Optimized Extreme Learning Machine (ELM) Based on Genetic Algorithm (GA) To Predict Carbon Price Under The Influence of Multiple Factors

Wei Sun
North China Electric Power University - Baoding Campus

Chumeng Ren (✉ 2251012206@qq.com)
North China Electric Power University - Baoding Campus

Research Article

Keywords: Carbon price forecasting, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, Extreme Learning Machine Optimized by Genetic Algorithm, Multiple factors affect prediction

DOI: https://doi.org/10.21203/rs.3.rs-702953/v1

License: ☺️ ☀️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Optimized Extreme Learning Machine (ELM) based on Genetic Algorithm (GA) to predict carbon price under the influence of multiple factors

Wei Sun\textsuperscript{a}, Chumeng Ren\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a}Department of Economic Management, North China Electric Power University, Baoding 071000, China

\textsuperscript{*}Corresponding author. Department of Economic Management, North China Electric Power University, Baoding 071000, China. Tel.: +8618683551016

E-mail address: 2251012206@qq.com (CM. Ren)

Abstract: The promotion of carbon market can accelerate the pace of low-carbon transformation of China's economic structure and achieve more efficient carbon emission reduction. Accurate carbon price prediction is conducive to improving the risk management of carbon market and the decision-making of investors, but it also brings great challenges to relevant industry practitioners and the government. In this paper, a new hybrid model is proposed, which combines complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and genetic algorithm (GA) optimized extreme learning machine (ELM). The application of GA-ELM in carbon price prediction is firstly studied in this paper. Eight intrinsic mode functions and one residual can be obtained by CEEMDAN, and then partial autocorrelation (PACF) is used to determine the partial correlation between each sequence and its lag data, and they were taken as internal factors affecting
the prediction. At the same time, energy, economic and social factors are selected as the external factors affecting the prediction, and the carbon price prediction is realized through internal and external factors. It has been proved that the model successfully overcomes the challenge of carbon price prediction based on multiple influencing factors. The hybrid model shows superiority in Beijing, Shanghai and Guangdong. The results show that the prediction performance of the proposed model is the best among the 15 models, and the prediction accuracy will be improved due to the decomposition of the carbon price. Besides, the CEEMDAN-GA-ELM model better overcomes the challenge of carbon price prediction with multiple influencing factors. This model provides a novel and effective tool for the government and enterprises to predict the carbon price. **Keywords:** Carbon price forecasting; Complete Ensemble Empirical Mode Decomposition with Adaptive Noise; Extreme Learning Machine Optimized by Genetic Algorithm; Multiple factors affect prediction

1. **Introduction**

The increase in carbon dioxide emissions and the associated environmental pressures have caused worldwide concern. It has caused the emergence of global warming, the greenhouse effect and extreme weather. As global warming accelerates, glaciers are melting faster and sea levels are rising every year, causing serious damage to the environment. Besides, international tensions and local conflicts will also become more frequent. In 1968, J.H.Dales, first proposed the concept of "emission trading" in an article published in Pollution, Wealth and Price, that is, the establishment of a legal right to discharge pollutants and the expression of it in the form of emission permits, so that environmental resources can be bought and sold like commodities. The U.S. Environmental Protection Agency first applied it to air pollution and river pollution management. Since then,
Germany, Australia, the United Kingdom and other countries have also implemented emission trading policy measures. In 1997, more than 100 countries in the world signed the "Kyoto Protocol" because of global warming. The treaty stipulates the emission reduction obligations of developed countries and puts forward three flexible emission reduction mechanisms at the same time. Carbon emission trading is one of them. On January 1, 2005, the European Union established the Emissions Trading Scheme (ETS) (Zhang et al., 2018). It is the world's first multinational emissions trading system, based on cap-and-trade, which provides a way to reduce emissions at the lowest economic cost. Carbon emissions can be effectively reduced by buying and selling carbon permits (Zhang et al., 2019). China is the world's second-largest emitter of greenhouse gases and is seen by many countries as the most promising market for reducing emissions, even though it is not constrained to do so. In October 2011, the National Development and Reform Commission of China issued the Notice on the Pilot Work of Carbon Emission Trading and approved seven provinces and cities to carry out carbon trading pilot work, including Beijing, Shanghai, Tianjin, Chongqing, Hubei, Guangdong and Shenzhen. The first carbon emission trading market in China was launched in Shenzhen on June 18, 2013, marking the beginning of carbon emission trading in China. Carbon price forecasts play an important role in trading. The concept that global warming emissions can be reduced by setting a reasonable price for carbon emissions is generally accepted by people (Zhang and Wei, 2010), and accurate carbon price prediction is beneficial to investors' decision-making. Therefore, the research on improving the prediction accuracy is of great significance to China's carbon market at the present stage.

Direct forecasts based on historical price data and forecasts based on impact factor analysis are the two main forecasting ideas. At present, scholars prefer to use the former in the research of
carbon price prediction in China.

(Zhu et al., 2017), (Ji et al., 2019), (Sun and Duan, 2019) and (Zhu et al., 2019) used historical price data for direct prediction in their studies. First, the carbon price was decompressed and the final prediction results were determined by predicting various internal model functions and residuals. The biggest difference in studies that use historical price forecasts is in the combination of decomposition and forecasting methods, (Zhang et al., 2018) combined the complete set empirical pattern decomposition (CEEMD) with the optimized grey neural network (GNN) to get a new hybrid model, (Sun and Li, 2020) used the same decomposition method, but the difference was that she chose the Long and Short Term Memory Network (LSTM) as the prediction method. Also (Wu and Liu, 2020) used EEMD and BA-LSSVM for combination, similar to the research of (Sun and Xu, 2021), but she used wavelet least squares support vector machine (WLSSVM) to build a new model. Besides, to improve the prediction accuracy and adaptability of the model, scholars conducted further research on the internal model function obtained by decomposition. (Liu and Shen, 2020) used fuzzy C-means clustering algorithm to divide these sub-components into trend, low frequency and high-frequency components for prediction to improve the prediction accuracy. (Wang et al., 2021) and (S. Yang et al., 2020) both use random forest (RF) to determine the prediction input. (Sun and Huang, 2020) used variational mode decomposition (VMD) to decompose the first inherent mode function (IMF1), and improved the prediction effect through secondary decomposition. In (Zhang et al., 2019)’ research, the carbon price is decomposed into approximate sequence and detailed sequence after the high-frequency components of previous carbon price data is eliminated by the multi-resolution singular value decomposition method, and the two-time series were used as the prediction input.
The experimental results of these studies all prove that the decomposition of a carbon price and the choice of prediction input can improve the prediction performance, and the prediction effect of the combined model is better than that of the single prediction model.

However, Changes in energy, economy, and weather conditions all cause fluctuations in carbon prices (Rickels et al., 2007). (Hao et al., 2020) mentioned in the study that most previous studies did not focus on the role of selecting appropriate input features, but only aimed at improving the accuracy of prediction. Blindly improving the decomposition and prediction method, neglecting the importance of various influencing factors, feature selection, carbon price and the pretreatment of exogenous variables in improving the prediction effectiveness, maybe the reason for the poor prediction effect and poor adaptability of the model.

(Sun et al., 2020) considers both external and internal influencing factors when selecting factors, and the experiment finds that considering external and internal influencing factors can significantly improve the performance of carbon price prediction, with a wider range of forecasting potential. (Yahşi et al., 2019) also found that Brent crude oil futures, coal, electricity and natural gas prices all had an impact on carbon price forecasts. (Hao and Tian, 2020) also emphasized that carbon price is more sensitive to coal than other factors. Considering and analyzing the influence of exogenous variables makes it a challenging and critical problem to study the influence of multiple factors on carbon price prediction (Hao and Tian, 2020). In the face of these challenges, the research on carbon price prediction considering various influencing factors is extremely insufficient and urgent. Providing more valuable information to the carbon market is of vital importance, which can provide support for the improvement of carbon price analysis and prediction performance.
Different influencing factors may lead to different prediction effects, so it is very important to select the appropriate influencing factors and deal with them, which also increases the difficulty of research (Zhu et al., 2017). In the prediction research based on impact factor analysis, the grey correlation method can be used to screen out the factors with high correlation with the explained variables, while factor analysis is often used to deal with the situation of high data repeatability, which can well reduce the data dimension. (Zhu et al., 2021) analyzed the influence of two concrete specimens on concrete slump and compressive strength through grey correlation. The selection of manufacturing enterprises also involves the use of grey correlation (Tian et al., 2018). In the study of (Sun et al., 2017), factor analysis was used for dimensionality reduction and 8 special factors were effectively extracted from 22 influencing factors. (Huang and He, 2020) adopted grey correlation method and factor analysis method as the tools to determine the input of the prediction. The factor analysis method is also used (Sun and Wang, 2020), which takes the extracted special factors as input variables and uses the least square support vector machine improved particle swarm optimization to make predictions. From these studies, it is found that factor analysis and grey correlation can well screen out the target factors, and reduce the redundancy of data, to reduce the difficulty of prediction and improve the accuracy of prediction. This is also an important reason why they are used in this paper to select the prediction input.

CEEMDAN is an advanced data denoising method (Lu et al., 2020). When separating Electroencephalogram (EGG) data, (Wu et al., 2021) found that CEEMDAN solved the modal aliasing problem while retaining most of the original EEG signal components, and the results showed that the separation effect of this method on EEG artefacts was better than previous studies. (Li et al., 2020) also proves that this method can improve the prediction effect of the model when
conducting short-term load forecasting. This method can well solve the problem that conventional EMD will be affected by mode aliasing (Cao et al., 2019). CEEMDAN has been proved to be an important improvement on EEMD, achieving negligible reconstruction errors and solving the problem of different forms of signal-plus noise implementation (Torres et al., 2011), and based on EEMD, the number of iterations is further reduced, the frequency aliasing region is compressed, the convergence performance is improved, and the non-stationary signal has a higher resolution ability for different frequency components (Wang et al., 2012). Based on the above reasons, the carbon price was decomposed by CEEMDAN in this paper.

Extreme Learning Machine algorithm (ELM) is a commonly used tool for predicting carbon price, which can be proved in studies (Taormina and Chau, 2015), (Yadav et al., 2016), (Adnan et al., 2019) and (Xu et al., 2020). However, the randomness of the input weight matrix initialized by ELM and hidden layer bias will reduce the accuracy of the model (Zheng et al., 2020). It is a good choice to use the global searching ability of the genetic algorithm to optimize the parameters in the model. Compared with the Online Sequential Extreme Learning Machine (OS-ELM), GA-ELM effectively improves the accuracy, detection rate and accuracy of intrusion detection, and reduces the false alarm rate and missing alarm rate (Zhang et al., 2020). The study of (Yi et al., 2017) compared the prediction results of the GA-ELM model with the GA-BP neural network model and the original ELM model, and finally verified that the prediction results of the GA-ELM model were the best. Although GA-ELM has a good performance in forecasting, no one has applied it to carbon price prediction for the time being. In this paper, the GA-ELM model will face the challenge of multiple factors for carbon price prediction for the first time, and we will explore whether the GA-ELM model has an excellent performance in predicting nonlinear and unstable
carbon price.

Given the shortcomings of existing researches, this paper chooses three aspects of energy, economy and society to study the effect of the CEEMDAN-GA-ELM model in carbon price prediction under the effect of these factors. The abbreviations used in this paper are explained in Table 1.

The innovation and contribution of this paper are mainly reflected in the following aspects:

(1) The carbon price prediction model based on influence factor analysis and CEEMDAN-GA-ELM model is proposed for the first time. The ELM optimized by GA can improve the stability of the prediction effect and the accuracy of the prediction. The model has been tested in Beijing, Guangdong and Shanghai and proved to have the best prediction effect. The application of the CEEMDAN-GA-ELM model in carbon price prediction can overcome the difficulties caused by many factors, and it has good applicability and can be extended to other pilot projects.

(2) The decomposition of carbon price series by CEEMDAN can improve the accuracy of prediction. In the Beijing pilot, the decomposition effect of CEEDAN is better than that of EEMD. When combined with different prediction methods, the prediction effect can be improved very well, among which, when combined with GA-ELM, the best prediction effect can be achieved. This study enriches the practice of the CEEMDAN algorithm and can provide a valuable reference for the research of decompose-prediction.

(3) Combine the internal and external influencing factors to realize the prediction of the carbon price. PACF was used in this paper to determine the partial correlation between each sequence decomposed by CEEMDAN and its lag data, which was used as an internal factor affecting the prediction. The energy factor composed of coal price, oil price and natural gas price; The
economic factor composed of exchange rate; The social factor composed of Baidu index are selected as the external factors affecting the prediction, and the internal and external factors are both taken as the input of the prediction. The good prediction effect reflects the success of introducing external factors into the carbon price prediction and has the potential for a wider range of prediction. The research results enrich the prediction theory based on influencing factors in the field of carbon price prediction, make up the gap of existing research, and provide a reference for further improving the accuracy and stability of carbon price prediction and expanding the application range of the model.

The rest structure of this paper is as follows: The theories and methods used in this paper are introduced in Section 2. The construction process of the hybrid model is explained in section 3. The empirical research and the analysis of the results are realized in the fourth section. The research conclusion of this paper is given in section 5.

2. Methodology

2.1 CEEMDAN

Empirical Mode Decomposition (EMD) is a new adaptive time-frequency signal processing method, which is often used in the analysis and processing of nonlinear non-stationary signals (Huang et al., 1998). Signal $x(t)$ needs to meet two conditions: 1) The number of extreme points equals or differs by no more than one from the number of zero crossings; 2) The mean of the upper and lower envelopes is 0 everywhere. For actual non-stationary time-varying signals, conventional EMD will be affected by mode aliasing. By introducing white noise with normal distribution, the noise auxiliary signal processing method is formed, which is called EEMD. In this paper, the CEEMDAN algorithm, based on the improved EEMD method, is adopted to obtain
the set of noisy signals by adding multiple groups of independent identically distributed adaptive
white noise with finite variance constraint to the original signal. The improved method can further
reduce the number of iterations, compress the frequency aliasing region, improve the convergence
performance, and have higher resolution ability for different frequency components of non-
stationary signals.

The implementation steps of CEEMDAN for signal $x(n)$ are as follows.

1. Generate signal sets containing noise

   $x_i(n) = x(n) + \omega_i(n)$ \hspace{1cm} (2.1.1)

   Where $\omega_i(n)$ ($i = 1, 2, 3, ..., I$) is the noise satisfying the Gaussian distribution, $V$ is the
   number of set samples;

2. The first-order IMF ($F_{1i}$) of each sample was obtained by EMD on $x_i(n)$, and then its mean
   value was calculated as the first-order IMF of $x(n)$.

   $F_1(n) = \frac{1}{I} \sum_{i=1}^{I} F_{1i}$ \hspace{1cm} (2.1.2)

3. Calculate first-order residuals

   $r_1(n) = x(n) - F_1(n)$ \hspace{1cm} (2.1.3)

4. Calculate the second-order IMF

   $F_2(n) = \frac{1}{I} \sum_{i=1}^{I} E_1[r_1(n) + \varepsilon_1 E_1[\omega_i(n)]]$ \hspace{1cm} (2.1.4)

   Where, $E_j(\cdot)$ represents the $j$-order IMF of the signal; $\varepsilon_j$ is the parameter to control the
   white noise energy.

5. Calculate $k$ order residuals

   $r_k(n) = r_{k-1}(n) - F_k(n)$ \hspace{1cm} (2.1.5)
(6) Calculate the $k + 1$ component

$$\tilde{F}_{k+1}(n) = \frac{1}{l} \sum_{i=1}^{l} E_{1}\{r_{k}(n) + \varepsilon_{k}E_{k}[\omega_{i}(n)]\}$$

(2.1.6)

(7) Repeat steps (5) and (6) until the residuals can no longer be decomposed or the highest IMF order is reached

The final residual is as follows:

$$R(n) = x(n) - \sum_{k=1}^{K} \tilde{F}_{k}(n)$$

(2.1.7)

The signal can be expressed as:

$$x(n) = \sum_{k=1}^{K} \tilde{F}_{k}(n) + R(n)$$

(2.1.8)

The CEEMDAN has different representations of the original signal frequency components of each order IMF component. The energy difference among the components of traditional EEMD is large, and the frequency mixing region is large. However, the energy balance among IMF components obtained by the CEEMDAN method has a narrow frequency aliasing region, and different frequency components have higher resolution ability for non-stationary signals.

2.2 ELM

ELM (Extreme Learning Machine) is an improved feed-forward neural network Learning algorithm proposed by (Huang et al., 2004) which has fast practice speed, global unique optimal solution and good generalization ability.

There are $Q$ learning samples, $\{(x_{i}, y_{i})\}$ $Q_{l} = 1$, and $x_{i} \in R^{T}$, $y_{i} \in R^{\psi}$. Assuming that the number of mythic elements in the hidden layer is $M$, Assume that the excitation function of neurons in the hidden layer is Sigmoid, and its expression is as follows:
\[ \sum_{j=1}^{M} a_j \phi(\omega_j x_i + b_j) = f_M(x) \]  

(2.2.1)

In the formula, \( \omega_j = [\omega_{j1}, \omega_{j2}, \ldots, \omega_{jT}] \) is the correlation weight between the input layer and the hidden layer; \( a_j = [a_{j1}, a_{j2}, \ldots, a_{j\phi}]^T \) is the correlation weight of hidden layer and output layer; \( \omega_j \) is the bias of the jth node in the neuron nodes of the hidden layer.

According to the zero-error approximation principle, there are \( b_j, a_j \) and \( \omega_j \), so that the standardized form is simplified as \( Ha = Y \)

\[
H(\omega_1, \ldots, \omega_M, b_1, \ldots, b_M, x_1, \ldots, x_Q) = \begin{bmatrix}
\phi(\omega_1 x_1 + b_1) & \ldots & \phi(\omega_M x_1 + b_M) \\
\vdots & \ddots & \vdots \\
\phi(\omega_1 x_Q + b_1) & \ldots & \phi(\omega_M x_Q + b_M)
\end{bmatrix}_{(Q \times M)}
\]  

(2.2.2)

\[
a = \begin{bmatrix}
a_1^T \\
a_2^T \\
\vdots \\
a_M^T
\end{bmatrix}_{(M \times \psi)} \quad Y = \begin{bmatrix}
y_1^T \\
y_2^T \\
\vdots \\
y_M^T
\end{bmatrix}_{(M \times \psi)}
\]

Where \( H \) is the output matrix. Once the input weight \( \omega_j \) and the bias \( b_j \) of the hidden layer are randomly determined, the output matrix \( H \) of the hidden layer is uniquely determined.

\[ Ha = Y \]  

(2.2.3)

\[ \hat{a} = H^+ Y \]  

(2.2.4)

**2.3 GA-ELM**

In 1967, Bagley, a student of Professor J. Holland, put forward the term genetic algorithm for the first time in his doctoral thesis. The genetic optimization algorithm is a random search method that evolved according to the evolutionary laws of the biological world. It has great advantages in implicit parallelism, practicability and global optimization ability.

Since the hidden layer input weights and deviations in the ELM model are randomly given, there may be some random set values of 0, leading to the failure of some hidden layer nodes. Therefore,
this paper adopts the genetic algorithm to optimize the input weights and deviations, and obtains
the optimal initial weights and thresholds through the selection, crossover and mutation operations
of the genetic algorithm, and then obtains the optimal ELM model.

The training steps are as follows:

(1) The fitness function, population number \( k \) and evolutionary times \( p \) were set. In this paper, the
mean square deviation of the sample data of the test set was selected as the fitness function.
The smaller the fitness function value, the more accurate the model.

(2) Set the population.
The population number \( K \) is generally set at 20–40. The weight matrix and deviation vector
of the hidden layer can be used to determine the individual length, that is, \( T = l(n + 1) \), where \( l \)
is the number of nodes in the hidden layer and \( M \) is the number of neurons in the input layer,
namely the dimension of the input vector.

\[ Q = [a_{\gamma 1}, a_{\gamma 12}, \ldots, a_{\gamma 1L}, a_{\gamma 2}, a_{\gamma 22}, \ldots, a_{\gamma 2L}, \ldots, a_{\gamma L}, a_{\gamma L1}, a_{\gamma L2}, \ldots, a_{\gamma LL}, b_{\gamma 1}, b_{\gamma 2}, \ldots, b_{\gamma L}] \] (2.3.1)

Where \( Q \) is the \( \gamma \)th individual in the population

(3) The optimal fitness function \( abest \) is locally solved

The initial value of \( abest \) is 10, and the initial value of \( \theta \) and \( \gamma \) is 0. The training data are
normalized into the model, and the fitness function of each individual is obtained successively
until the cycle terminates when \( \gamma = k \), and the value of \( abest \) is solved.

(4) The optimal fitness function \( abest \) is solved globally.

After each solution to the optimal fitness function, adopting the crossover and mutation to
population optimization, examine times \( \theta \) of evolution when \( \theta \) less than or equal to \( p \), Start the
value of \( \gamma \) as 0 and return to step (3) until \( \theta \) is greater than \( p \) and stop the operation, and \( abest \)
is the optimal fitness function, and the optimized ELM model can be obtained according to the corresponding parameters.

### 3. Process of Model construction

The research ideas of this paper and the construction process of the CEEMDAN-GA-ELM model are shown in the Figure 1. The prediction steps are carried out as follows:

**Step 1:** Determination of correlation, represented by an orange wireframe in the graphic summary. GRA is used to analyze the correlation between energy, economic and social factors and carbon price.

**Step 2:** Data preprocessing, represented by purple wireframe in the graphic abstract. Stock trading time is from Monday to Friday, and the market is closed on weekends and national holidays. Therefore, the average of adjacent values is used to make up the missing data of carbon price, coal price, oil price and natural gas price, to obtain complete and uninterrupted data that does not change the original trend, which is regarded as one of the external influencing factors.

**Step 3:** Factor analysis, shown in yellow in the graphic summary. Too many input ends of prediction will affect the prediction accuracy. Factor analysis is used to reduce the dimension of social factors to reduce the redundancy and repeatability of information. Use the new factor to replace the original factor as one of the external influence factors.

**Step 4:** Decomposition of the carbon price, which is represented in blue in the graphic abstract. To reduce the difficulty of prediction, CEEMDAN is used to decompose the carbon price, and finally, 8 internal module functions and residuals can be obtained.

**Step 5:** Select the predicted input and indicate it with a black wire box in the graphic summary.
PACF was used to determine the relationship between each internal model function and historical data, and data highly correlated with the analysis sequence was selected as internal input for impact prediction.

Step 6: Forecast, shown in green in the graphic summary. GA-ELM was used to predict the internal model functions respectively. The final estimate of the carbon price is obtained by adding up all the predictions.

4. Empirical analysis

4.1 Data

4.1.1 Carbon price data

So far, China has set up carbon trading markets in Shenzhen, Guangdong, Hubei, Tianjin, Shanghai, Chongqing, Beijing and Fujian. Of China's eight major carbon markets, in 2019, more than 7 million tons of carbon market quotas were traded in Beijing, accounting for approximately 8% of the national market. The price of carbon allowances in Beijing is the highest overall, maintaining a high level throughout the year, with the highest being more than 87 yuan and the lowest being 48.4 yuan. It is the carbon trading market with the highest carbon emission price and the most active. In this paper, Beijing carbon price is selected as a specific case for empirical analysis, which has the value of in-depth research. The data interval is from January 9, 2019, to December 31, 2019. Figure 2 shows the comparison before and after the carbon price data supplement. Highly nonlinear and fluctuating are features of the carbon price series, which can be found in Figure 2. Therefore, we take Shanghai and Guangdong as supplementary cases to fully prove the superiority and applicability of the model. The relevant data comes from the China Carbon Trading website.
4.1.2 External factors

(1) Energy

The price of domestic carbon trading is determined by the demand for fossil fuels (Adekoya, 2021). Therefore, thermal coal futures prices, WTI crude oil prices and natural gas futures prices are selected as energy factors that affect carbon prices. Thermal coal futures prices come from Zhengzhou Commodity Exchange, WTI crude oil prices and natural gas futures prices come from Yingwei Financial Information Portal.

(2) Economy

Exchange rates have an impact on carbon price fluctuations (X. Yang et al., 2020). The fluctuation of the exchange rate represents the appreciation and depreciation of RMB, which changes the purchasing power of consumers, and the change of consumer demand for enterprise products affects the economic benefits of enterprises. Secondly, the fluctuation of the RMB exchange rate will directly change the energy price through import and export channels, which will increase or decrease the production cost of enterprises. Finally, to deal with the fluctuation of the RMB exchange rate, the government adopts different monetary policies, and the government macro-control changes the behaviour of enterprises. In the carbon market, through the influence of these three levels, enterprises choose to sell or buy carbon emissions for the maximization of their interests, which affects the change of carbon price from the two aspects of supply and demand. The data comes from Fred Economic Data.

(3) Society

The increasing importance of Internet search big data in the eyes of investors has become an
important indicator to reflect the degree of social attention to the market. Therefore, according to the Baidu demand graph and the popularity of related words, the five Baidu search index data were selected from the top 10 keywords: carbon neutrality, carbon trading, climate change, carbon footprint, carbon sink, as the social factors affecting the carbon price.

Table 2 is all the data sources used in this article, Table 3 shows the carbon price data description of the three-carbon trading pilots.

4.2 Data preprocessing and inputs selection

4.2.1 Grey correlation analysis

Grey correlation analysis (GRA) proposed by (Julong, 1989) is an important part of the theory of the grey system. It can reflect the degree of connection between things well. The more similar the shape of the two sequence curves is, the greater the correlation degree is; otherwise, the smaller the correlation degree is (Liu and Lin, 2010). Through grey correlation analysis, we use it as a tool to choose factors that are highly correlated with carbon prices. Table 4 shows the grey correlation analysis results.

According to the grey correlation analysis results, the selected 9 related factors have a very high correlation with the carbon price, and it is reasonable to take them as the influencing factors of the carbon price prediction. Among them, the exchange rate has the highest correlation, followed by the thermal coal futures price. The fluctuation of the RMB exchange rate directly changes the energy price through the import and export channel. Among them, "carbon neutral" has the lowest correlation, and its correlation also reaches 0.7938. It is believed that the selected factors can be used in the study of carbon price analysis.

4.2.2 Factor analysis
By extracting some common factors to replace the original indicators, factor analysis can reduce the redundancy of influencing factors and achieve the purpose of dimension reduction. The extraction of common factors can effectively reduce the prediction error (Mingxing et al., 2009). The characteristics of high information dimension, high redundancy and high repeatability will be brought about by the selected nine factors in this paper. Therefore, we use factor analysis to reduce the dimension of relevant data. This paper will conduct a factor analysis of the five Baidu search indexes. Firstly, KMO and Bartlett tests will be used to determine whether the conditions for factor analysis are met. KMO can be used to reflect the relationship between the factors. The closer the KMO value is to 1, the closer the relationship between the factors is. In general, when KMO is greater than 0.7, it indicates that the original variable is suitable for factor analysis. Whether the variables are independent or not can be reflected by the Bartlett test. When the significance level is less than 0.01, it means that the zero hypotheses of the Bartlett test is rejected and factor analysis can be carried out. Table 5 shows the results of the KMO and Bartlett tests. Table 6 shows the results of the factor analysis.

Factor analysis by SPSS shows that KMO=0.774 and significance level is less than 0.01, indicating that the five Baidu indexes are not completely independent and suitable for dimension reduction in factor analysis. The variance value of the common factor reflects that the common degree obtained from the two extracted factors is greater than 80%, indicating that the extracted factors have high representativeness to the original variable. The two extracted factors accounted for 85.804% of the five indices. It shows that $F_1$ and $F_2$ are suitable as common factors of the Baidu index.

According to the component scoring coefficient, the expressions of $F_1$ and $F_2$ can be
obtained:

\[ F_1 = -0.266x_1 + 0.320x_2 + 0.475x_3 + 0.469x_4 - 0.053x_5 \]  \hfill (4.2.1)

\[ F_2 = 0.642x_1 + 0.166x_2 - 0.189x_3 - 0.183x_4 + 0.459x_5 \]  \hfill (4.2.2)

The resulting \( F_1 \) and \( F_2 \) will be used as one of the prediction inputs.

**4.2.3 Carbon price decomposition**

The unprocessed data presents obvious volatility, and the large volatility of the data will affect the prediction effect, so the original sequence is decomposed by CEEMAND, and eight intrinsic mode functions and one residual are obtained. Compared with the original signal, the structure of the decomposed internal mode function is simpler and smoother. PACF is then used to determine the number of lag periods for each modal function, and the data corresponding to the number of lag periods that are highly correlated with historical data is selected as one of the inputs to the prediction model. The decomposition results of Beijing carbon price are shown in Figure 3. Table 7 shows the number of lag periods for each internal module function and residuals.

**4.3 Model accuracy assessment**

\( R^2 \), \( RMSE \), and \( MAPE \) are widely used as indicators to evaluate the effect of predictive models. \( R^2 \) is between 0 and 1. The larger the ratio, the more accurate the model and the better the prediction effect. \( RMSE \) stands for root mean square error, and \( MAPE \) represents the average absolute percentage error, and their range is \([0, +\infty]\). The larger the value, the larger the error. When the predicted value is completely consistent with the true value, it is equal to 0. Their calculation formula is as follows:

\[ R^2 = 1 - \frac{\sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{N} (y_t - \bar{y}_t)^2} \]  \hfill (4.3.1)
\[
RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2} 
\]

(4.3.2)

\[
MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\%
\]

(4.3.3)

4.4 Results and discussion

In this paper, CEEMDAN-GA-ELM is used to forecast the carbon price of Beijing from January 9, 2019, to December 31, 2019. To prove the superiority of the model used, several other models are selected for comparison, which includes no carbon price decomposition and carbon price decomposition using EEMD, and five prediction methods (BP, PSO-BP, LSSVM, ELM and GA-ELM). Table 8 and Figure 4 show the prediction results of all models in Beijing pilot project, and Table 9 shows the forecast results of Guangdong and Shanghai.

Through the prediction results, we can find that:

1. From any evaluation criteria, the prediction effect of the CEEMDAN-GA-ELM model used in this paper is the best ($R^2=0.9898$, $RMSE=1.1294$, $MAPE=0.0120$).

2. Optimization of BP and ELM can improve prediction accuracy. GA-ELM is better than ELM in $R^2$, $RMSE$ and $MAPE$. The prediction effect of EEMD-GA-ELM is better than that of EEMD-ELM, and the CEEMDAN-GA-ELM performs better than the CEEMDAN-ELM. The same result is also found for PSO-BP and BP. The optimized model has a higher prediction accuracy than the original one, regardless of whether it is decomposed or not.

3. The prediction accuracy can be improved by the decomposition of the original sequence, and the ability to improve the prediction accuracy of CEEMDAN is stronger than that of EEMD. By comparing the prediction effects of models 1, 7, 12; 2, 9, 14;
3, 8, 13; 4, 6, 11, and 5, 10, 15 can find that models that break down the carbon price get better predictions. It is also found that when the prediction methods are BP, LSSVM, PSO-BP, ELM and GA-ELM, the prediction effect obtained by using CEEMDAN to decomposition the original sequence is better than that by using EEMD.

(4) By comparing the observation model prediction of 6 ~ 15, can be found after the decomposition process, all models can achieve very high prediction accuracy, this shows that this paper choose the internal and external factors as input can get high precision of prediction output, factors influencing the choice is reasonable and effective.

(5) LSSVM has better prediction performance for decomposed sequences. When the carbon price was not decomposed, the prediction performance of LSSVM \( (R^2=0.7580, \text{RMSE}=4.3848, \text{MAPE}=0.0442) \) was only superior to that of BP with the worst effect. After EEMD or CEEMDAN decomposition, the prediction performance was significantly improved, which was second only to GA-ELM.

(6) GA-ELM performs better than the other four prediction methods. By comparing models 1 to 5, it is found that GA-ELM has the best prediction performance among the five prediction methods without any decomposition method. Compared with models 6-10, it was found that when the original sequence was decomposed by EEMD, the prediction performance of EEMD-GA-ELM was the best. Comparing models 11 to 15, it was found that when the original sequence was decomposed by CEEMDAN, the prediction performance of CEEMDAN-GA-ELM was the best.

According to Table 9, supplementary experiments in Shanghai and Guangdong can once
again prove the superiority of the CEEMDAN-GA-ELM model in predicting carbon price. Other forecasting models have a lot of variation in their performance in different places, For example, EEMD-PSO-BP in Shanghai has a high degree of fit ($R^2=0.7796$), but in Shanghai, the fit is not ideal ($R^2=0.7796$). This also shows that the CEEMDAN-GA-ELM model not only has higher prediction accuracy but also a more stable prediction effect.

5. conclusion

A new combined forecasting method realizes the prediction of non-linear and unstable carbon prices in this paper. Since researchers mostly use time series to predict the price of carbon and ignore the impact of other factors on the carbon price, causing the problem of low prediction accuracy and low applicability, this paper studies the internal and external influences of the carbon price. Firstly, five Baidu indexes were selected as the social factors affecting the prediction of the carbon price. Coal prices, oil prices and natural gas futures prices are energy factors; the exchange rate is an economic factor, these three factors together constitute the external factors affecting the carbon price. After analyzing the correlation between them and the carbon price using the grey correlation analysis method, the social factors are analyzed by factor analysis to reduce the duplication and redundancy of data and reduce the dimensionality of the input terminal. Two new factors are extracted as new social factors, and then together with energy factors and economic factors are used as one of the prediction inputs.

After testing the prediction performance of the other 14 comparison models, It turns out that the prediction model proposed in this paper has the best performance in carbon price prediction. This model can overcome the difficulties caused by multiple factors in carbon emission prediction, and greatly improve the accuracy of carbon price prediction. It is also found that even in different
carbon trading markets, the model can achieve very good prediction results, the prediction performance is stable, and it has the potential for popularization and application. Secondly, even under the influence of multiple factors, the decomposition of a carbon price can still effectively improve the prediction accuracy, and the prediction effect of LSSVM will be greatly improved due to the decomposition processing. Compared with EEMD, the internal model function and residuals of CEEMDAN decomposition are more stable, which is more conducive to the prediction of the carbon price. Besides, it can be found that after decomposition, all the models can achieve very high prediction accuracy, which indicates that the internal and external influencing factors selected in this paper can be used as the prediction input to obtain high-precision prediction output. Among the 5 prediction methods, GA-ELM has high prediction accuracy and stable effect. GA-ELM can provide high prediction accuracy no matter whether carbon price is decomposed or EEMD or CEEMDAN is decomposed.

Accurate prediction of a carbon price is helpful to avoid the risk of the carbon market and can provide a reference for the government to make policy. The prediction of a carbon price based on the CEEMDAN-GA-ELM proposed in this paper under the influence of multiple factors provides a new idea for the prediction of carbon market transaction prices and also provides a reference for subsequent research. The impact of qualitative factors such as national policies and the international situation on carbon trading price prediction will be discussed in future research.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**
Availability of data and materials

The datasets generated during and/or analysed during the current study are available in:

The carbon price: China Emissions Trading Website (http://www.tanpaifang.com/)

The price of coal: China zhengzhou Commodity Exchange (http://www.czce.com.cn/)

WTI crude oil prices: Yingwei Caiqing Financial Information Portal (https://cn.investing.com/)

Futures price of natural gas: Yingwei Caiqing Financial Information Portal (https://cn.investing.com/)

The exchange rate: Fred Economic Data (https://fred.stlouisfed.org/)

Carbon neutral: Baidu index (http://index.baidu.com/v2/index.html#/)

Carbon trading: Baidu index (http://index.baidu.com/v2/index.html#/)

Climate change: Baidu index (http://index.baidu.com/v2/index.html#/)

Carbon footprint: Baidu index (http://index.baidu.com/v2/index.html#/)

Carbon sequestration: Baidu index (http://index.baidu.com/v2/index.html#/)

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

Sun Wei: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision. Ren Chumeng: Data curation, Software, Validation, Visualization, Writing- original
Acknowledgements

Thanks are due to Sun Wei for valuable discussion.

References

Adekoya, O.B., 2021. Predicting carbon allowance prices with energy prices: A new approach. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2020.124519

Adnan, R.M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B., Kisi, O., 2019. Daily streamflow prediction using optimally pruned extreme learning machine. J. Hydrol. 577, 123981. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.123981

Cao, J., Li, Z., Li, J., 2019. Financial time series forecasting model based on CEEMDAN and LSTM. Phys. A Stat. Mech. its Appl. 519, 127-139. https://doi.org/https://doi.org/10.1016/j.physa.2018.11.061

Hao, Y., Tian, C., 2020. A hybrid framework for carbon trading price forecasting: The role of multiple influence factor. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2020.120378

Hao, Y., Tian, C., Wu, C., 2020. Modelling of carbon price in two real carbon trading markets. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2019.118556

Huang, G. Bin, Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: A new learning scheme of feedforward neural networks, in: IEEE International Conference on Neural Networks - Conference Proceedings. https://doi.org/10.1109/IJCNN.2004.1380068
Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Snin, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hubei spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. A Math. Phys. Eng. Sci. 454, 903–995. https://doi.org/10.1098/rspa.1998.0193

Huang, Y., He, Z., 2020. Carbon price forecasting with optimization prediction method based on unstructured combination. Sci. Total Environ. 725. https://doi.org/10.1016/j.scitotenv.2020.138350

Ji, L., Zou, Y., He, K., Zhu, B., 2019. Carbon futures price forecasting based with ARIMA–CNN–LSTM model, in: Procedia Computer Science. https://doi.org/10.1016/j.procs.2019.11.254

Julong, D., 1989. Introduction to grey system. J. Grey Syst.

Li, T., Qian, Z., He, T., 2020. Short-Term Load Forecasting with Improved CEEMDAN and GWO-Based Multiple Kernel ELM. Complexity 2020. https://doi.org/10.1155/2020/1209547

Liu, H., Shen, L., 2020. Forecasting carbon price using empirical wavelet transform and gated recurrent unit neural network. CARBON Manag. 11, 25–37. https://doi.org/10.1080/17583004.2019.1686930

Liu, S., Lin, Y., 2010. Introduction to grey systems theory. Underst. Complex Syst. https://doi.org/10.1007/978-3-642-16158-2_1

Lu, H., Ma, X., Huang, K., Azimi, M., 2020. Carbon trading volume and price forecasting in China using multiple machine learning models. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2019.119386
Mingxing, C., Dadao, L., Hua, Z., 2009. Comprehensive evaluation and the driving factors of China’s urbanization. Dili Xuebao/Acta Geogr. Sin.

Rickels, W., Duscha, V., Keller, A., Peterson, S., 2007. The Determinants of Allowance Prices in the European Emissions Trading Scheme - Can We Expect an Efficient Allowance Market.

Sun, W., Duan, M., 2019. Analysis and forecasting of the carbon price in China’s regional carbon markets based on fast ensemble empirical mode decomposition, phase space reconstruction, and an improved extreme learning machine. Energies. https://doi.org/10.3390/en12020277

Sun, W., Huang, C., 2020. A novel carbon price prediction model combines the secondary decomposition algorithm and the long short-term memory network. Energy. https://doi.org/10.1016/j.energy.2020.118294

Sun, W., Li, Z., 2020. An ensemble-driven long short-term memory model based on mode decomposition for carbon price forecasting of all eight carbon trading pilots in China. Energy Sci. Eng. 8, 4094–4115. https://doi.org/10.1002/ese3.799

Sun, W., Sun, C., Li, Z., 2020. A hybrid carbon price forecasting model with external and internal influencing factors considered comprehensively: A case study from China. Polish J. Environ. Stud. https://doi.org/10.15244/pjoes/114261

Sun, W., Wang, C., Zhang, C., 2017. Factor analysis and forecasting of CO2 emissions in Hebei, using extreme learning machine based on particle swarm optimization. J. Clean. Prod. 162, 1095–1101. https://doi.org/10.1016/j.jclepro.2017.06.016
Sun, W., Wang, Y.W., 2020. Factor analysis and carbon price prediction based on empirical mode decomposition and least squares support vector machine optimized by improved particle swarm optimization. Carbon Manag. https://doi.org/10.1080/17583004.2020.1755597

Sun, W., Xu, C., 2021. Carbon price prediction based on modified wavelet least square support vector machine. Sci. Total Environ. 754, 142052. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.142052

Taormina, R., Chau, K., 2015. Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. J. Hydrol. 529, 1617–1632. https://doi.org/10.1016/j.jhydrol.2015.08.022

Tian, G., Zhang, H., Feng, Y., Wang, D., Peng, Y., Jia, H., 2018. Green decoration materials selection under interior environment characteristics: A grey-correlation based hybrid MCDM method. Renew. Sustain. Energy Rev. 81, 682–692. https://doi.org/https://doi.org/10.1016/j.rser.2017.08.050

Torres, M.E., Colominas, M.A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings. pp. 4144–4147. https://doi.org/10.1109/ICASSP.2011.5947265

Wang, J.-J., Zhang, W.-Y., Liu, X., Wang, C.-Y., 2012. Modifying Wind Speed Data Observed from Manual Observation System to Automatic Observation System Using
Wavelet Neural Network. Phys. Procedia 25, 1980–1987.

https://doi.org/10.1016/j.phpro.2012.03.338

Wang, J., Sun, X., Cheng, Q., Cui, Q., 2021. An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. Sci. Total Environ.

https://doi.org/10.1016/j.scitotenv.2020.143099

Wu, Q., Liu, Z., 2020. Forecasting the carbon price sequence in the Hubei emissions exchange using a hybrid model based on ensemble empirical mode decomposition. Energy Sci. Eng. https://doi.org/10.1002/ese3.703

Wu, Q., Zhang, WenQiang, Wang, Y., Zhang, Wenxi, Liu, X., 2021. Research on removal algorithm of EOG artifacts in single-channel EEG signals based on CEEMDN–BD. Comput. Methods Biomech. Biomed. Engin. 1–12.

https://doi.org/10.1080/10255842.2021.1889525

Xu, H., Wang, M., Jiang, S., Yang, W., 2020. Carbon price forecasting with complex network and extreme learning machine. Phys. A Stat. Mech. its Appl.

https://doi.org/10.1016/j.physa.2019.122830

Yadav, B., Ch, S., Mathur, S., Adamowski, J., 2016. Discharge forecasting using an Online Sequential Extreme Learning Machine (OS-ELM) model: A case study in Neckar River, Germany. Meas. J. Int. Meas. Confed. 92, 433–445.

https://doi.org/10.1016/j.measurement.2016.06.042

Yahşi, M., Çanakoğlu, E., Ağralı, S., 2019. Carbon price forecasting models based on big data analytics. Carbon Manag. https://doi.org/10.1080/17583004.2019.1568138
Yang, S., Chen, D., Li, S., Wang, W., 2020. Carbon price forecasting based on modified ensemble empirical mode decomposition and long short-term memory optimized by improved whale optimization algorithm. Sci. Total Environ. 716. https://doi.org/10.1016/j.scitotenv.2020.137117

Yang, X., Zhang, C., Yang, Y., Wu, Y., Yun, P., Wagan, Z.A., 2020. China’s carbon pricing based on heterogeneous tail distribution. Sustain. https://doi.org/10.3390/su12072754

Yi, M., Quanlong, S., Lihua, Y., Chuanrong, W., Huaqiang, X., 2017. Grain Size Prediction of Aluminum Alloy Dies Castings Based on GA-ELM. ACTA Metall. Sin. 53, 1125–1132. https://doi.org/10.11900/0412.1961.2016.00573

Zhang, J., Li, D., Hao, Y., Tan, Z., 2018. A hybrid model using signal processing technology, econometric models and neural network for carbon spot price forecasting. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2018.09.071

Zhang, K., Hu, Z., Zhan, Y., Wang, X., Guo, K., 2020. A smart grid AMI intrusion detection strategy based on extreme learning machine. Energies 13, 1–19. https://doi.org/10.3390/en13184907

Zhang, X., Zhang, C., Wei, Z., 2019. Carbon Price Forecasting Based on Multi-Resolution Singular Value Decomposition and Extreme Learning Machine Optimized by the Moth-Flame Optimization Algorithm Considering Energy and Economic Factors. ENERGIES 12. https://doi.org/10.3390/en12224283
Zhang, Y.J., Wei, Y.M., 2010. An overview of current research on EU ETS: Evidence from its operating mechanism and economic effect. Appl. Energy. https://doi.org/10.1016/j.apenergy.2009.12.019

Zheng, G., Hua, W., Qiu, Z., Gong, Z., 2020. Detecting Water Depth from Remotely Sensed Imagery Based on ELM and GA-ELM. J. Indian Soc. Remote Sens. https://doi.org/10.1007/s12524-020-01270-w

Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T., Wei, Y.M., 2017. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. Appl. Energy. https://doi.org/10.1016/j.apenergy.2017.01.076

Zhu, J., Wu, P., Chen, H., Liu, J., Zhou, L., 2019. Carbon price forecasting with variational mode decomposition and optimal combined model. Phys. A Stat. Mech. its Appl. https://doi.org/10.1016/j.physa.2018.12.017

Zhu, L., Zhao, C., Dai, J., 2021. Prediction of compressive strength of recycled aggregate concrete based on gray correlation analysis. Constr. Build. Mater. 273, 121750. https://doi.org/10.1016/j.conbuildmat.2020.121750

Figure Legends

Figure 1 Flowchart

Figure 2 Comparison before and after the carbon price data supplement

Figure 3 The decomposition results of Beijing carbon price

Figure 4 Prediction results of all models in Beijing
Figures

Figure 1
Flowchart
Figure 2

Comparison before and after the carbon price data supplement
Figure 3

The decomposition results of Beijing carbon price
Figure 4

Prediction results of all models in Beijing