.219\varepsilon_{t-4}
 \end{aligned} \tag{C3}$$

And the optimal ARMA model for the low frequency series of the Dec17 is:

$$\begin{aligned}
 Low_t = & 2.323Low_{t-1} - 1.099Low_{t-2} - 1.606Low_{t-3} + 2.0656Low_{t-4} \\
 & - 0.712Low_{t-5} + \varepsilon_t - 2.026\varepsilon_{t-1} - 1.854\varepsilon_{t-2} - 1.595\varepsilon_{t-3} - 1.306\varepsilon_{t-4} \\
 & - 0.726\varepsilon_{t-5} - 0.1887\varepsilon_{t-6}
 \end{aligned} \tag{C4}$$

Result of identification of ARMA model during the pandemic

Following the approach in Sect. 5.3, we identify the optimal ARMA models for the high and low frequency series of the Dec20 by the AIC criteria, respectively. The high and low frequency series of the training sample of the Dec20 are as follows:

$$\begin{aligned} High_t = & -0.5846High_{t-1} - 0.357High_{t-2} + 0.329High_{t-3} - 0.341High_{t-4} \\ & - 0.131High_{t-5} + 0.065High_{t-6} + \varepsilon_t - 0.260\varepsilon_{t-1} - 0.653\varepsilon_{t-3} + 0.264\varepsilon_{t-4} \end{aligned} \quad (C5)$$

$$\begin{aligned} Low_t = & 4.409Low_{t-1} - 8.604Low_{t-2} + 9.531Low_{t-3} - 6.343Low_{t-4} \\ & + 2.418Low_{t-5} - 0.417Low_{t-6} + \varepsilon_t - 1.1219\varepsilon_{t-1} - 0.253\varepsilon_{t-2} \\ & + 0.116\varepsilon_{t-3} - 0.195\varepsilon_{t-4} \end{aligned} \quad (C6)$$

References

- Arouri, M. E. H., Jawadi, F., & Nguyen, D. K. (2012). Nonlinearities in carbon spot-futures price relationships during Phase II of the EU ETS. *Economic Modelling*, 29(3), 884–892.
- Benz, E., & Trück, S. (2009). Modeling the price dynamics of CO 2 emission allowances. *Energy Economics*, 31(1), 4–15.
- Byun, S. J., & Cho, H. (2013). Forecasting carbon futures volatility using GARCH models with energy volatilities. *Energy Economics*, 40, 207–221.
- Chen, W., Wang, Z., Xie, H., et al. (2007). Characterization of surface EMG signal based on fuzzy entropy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(2), 266–272.
- Chen, W., Zhuang, J., Yu, W., et al. (2009). Measuring complexity using FuzzyEn, ApEn, and SampEn. *Medical Engineering & Physics*, 31(1), 61–68.
- Chevallier, J. (2011). Nonparametric modeling of carbon prices. *Energy Economics*, 33(6), 1267–1282.
- de Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, 40(18), 7596–7606.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Fan, X., Li, S., & Tian, L. (2015). Chaotic characteristic identification for carbon price and a multi-layer perceptron network prediction model. *Expert Systems with Applications*, 42(8), 3945–3952.
- Feng, Z. H., Wei, Y. M., & Wang, K. (2012). Estimating risk for the carbon market via extreme value theory: An empirical analysis of the EU ETS. *Applied Energy*, 99, 97–108.
- Feng, Z. H., Zou, L. L., & Wei, Y. M. (2011). Carbon price volatility: Evidence from EU ETS. *Applied Energy*, 88(3), 590–598.
- Guo, Z., Zhao, W., Lu, H., et al. (2012). Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renewable Energy*, 37(1), 241–249.
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497.
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291.
- Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1), 489–501.
- Huang, N. E., Shen, Z., Long, S. R., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903–995.
- Huang, Y., Dai, X., Wang, Q., & Zhou, D. (2021). A hybrid model for carbon price forecasting using GARCH and long short-term memory network. *Applied Energy*, 285, 116485.
- Jiao, L., Liao, Y., & Zhou, Q. (2018). Predicting carbon market risk using information from macroeconomic fundamentals. *Energy Economics*, 73, 212–227.
- Jiang, L., & Wu, P. (2015). International carbon market price forecasting using an integration model based on SVR. In *International Conference on Engineering Management, Engineering Education and Information Technology*.

- Jiang, M., Jia, L., Chen, Z., & Chen, W. (2020). The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm. *Annals of Operations Research*, 1–33. <https://doi.org/10.1007/s10479-020-03690-w>
- Li, M. J., He, Y. L., & Tao, W. Q. (2017). Modeling a hybrid methodology for evaluating and forecasting regional energy efficiency in China. *Applied Energy*, 185, 1769–1777.
- Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, 37, 152–166.
- Liu, H., Tian, H., & Li, Y. (2012). Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. *Applied Energy*, 98, 415–424.
- Meng, A., Ge, J., Yin, H., et al. (2016). Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. *Energy Conversion and Management*, 114, 75–88.
- Montagnoli, A., & De Vries, F. P. (2010). Carbon trading thickness and market efficiency. *Energy Economics*, 32(6), 1331–1336.
- Ortiz-Cruz, A., Rodriguez, E., Ibarra-Valdez, C., et al. (2012). Efficiency of crude oil markets: Evidences from informational entropy analysis. *Energy Policy*, 41, 365–373.
- Ren, Y., Suganthan, P. N., & Srikanth, N. (2015). A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods. *IEEE Transactions on Sustainable Energy*, 6(1), 236–244.
- Rootzén, J., & Johnsson, F. (2016). Paying the full price of steel—Perspectives on the cost of reducing carbon dioxide emissions from the steel industry. *Energy Policy*, 98, 459–469.
- Sanin, M. E., Violante, F., & Mansanet-Bataller, M. (2015). Understanding volatility dynamics in the EU-ETS market. *Energy Policy*, 82, 321–331.
- Segnon, M., Lux, T., & Gupta, R. (2017). Modeling and forecasting the volatility of carbon dioxide emission allowance prices: A review and comparison of modern volatility models. *Renewable and Sustainable Energy Reviews*, 69, 692–704.
- Song, Y., Liu, T., Liang, D., Li, Y., & Song, X. (2019). A Fuzzy Stochastic Model for Carbon Price Prediction Under the Effect of Demand-related Policy in China's Carbon Market. *Ecological Economics*, 157, 253–265.
- Sun, W., Wang, C. F., & Zhang, C. (2017). Factor analysis and forecasting of CO₂ emissions in Hebei, using extreme learning machine based on particle swarm optimization. *Journal of Cleaner Production*, 162, 1095–1101.
- Taormina, R., & Chau, K. W. (2015). Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. *Journal of Hydrology*, 529, 1617–1632.
- Tan, X., Sirichand, K., Vivian, A., & Wang, X. (2020). How connected is the carbon market to energy and financial markets? A systematic analysis of spillovers and dynamics. *Energy Economics*, 90, 104870.
- Tang, B. J., Gong, P. Q., & Shen, C. (2017). Factors of carbon price volatility in a comparative analysis of the EUA and sCER. *Annals of Operations Research*, 255(1), 157–168.
- Tsai, M. T., & Kuo, Y. T. (2013). A forecasting system of carbon price in the carbon trading markets using artificial neural network. *International Journal of Environmental Science and Development*, 4(2), 163.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer.
- Wang, W., Chau, K., Xu, D., et al. (2015). Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management*, 29(8), 2655–2675.
- Wang, S., Zhang, N., Wu, L., et al. (2016). Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renewable Energy*, 94, 629–636.
- Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1), 1–41.
- Yu, L., Wang, Z., & Tang, L. (2015). A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. *Applied Energy*, 156, 251–267.
- Zhang, N., Lin, A., & Shang, P. (2017). Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. *Physica a: Statistical Mechanics and Its Applications*, 477, 161–173.
- Zhang, X., Karplus, V. J., Qi, T., et al. (2016). Carbon emissions in China: How far can new efforts bend the curve? *Energy Economics*, 54, 388–395.
- Zhang, X., Lai, K. K., & Wang, S. Y. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics*, 30(3), 905–918.
- Zhang, Y. J., & Wei, Y. M. (2010). An overview of current research on EU ETS: Evidence from its operating mechanism and economic effect. *Applied Energy*, 87(6), 1804–1814.
- Zhu, B. (2012). A novel multiscale ensemble carbon price prediction model integrating empirical mode decomposition, genetic algorithm and artificial neural network. *Energies*, 5(2), 355–370.

- Zhu, B., Han, D., Wang, P., & Wu, Z. C. (2017). Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Applied Energy*, *191*, 521–530.
- Zhu, B., & Wei, Y. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega*, *41*(3), 517–524.
- Zhu, B., Ye, S., He, K., Chevallier, J., & Xie, R. (2019). Measuring the risk of European carbon market: An empirical mode decomposition-based value at risk approach. *Annals of Operations Research*, *281*(1), 373–395.
- Zhu, B., Ye, S., Wang, P., He, K., Zhang, T., & Wei, Y. M. (2018). A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. *Energy Economics*, *70*, 143–157.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.