



# Predicting China's thermal coal price: Does multivariate decomposition-integrated forecasting model with window rolling work?\*

Qihui Shao<sup>a,b</sup>, Yongqiang Du<sup>a,\*</sup>, Wenxuan Xue<sup>a</sup>, Zhiyuan Yang<sup>a,c</sup>, Zhenxin Jia<sup>a</sup>, Xianzhu Shao<sup>a</sup>, Xue Xu<sup>a</sup>, Hongbo Duan<sup>d</sup>, Zhipeng Zhu<sup>a</sup>

<sup>a</sup> School of Science, Tianjin University of Commerce, Tianjin, 300134, China

<sup>b</sup> School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, 102206, China

<sup>c</sup> Institute for Financial Studies, Shandong University, Jinan, 250100, China

<sup>d</sup> School of Economics and Management, University of Chinese Academy of Sciences, Beijing, 100190, China

## ARTICLE INFO

### Keywords:

China coal  
Price forecasting  
Rolling ICEEMDAN-Methods  
Influencing factors

## ABSTRACT

Coal, as the primary energy source in China, significantly affects the country's energy security and national economic stability. However, the highly nonlinear and non-stationary nature of coal prices poses challenges for accurate forecasting. In this study, we propose the Rolling ICEEMDAN-Methods series model based on the "divide and conquer" approach to predict the Bohai-Rim Steam-Coal Price Index (BSP), involving the integration of multiple methods, including ANN, CNN, LSTM, GRU, LightGBM, and ERT. Unlike conventional univariate forecasting, we comprehensively summarise the factors influencing coal prices into eight categories, totalling 27 variables, with the aim of capturing more meaningful information. By employing the window-rolling decomposition-ensemble forecasting method, we effectively avoided information leakage and boundary effects, leading to a significant improvement in prediction accuracy. Experimental results demonstrate that the proposed Rolling ICEEMDAN-Methods outperforms other Rolling Methods in terms of accuracy and stability. Novel variables, such as attention, and the other seven categories of influencing factors contribute to enhanced prediction accuracy, among which past coal prices exhibit higher importance in determining forecast results. The findings offer valuable guidance to coal enterprises in making production decisions and provide a basis for the government to formulate macroeconomic energy policies.

## 1. Introduction

On September 22, 2020 General Secretary Xi Jinping announced at the 75th United Nations General Assembly that China aimed to peak its carbon dioxide emissions before 2030 and achieve carbon neutrality by 2060 (Ding and Wen, 2022). The "dual-carbon" strategy advocates for a green, environmentally friendly, and low-carbon lifestyle, showcasing China's strong determination to address climate change and promote the development of all humankind. It is noteworthy that China is currently the world's largest emitter of carbon dioxide (Lv et al., 2020). In 2020, China's carbon dioxide emissions reached 9.8935 billion tons, accounting for 30.93% of the global total, the highest proportion globally. Coal is the primary source of carbon dioxide emissions, accounting for

nearly 80% (Zhang and Chen, 2022). Taking 2018 as an example, the majority of China's energy-related carbon dioxide emissions came from fossil fuels, with coal accounting for 79.89%, oil for 14.32%, and natural gas and other sources accounting for only 5.45% and 0.35%, respectively. Although the proportion of new energy has increased annually in recent years, the intermittent problem and high user-side costs make it difficult for coal to significantly change its position in China's energy consumption and supply patterns in the short term (Tong et al., 2022). Coal will continue to play an indispensable role in China's future economic development. The change of this role will also inevitably affect the realisation of the national "dual carbon" strategic goal.

Coal and electricity are foundational industries for the development of the national economy and are closely interconnected upstream and

\* This work was supported by the National Natural Science Foundation of China (Nos.72022019, 72243011), and the National Key Research and Development Program of China (2020YFA0608603).

\* Corresponding author. School of Science, Tianjin University of Commerce, Tianjin, 300134, China.

E-mail addresses: [shaqihui2805@163.com](mailto:shaqihui2805@163.com) (Q. Shao), [duyongqiang198410@163.com](mailto:duyongqiang198410@163.com) (Y. Du), [xwx\\_0105@163.com](mailto:xwx_0105@163.com) (W. Xue), [hayangzy@gmail.com](mailto:hayangzy@gmail.com) (Z. Yang), [1447536783@qq.com](mailto:1447536783@qq.com) (Z. Jia), [charon0328@163.com](mailto:charon0328@163.com) (X. Shao), [xuxue@tjcu.edu.cn](mailto:xuxue@tjcu.edu.cn) (X. Xu), [hbduan@ucas.ac.cn](mailto:hbduan@ucas.ac.cn) (H. Duan), [15237518575@163.com](mailto:15237518575@163.com) (Z. Zhu).

downstream. The harmonious stability of these relationships is essential for sustainable economic growth (Corey et al., 2013). In recent years, with the increase in coal prices, downstream industries in the consumption sector have experienced an increase in production costs, indirectly triggering cost-driven inflation, which hampers the sustainable development of the economy (Zhangb et al., 2022). According to the Qinhuangdao Thermal Coal Comprehensive Trading Price, the price of 5500 kcal of coal reached 1079 yuan/ton in September 2021, which was 1.9 times higher than the price during the same period last year (563 yuan/ton in September 2020). The continuous increase in domestic coal prices has caused a sharp increase in generation costs for thermal power enterprises. The contradiction between "market coal and planned electricity" has become prominent, and thermal power enterprises are facing the dilemma of "generating electricity equals loss". This has resulted in a nationwide electricity shortage. In contrast to the sluggish supply side, overseas manufacturing has been significantly limited since the onset of the COVID-19 pandemic, while China's exports have continued to improve. Rapid growth in industrial production has contributed to a rapid increase in electricity consumption and intensified the imbalance between electricity supply and demand. From January to August 2021, the total electricity consumption in the whole society reached 4285.1 billion kilowatt-hours, representing an annual increase of 13.7%. "Power rationing" has become the last resort to fill this gap and ensure the safety of the power system. Since the end of September 2021, the decoupling of coal prices from electricity prices has led to power rationing in multiple provinces in China, which not only affects industrial production, but also impacts the daily lives of residents, triggering sensitive nerves throughout society.

Therefore, the assessment and prediction of coal price trends are crucial. Such predictions can not only assist businesses in estimating costs and stabilising their production decisions but also provide a scientific basis for formulating energy and economic policies in the coal-based market (Zhangb et al., 2022).

Owing to the nonlinear and non-stationary characteristics of coal prices, accurate prediction has become a challenge. Single models often fail to satisfy high-precision requirements. To address the limitations of the individual models, various hybrid models have been developed based on data pattern decomposition and ensemble mechanisms. Although the accuracy has improved, some issues persist, such as limited information, information leakage, and boundary effects. In this study, we introduce 27 influencing factors, including Google Trends, and construct a Rolling ICEEMDAN-Methods series model to provide a more accurate forecast of China's dynamic coal prices.

The contributions of this study are as follows.

- (1) Compared with existing research, the Rolling-Decomposition approach avoids the drawbacks of information leakage and boundary effects. This study utilises the Rolling ICEEMDAN-Methods series models to effectively capture the nonlinear and non-stationary characteristics of Chinese dynamic coal prices. This improves the prediction accuracy compared to Rolling Methods series models.
- (2) This study considers multiple factors influencing coal prices, addressing the limitations of the existing research, which relies solely on a single time series of coal prices for prediction. Multivariate prediction often yields higher accuracy than univariate prediction and ensures the stability of the relationship between the response and predictor variables over time.
- (3) To the best of our knowledge, existing literature does not consider the impact of attention factors on coal prices. This study, for the first time, incorporated attention factors into the modelling process. The empirical results demonstrated that the introduction of attention factors improved the prediction accuracy of the model.
- (4) This study ranks the importance of the eight proposed factors and provides policy recommendations that can guide the relevant departments in policy formulation.

The remainder of this study is organised as follows. Section 2 provides a synthesis of the existing literature. Section 3 provides a brief introduction to the methods and overall framework used in this study. Section 4 presents a detailed analysis of the results, including comparative analysis and model validation. Finally, Section 5 summarizes the findings and provides policy recommendations.

## 2. Literature review

The prediction of coal prices can be regarded as a challenging task in time-series modelling (Zhangb et al., 2022). Coal price series often exhibit characteristics such as nonlinearity, nonstationarity, high noise, and susceptibility to external environmental factors (Ding et al., 2017). This poses a challenge in establishing accurate coal price prediction models (Zhangb et al., 2022).

Coal price prediction methods can be broadly divided into two categories: statistical and machine-learning methods. Representative models in the former category include Vector Autoregression (VAR), Autoregressive Integrated Moving Average (ARIMA), and Generalised Autoregressive Conditional Heteroskedasticity (GARCH). However, owing to the strict application conditions and limited applicability of these methods, they fail to capture the nonlinear characteristics of the coal price series, resulting in lower predictive accuracy of the models (Zhangb et al., 2022).

In recent years, with the vigorous development of computer science and big data technology, the application of machine learning methods to coal price prediction has attracted attention. For example, Ding et al. (2021) developed a C-MIDAS-X model to capture the nonlinear relationship between thermal coal prices and used the XGBoost model to correct prediction errors of the C-MIDAS model. XGBoost performed well in capturing the nonlinearity of coal prices, improving the prediction accuracy of the C-MIDAS model by approximately 22%. Matyjaszek et al. (2019) examined the performance of traditional time-series models, generalised regression neural networks, and multilayer feed-forward networks in predicting coking coal prices and found that the generalised regression neural network model provided more accurate predictions. Ming et al. (2016) used a generalised regression neural network to select the key factors influencing coal prices and determined the core parameters of a least-squares support vector machine for fast and reasonable coal price prediction using the stochastic Nelder-Mead optimisation-based Artificial Bee Colony Algorithm. Herrera et al. (2019) compared the long-term predictive performance of traditional econometric models and machine learning methods (Neural Networks and Random Forest) for major global energy commodities. The results show that machine learning methods outperform traditional econometric methods, providing further evidence for the use of machine learning algorithms for more accurate predictions.

Models built using machine learning and deep learning methods have been widely applied, particularly to improve prediction accuracy (Zhou et al., 2022). This can be attributed to the ability of these methods to effectively distinguish random factors and capture hidden nonlinear features (Herrera et al., 2019). The Light Gradient Boosting Machine (LightGBM) is a typical representative in this regard. Some researchers believe that in most cases, LightGBM outperforms other machine-learning algorithms in terms of prediction accuracy (Wang et al., 2023). Additionally, the computational speed of the LightGBM model was ten times faster than that of the original Gradient Boosting Decision Tree (GBDT) method, with a memory requirement of only one-third that of the GBDT method (Ren et al., 2022). In the context of coal price prediction, no studies utilizing this method have been found. Moreover, Extremely Randomized Trees (ERT) is a new method based on decision trees, which is a variant of random forests. Their excellent prediction accuracy and computational efficiency have made them popular and widely used in medical diagnosis, biochemistry, and other fields (Sharma et al., 2022). Similar to the LightGBM, ERT has not been extensively applied in coal price prediction research. Considering their

excellent performance, this study incorporated the two aforementioned methods for modelling and analysis.

Existing research indicates that neural network methods can achieve excellent prediction accuracy, which may be attributed to their self-learning ability (Xu and Zhang, 2022). Ahmed and Chen (2023) consider Artificial Neural Network (ANN) as the most effective tool for revealing the correlations between system inputs and outputs. The approximation capability of ANN makes it highly suitable for modelling problems. Herrera et al. (2019) utilized ANN to study energy prices, such as coal, and achieved high prediction accuracy. Lu et al. (2020) suggested that Convolutional Neural Networks (CNN) can effectively uncover data features and exhibit high accuracy in financial time forecasting.

Although these models have strong nonlinear processing capabilities, they cannot capture temporal dynamics within time-series data (Zhang et al., 2022).

To address this issue, a class of intelligent algorithms specifically designed for time-series modelling has emerged Long Short-Term Memory (LSTM) neural network. LSTM networks exhibit the ability to remember long- and short-term information and effectively overcome the problems of vanishing and explosion gradients commonly encountered in traditional recurrent neural network models owing to their unique architectural design. They are particularly well-suited for handling time-series data (Donahue et al., 2015). The LSTM has been widely researched and applied in various fields (Zhou et al., 2022). These studies demonstrate that LSTM models exhibit favourable predictive accuracy.

In 2014, Cho et al. proposed a structurally simpler variant of the LSTM called the Gated Recurrent Unit (GRU). The gate mechanism in GRU can scale the model, facilitate data convergence, and reduce the risk of overfitting using limited data (Faraji et al., 2022). Wu et al. (2023) studied the Bohai-Rim Steam-Coal Price index using the GRU model, and the results showed high accuracy.

Although the aforementioned models can capture temporal dependencies in time-series data, a single model is still insufficient to handle the complex nonstationary characteristics of coal price sequences (Zhang et al., 2022). To alleviate the limitations of single models, researchers have developed various hybrid models based on data pattern decomposition and ensemble mechanisms (Wu et al., 2019; Guo et al., 2022). These studies are beneficial for improving the forecasting capabilities of coal prices.

The main steps of the decomposition and ensemble models are as follows: Based on the "divide and conquer" approach, the original sequence is first decomposed into several sub-sequences. The purpose is to decompose the original non-stationary time series into multiple relatively stationary components that can better reflect the local characteristics of the original sequence. Second, each subsequence is predicted individually. Finally, the predictions from each subsequence were reconstructed to obtain the final results.

Among many decomposition techniques, selecting the appropriate decomposition technology is crucial for constructing the hybrid model. In recent years, timing decomposition technology has undergone rapid development, involving the classical addition model, multiplication model, AR, MA and other models. Actually, the more prevalent decomposition techniques encompass Seasonal Extraction in ARIMA Time Series (SEATS), Seasonal Trend Decomposition using LOESS (STL), Wavelet Transform (WT), Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), Complementary Ensemble Empirical Mode Decomposition (CEEMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), and Variational Mode Decomposition (VMD).

Within these decomposition techniques, SEATS can accurately extract seasonal components from the time series. However, its application is restricted to specific types of time series data, namely, quarterly

data and monthly data, rendering it potentially unsuitable for other types of data. STL decomposition constitutes one of the most widely adopted methods within the field of time series decomposition. Yuan et al. (2024) established the STL-ANDGM-GPR model to forecast CO<sub>2</sub> emissions with obvious trends and seasonal characteristics. The results demonstrated that the model had good prediction accuracy. However, this method is confined to providing only additive decomposition and might not be optimal for certain specific time series data. Similar to STL, MA also decomposes time series into trend, seasonal, and residual components, and the application of MA is generally confined to the linear and stationary time series. Despite the WT has become a decomposition method integrating various machine learning and deep learning models, for example, Fantini et al. (2024) used WT, RNN, GRU and other methods to predict wind speed in the short term, it suffers from the "non-adaptiveness" drawback and requires the original time series to be stationary, relying heavily on the choice of the mother wavelet.

VMD is capable of handling nonlinear and non-stationary signals and demonstrates a high level of accuracy. Zeng et al. (2024) constructed a WOA-VMD-BiLSTM hybrid model aimed at predicting PM2.5 concentration. The findings indicated that VMD significantly enhanced the prediction accuracy. It is worth noting that, the results of VMD cannot be fully reconstructed (Zuo et al., 2020). EMD and its extended versions have powerful capabilities in handling nonlinear and non-stationary time series, and the decomposition results can be reconstructed well. Nonetheless, EMD is plagued by a significant mode aliasing issue (Wu and Huang, 2004). Unfortunately, EEMD has not adequately resolved the mode aliasing problem, prompting researchers to develop CEEMD as a more effective solution. It has gained popularity in the field of price series forecasting. For instance, Huang et al. (2024) utilized CEEMD and GRU to forecast the price of power transformer materials, while Aggarwal et al. (2020) employed CEEMD and SVM to predict the price of bitcoin. To enhance the accuracy and stability of signal decomposition, an adaptive control strategy is implemented during the addition of white noise, allowing for better control over the noise's magnitude and distribution, thereby leading to the development of CEEMDAN. The distinction between ICEEMDAN and CEEMDAN lies in that ICEEMDAN refrains from directly adding Gaussian white noise during the decomposition process, instead selecting the Kth intrinsic mode function (*IMF*(t)) component resulting from the decomposition of white noise by EMD, thereby achieving superior noise control. This technique is currently widely adopted. For example, Xiao et al. (2021) employed the ICEEMDAN-ICA method to predict air pollutant concentrations, and their results demonstrated that the mixed model incorporating ICE-EMDAN exhibited the highest prediction performance.

However, three crucial issues still exist in decomposition and ensemble models based on the EMD family. First, there are problems with obtaining results with high prediction accuracy. Many studies (e.g., Wu et al., 2019) perform the decomposition on the original dataset in which some information belonging to the test dataset is leaked in advance (Zuo et al., 2020; Teng et al., 2022). Consequently, decomposition and ensemble models built in this manner are considered unreliable (Zhang et al., 2015). Second, if the original time series is decomposed only on the training set, it leads to severe boundary effects in the modelling results (Zuo et al., 2020; Xiao et al., 2021), resulting in a significant degradation in the model's generalisation ability (Xiao et al., 2021; Zhang et al., 2015). To address these issues, Xiao et al. (2021) developed an adaptive decomposition and ensemble scheme to adapt to boundary effects by continuously adjusting the time series to be decomposed. Third, it is essential to consider the factors influencing coal prices rather than relying solely on a single time series of coal prices for prediction. Most existing studies utilize only a single time series of coal prices and employ statistical or machine-learning methods for price forecasting (Herrera et al., 2019). Only a few studies have considered the fluctuations in coal prices influenced by fundamental factors (such as coal consumption, supply and demand, and inventory) and non-fundamental factors (such as economic policies and other

unpredictable factors) (Ding et al., 2021). Other energy prices have been extensively studied, providing some references for our research on coal prices. For example, Guo et al. (2022) studied crude oil prices and considered the impact of Google Trends, Baidu Index, and textual data on crude oil prices. They found that Google Trends and the keyword-based Baidu Index serve as the best proxy variables for market Internet attention. Therefore, this study incorporates alternative energy prices, supply and demand factors, transportation schedules, commodity currency factors, macroeconomic factors, international coal price levels, historical coal prices, and Google Trends as factors influencing coal prices. This makes forecasting coal prices challenging.

Based on an extensive review of existing research, despite the effectiveness of decomposition-ensemble prediction models in addressing the prediction challenges caused by high nonlinearity and non-stationarity in other time-series data, there is still limited research on coal price forecasting, particularly in the literature related to Chinese coal prices (Ding et al., 2021).

### 3. Methodology

This section briefly describes the models used in this study, including ANN, CNN, LSTM, GRU, LightGBM, ERT, and ICEEMDAN. The predictive process is also presented, summarising the specific research steps and framework diagram of this study.

#### 3.1. Artificial Neural Network

ANN is a parallel processing network used to establish complex nonlinear relationships between independent and dependent variables. It comprises an input layer, hidden layers, and an output layer. After determining the input layer, weights were assigned to each variable, which helped determine the contribution of each variable to the output. The input values are multiplied by their corresponding weights, summed, and processed using an activation function that determines the output. If the output exceeds a given threshold, the node is triggered and the data are passed to the next layer. This continuous propagation results in the output of one node becoming the input of the next. Owing to their high-precision performance, ANN has become one of the most commonly used nonlinear analysis models (Hema et al., 2023).

#### 3.2. Convolutional neural Network

CNN is a common type of feedforward neural network that processes input data through multiple layers of convolutional operations to achieve efficient feature extraction and classification. The one-dimensional Convolutional Neural Network (1D-CNN) is a subclass of CNN in which the convolutional kernel runs in only one direction. 1D-CNN is commonly used for processing time-series data and possesses strong feature extraction capabilities (Guo et al., 2023).

#### 3.3. Long Short-Term Memory

LSTM is an improved model based on Recurrent Neural Networks (RNN). Its recurrent structure allows hidden layer nodes to learn from the current input and retain information from previous nodes. Unlike RNN, this model maintains long-term dependencies in sequence through forget, input, and output gates. Additionally, LSTM can address the issues of vanishing and exploding gradients in RNN. As a result, LSTM has the advantage of learning long-term correlations and time-series modelling (Yuan and Jiao, 2023).

#### 3.4. Gate recurrent unit

GRU is a variant of LSTM. It retains the predictive power of LSTM while effectively addressing the issues of gradient explosion and vanishing in RNN, while simplifying the structure (Zhao et al., 2023). In

GRU, the update gate controls the number of historical states retained in the current output state and the number of candidate states retained at the current time step. The reset gate determines whether the candidate state at the current time step depends on the network state of the previous time step, and to what extent. Compared to LSTM, the GRU model has a simpler structure and reduces many matrix multiplications by eliminating one gating unit, making it more efficient.

#### 3.5. Light gradient Boosting machine

LightGBM is an open-source framework developed by Microsoft that implements the Gradient Boosting Decision Tree (GBDT) algorithm, which supports efficient parallel training. The selection of LightGBM in this study was primarily driven by the requirements for accuracy and scalability in dynamic coal price prediction. Its high accuracy is attributed to the definition of the loss function, which expands the Taylor objective function to the second order, retaining more information about the objective function, and thereby improving accuracy. Scalability was achieved using three main techniques: Gradient-based One-Side Sampling (GOSS), histogram-based algorithms, and Exclusive Feature Bundling (EFB). Using GOSS allows the exclusion of most small-gradient data, utilizing only the remaining data to estimate the information gain, significantly reducing computational and time costs. The histogram-based algorithm reduces the time complexity and memory consumption. The EFB algorithm binds mutually exclusive features to new features to reduce feature dimensionality.

#### 3.6. Extremely Randomized Trees

ERT, also known as Extra Trees, are an ensemble method built on tree models. It differs from Random Forest (RF) in two ways. First, while RF selects the best splitting nodes based on criteria such as information gain or the Gini coefficient, ERT selects splitting nodes completely at random to reduce the influence of variance. Second, ERT uses the entire training dataset to build trees, thereby preventing bias. After generating a predefined number of trees, their individual predictions were aggregated, and the final prediction result was obtained by taking the average. Extensive experiments have demonstrated that the ERT algorithm outperforms other tree-based ensemble methods, including Random Forest.

#### 3.7. Improved CEEMD with Adaptive Noise

This study used ICEEMDAN to decompose a coal price series with nonlinear and non-stationary features. ICEEMDAN, developed by Colominas et al. (2014), is a member of the EMD family of methods that can decompose non-stationary and nonlinear signals into a finite number of  $IMF(t)$  and residues. Each  $IMF(t)$  contains the local characteristic signals of the original signal at different time scales. Compared to other methods in the EMD family, ICEEMDAN can further reduce noise residue and mode mixing in  $IMF(t)$ , resulting in better decomposition performance (Du et al., 2020; Xiao et al., 2021).

Assuming  $y(t)$  is the original sequence,  $M(\cdot)$  is the operator for calculating the local mean of the original sequence, and  $E_k(\cdot)$  represents the operator for the  $k$  th EMD operation ( $k = 1, 2, \dots, K$ ), the specific steps of ICEEMDAN are as follows.

##### Step 1. Constructing the noisy signal:

$$y^i(t) = y(t) + \beta_0 E_1(\eta^i(t)), \quad (1)$$

where  $\beta_0 = \varphi_0 \text{std}(y(t)) / \text{std}(E_k(\eta^i(t)))$  is used for noise removal,  $\varphi_0$  represents the reciprocal of the desired signal-to-noise ratio between the initially noisy signal and the analysed signal,  $\text{std}$  denotes the standard deviation,  $\eta^i(t)$  represents the Gaussian white noise at time  $t$  for the  $i$  th iteration of noise addition, and  $i$  is the number of noise additions,  $i = 1, 2, \dots, I$ . Consequently, the expression for the first residue  $r_1(t)$  is

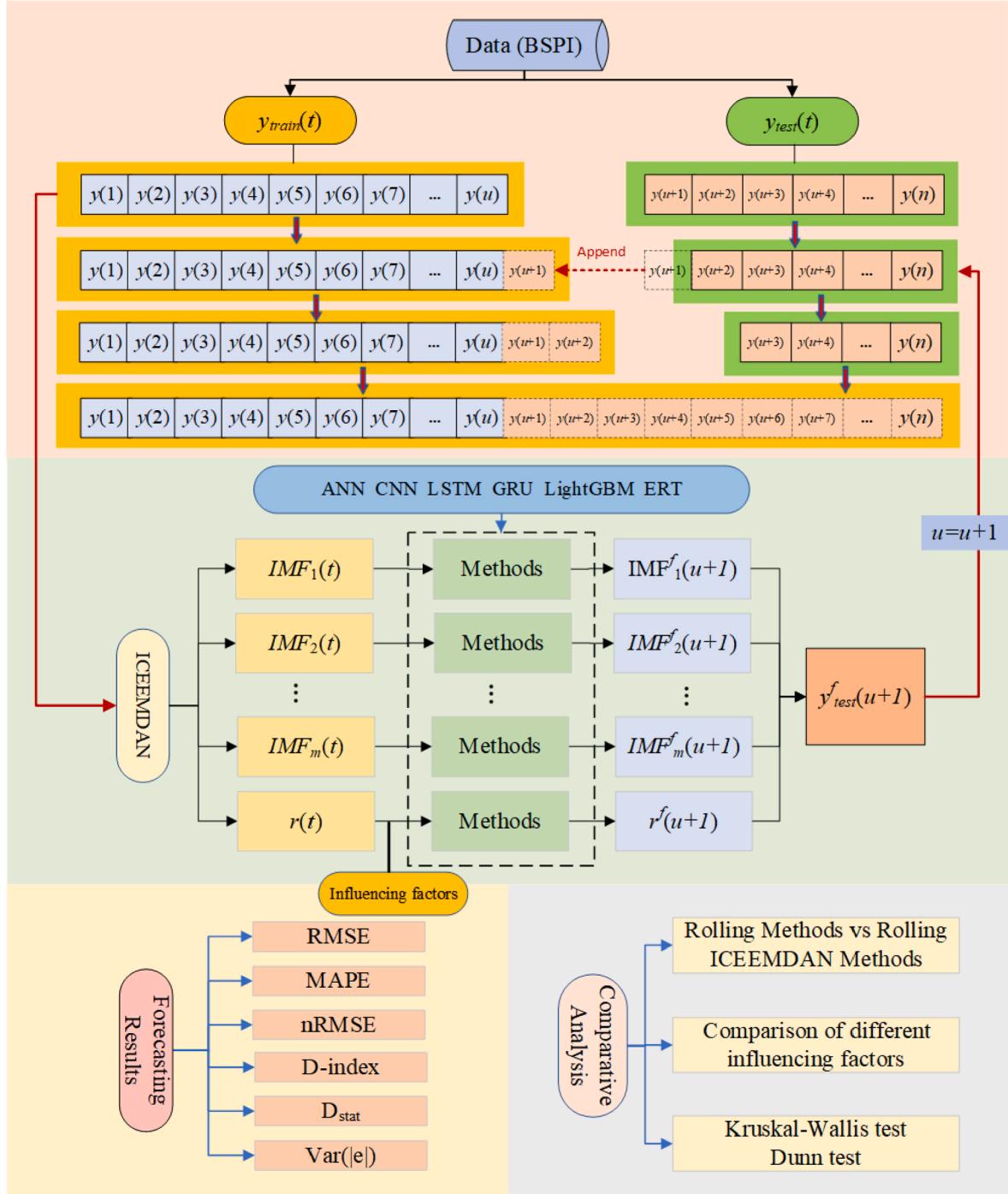


Fig. 1. Framework of model.

derived as follows:

$$r_1(t) = \frac{1}{I} \sum_{i=1}^I M(y^i(t)), \quad (2)$$

**Step 2.** Computing the first intrinsic mode function  $IMF_1(t)$ :

$$IMF_1(t) = y(t) - r_1(t), \quad (3)$$

**Step 3.** Utilizing the average of the local means of  $r_1(t) + \beta_1 E_2(\eta^i(t))$ , we obtain the second residue  $r_2(t)$  and the second intrinsic mode function  $IMF_2(t)$ , expressed as Equations (4) and (5), respectively:

$$r_2(t) = \frac{1}{I} \sum_{i=1}^I M(r_1(t) + \beta_1 E_2(\eta^i(t))), \quad (4)$$

$$IMF_2(t) = r_1(t) - r_2(t), \quad (5)$$

**Step 4.** For  $k = 3, 4, \dots, K$ , the  $k$  th residue  $r_k(t)$  and  $k$  th intrinsic mode function  $IMF_k(t)$  are calculated.

$$r_k(t) = \frac{1}{I} \sum_{i=1}^I M(r_{k-1}(t) + \beta_{k-1} E_k(\eta^i(t))), \quad (6)$$

$$IMF_k(t) = r_{k-1}(t) - r_k(t), \quad (7)$$

**Step 5.** Repeating Step 4, we obtain all the intrinsic mode functions and residues.

The original intention of using ICEEMDAN for time series decomposition is based on the "divide and conquer" approach, where non-stationary and nonlinear data are decomposed into several stationary sub-sequences. This allows for a better reflection of the local characteristics of the original sequence, which is beneficial for improving the accuracy of price prediction (Xiao et al., 2021). The average price index of Bohai Sea thermal coal in this study is a non-stationary and nonlinear time series. Therefore, applying ICEEMDAN to decompose the coal price series can effectively reduce the complexity of the sequence and lay the foundation for subsequent predictions.

### 3.8. The proposed framework

This section provides an overview of the prediction process, which consists of seven steps, as shown in Fig. 1. A detailed description of each step is provided below.

**Step 1.** Data collection. Weekly data of the Chinese Bohai Sea thermal coal price index from October 14, 2013, to March 5, 2023,  $y(t)$  was collected as the response variable. A total of 27 variables, including coal mining industry index, coal shipping price index, and indicators reflecting coal price attention, were selected as predictor variables.

**Step 2.** Data partitioning. The entire time-series  $y(t)$  was divided into a training set  $y_{train}(t)(t=1, 2, \dots, u)$  and a test set  $y_{test}(t)(t=u+1, u+2, \dots, n)$  in chronological order. Specifically, 343 weekly data points from October 14, 2013 to May 10, 2020 were allocated to the training set, whereas 147 weekly data points from May 11, 2020 to March 5, 2023 were allocated to the test set.

**Step 3.** Data decomposition. To capture the intrinsic features of the Bohai Sea thermal coal price index, the training set data  $y_{train}(t)$  was decomposed into  $m$  intrinsic mode functions  $IMFs(t)$  and a residue  $r(t)$  using ICEEMDAN, namely  $IMF_1(t), IMF_2(t), \dots, IMF_m(t), r(t)$ .

**Step 4.** Prediction of submodes and residues. Each submode and residue obtained from the decomposition in Step 3 was modelled using ANN, CNN, LSTM, GRU, LightGBM, and ERT methods. Using the trained models and 27 predictor variables (such as coal mining industry index and coal shipping price index) at time point  $u$ , the predicted values of each submode and residue at time point  $u+1$ , denoted as  $IMF_1(u+1), IMF_2(u+1), \dots, IMF_m(u+1)$  and residue  $r^f(u+1)$ , were obtained.

**Step 5.** Reconstruction of prediction results. The predicted results of the  $m+1$  submodes and residue were summed to obtain the predicted value of the Bohai Rim steam-coal price index at time point  $u+1$ , which is denoted as  $y_{test}^f(u+1)$ .

**Step 6.** Rolling prediction. The true value  $y_{test}(u+1)$  was combined with the previous step's training set  $y_{train}(t)(t=1, 2, \dots, u)$  to form a new training set  $y_{train}(t)(t=1, 2, \dots, u+1)$ . Steps 3 to 5 were repeated on the new training set to obtain the predicted value of the whole test set  $y_{test}^f(u+2), y_{test}^f(u+3), \dots, y_{test}^f(T)$ .

**Step 7.** Comparative analysis. To validate the effectiveness of the proposed "decompose-predict-ensemble" modelling approach, a comparison was conducted between Rolling ICEEMDAN-Methods and Rolling Methods. Comparative experiments were performed by excluding a specific influencing factor, and the importance of each factor was ranked.

### 3.9. Model evaluation metrics

To verify the accuracy and stability of the models, six metrics were employed as evaluation criteria: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), normalised Root Mean Square Error ( $nRMSE$ ), Agreement Index ( $D$ -index), Directional Statistics ( $D_{stat}$ ), and Variance of Absolute Error (Var(| $e$ |)).

The Root Mean Square Error (RMSE) is a measure of the goodness of fit and is defined by Equation (8).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}, \quad (8)$$

where  $\hat{y}_i$  is the  $i$  th prediction value,  $y_i$  is the  $i$  th true value,  $n$  represents the length of the test set, and  $i$  is the index of the sample in the test set.

Mean Absolute Percentage Error (MAPE) is frequently used in finance and economics because returns and losses are often measured in relative terms, and customers are sometimes more sensitive to relative changes than absolute changes (De Myttenaere et al., 2016). This is defined as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\%, \quad (9)$$

the symbols have the same meaning as in Equation (8).

Normalised Root Mean Square Error ( $nRMSE$ ) is a measure of the relative difference between predicted and observed values. The simulation was considered excellent when  $nRMSE$  was  $<10\%$ . If  $nRMSE$  was greater than 10%, but less than 20%, the prediction was considered good. If  $nRMSE$  was greater than 20%, but less than 30%, the prediction was considered fair. If  $nRMSE$  is greater than 30%, the prediction performance is considered poor (Dettori et al., 2011). The specific formula is shown in Equation (10).

$$nRMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \times \frac{100\%}{\bar{y}}, \quad (10)$$

where  $\bar{y}$  represents the mean of the true values, and the other symbols are the same as those in Equation (8).

The Agreement Index,  $D$ -index (Willmott, 1981), is defined as

$$D - index = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2}, \quad (11)$$

where the numerator is the sum of squared errors, and the denominator is related to the variability of the true and predicted values. If the model is perfect, the predicted values would be equal to the model's forecast,  $D - index = 1$ . If the model predicts the same value for all cases and is equal to the average of the observed values,  $\hat{y}_i = \bar{y}$ , then  $D - index = 0$ .

Accuracy is one of the most important criteria for prediction models, and another criterion is the improvement in decision-making based on directional forecasting. From a business perspective, the latter is more important than the former (Yu et al., 2008). The ability to predict direction changes can be measured using Directional Statistics ( $D_{stat}$ ), which can be expressed as

$$D_{stat} = \frac{1}{n} \sum_{i=1}^n \beta_i \times 100\%, \quad (12)$$

where if  $(y_{i+1} - y_i)(\hat{y}_{i+1} - y_i) \geq 0$ , then  $\beta_i = 1$ ; otherwise  $\beta_i = 0$ .

The variance of the absolute error (Var(| $e$ |)) directly calculates the variance of the absolute error  $|e_i| = |\hat{y}_i - y_i|$ . This effectively measured the strength or weakness of the stability of the prediction model. The specific formula for this is given in Equation (13).

$$Var(|e|) = \frac{\sum_{i=1}^n (|e_i| - \bar{|e|})^2}{n-1}, \quad (13)$$

**Table 1**  
Data description.

Variable Type	Class	Variable Name	Frequency	Unit	Source
Response Variable Influencing Factors	Alternative Energy	Bohai-Rim Steam Coal Price Index (BSPI)	Weekly	Yuan/ton	Choice
		WTI Crude Oil Futures Price	Weekly	USD/barrel	Investing
		Natural Gas Futures Price	Weekly	USD/million British thermal units	Investing
		Gold Futures Price	Weekly	USD/troy ounce	Investing
		Silver Futures Price	Weekly	USD/troy ounce	Investing
	Supply and Demand Factors	Copper Futures Price	Weekly	USD/pound	Investing
		Iron Ore Futures Price	Daily	Yuan/ton	Choice
		New Energy Price Index	Weekly	–	Choice
		Coal Mining and Washing	Daily	Ten thousand Yuan	Choice
		National Coal Inventory	Monthly	Ten thousand tons	Ressel
Transportation and Dispatch International Coal Price Level Commodity Currency Factors Macroeconomic Factor Attention Historical data of coal prices	International Coal Price Level	National Coal Sales Volume	Monthly	Ten thousand tons	Ressel
		Industrial Electricity Consumption	Monthly	Hundred million kilowatt-hours	Choice
		Power Index	Weekly	Billion Yuan	CSMAR
		Overseas Coal Freight Index	Daily	–	Choice
	Commodity Currency Factors	National Key State-Owned Coal Mine Railway Volume	Monthly	Ten thousand tons	Ressel
		Australia Newcastle Thermal Coal Price Index	Weekly	–	Investing
		USD/CNY Exchange Rate	Daily	–	Choice
	Macroeconomic Factor Attention	SHSE Composite Index	Weekly	–	CSMAR
		coal, coal price, thermal coal, coal mining, coal share price, thermal power, energy, oil price, iron ore price	Weekly	–	Google
	Historical data of coal prices	First-lagged Coal Price Index	Weekly	Yuan/ton	Choice

where  $|\bar{e}|$  represents the mean of the absolute error.

Prediction Interval Coverage Probability (*PICP*) is used to evaluate the reliability of the forecast interval, indicating the probability that the actual observed value falls within the upper and lower bounds of the forecast interval. The specific formula is given as below:

$$PICP = \frac{1}{n} \sum_{i=1}^n \kappa_i, \quad (14)$$

where  $n$  denotes the number of predicted samples,  $\kappa$  is the boolean quantity, if the predicted value includes the upper and lower limits of the interval prediction,  $\kappa = 1$ , otherwise  $\kappa = 0$ .

Prediction Interval Normalised Average Width (*PINAW*) measures

the clarity of the prediction, and avoids the prediction interval being too wide to give effective uncertainty information of the predicted value due to the simple pursuit of reliability. That is,

$$PINAW = \frac{1}{nR} \sum_{i=1}^n [U_i - L_i], \quad (15)$$

where  $n$  is the number of predicted samples and  $R$  is the variation range of predicted values. When *PICP* is constant, the smaller the value of *PINAW*, the narrower the prediction interval, and more accurate the estimation of the overall parameter.

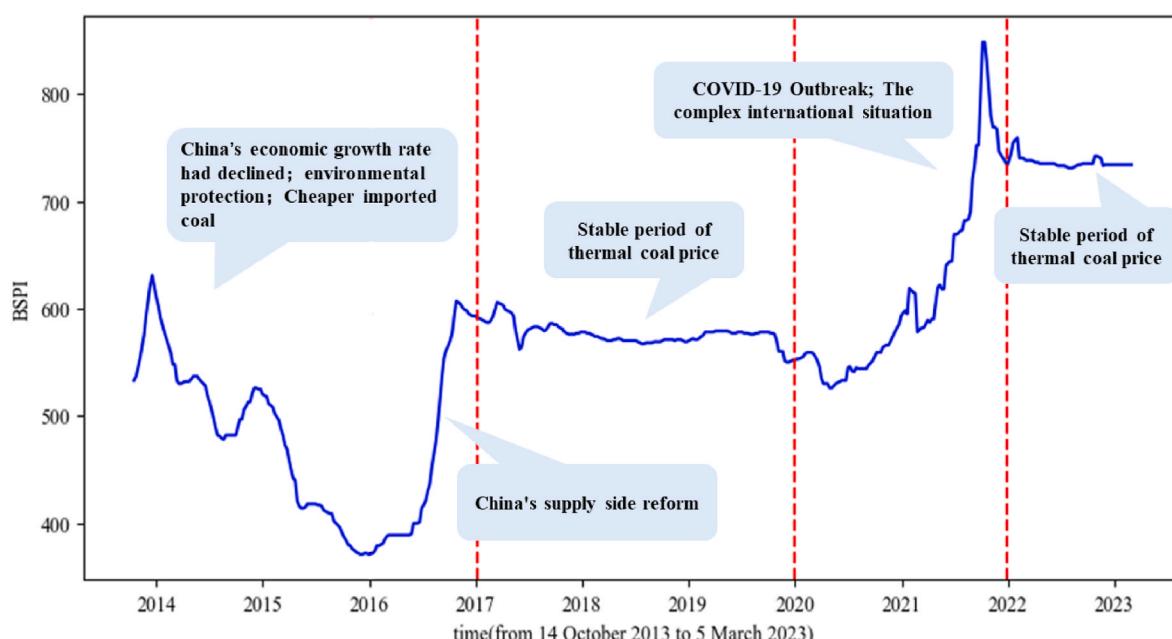


Fig. 2. Original BSPI.

**Table 2**

Compilation of literature on multivariate coal price forecasting.

Study	Frequency	Model	Metrics	Dependent variables	Independent variables
Krzemien et al. (2015)	Monthly	AR(2)	RMSE, MAPE	ICE Brent Crude Oil Front Month Futures Index, Henry Hub Natural Gas Front Month Futures Index, New York Stock Exchange Composite Index, Shanghai Stock Exchange Composite Index, USD/CNY Exchange Rate, Australian Thermal Coal Price Index	McCloskey NW Europe Steam Coal Marker (MCIS)
Alameer et al. (2020)	Monthly	LSTM-DNN	MSE, RMSE, MAE	Price of crude oil, natural gas, copper, iron, silver and gold, Exchange Rate of Australian Dollar, Indonesian Rupiah and Chinese Yuan	Australian steam coal price
Liu and Li (2021)	Monthly	LSTM with improved Adam optimizer	MSE, RMSE, MAE	Daily average shipping price, Number of anchored vessels, Last month's coal price, International coal import price, Average inventory of the six major power generation groups, Daily coal consumption of the six major power generation groups	Settlement price of steam coal in the Bohai region
Ding et al. (2021)	Weekly	C-MIDAS- XGBoost	RMSE, MAE	Renewable energy, Daqing Oilfield, Natural gas in Japan, Australian thermal coal price, Coal mining industry index, A-share power industry index, A-share index, Coal industry index, Temperature	Free-on-Board (FOB) price of thermal coal at Qinhuangdao Port in China
Wua et al. (2023)	Weekly	AOA-VMD-DeepTCN	MAPE, SMAPE, MASE	Daqing Crude Oil Spot Price(DCOP), New York Natural Gas Spot Price (NPNGSP), Hydropower Industry Index(HII), Wind Power Industry Index (WPPI), Coal Industry Index(CMII), Industrial Index(II), A-share Power Industry Index(ASEI), Newcastle Thermal Coal Spot Price in Australia (NPCSP), Average Maximum Temperature in Major Cities Across the Country(Tem), Coal Industry Index(CII), A-share Index(ASI), US Dollar to Chinese Yuan Exchange Rate(U-C), Australian Dollar to Chinese Yuan Exchange Rate(A-C), Shanghai Interbank Offered Rate (Overnight) (SHIBOR), 6-Month Government Bond Yield(SMTY)	China Bohai-Rim Steam Coal Average Price Index (BSPI)

#### 4. Empirical analysis

The models in this study were implemented on the Python 3 and Matlab R2021a platform using the deep learning framework TensorFlow GPU 2.2.0.

##### 4.1. Data description

This study employs the techniques of database querying and online searching to procure data from Choice, Rerset, and CSMAR databases, along with Investing and Google Trend websites. The dataset utilized covers the period from October 14, 2013, to March 5, 2023. The fundamental details of the response variables and influencing factors are presented in Table 1. The rationale behind the selection of these variables and the manner in which these influencing factors impact the response variables will be elaborated upon in Sections 4.1.1 to 4.1.2.

##### 4.1.1. Response variable

This study uses the Bohai-Rim Steam-Coal Price index (BSPI) as the response variable, which is a comprehensive index system representing the level and fluctuations in offshore closing prices for thermal coal at ports around the Bohai Sea in China. BSPI is acclaimed as the "coal price barometer" and serves as a vital window for research institutions to conduct analysis and studies in the coal industry.

Fig. 2 shows the weekly price index of Bohai-Rim steam-coal from October 14, 2013 to March 5, 2023. This dataset experienced significant fluctuations from October 2013 to December 2016 and January 2020 to December 2021. The reasons for these fluctuations can be traced back to the continuous impact of the "Golden Decade" in the coal industry, which led to an increase in coal production with the release of additional capacity. In 2013, as China's economic growth rate began to decline, the growth rate of downstream thermal power generation decreased by a single digit. This, coupled with an increase in supply and a decrease in demand, has resulted in a continuous decline in coal prices. Additionally, changes in environmental regulations and low international coal prices have significantly impacted China's coal market, with excess capacity. It was not until 2016 that the national supply-side reform ended the downward trend in coal prices.

In late 2019, the COVID-19 pandemic caused a global shutdown of manufacturing, which led to a dramatic reduction in electricity consumption. Ports and power companies experienced a substantial increase in coal inventories, resulting in a decline in coal prices. As the

COVID-19 situation gradually came under control, downstream enterprises accelerated the resumption of operations, leading to a gradual recovery in industrial power demand (Li et al., 2022). The coal market faced an imbalance between supply and demand, which led to a surge in prices. Such changes clearly pose challenges to macroeconomic regulation; one of their impacts is power rationing in China's northeastern region. Simultaneously, the deterioration of trade relations between China and Australia contributed to coal price fluctuations.

Based on the above analysis, it is evident that policy changes and unforeseen events can cause significant fluctuations in coal prices. Predicting the nonlinear and non-stationary price dynamics of thermal coal in China has become challenging.

##### 4.1.2. Analysis of influencing factors

Zhao et al. (2016) conducted a study on the offshore price (FOB) of thermal coal at the Qinhuangdao Port in China using multifractal detrended fluctuation analysis (MFdfa). Xu and Zhang (2022) analysed the daily closing price data of thermal coal trading on the Zhengzhou Commodity Exchange in China using nonlinear autoregressive neural networks. Zhangb et al. (2022) predicted coal prices at the Qinhuangdao Port in China using VMD-A-LSTM-SVR. The aforementioned studies used univariate data to predict coal prices but did not consider the impact of exogenous variables on coal prices. Lin and Shi (2022) demonstrated that multivariate predictions often yield higher accuracy than univariate predictions and that multivariate forecasting can ensure the stability of the relationship between the response variable and predictor variables over time. Selection of appropriate predictor variables for multivariate forecasting is a crucial and challenging task.

The predictor variables selected by most models can be broadly categorized into two types. For example, Krzemien et al. (2015) predicted European thermal coal prices by considering energy-driven factors, demand-driven factors, commodity currency factors, and supply-driven factors. Ding et al. (2021) predicted the weekly prices of thermal coal in China from five perspectives: energy, supply, demand, macroeconomics, and weather. Wua et al. (2023) studied the factors that influence the average price index of thermal coal in the Bohai Rim region of China from six perspectives: energy, supply, demand, international markets, weather, and macroeconomics. We collates other literatures on multivariate coal price prediction, and summarizes the influencing factors selected by predecessors as shown in Table 2, which plays a key role in variables' selection of this study. Another type of predictor variable considers attention factors. For instance, Wub et al.

**Table 3**

Mechanism keywords of the influence for each variable on coal price.

Variable Type	Variable Name	Influencing Mechanism Keywords
Alternative Energy	WTI Crude Oil Futures Price, Natural Gas Futures Price Gold Futures Price, Silver Futures Price, Copper Futures Price, Iron Ore Futures Price New Energy Price Index	Substitution Effect, Industrial Chain Correlation Overall Trend of Energy Market, Macroeconomy
Supply and Demand Factors	Coal Mining and Washing	Substitution Effect
	National Coal Inventory, National Coal Sales Volume Industrial Electricity Consumption, Power Index	Coal Mining Efficiency, Mining Cost, Mining Cost Quantity Direct Impact
Transportation and Dispatch	Overseas Coal Freight Index, National Key State-Owned Coal Mine Railway Volume	Electricity Demand
International Coal Price Level	Australia Newcastle Thermal Coal Price Index	Transportation Cost
Commodity Currency Factors	USD/CNY Exchange Rate	Price Transmission Mechanism, Market Expectations and Psychology, Policy and Supervision
Macroeconomic Factor	SHSE Composite Index	Import Cost, Market Competition
Attention	Google Trends-coal, coal price, thermal coal, coal mining, coal share price, thermal power, energy, oil price, iron ore price	Macroeconomic Expectations, Market Sentiment and Capital flows
Historical Data of Coal Prices	First-lagged BSPI	Market Attention, Market Expectations, Investor Behavior
		Market Expectations and Psychology, Supply-Demand Adjustment, Inventory and Logistics, Policy and Regulatory Environment

(2023) forecast tourist flow from mainland China to Hong Kong based on search query data, the Baidu Index, online news data, and others. Wu et al. (2021) investigate crude oil prices using Google Trends and online crude oil news. These studies indicate that both types of indicators can improve the predictive performance of the models.

Supply and demand are the basic drivers of price formation, and the price of substitutes, transportation costs of goods, global commodity prices, exchange rate fluctuations, macroeconomy, public attention to the commodity, and the historical price of the commodity all have impacts on the price of the commodity. At present, this study has not found any scholars using variables such as Google trends to analyze its impact on coal price prediction. In constructing the coal price prediction model, this study comprehensively considered both types of data and classified the influencing factors into eight categories: supply and demand factors, alternative energy, transportation scheduling, international coal price level, commodity currency factors, macroeconomic factors, attention factors, and historical coal prices. Specific information is provided in Table 3.

(1) **Alternative Energy:** Coal, petroleum, and natural gas are the major fuels used in industrial boilers and power generation, and they can be used interchangeably in certain circumstances, with mutual conversions, such as synthetic natural gas (SNG) and gas-to-liquids (GTL) (Li et al., 2021). Studies have shown that their prices are interconnected (Alameer et al., 2020). Since 2000, the correlation coefficient between coal and petroleum prices has

consistently been above 0.8, which is higher than in any period since 1980, and coal and petroleum prices have exhibited a parallel trend (Yang et al., 2012). Compared with coal and petroleum, natural gas can effectively reduce the emissions of particulate matter, sulfur dioxide, carbon dioxide, and nitrogen oxides, thereby fundamentally improving environmental quality. In the face of increasingly severe climatic conditions, natural gas has been rapidly developed as a substitute for coal. Scholars have pointed out that natural gas consumption is highly sensitive to fluctuations in coal prices, especially in the power generation, heating, and transportation sectors (Li et al., 2021). Therefore, WTI crude oil and natural gas futures prices were chosen to reflect coal prices.

Given the influence of advancements in new energy technologies and decreasing costs, alongside policy orientations and transformations in energy structures, nations will decrease their reliance on fossil fuels, resulting in an accelerated adoption of alternative energy. Such trend will have significant and enduring ramifications for coal pricing. The proportion of renewable energy generation in China's national energy mix has reached 31.5%, making renewable energy an indispensable component of energy generation. Owing to its advantages such as low pollution, its contribution is expected to increase in the future. The higher the proportion of renewable energy generation, the lower the coal consumption for power generation, leading to price fluctuations. Hence, the new energy price index was selected to reflect coal price (Ding et al., 2021).

How do these factors influence coal prices? Firstly, there is substitution effect, where oil, natural gas, and new energy serve as mutual substitutes for coal. When the price of one of these substitutes rises, some consumers or producers who initially relied on oil and natural gas may switch to coal as an alternative, consequently increasing the demand for coal and driving up its price. Secondly, these energy sources are interconnected within the industrial chain. For instance, petrochemical, natural gas chemical, and coal chemical industries are distinct fields, yet they all depend on their respective energy raw materials. If the price of oil or natural gas increases, the cost of the associated chemical products may rise, potentially prompting some producers to shift towards coal-based chemical products, which in turn increases the demand for coal and subsequently leads to an increase in coal prices. Due to advancements in new energy technologies, cost reductions, policy orientations, and carbon mitigation pressures, countries will decrease their reliance on fossil fuels, which will have long-lasting implications for coal prices.

Research suggests that there is a high correlation between gold, silver, copper, and coal prices (Krzemien et al., 2015). Therefore, this study selected the prices of gold, silver, copper, and iron ore, which exhibit a high correlation with coal prices, as influencing factors.

These energy sources are important parts of the energy market, and their price trends are often affected by a variety of factors such as the overall supply and demand relationship, macroeconomic conditions, and policy changes. Fluctuations in these energy prices may indirectly affect coal prices through the overall movement of the energy market. For instance, from a macroeconomic perspective, escalating prices of precious metals like gold and silver, as well as industrial metals such as copper and iron ore, can result in heightened costs for specific industries, which may subsequently drive up the overall price level and potentially exacerbate inflationary pressures. The surge in inflation will influence the prices of commodities and equipment essential for coal production, ultimately leading to a rise in coal production costs and, consequently, an increase in coal prices.

(2) **Supply and demand factors:** Commodity prices often respond to new inventory data (Plante and Dhaliwal, 2017). When inventory is abundant, coal prices tend to decrease, whereas they increase

when supply is limited. Therefore, coal mine inventory was chosen as a supply side influencing factor.

The coal mining and washing industry index was selected to represent the level of coal production (Ding et al., 2021). Coal mining cost, mining efficiency and mining volume directly affect the coal price. Higher coal industry mining index will lead to larger coal production, thereby increasing the market supply, which in turn affects the coal price.

On the demand side, energy consumption in the spot market is closely related to energy prices (Gil-Alana et al., 2020). An increase in coal prices may reduce the output of enterprises with high coal consumption and worsen their financial condition (Kong et al., 2020), thus affecting coal prices. In short, consumption directly affects the balance of market supply and demand, and when coal consumption increases, coal prices rise. Therefore, the national coal sales volume was selected as the indicator of demand.

Considering that thermal coal is mainly used for power generation, and electricity demand is mainly driven by the industry (Yue et al., 2021), the national coal-fired power generation in 2021 increased by 8.6% compared to the previous year, becoming the main driving force for the growth of coal consumption. The increase in demand will cause coal prices to rise; therefore, industrial electricity consumption was chosen to reflect the demand factor. The increase in industrial electricity consumption will lead to an increase in coal demand, which will push up coal prices and enhance the activity of the coal market.

Among the four major coal-consuming industries, the power industry accounted for the largest proportion of coal consumption. China's thermal power plants consume more than half of its annual coal, accounting for over 70% of its annual power supply (Lin and Wang, 2021), and steam power generation relies on thermal coal. Therefore, the power index was selected to reflect the demand for coal. The rising of power index means the increasing of demand for electricity, which leads to the growth of coal demand, which in turn increases the price of coal.

**(3) Transportation and dispatch:** The main transportation modes for coal in China are railways and waterways. The data show that in 2021, cumulative coal transportation by railway in China exceeded 2.58 billion tons, accounting for 62.46% of the national coal production. Based on this, this study introduces the national key state-owned coal mine railway volume to reflect the coal transportation volume. Transportation costs significantly affect the final prices of transported goods (Volpe et al., 2013). Melas and Michail (2021) suggested a positive nonlinear relationship between commodity prices and charter rates. Therefore, the Ocean Freight Coal Index (OCFI) was chosen to reflect the freight rate trends in the coastal coal transportation market.

If there is an increase in rail or sea traffic, the transportation cost per unit of coal may decrease as a result of scale effects and enhanced transportation efficiency, thereby contributing to a reduction in the final selling price of coal.

**(4) International coal price level:** China is a net coal-importing country, and one of the main sources of imports is Australia (Ding et al., 2021). Australia is one of the largest coal-exporting countries worldwide. Therefore, the Australian newcastle coal price index was selected to reflect the trends in international coal prices.

There is a certain price transmission mechanism between the international coal market and the domestic coal market. When the international coal price rises, the domestic coal price may also be subject to a certain degree of upward pressure due to the relatively close connection between the two markets. When international coal prices persist in rising, the market may develop an expectation of increasing prices,

potentially prompting participants in the domestic market to adjust their purchasing and sales strategies, thereby impacting the supply and demand relationship and price levels. In addition, changes in government policies and the regulatory environment will also have an impact on the price relationship between the two markets. For instance, if the government tightens the supervision of coal imports or modifies policies such as coal import tariffs, the impact of international coal prices on domestic coal prices may be constrained or altered to a certain extent.

**(5) Commodity currency factor:** Exchange rates play an important role in predicting commodity price fluctuations (Chen et al., 2010). Krzemień et al. (2015) indicate that changes in the USD/CNY exchange rate are a major influencing factor affecting European spot coal prices. Positive exchange rate fluctuations attract foreign investors, thereby increasing the demand for natural resource commodities and ultimately raising their prices, contributing to the country's economic development (Huawei, 2022). The United States is the second-largest coal-producing country after China (Krzemień et al., 2015), and the US dollar is one of the world's reserve currencies. Therefore, this study introduced the USD/CNY exchange rate as a predictive factor.

If the US dollar appreciates relative to the Chinese yuan, resulting in the depreciation of the yuan, this will elevate the costs of importing coal for domestic enterprises, subsequently driving up coal prices. In addition, there is a certain competitive relationship between the domestic coal market and the imported coal market. When the depreciation of RMB causes the cost of imported coal to rise, then the competitiveness of imported coal decreases, and the demand of domestic coal market increases, which plays a rising impact on the domestic coal price.

**(6) Macroeconomic factor:** Macroeconomics can influence the prices of natural resources. Ekananda (2022) found a positive correlation between macroeconomic factors such as exports, GDP growth, foreign direct investment, and natural resource prices. As coal is China's primary energy source, meeting three-fifths of China's energy demand, its price fluctuations can exert pressure on the Chinese economy (Li et al., 2019). Drawing on the findings of Ding et al. (2021), we selected the SHSE composite index to reflect China's macroeconomic conditions.

When the SHSE composite index sustains an upward trend, it may signify an improvement in the macroeconomic landscape, potentially leading to an increase in industrial production and energy demand, among other factors. Since domestic industrial production is one of the important sources of coal demand, this change in macroeconomic expectations could have a positive impact on coal prices. Moreover, the rise and fall of the SHSE composite index often reflect the changes in sentiment and capital flows of market participants. During periods of rising SHSE composite index, market sentiment tends to be more optimistic, prompting an influx of funds into the stock market, with numerous industries, including the coal sector, poised to reap the benefits. These inflows could elevate the prices of coal stocks, thereby influencing market expectations for future coal prices.

**(7) Attention:** This study incorporates Google Trends to reflect user interests in coal-related topics. Among all search engines, Google Search ranked first in terms of traffic. As of April 2023, the global market share was 63.51%. Google Trends analyses billions of daily search data from Google search engines, and provides users with search frequencies and related statistics for a given keyword across different periods. Wu et al. (2021), based on Google Trends and online media text mining technology, used BPNN, MLR, SVM, LSTM, GRU, and RNN to build models to predict crude oil prices and confirmed that Google Trends is a very effective predictor. In this study, four keywords ("coal", "coal price", "thermal coal",

- "thermal coal price") are selected to represent Internet users' interests in the coal market. The specific process is as follows.
- (1) Each of the four keywords was used as a seed keyword in the search.
  - (2) The recommended keywords from the previous round were used as second-round search keywords.
  - (3) Step (2) is repeated until no new keywords remain in the recommended list.

After completing the above steps, keywords with missing values were removed, and Google Trends data for the nine keywords were retained. They are "coal", "coal price", "thermal coal", "coal mining", "coal share price", "thermal power", "energy", "oil price", and "iron ore price".

The influence of Google trends on coal prices is subtle, intricate, and fluctuating. Google trends can, in some instances, mirror the market's focus and anticipated shifts within the coal and related industries, potentially influencing coal prices through market attention, expectations, and investor behavior. On the one hand, if Google trends shows a significant increase in searches for "coal" or related keywords, it implies that the market's attention to the industry is rising, at which point investors and traders may pay more attention to the dynamics of the coal market, which could influence their trading behavior and decisions, which in turn have an impact on coal prices. On the other hand, Google trends' search data can also be used as a guide to market expectations. For instance, if rumors of constrained coal supply or heightened market demand spark an increase in searches for related keywords, this may mirror optimistic market forecasts for coal prices. Such expectations could incite investors to purchase coal futures or related stocks, thereby pushing prices upwards.

- (8) **Historical data of coal prices:** Historical prices can impact current economic outcomes (Campbell, 2020). Many economic variables are related to their lagged values, indicating lags in the economic variables. Regardless of whether it is pricing by manufacturers or purchasing decisions by customers, historical coal prices serve as important references.

So how does the previous price of coal affect the current price? First of all, the trend of coal prices in the early stage will affect the expectations and psychology of market participants. If the price of coal continues to rise in the early period, the market may form expectations of price increases, which will promote the rise of coal prices in the current period. Conversely, the current coal prices may face downward pressure. Secondly, fluctuations in coal prices during the early phase will induce adjustments in behavior among both suppliers and demanders. For example, if the price of coal rises, suppliers may increase mining and production to meet market demand and capture higher profits. And the demand side may purchase or look for alternative energy sources in advance to cope with the cost pressure caused by higher prices. Furthermore, variations in coal prices during the early phase will also impact the inventory management and logistics of coal. If the historical coal price is high, suppliers may increase inventories to cope with possible price increases in the future. Further, the demand side may decrease their inventories to minimize costs. This inventory adjustment will affect the supply and demand balance and price level of the coal market. At the same time, the change of logistics costs will also affect the transportation and distribution costs of coal, and then affect the price of coal. Changes in government policy and the regulatory environment have a direct impact on coal prices. In the event of abnormal fluctuations in coal prices during the initial phase, the government may enact pertinent policies to intervene and regulate the market. Such policies may impose restrictions on coal mining and production, modify coal import and export regulations, enhance environmental protection and safety oversight, among other measures, thereby influencing the supply, cost, and ultimately the current price of coal.

Therefore, in constructing the coal price forecasting model, this

**Table 4**

The correlation results between the lagged variables and the current variables of BSPI.

Metrics	First-order lag	Second-order lag	Third-order lag	Fourth-order lag
$D_{stat}$	<b>0.985</b>	0.962	0.935	0.900
Pearson correlation coefficient	<b>0.998</b>	0.992	0.983	0.972
Spearman correlation coefficient	<b>0.991</b>	0.971	0.946	0.919

study considers the influence of the coal prices of the previous period and calculates the  $D_{stat}$ , Pearson correlation coefficient, and Spearman correlation coefficient between the average price index of Bohai-Rim steam coal on the training set  $y(t)$  and its first-order lag  $y(t - 1)$ , second-order lag  $y(t - 2)$ , third-order lag  $y(t - 3)$ , and fourth-order lag  $y(t - 4)$ . Table 4 presents the results of the study.

As shown in Table 4, the  $D_{stat}$  between BSPI  $y(t)$  and its lagged variable  $y(t - 1)$  is closest to 1, indicating a high degree of positive correlation and strong alignment in their changes. The Pearson and Spearman correlation coefficients were also the highest, further confirming a significant correlation between the two. Therefore, it is reasonable to introduce the lagged variable  $y(t - 1)$  as an influencing factor.

Although the current coal price  $y(t)$  has  $D_{stat}$ , the Pearson correlation coefficient, and the Spearman correlation coefficient above 0.9, with the lagged variables of the second-order  $y(t - 2)$ , third-order  $y(t - 3)$ , and fourth-order  $y(t - 4)$ , this is mainly due to the small differences in the adjacent four prices of the BSPI, leading to a strong correlation between the lagged variables and the current variable. Considering the principle of model simplicity, these three variables were excluded as influencing factors.

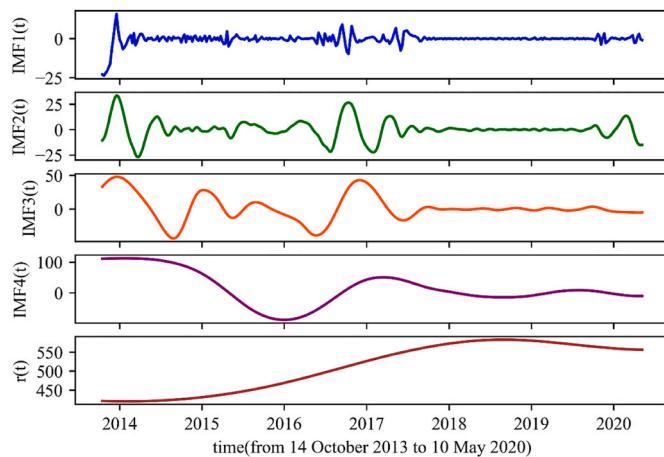
#### 4.2. Data processing

In this study, we modelled weekly data of the Bohai-Rim Steam Coal Price Index (BSPI) and its influencing factors. There were three reasons for this finding. First, the planning cycle of power plants typically spans weeks, making weekly coal price predictions more relevant to power plant operations (Ding et al., 2021). Second, weekly prices capture the dynamic characteristics of the coal market better than monthly prices because monthly fluctuations can be masked by temporal aggregation (Ding et al., 2021). Weekly data contain useful forecasting information (Hu et al., 2022; Wua et al., 2023). Third, the impact of relevant driving factors on coal prices may be short-term, making it difficult to determine using monthly data alone (Ding et al., 2021).

An investigation found that iron ore futures prices and coal mining industry indices are reported daily, whereas coal inventory, national coal consumption, and industrial electricity consumption are reported monthly. These influencing factors are heterogeneous data from multiple sources and frequencies, requiring them to be unified into weekly data. For daily data, the average value for each week was calculated from the available data. Additionally, the study utilized the quadratic-match sum method (Borjigin et al., 2018) to convert monthly data into weekly data using the Eviews 7.2 software for this process.

The dataset used in this study spanned October 14, 2013 to May 10, 2020 comprising 343 data points, which were divided into a training set. The period from May 11, 2020 to March 5, 2023 consisting of 147 data points, was designated as the testing set, accounting for 30% of the original data.

Owing to the needs of the model, the data were standardised. For example, Alameer et al. (2020) and Wu<sup>b</sup> et al. (2023) processed original data using the max-min normalisation method. However, these two studies substituted the maximum and minimum values of all the data into the normalisation formula, leading to information leakage. In this study, only information from the training set was used for normalisation. Given the scenario predicted by the rolling window in this study, the



**Fig. 3.** Decomposition results of ICEEMDAN

maximum and minimum values of the training set must be constantly updated.

#### 4.3. Data decomposition

Building on the model introduction in Section 3.8, this study first decomposed the training set data to obtain a relatively stationary subseries. The ICEEMDAN algorithm was employed for this purpose, with

an added white noise standard deviation of 0.05 and a set iteration count of 100. The resulting  $IMFs(t)$  and residual term  $r(t)$  after decomposition are illustrated in Fig. 3.

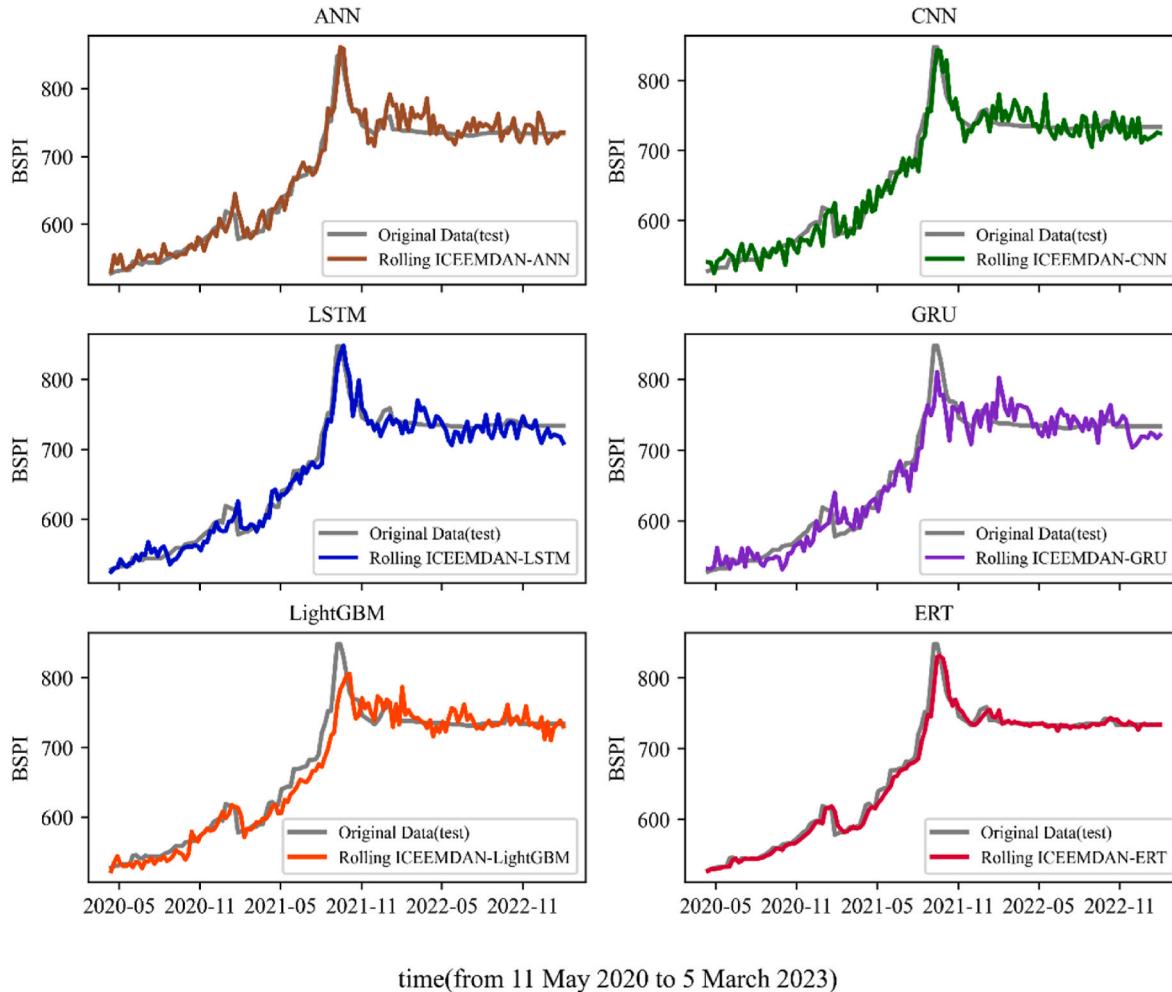
#### 4.4. Model prediction

Each method has its advantages and disadvantages. Numerous forecasting methods have been proposed, and research results regarding the superiority of these methods vary (Herrera et al., 2019). In this study, we chose the classic neural network ANN, convolutional neural network CNN, recurrent neural networks LSTM and GRU, and tree-based models LightGBM and ERT to predict thermal coal prices.

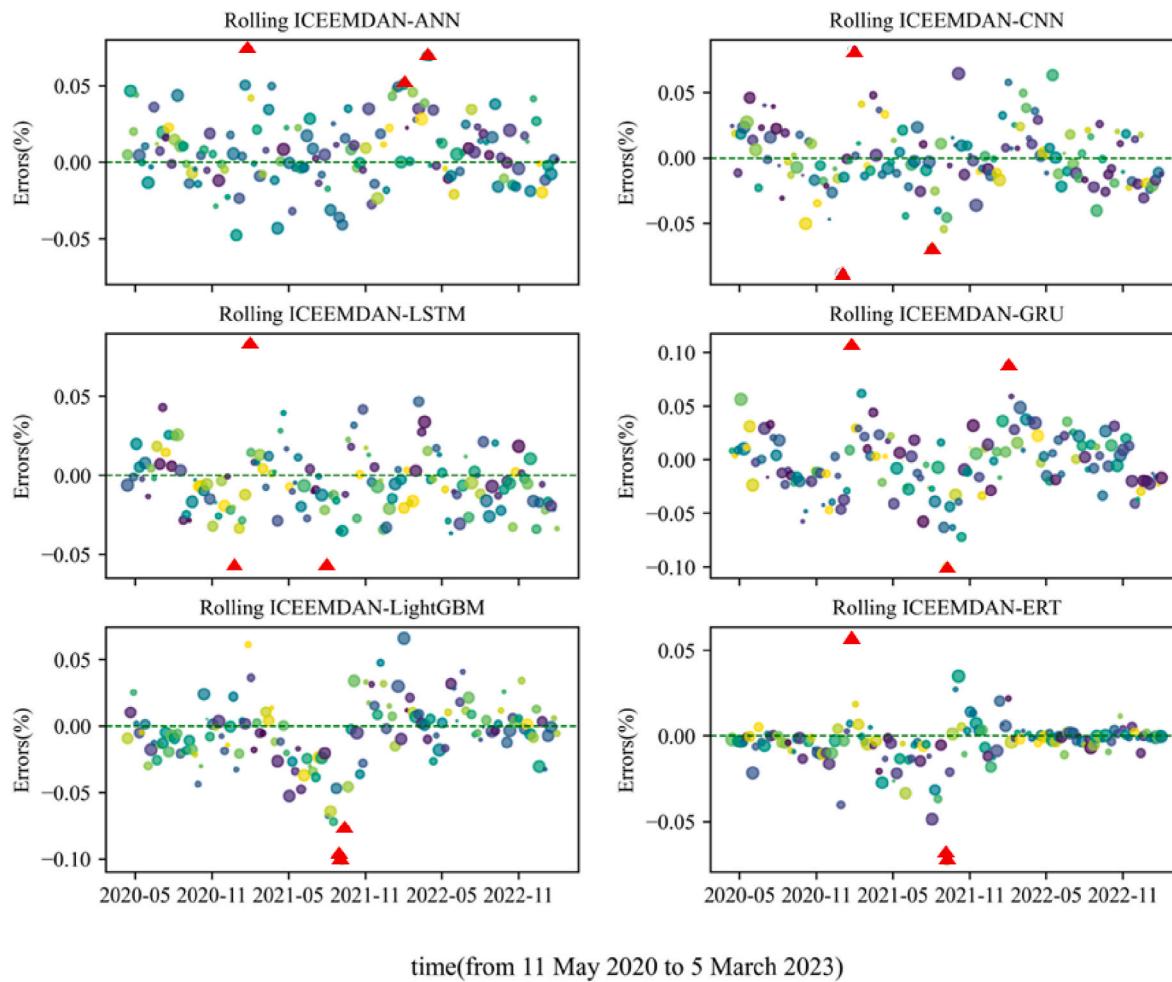
First, we constructed models using the decomposed  $IMFs(t)$  and  $r(t)$  obtained from Section 4.3 for prediction. The predictions from each submodel were then aggregated to obtain the final prediction results. When building the models, the factors influencing the thermal coal prices provided in Section 4.1.2 were used as inputs, and the  $IMFs(t+1)$  and  $r(t+1)$  derived from the ICEEMDAN decomposition were used as outputs, thus constructing the thermal coal price prediction models.

For the adaptive forecasting approach, the Rolling ICEEMDAN-Methods model was updated whenever new data were added to the training set, resulting in weekly predictions. In this study, the entire process was repeated 147 times to obtain 147 predicted data points from the test set.

The prediction results of Rolling ICEEMDAN-ANN, Rolling ICEEMDAN-CNN, Rolling ICEEMDAN-LSTM, Rolling ICEEMDAN-GRU, Rolling ICEEMDAN-LightGBM, and Rolling ICEEMDAN-ERT are shown



**Fig. 4.** Rolling ICEEMDAN-Methods prediction results.



**Fig. 5.** Bubble chart of relative errors for rolling ICEEMDAN-Methods.

in Fig. 4. Overall, all six models demonstrated good predictive performance, capturing the general trends in thermal coal prices. The curve of Rolling ICEEMDAN-ERT closely matches the original data, followed closely by that of Rolling ICEEMDAN-LSTM.

Fig. 5 presents a bubble chart of the relative errors for the Rolling ICEEMDAN-Methods on the test set. Data points with larger relative errors are marked with red triangles, primarily during the three periods of high volatility (February 2021, October 2021, and February 2022).

The traditional Chinese Spring Festival will occur in February 2021 and February 2022. To ensure adequate coal supply during the Spring Festival and major events, large- and medium-sized coal enterprises arranged for coal supply before the festival. As the Spring Festival approached, overall coal demand declined, leading to a slight decrease in prices. In October 2021, power rationing will be enforced in north-eastern China. During this period, coal prices fluctuated abnormally, which resulted in relatively large prediction errors. However, from a macro perspective, the Rolling ICEEMDAN-Methods still captured these changes effectively.

From Fig. 5, it can be observed that the data points of relative errors for the Rolling ICEEMDAN-Methods are randomly distributed around zero and fluctuate within the range of (-0.1, 0.1). The Rolling ICEEMDAN-ERT exhibited the smallest relative error.

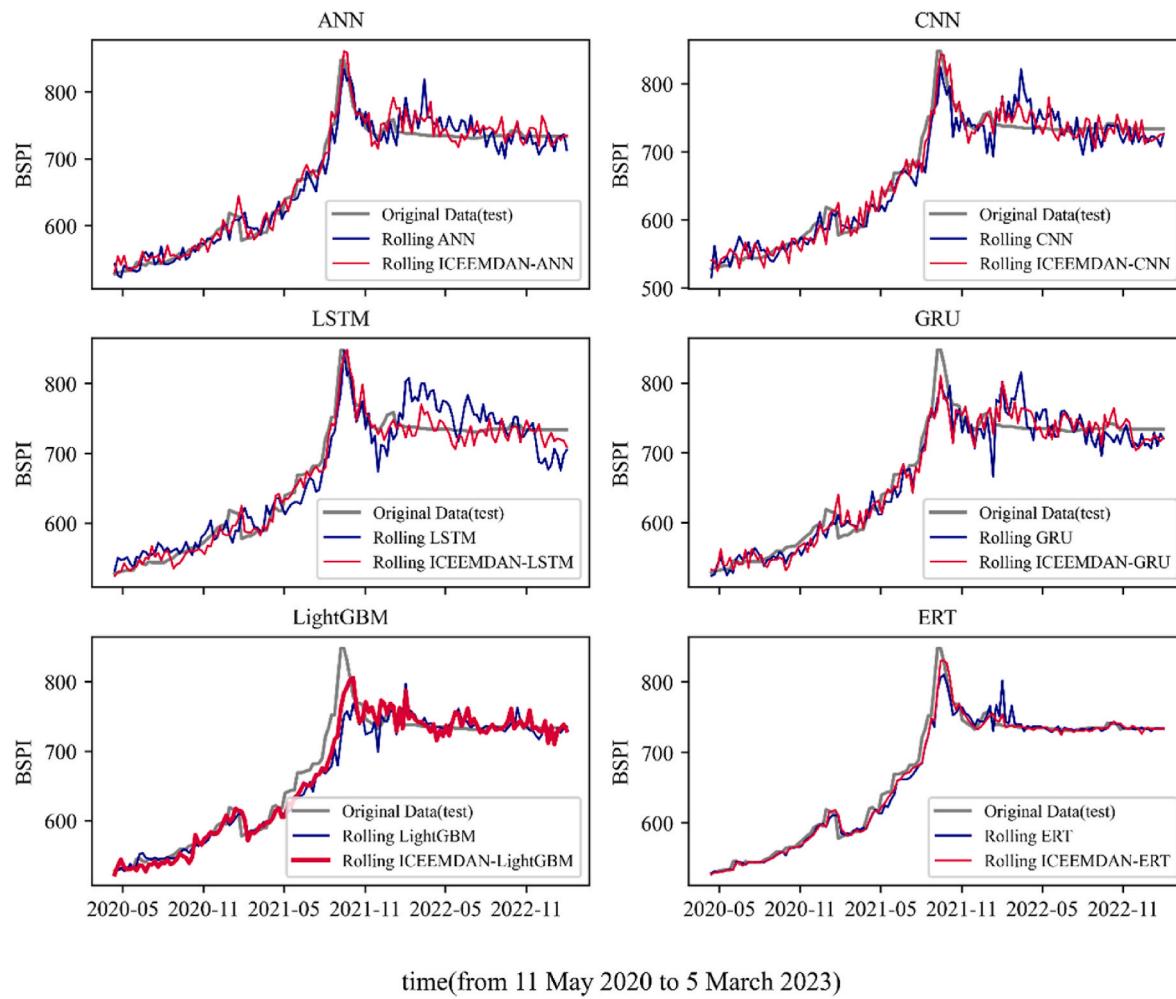
#### 4.5. Comparative experiments

In this section, two sets of comparative experiments are conducted to validate the high performance of the Rolling ICEEMDAN-Methods model.

##### 4.5.1. Comparative experiment I

To demonstrate the ability of the proposed decomposition-prediction-ensemble model to improve the prediction performance, six models—Rolling ICEEMDAN-ANN, Rolling ICEEMDAN-CNN, Rolling ICEEMDAN-LSTM, Rolling ICEEMDAN-GRU, Rolling ICEEMDAN-LightGBM, and Rolling ICEEMDAN-ERT—were compared with non-decomposition-based models, namely Rolling ANN, Rolling CNN, Rolling LSTM, Rolling GRU, Rolling LightGBM, and Rolling ERT, including the prediction accuracy, stability and running time.

The curves in Fig. 6 indicate that both categories of models under the six methods exhibited good predictive performance. The predicted results of Rolling Methods and Rolling ICEEMDAN-Methods from May 2020 to May 2021 were relatively close, and the decomposition advantage was not evident. The price of thermal coal began to soar in May 2021, reaching approximately 136.78% of the May price by mid-October 2021, and then decreased by approximately 13.67% by the end of October. During this period, compared to the Rolling Methods, the Rolling ICEEMDAN-Methods series models were closer to the original data, indicating that the decomposition-prediction-ensemble models were more capable of capturing drastic changes in thermal coal prices, thus confirming their excellent predictive performance. Dramatic price fluctuations during this period also pose a greater challenge for forecasting in later periods. Although the thermal coal price plateaued after 2022, the prediction effect of the decomposition-based models is obviously better than that of the non-decomposition-based models because Rolling Methods cannot better adapt to the soaring price in the previous period. The Rolling ICEEMDAN-Methods models proposed in this study are more suitable for nonlinear and non-



**Fig. 6.** Comparison of results between rolling ICEEMDAN-Methods and rolling methods.

stationary data prediction.

Fig. 7 presents a scatter plot of the relative errors under Rolling Methods and Rolling ICEEMDAN-Methods models. Overall, the error points were randomly distributed around zero, and the error points of Rolling Methods exhibited a wider distribution than those of the Rolling ICEEMDAN-Methods series models. Additionally, there were individual values with relative errors exceeding 0.1 in Rolling Methods, indicating an improvement in model accuracy after applying ICEEMDAN.

To objectively evaluate the predictive accuracy and stability of the proposed models, various evaluation metrics, including *RMSE*, *MAPE*, *nRMSE*, *D-index*, *Dstat*, and error variance, were calculated. The results are shown in Fig. 8, which indicates that the models based on ICEEMDAN performed better. According to Table 5, compared with Rolling Methods, the Rolling ICEEMDAN-Methods models achieved a reduction in *RMSE* of 19.91%, 32.76%, 67.26%, 11.44%, 27.74%, and 25.61%, and a reduction in *MAPE* of 16.95%, 26.29%, 62.73%, 10.00%, 2.11%, and 25.88%, respectively. By analysing the *nRMSE* values for each model, it was evident that all the models in this study had *nRMSE* values less than 0.1, indicating excellent predictive performance, with Rolling ICEEMDAN-Methods consistently outperforming Rolling Methods. Notably, the LSTM model had an *nRMSE* value of 0.0351, indicating high accuracy. However, after decomposition, the *nRMSE* decreased to 0.0201, confirming the superiority of the divide-and-conquer approach. This observation holds true for the ERT, ANN, CNN, and other models.

Furthermore, the variance of the absolute errors,  $\text{Var}(|e|)$ , was introduced to evaluate the stability of the models. The results reveal that the Rolling ICEEMDAN-LSTM has the smallest variance in absolute

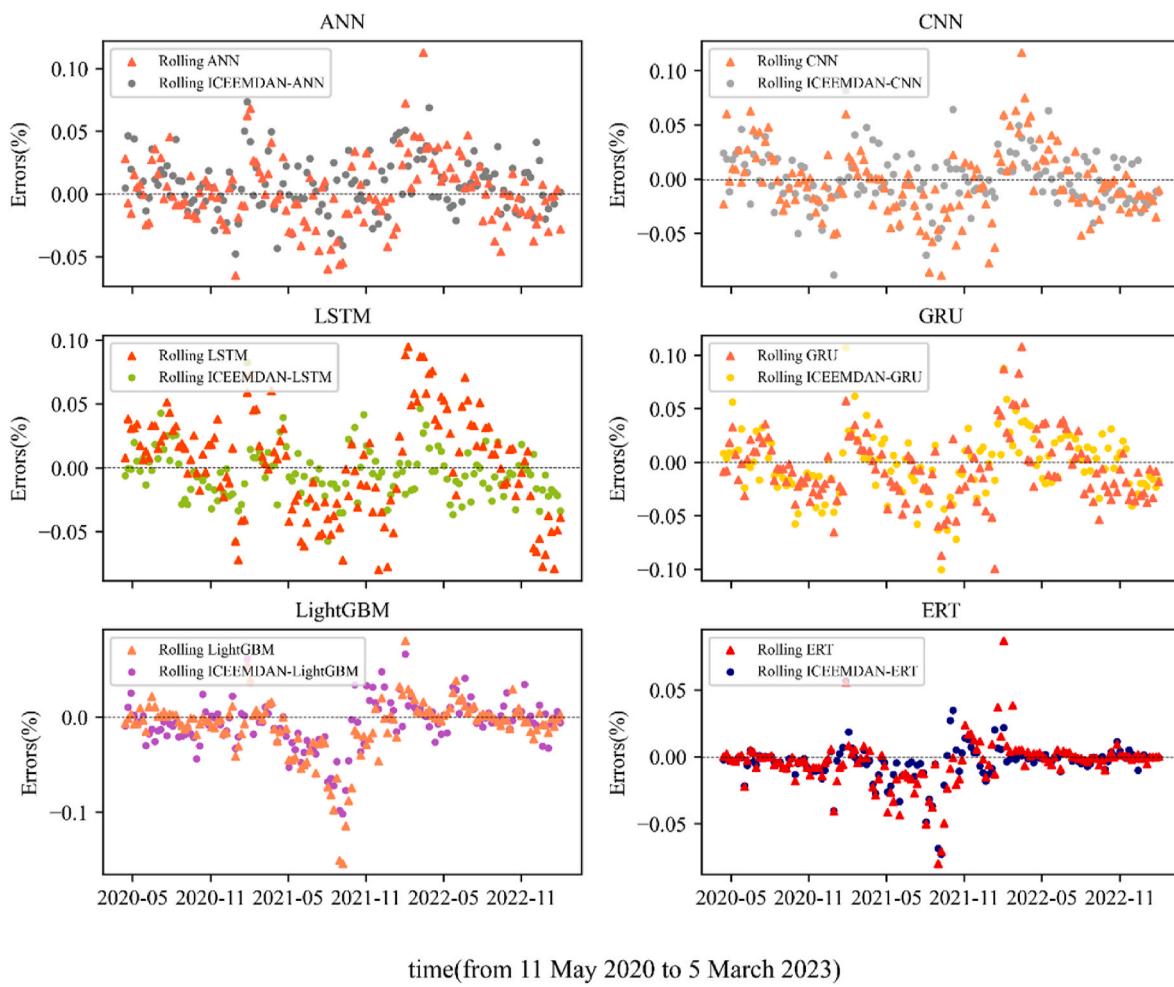
errors, indicating the strongest stability among the models. Upon examining the remaining models, it can be observed that the stability of the Rolling ICEEMDAN-Methods is superior to that of the Rolling Methods.

The scheme of window rolling prediction in this study can largely solve both the problem of information leakage and boundary effect, but the running time and resource consumption of the model should also be paid attention under this framework. Table 6 shows the running time of the Rolling ICEEMDAN-Methods series of models.

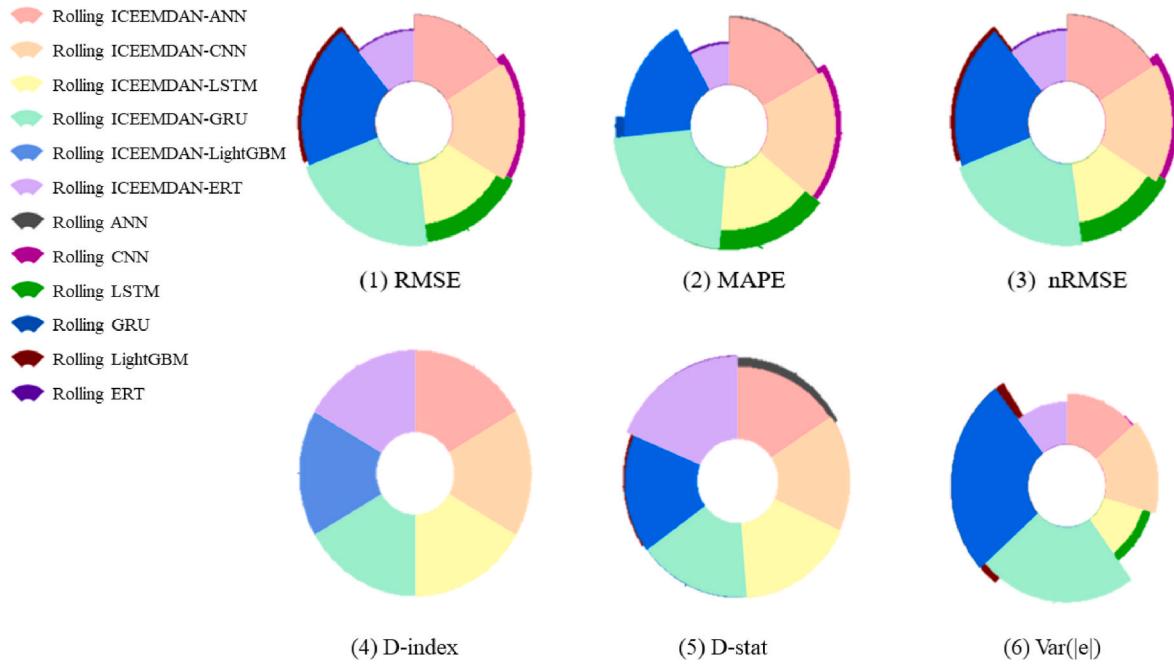
As indicated in Table 6, the Rolling-ICEEMDAN-LightGBM model demonstrates the most notable advantage in running time, with the Rolling-ICEEMDAN-ERT model trailing closely behind. It is evident that the window rolling prediction scheme holds a distinct advantage in operational efficiency.

Users frequently encounter training failures or performance degradation due to inadequate memory. Consequently, ensuring sufficient memory to accommodate the processing of large datasets and the training of complex models is crucial. When calculating complexity, precise figures for the number of runs or additional run space are not necessary; rather, a number that significantly influences the overall complexity calculation suffices.

In terms of space complexity, the number of outer loops is  $n$ , where  $n$  is the length of the whole data set minus the length of the training set.  $k$  is the number of  $IMF(t)$  obtained by each decomposition; Then the set used to store all  $IMF(t)$  results is  $n \cdot k$ . On each outer loop, the generated  $IMF(t)$  gradually copies the data into memory, marking the cost of an outer loop as  $O(i)$ ,  $i$  is used to temporarily store data set length. Each



**Fig. 7.** Scatter plot of errors between Rolling ICEEMDAN-Methods and Rolling Methods.



**Fig. 8.** Rose diagram of evaluation metrics.

**Table 5**  
Results of Model evaluation metrics.

Methods	Metrics					
	RMSE	MAPE	nRMSE	D-index	D-stat	Var( e )
Rolling-ANN	18.9737	0.0207	0.0282	0.9876	0.8639	160.2417
Rolling-ICEEMDAN-ANN	15.8237	0.0177	0.0235	0.9915	0.7687	107.7367
Rolling-CNN	22.7139	0.0245	0.0338	0.9816	0.8163	234.1333
Rolling-ICEEMDAN-CNN	17.1092	0.0194	0.0254	0.9899	0.7959	124.2190
Rolling-LSTM	23.5794	0.0262	0.0351	0.9808	0.8095	229.3436
Rolling-ICEEMDAN-LSTM	14.0975	0.0161	0.0201	0.9931	0.8163	<b>80.8379</b>
Rolling-GRU	23.0382	0.0253	0.0343	0.9814	0.8163	225.4400
Rolling-ICEEMDAN-GRU	20.6723	0.0230	0.0307	0.9851	0.7959	183.3188
Rolling-LightGBM	24.7948	0.0194	0.0366	0.9777	0.8503	426.4183
Rolling-ICEEMDAN-LightGBM	19.4100	0.0190	0.0286	0.9871	0.8027	204.9600
Rolling-ERT	13.5342	0.0107	0.0201	0.9937	0.8844	128.7089
Rolling-ICEEMDAN-ERT	<b>10.7751</b>	<b>0.0085</b>	<b>0.0160</b>	<b>0.9960</b>	<b>0.9116</b>	82.2321

Note: Bold black indicates the best results of a certain model for that metric.

**Table 6**  
Comparison of running time of Rolling ICEEMDAN-Methods series models.

Models	Rolling ICEEMDAN-ANN	Rolling ICEEMDAN-CNN	Rolling ICEEMDAN-LSTM	Rolling ICEEMDAN-GRU	Rolling ICEEMDAN-LightGBM	Rolling ICEEMDAN-ERT
Running Time	53.38min	68.2min	100.1min	90.47min	39s	4min

**Table 7**  
Correspondence of independent variables for Model A to Model I.

Model	Influencing Factors							
	Attention	Alternative Energy	Supply and Demand Factors	Transportation and Dispatch	International Coal Price Level	Commodity Currency Factors	Macro economic Factors	Historical Data of Coal Prices
Model A	✓	✓	✓	✓	✓	✓	✓	✓
Model B		✓	✓	✓	✓	✓	✓	✓
Model C	✓		✓	✓	✓	✓	✓	✓
Model D	✓	✓		✓	✓	✓	✓	✓
Model E	✓	✓	✓		✓	✓	✓	✓
Model F	✓	✓	✓	✓		✓	✓	✓
Model G	✓	✓	✓	✓	✓		✓	✓
Model H	✓	✓	✓	✓	✓	✓		✓
Model I	✓	✓	✓	✓	✓	✓	✓	

outer loop stores  $k$   $IMF(t)$  results, denoted by  $O(k)$ . Training set features (model parameters) also take up some space, set to  $O(p)$ , where  $p$  is the number of parameters in the model. In summary, the space complexity is  $O(n \cdot k + i + k + p)$ . And the complexity of Rolling ICEEMDAN-Methods is directly related to the parameters  $p$ . The configuration used in this study is RTX 4090D, and its performance can support the construction and training of the described model. Of course, higher specifications of the hardware configuration will further improve the computational efficiency, significantly reducing the training time of the model.

#### 4.5.2. Comparative experiment II

To investigate the contributions of various factors influencing the prediction performance of Chinese thermal coal prices, Rolling ICEEMDAN-ERT was taken as an example, and one of the eight influencing factors was removed individually for modelling. Finally, the importance of each factor was ranked based on the results. Table 7 presents the independent variables used in the model.

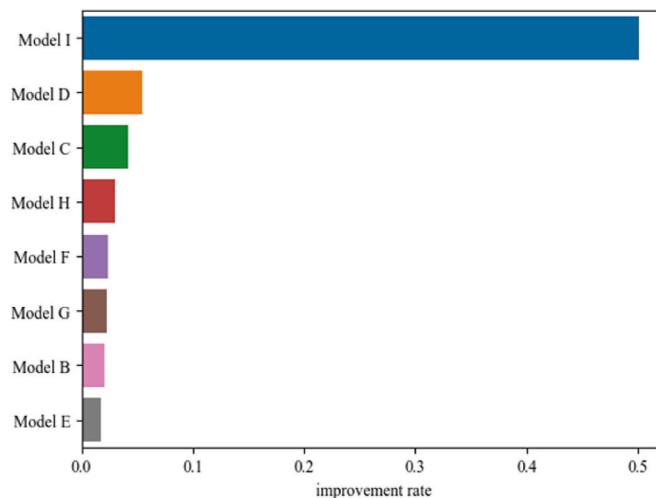
According to Table 8, the accuracy and stability of Models B–I are inferior to those of Model A. Model I performs the worst, indicating that

**Table 8**  
Evaluation metrics results for Model A to Model I.

Model	Metrics					
	RMSE	MAPE	nRMSE	D-index	D-stat	Var( e )
Model A	<b>10.7751</b>	<b>0.0085</b>	<b>0.0160</b>	<b>0.9960</b>	<b>0.9116</b>	<b>82.2321</b>
Model B	10.9949	0.0087	0.0164	0.9958	0.9048	84.9696
Model C	11.2254	0.0086	0.0167	0.9957	0.9048	91.2788
Model D	11.3624	0.0090	0.0169	0.9955	0.8980	91.0276
Model E	10.9653	0.0087	0.0163	0.9959	0.8980	84.3036
Model F	11.0294	0.0089	0.0164	0.9958	0.8707	84.4166
Model G	11.0152	0.0086	0.0164	0.9958	0.8844	86.9618
Model H	11.0970	0.0090	0.0160	0.9957	0.8911	83.3782
Model I	16.1737	0.0142	0.0241	0.9909	0.8367	167.1890

the inclusion of the previous coal price variable significantly improves the model's performance. This variable plays an important role in predicting thermal coal prices in China.

Next, we ranked the importance of the influencing factors by



**Fig. 9.** Ranking the importance of influencing factors.

measuring the improvement rate of the RMSE from Models A to B to Model I. This helps us understand the contributions of the eight influencing factors to the prediction of thermal coal prices. As shown in Fig. 9, the coal-lagged price is the most important factor, as its inclusion leads to a 50.45% reduction in RMSE. Supply and demand factors ranked second, followed by alternative energy, macroeconomic factors, international coal price levels, commodity currency factors, and attention. The contributions of transportation and dispatch were the smallest; however, their inclusion improved the accuracy of the model, validating the rationality of the variables used in this study.

Model B included all influencing factors, except for attention. From Fig. 10, it can be observed that Models A and B yield good results. However, when the BSPI exhibits significant fluctuations, Model A performs better. Compared with Model B, Model A showed an improvement in predictive performance, with a decrease of 2.04% in RMSE, a decrease of 2.74% in the variance of absolute errors, and an increase of 0.75% in the direction statistic. This indicates that Model A can capture useful information, highlighting the superiority of attention in coal price prediction models. Although attention was not the most

important factor in ranking importance, it is clear that variables of this nature can enhance the predictive performance of the model.

#### 4.6. Significance test of Model differences

To verify whether there was a significant difference in performance between the Rolling ICEEMDAN-Methods and Rolling Methods series models, Kruskal-Wallis and Dunn tests were conducted on the two groups.

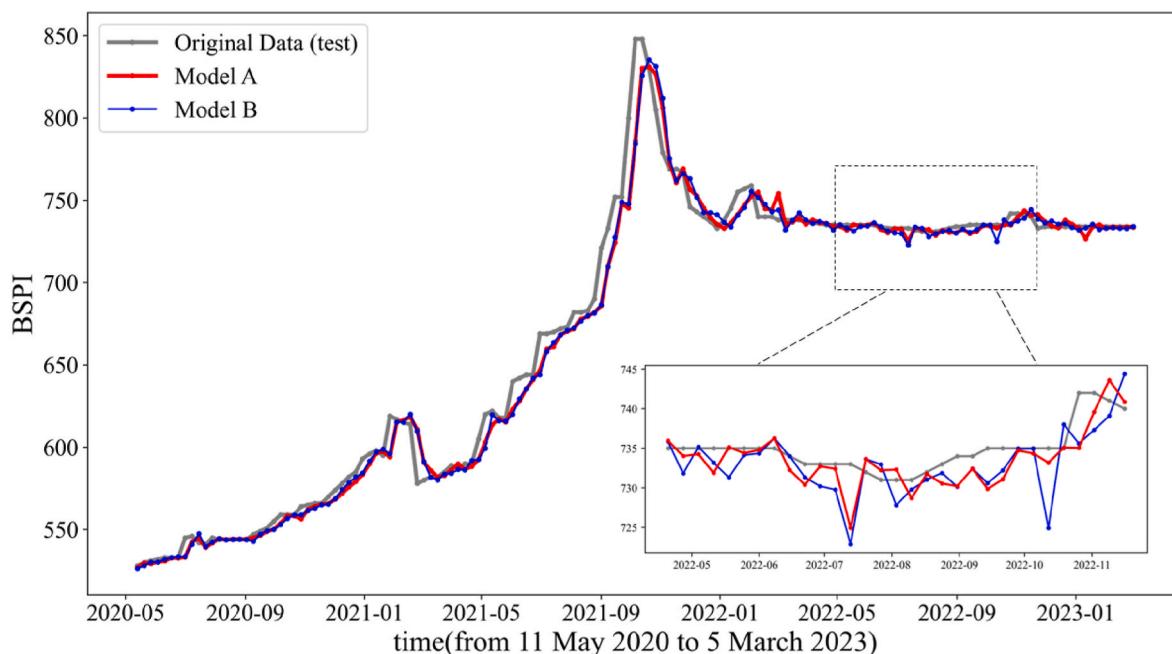
First, the central tendencies and dispersion ranges of the absolute errors of the models were explored. As shown in Figs. 11 and 12, the box sizes and positions of the Rolling ICEEMDAN-Methods and Rolling Methods models are different, and regardless of whether decomposition is performed, the median absolute errors of the ERT method are the smallest.

For the absolute errors of the Rolling ICEEMDAN-Methods models, a Kruskal-Wallis test was conducted, and the results are shown in Table 9. The test yielded a  $p$ -value of 0.000, which was less than the significance level  $\alpha = 0.05$ , indicating a significant difference in the predictive performance among the Rolling ICEEMDAN-method models. Further analysis using Dunn's test was performed to determine significant model differences. As shown in Table 10, significant differences were observed between Rolling ICEEMDAN-ERT and the other models, confirming the superiority of this method.

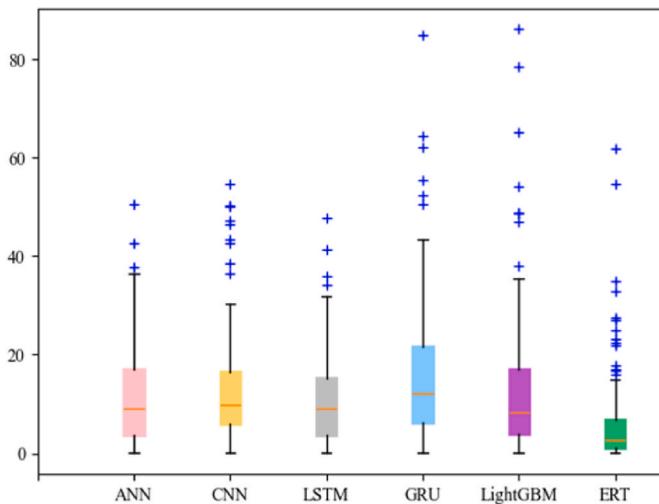
Similarly, a Kruskal-Wallis test was conducted for the Rolling Methods models, and the results are presented in Table 11. The test yielded a  $p$ -value of 0.000, indicating a significant difference in the results between the Rolling Methods models. As shown in Table 12, significant differences were observed between the tree-based and neural network models. In particular, Rolling ERT exhibited significant differences from the other models and demonstrated the lowest relative error, indicating its superiority, even without decomposition.

#### 4.7. The effect of different components on the predicted performance

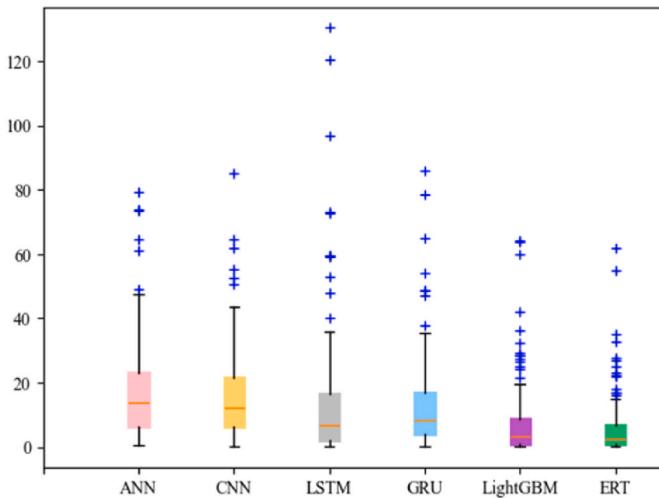
In this section, the  $IMFs(t)$  are categorized into high-frequency, low-frequency, and residual components, with an exploration of their respective contributions to the final forecasting outcomes. Furthermore, these three components possess significant economic implications and



**Fig. 10.** Comparison of results between Model A and Model B.



**Fig. 11.** Boxplot of absolute errors for rolling ICEEMDAN-Methods.



**Fig. 12.** Boxplot of absolute errors for rolling methods.

**Table 9**  
Kruskal-Wallis test results for the rolling ICEEMDAN-Methods models.

Analysis Item	Model	Median	Standard Deviation	Statistic	p-value
Absolute Error	Rolling ICEEMDAN-ANN	9.082	10.38	105.216	0.000***
	Rolling ICEEMDAN-CNN	9.761	11.145		
	Rolling ICEEMDAN-LSTM	9	8.991		
	Rolling ICEEMDAN-GRU	12.072	13.54		
	Rolling ICEEMDAN-LightGBM	8.452	14.316		
	Rolling ICEEMDAN-ERT	2.627	9.068		
	Sum	8.505	11.784		

**Table 10**  
Dunn Test p-values for Rolling ICEEMDAN-Methods Models.

Methods	ANN	CNN	LSTM	GRU	LightGBM	ERT
ANN	1	1	1	0.067	1	0
CNN	1	1	0.875	0.719	1	0
LSTM	1	0.875	1	0.020	1	0
GRU	0.067	0.719	0.020	1	0.067	0
LightGBM	1	1	1	0.067	1	0
ERT	0	0	0	0	0	1

**Table 11**  
Kruskal-Wallis test results for the rolling methods models.

Analysis Item	Model	Median	Standard Deviation	Statistic	p-value
Absolute Error	Rolling ANN	10.488	12.659	138.831	0.000***
	Rolling CNN	12.055	15.301		
	Rolling LSTM	18.227	16.608		
	Rolling GRU	13.747	15.015		
	Rolling LightGBM	6.653	20.65		
	Rolling ERT	3.19	11.345		
	Sum	10.215	16.168		

**Table 12**  
Dunn Test p-values for Rolling Methods model.

Methods	ANN	CNN	LSTM	GRU	LightGBM	ERT
ANN	1	0.447	0	0.139	0.062	0
CNN	0.447	1	0.010	0.447	0.001	0
LSTM	0	0.010	1	0.074	0	0
GRU	0.139	0.447	0.074	1	0	0
LightGBM	0.062	0.001	0	0	1	0.001
ERT	0	0	0	0	0.001	1

have the potential to uncover novel characteristics of coal prices.

Zhang et al. (2008) pointed out that EMD can serve as a filter for separating high-frequency (fluctuating) and low-frequency (slow) modes. The  $IMFs(t)$  undergo reconstruction from fine to coarse scales, involving high-pass filtering those transitions from rapid oscillations ( $IMFs(t)$  with smaller indices) to slower ones ( $IMFs(t)$  with larger indices). The specific algorithm is outlined as follows.

- (1) Calculate the mean of the sum of the components  $IMF_1(t)$  through  $IMF_k(t)$  (except  $r(t)$ );
- (2) Use the t-test to determine which  $k$ 's mean deviates significantly from zero;
- (3) Upon identifying  $k$  as a significant change point, the partial reconstruction spanning from this point to the endpoint is classified as a slow-varying component, while the partial reconstruction employing the remaining components is designated as a fluctuating process.

$IMFs(t)$  are categorized according to the above algorithm, and  $IMF_k^f(t)$  is reconstructed according to this grouping method. Each decomposition is grouped and the result on the test set is presented in Fig. 13.

As evident from Fig. 13, the  $r(t)$  fluctuated around its long-term mean. Therefore, it can be regarded as a long-term trend in the evolution of coal prices. Historically, while coal prices have fluctuated wildly in response to major events, they have eventually returned to the general trend. High-frequency components have the characteristics of small amplitude, including the impact of short-term market fluctuations. Each sharp rise or fall in the low-frequency component corresponds to an important event, and it should represent the impact of those events (Zhang et al., 2008). According to the calculation, the average

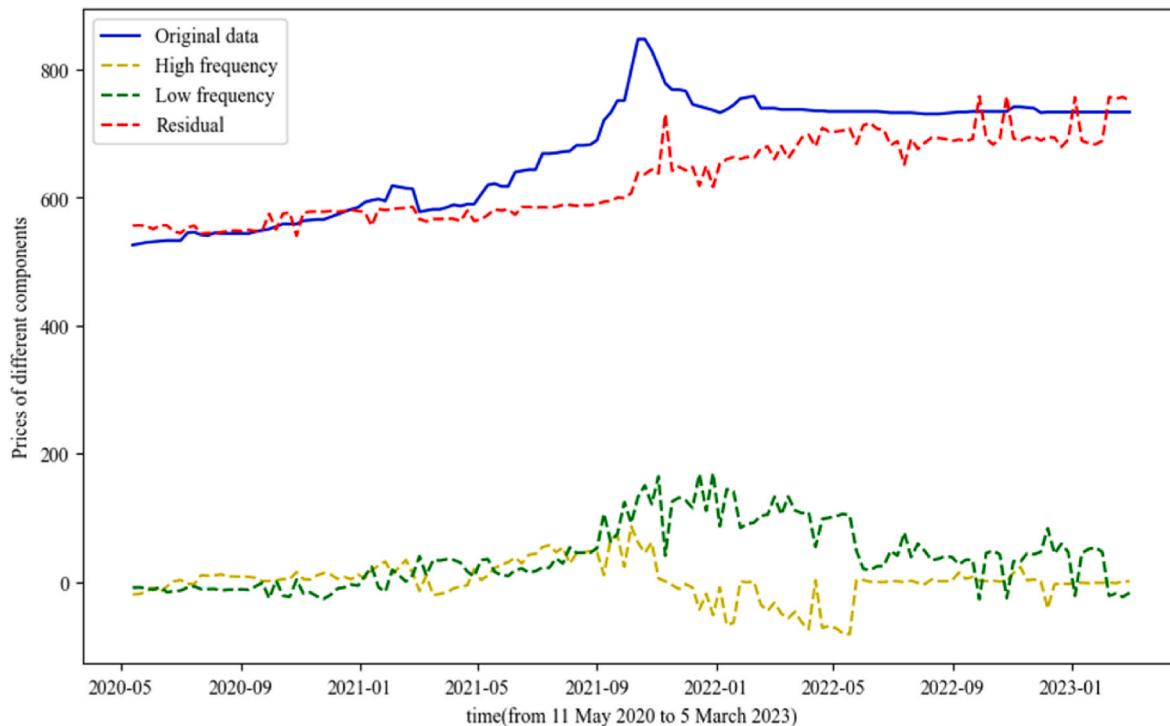


Fig. 13. Grouping results of Rolling ICEEMDAN-Methods.

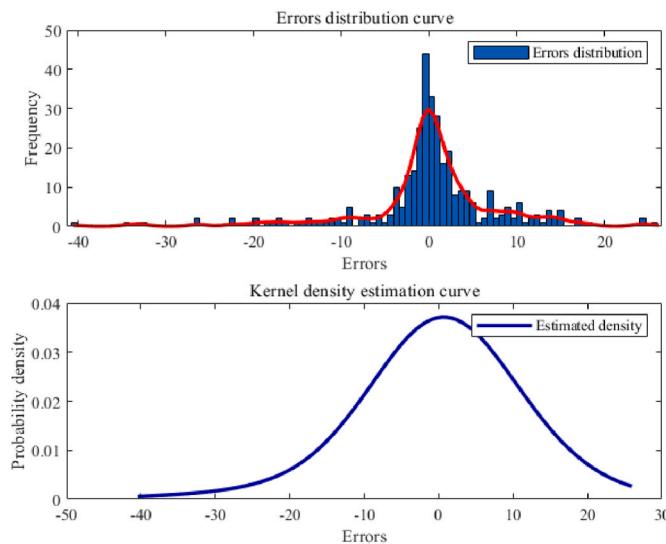


Fig. 14. Distribution of errors.

contribution rates of the high frequency term, low frequency term and residual term to the prediction results are 0.302%, 5.278% and 94.420%, respectively. It is evident that within the combined prediction model utilizing ICEEMDAN decomposition technology, the residual term makes the greatest contribution to the prediction. Despite the relatively minor contributions of the high-frequency and low-frequency terms to the predicted value, they remain integral components.

#### 4.8. Uncertainty estimation

Although the Rolling ICEEMDAN-ERT has achieved high accuracy on the test set of the BSPI, coal prices are closely related to political, economic and other factors, with great uncertainty. Kernel Density Estimation (KDE) serves as an effective interval analysis method (Wuc et al.,

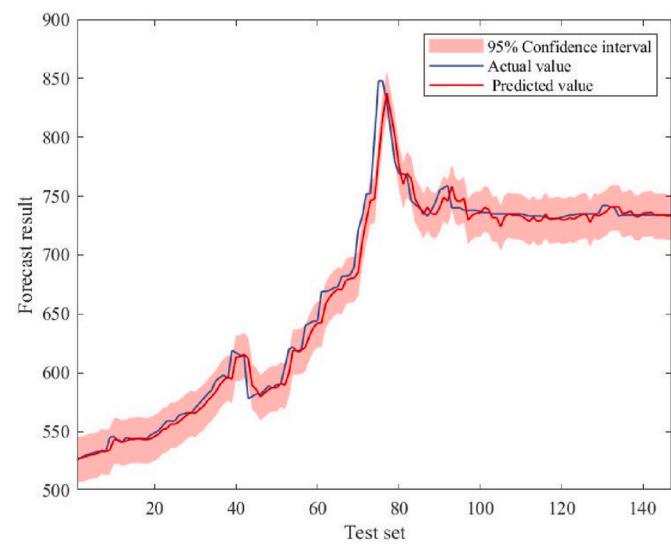


Fig. 15. Results of interval estimation based on KDE.

2023). It does not require any additional assumptions about the sample distribution. Illustrated in Fig. 14 are the error distribution curve and the kernel density estimation curve for the test set, and the interval estimation results of Rolling ICEEMDAN-ERT model on the test set are shown in Fig. 15.

Then, we performed interval estimation on the point prediction

**Table 13**  
Values of interval estimation indicators.

Confidence Level	PICP	PINAW
95%	0.918	0.112
90%	0.884	0.090
85%	0.871	0.071

results at 95%, 90% and 85% confidence intervals, and used *PICP* and *PINAW* to evaluate the performance of the model, as shown in [Table 13](#). The *PICP* of Rolling ICEEMDAN-ERT model at 95% confidence level is 0.918, which is higher than 90% and 85% confidence level, indicating the validity of interval estimation.

## 5. Conclusion

Coal is a crucial energy source that has supported China's long-term economic and social stability. Despite efforts to transition away from coal, owing to the carbon neutrality goals, it will remain significant for energy security and economic development. Therefore, the accurate forecasting of coal prices is essential. This study analyses periods of significant fluctuations in the Bohai-Rim Steam-Coal Price index in China and identifies 27 key factors influencing coal prices, including alternative energy sources, supply and demand dynamics, transportation, macroeconomic factors, international coal price levels, public attention, and historical trends. To establish high-performance prediction models, this study adopts the "divide and conquer" approach and proposes the Rolling ICEEMDAN-Methods series models to predict the weekly average price index of BSPI. These Methods include ANN, CNN, LSTM, GRU, LightGBM, and ERT.

Specifically, ICEEMDAN is used to decompose nonlinear and non-stationary price series into several subseries, and multiple submodels are constructed. The results of the submodels were aggregated to obtain the predicted coal prices for the next time step. The rolling window approach serves two purposes: preventing information leakage by dividing the training and testing sets before decomposition, ensuring the feasibility of the models, adapting to boundary effects, and providing prediction accuracy. To highlight the superiority of the proposed models, a comparison was made with models that did not include ICEEMDAN, and the results demonstrated that the Rolling ICEEMDAN-Methods series models exhibited higher accuracy and stability. To validate the importance of the selected variables, each influencing factor was individually removed, and modelling was conducted to determine the contributions of the novel variables, including attention level and eight influencing factors, to the improvement of prediction accuracy. The importance ranking of each factor was provided.

This study provides a practical and effective approach for short-term forecasting of thermal coal prices in China. The Bohai-Rim Steam-Coal Price Index (BSPI), renowned as a "coal price barometer", has been shown to significantly influence the prediction of thermal coal prices, especially during periods of market volatility and evident demand fluctuations. When conducting analyses and making decisions in the coal-trading market, early-stage coal prices should be considered crucial reference factors. This study offers methodological support for predicting early-stage coal prices.

Supply and demand should not be overlooked as influencing factors. The nonlinearity and nonstationarity of coal prices indicate that achieving a short-term supply-demand balance in the Chinese coal market is challenging. Furthermore, owing to energy conservation, emission reduction, and environmental protection policies, coal enterprises face significant pressure, and coal prices are difficult to decrease. The government's "visible hand" may help stabilise coal prices to some extent.

In addition, this study introduced a novel big data source, Google Trends, as one of the influencing factors. The empirical results demonstrated its superiority in predicting thermal coal prices. Governments can use Google Trends to understand public attention and responses to thermal coal, enabling them to make more informed and rational economic decisions. Moreover, businesses can leverage these data to gain insights into consumer interests and demand for thermal coal, thereby formulating more precise marketing strategies.

In summary, the establishment of a coal price forecasting model in China holds considerable practical and applied value. For instance, taking into account operational expenses, some research prefers tree

models when employing window-rolling techniques for prediction. In the absence of unforeseen circumstances, power generation enterprises can adjust their fuel procurement strategies in accordance with the outcomes of short-term coal price forecasts so as to reduce power generation costs and improves economic benefits. Coal production enterprises can revise their production plans based on the predictions generated by the short-term forecast model, thereby minimizing economic losses resulting from overcapacity or undersupply. Coal traders can rationally allocate their purchasing and sales channels, optimize their inventory management, and augment the efficiency of capital utilization. Financial market investors can utilize the coal price short-term forecast model to examine trends in the coal market, futures, options, and other financial product transactions, thereby ensuring investment returns. The government can employ the coal price short-term forecast model to evaluate the supply and demand dynamics within the coal market. The short-term forecast results of this model can serve as a reference for the formulation of prudent energy policies and price adjustment measures. Furthermore, coal plays a pivotal role in China's energy mix, and research on its price can also offer insights into other energy sources (such as oil, natural gas, etc.), thereby facilitating the holistic planning of China's energy market.

## CRediT authorship contribution statement

**Qihui Shao:** Writing - original draft, Data curation, Software, Visualization. **Yongqiang Du:** Methodology, Writing – original draft, Supervision. **Wenxuan Xue:** Data curation, Software. **Zhiyuan Yang:** Software, Visualization. **Zhenxin Jia:** Data curation, Software. **Xianzhu Shao:** Software, Validation. **Xue Xu:** Project administration, Supervision. **Hongbo Duan:** Project administration, Writing – review & editing. **Zhipeng Zhu:** Software, Validation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## References

- Aggarwal, D., Chandrasekaran, S., Annamalai, B., 2020. A complete empirical ensemble mode decomposition and support vector machine-based approach to predict bitcoin prices. *Journal of Behavioral and Experimental Finance*, 100335.
- Ahmed, F., Chen, W., 2023. Investigation of steam ejector parameters under three optimization algorithm using ANN. *Appl. Therm. Eng.* 225, 120205.
- Alameer, Z., Fathalla, A., Li, K., Ye, H., Jianhua, Z., 2020. Multistep-ahead forecasting of coal prices using a hybrid deep learning model. *Ressour. Pol.* 65, 101588.
- Borjigin, S., Yang, Y., Yang, X., Sun, L., 2018. Econometric testing on linear and nonlinear dynamic relation between stock prices and macroeconomy in China. *Phys. Stat. Mech. Appl.* 493, 107–115.
- Campbell, D.L., 2020. Relative prices and hysteresis: evidence from US manufacturing. *Eur. Econ. Rev.* 129, 103474.
- Chen, Y.C., Rogoff, K.S., Rossi, B., 2010. Can exchange rates forecast commodity prices? *Q. J. Econ.* 125 (3), 1145–1194.
- Colominas, M.A., Schlotthauer, G., Torres, M.E., 2014. Improved complete ensemble EMD: a suitable tool for biomedical signal processing. *Biomed. Signal Process Control* 14, 19–29.
- Corey, G., R., 2013. The future of coal and electric power. *Electrical Engineering* 78 (3), 204–207.
- De Myttenaere, A., Golden, B., Le Grand, B., Rossi, F., 2016. Mean absolute percentage error for regression models. *Neurocomputing* 192, 38–48.
- Dettori, M., Cesariaccio, C., Motroni, A., Spano, D., Duce, P., 2011. Using CERES-Wheat to simulate durum wheat production and phenology in Southern Sardinia, Italy. *Field Crops Res.* 120 (1), 179–188.
- Ding, Z., Feng, C., Liu, Z., Wang, G., He, L., Liu, M., 2017. Coal price fluctuation mechanism in China based on system dynamics model. *Nat. Hazards* 85, 1151–1167.
- Ding, C.D., Wen, Z., 2022. China's green deal: can China's cement industry achieve carbon neutral emissions by 2060? *Renew. Sustain. Energy Rev.* 155, 111931.

- Ding, L., Zhao, Z., Han, M., 2021. Probability density forecasts for steam coal prices in China: the role of high-frequency factors. *Energy* 220 (1), 119758.
- Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T., 2015. Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2625–2634.
- Du, P., Wang, J., Hao, Y., Niu, T., Yang, W., 2020. A novel hybrid model based on multi-objective Harris hawks optimization algorithm for daily PM<sub>2.5</sub> and PM<sub>10</sub> forecasting. *Appl. Soft Comput.* 96, 106620.
- Ekananda, M., 2022. Role of macroeconomic determinants on the natural resource commodity prices: Indonesia futures volatility. *Resour. Pol.* 78, 102815.
- Fantini, D.G., Silva, R.N., Siqueira, M.B.B., Pinto, M.S.S., Guimaraes, M., Brasil, A.C.P., 2024. Wind speed short-term prediction using recurrent neural network GRU model and stationary wavelet transform GRU hybrid model. *Energy Convers. Manag.* 308, 118333.
- Farajji, M., Nadi, S., Ghaffarpasand, O., Homayoni, S., Downey, K., 2022. An integrated 3D CNN-GRU deep learning method for short-term prediction of PM<sub>2.5</sub> concentration in urban environment. *Sci. Total Environ.* 834, 155324.
- Gil-Alana, L.A., Martin-Valmayor, M., Wanke, P., 2020. The relationship between energy consumption and prices: evidence from futures and spot markets in Spain and Portugal. *Energy Strategy Rev.* 31, 100522.
- Guo, J., Zhao, Z., Sun, J., Sun, S., 2022. Multi-perspective crude oil price forecasting with a new decomposition-ensemble framework. *Resour. Pol.* 77, 102737.
- Guo, Z., Yang, C., Wang, D., Liu, H., 2023. A novel deep learning model integrating CNN and GRU to predict particulate matter concentrations. *Process Saf. Environ. Protect.* 604–613.
- Hema, M., Toghraili, D., Amoozad, F., 2023. Prediction of viscosity of MWCNT-Al2O3 (20: 80)/SAE40 nano-lubricant using multi-layer artificial neural network (MLP-ANN) modeling. *Eng. Appl. Artif. Intell.* 121, 105948.
- Herrera, G.P., Constantino, M., Tabak, B.M., Pistori, H., Su, J.J., Naranpanawa, A., 2019. Long-term forecast of energy commodities price using machine learning. *Energy* 179, 214–221.
- Hu, M., Li, H., Song, H., Li, X., Law, R., 2022. Tourism demand forecasting using tourist-generated online review data. *Tourism Manag.* 90, 104490.
- Huang, Y., Hu, Y.F., Wu, L.Z., Wen, S.Y., Wan, Z.D., 2024. Price prediction of power transformer materials based on CEEMD and GRU. *Global Energy Interconnection* 217–227.
- Huawei, T., 2022. Does gross domestic product, inflation, total investment, and exchanges rate matter in natural resources commodity prices volatility. *Resour. Pol.* 79, 103013.
- Kong, D., Yang, X., Xu, J., 2020. Energy price and cost induced innovation: evidence from China. *Energy* 192, 116586.
- Krzemieñ, A., Fernández, P.R., Sánchez, A.S., Lasheras, F.S., 2015. Forecasting European thermal coal spot prices. *Journal of Sustainable Mining* 14 (4), 203–210.
- Li, J., Xie, C., Long, H., 2019. The roles of inter-fuel substitution and inter-market contagion in driving energy prices: evidences from China's coal market. *Energy Econ.* 84, 104525.
- Li, F., Li, X., Zheng, H., Yang, F., Dang, R., 2021. How alternative energy competition shocks natural gas development in China: a novel time series analysis approach. *Resources Policy* 74.
- Li, Z.Z., Su, C.W., Chang, T., Lobont, O.R., 2022. Policy-driven or market-driven? Evidence from steam coal price bubbles in China. *Resour. Pol.* 78, 102878.
- Lin, B., Shi, L., 2022. New understanding of power generation structure transformation, based on a machine learning predictive model. *Sustain. Energy Technol. Assessments* 51, 101962.
- Lin, B., Wang, C., 2021. Impacts of coal prices on the performance of Chinese financial institutions: does electricity consumption matter? *Int. Rev. Econ. Finance* 76, 884–896.
- Liu, X., Li, S., 2021. A comparison analysis of the decoupling carbon emissions from economic growth in three industries of Heilongjiang province in China. *Environ. Sci. Pollut. Control Ser.* 28, 65200–65215.
- Lu, W., Li, J., Li, Y., Sun, A., Wang, J., 2020. A CNN-LSTM-based model to forecast stock prices. *Complexity* 2020, 1–10.
- Lv, J., Cao, M., Wu, K., Li, H., Mohi-Ud-Din, G., 2020. Price volatility in the carbon market in China. *J. Clean. Prod.* 255, 120171.
- Matyjaszek, M., Fernández, P.R., Krzemieñ, A., Wodarski, K., Valverde, G.F., 2019. Forecasting coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory. *Resour. Pol.* 61, 283–292.
- Melas, K.D., Michail, N.A., 2021. The relationship between commodity prices and freight rates in the dry bulk shipping segment: a threshold regression approach. *Maritime Transport Research* 2, 100025.
- Ming, Z., Shulei, L., Song, X., Xiaoli, Z., Lingyun, L., 2016. Prediction of China's coal price during twelfth five-year plan period. *Energy Sources B Energy Econ. Plann.* 11 (1–6), 511–517.
- Plante, M., Dhaliwal, N., 2017. Inventory shocks and the oil–ethanol–grain price nexus. *Econ. Lett.* 156, 58–60.
- Ren, J., Yu, Z., Gao, G., Yu, G., Yu, J., 2022. A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism. *Energy Rep.* 8, 437–443.
- Sharma, D., Kumar, R., Jain, A., 2022. Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. *Measurement: Sensors* 24, 100560.
- Teng, M., Li, S., Xing, J., Song, G., Yang, J., Dong, J., et al., 2022. 24-hour prediction of pm2.5 concentrations by combining empirical mode decomposition and bidirectional long short-term memory neural network. *Sci. Total Environ.* 821, 153276.
- Tong, M., Dong, J., Luo, X., Yin, D., Duan, H., 2022. Coal consumption forecasting using an optimized grey model: the case of the world's top three coal consumers. *Energy* 242.
- Volpe, R., Leibtag, E., Roeger, E., 2013. How transportation costs affect fresh fruit and vegetable prices. *Economic Research Report*.
- Wang, Chen, Lin, Mu, Miao, Hongchao, Shang, Yan, Yin, Hongchao, Dong, Ming, 2023. An innovative application of machine learning in prediction of the syngas properties of biomass chemical looping gasification based on extra trees regression algorithm. *Energy* 127438.
- Willmott, C.J., 1981. On the validation of models. *Phys. Geogr.* 2 (2), 184–194.
- Wu, Z., Huang, N.E., 2004. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Centre for Ocean-Land-Atmosphere Studies. Technical Report 193, 51.
- Wu, Y.X., Wu, Q.B., Zhu, J.Q., 2019. Improved EEMD-based crude oil price forecasting using LSTM networks. *Phys. Stat. Mech. Appl.* 516, 114–124.
- Wu, B., Wang, L., Lv, S.X., Zeng, Y.R., 2021. Effective crude oil price forecasting using new text-based and big-data-driven model. *Measurement* 168, 108468.
- Wu, S., Xia, G., Liu, L., 2023. A novel decomposition integration model for power coal price forecasting. *Resour. Pol.* 80, 103259.
- Wu, J., Li, M., Zhao, E., Sun, S., Wang, S., 2023. Can multi-source heterogeneous data improve the forecasting performance of tourist arrivals amid COVID-19? Mixed-data sampling approach. *Tourism Manag.* 98, 104759.
- Wu, J., Dong, J., Wang, Z., Hu, Y., Dou, W., 2023. A novel hybrid model based on deep learning and error correction for crude oil futures prices forecast. *Resour. Pol.* 83, 103602.
- Xiao, Y.J., Wang, X.K., Wang, J.Q., Zhang, H.Y., 2021. An adaptive decomposition and ensemble model for short-term air pollutant concentration forecast using iceemdanica. *Technol. Forecast. Soc. Change* 166, 120655.
- Xu, X., Zhang, Y., 2022. Thermal coal price forecasting via the neural network. *Intelligent Systems with Applications* 14, 200084.
- Yang, C.J., Xuan, X., Jackson, R.B., 2012. China's coal price disturbances: Observations, explanations, and implications for global energy economies. *Energy Pol.* 51, 720–727.
- Yu, L., Wang, S., Lai, K.K., 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Econ.* 30 (5), 2623–2635.
- Yuan, J., Jiao, Z., 2023. Faulty feeder detection for single phase-to-ground faults in distribution networks based on patch-to-patch CNN and feeder-to-feeder LSTM. *Int. J. Electr. Power Energy Syst.* 147, 108909.
- Yuan, H., Ma, X., Ma, M., Ma, J., 2024. Hybrid framework combining grey system model with Gaussian process and STL for CO<sub>2</sub> emissions forecasting in developed countries. *Appl. Energy* 360, 122824.
- Yue, H., Worrell, E., Crijns-Graus, W., Zhang, S., 2021. The potential of industrial electricity savings to reduce air pollution from coal-fired power generation in China. *J. Clean. Prod.* 7, 126978.
- Zeng, T., Xu, L., Liu, Y., Liu, R., Luo, Y., Xi, Y., 2024. A hybrid optimization prediction model for pm<sub>2.5</sub> based on vmd and deep learning. *Atmos. Pollut. Res.* 15 (7), 102152.
- Zhang, X., Lai, K.K., Wang, S.Y., 2008. A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Econ.* 30 (3), 905–918.
- Zhang, X., Peng, Y., Zhang, C., Wang, B., 2015. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. *J. Hydrol.* 530, 137–152.
- Zhang, Z., Chen, H., 2022. Dynamic interaction of renewable energy technological innovation, environmental regulation intensity and carbon pressure: evidence from China. *Renew. Energy* 192, 420430.
- Zhang, K., Cao, H., Thé, J., Yu, H., 2022. A hybrid model for multi-step coal price forecasting using decomposition technique and deep learning algorithms. *Appl. Energy* 306, 118011.
- Zhao, Z.Y., Zhu, J., Xia, B., 2016. Multi-fractal fluctuation features of thermal power coal price in China. *Energy* 117 (1), 10–18.
- Zhao, L., Li, Z., Qu, L., Zhang, J., Teng, B., 2023. A hybrid VMD-LSTM/GRU model to predict non-stationary and irregular waves on the east coast of China. *Ocean Engineering* 276, 114136.
- Zhou, F., Huang, Z., Zhang, C., Yan, J., 2022. Carbon price forecasting based on CEEMDAN and LSTM. *Appl. Energy* 311, 118601.
- Zuo, G., Luo, J., Wang, N., Lian, Y., He, X., 2020. Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. *J. Hydrol.* 585, 124776.