ITC 3/55
Information Technology and Control
Vol. 55 / No. 3/ 2026
pp. 1337-1357
DOI 10.5755/j01.itc.55.3.45066

Carbon Market Volatility and Its Response Mechanism Based on CNN-BiLSTM Model and Signal Decomposition

Received 2026/05/20 | Accepted after revision 2026/07/31

HOW TO CITE: Zhao, Y. (2026) Carbon Market Volatility and Its Response Mechanism Based on CNN-BiLSTM Model and Signal Decomposition. *Information Technology and Control*, 55(3), 1336-1356. https://doi.org/10.5755/j01.itc.55.3.45066

Carbon Market Volatility and Its Response Mechanism Based on CNN-BiLSTM Model and Signal Decomposition

Yang Zhao

School of Economy and Trade Management, Tianjin Sino-German University of Applied Sciences, Tianjin, 300350, China; e-mail: yz_yzy1000@163.com

Corresponding author: yz_yzy1000@163.com

The carbon market volatility series exhibits strong nonlinearity, non-stationarity, and noise interference, which limits the accuracy of both traditional econometric models and standalone deep learning approaches. To address this, a hybrid prediction framework is developed. Firstly, an improved bald eagle search algorithm is adopted to optimize the hyperparameters of variational mode decomposition (VMD). Through chaos mapping initialization, dynamic inertia weight adjustment, and Gaussian mutation mechanism, population diversity and global search capability are enhanced, and the optimal number of VMD and penalty factor combination are obtained. Secondly, VMD is utilized to decompose the volatility series into several intrinsic modal components. Finally, bidirectional long short-term memory neural networks, convolutional neural networks, and attention mechanisms are introduced to achieve high-precision prediction of volatility series. The results using the EU Emissions Trading System as an empirical object show that traditional generalized autoregressive conditional heteroskedasticity (GARCH) models have limited performance in predicting carbon futures volatility. The optimal GARCH model has an R2 of 92.6779% and an average absolute percentage error of 9.8358%. The suggested model outperforms all traditional econometric models significantly in four evaluation indicators: The average absolute error is reduced by 71.19%, the root mean square error is reduced by 70.13%, the average absolute percentage error is reduced by 63.80%, and the R^2 is enhanced by 6.11%. The designed framework provides a scalable technical path for predicting carbon market volatility and has practical application value for pricing carbon financial derivatives and market risk management.

KEYWORDS: Carbon market, Volatility, Deep learning, Variational mode decomposition, Prediction

1. Background

As global climate governance accelerates, the carbon market – a core policy instrument for achieving emission reduction targets – has drawn growing attention for its operational efficiency and price stability. The European union emissions trading system (EU-ETS), as the world's largest carbon trading market, has a relatively mature mechanism for price discovery and risk transmission [2]. The volatility of the carbon market can be evaluated through volatility, an important indicator for measuring financial asset risk. This indicator can meet economic policy formulation and carbon market expectations, therefore, studying volatility can help deepen our understanding of carbon market volatility [7]. The existing research on carbon market volatility mainly follows two paths. One is the econometric method based on generalized autoregressive conditional heteroskedasticity (GARCH) models. Although it can capture the clustering characteristics of volatility, it is difficult to effectively handle the nonlinearity and non stationarity of sequences [6]. Second, machine learning and deep learning (DL) methods are introduced to advance the prediction accuracy through the adaptive feature extraction ability of neural networks. However, the single model still has the problem of insufficient information extraction when dealing with multi-scale fluctuation patterns [18].

Existing research has systematically explored the prediction and trait mining of carbon market volatility from multiple dimensions, including model optimization, data decomposition, and multi frequency fusion. For example, Zhu et al. [20] utilized wavelet analysis to denoise carbon prices and selected genetic algorithm (GA) to optimize extreme gradient boosting (XGBoost) parameters for predicting carbon market volatility. The results implied that the overall prediction performance of this method was good [20]. Chai et al. [3] utilized VMD to analyze the characteristics of carbon pricing in China's carbon trading pilot program. They decomposed the original carbon pricing sequence, reconstructed the subsequence through sample entropy, and optimized the Extreme Learning Machine (ELM) using Particle Swarm Optimization (PSO) for prediction. The results implied that the model had good predictive performance [3]. Xiao and Liu [15] suggested a multi frequency data fusion carbon price prediction model (MFF-CPPM) to tackle the issues of data nonlinearity and inconsistent trait factor frequencies in carbon trading price prediction. The results implied that the model had generalization and robustness [15]. Liu's team constructed an extended model based on the exponential GARCH (EGARCH) model by combining mixed frequency data sampling techniques and time-varying volatility persistence (TVP) to explore the structural changes and time-varying persistence characteristics of carbon market volatility. The results implied that the model's internal fitting and out of sample prediction capabilities were superior to traditional models [9].

DL technology has shown significant advantages in predicting carbon markets and related financial time series, becoming one of the core means to advance prediction accuracy. For example, Sha et al. aimed to advance the management and operational efficiency of the carbon price market by analyzing social media texts through natural language processing technology and inputting the data into a CNN-BiLSTM-Attention model to generate initial predictions. Subsequently, the prediction bias was modeled in the error correction stage, and the results implied that this method has low trading risk and high trading return [11]. Fang et al. [5] utilized complete ensemble empirical mode decomposition (CEEMDAN) and VMD quadratic decomposition to analyze high-frequency components to advance the accuracy of carbon price prediction. They also selected coati optimization algorithm (COA) to optimize the parameters of BiLSTM. The results implied that the prediction error of the model in EU-ETS was only 0.173% [5]. Wang et al. [14] utilized adaptive symmetric geometric mode decomposition to solve the irregularity of carbon prices for carbon trading price prediction, and completed the prediction using BiLSTM optimized by Light Gradient Boosting Machine (LightGBM) and Bald Eagle Search (BES) algorithms. The results implied that this technology was practical [14]. To cope with the increasing volatility and complexity of the carbon trading market, Wang and Li [13] proposed an Random Forest - Long Short-Term Memory (RF-LSTM) hybrid prediction model based on CEEMDAN decomposition and multi-scale entropy recon-

struction. The results showed that this method has superior performance in both prediction accuracy and stability [13]. In response to the shortcomings of existing models in processing multi-source heterogeneous data fusion such as carbon market price data and policy texts, Zhang et al. [19] designed an AutoBERT-MCTS model, in which autoencoders are used to extract market data features, Bidirectional Encod-

Table 1
Comparison between existing carbon market time series forecasting models and the proposed method.

Signal Decomposition Strategy	Parameter Optimization Algorithm	Optimization Objective	Deep Network Architecture	Attention Mechanism	Customized for Carbon Volatility Series	Research Object	Literature
Wavelet decomposition	GA	Prediction error	XGBoost	No	No, generic GA	Carbon futures volatility	Zhu et al. [20]
VMD	PSO	Prediction error	ELM	No	No, generic PSO	China's pilot carbon prices	Chai et al. [3]
Multi-frequency data fusion, no decomposition	No intelligent optimization	Model fitting error	Multilayer Perceptron	No	/	Carbon price series	Xiao & Liu [15]
No signal decomposition	No parameter optimization	Trading loss & return	CNN-BiLSTM	Yes	No optimization algorithm adopted	Carbon prices (integrated social media text)	Sha et al. [11]
CEEMDAN + VMD secondary decomposition	COA	Prediction error	BiLSTM	No	Generic COA without carbon time series adaptation	EU-ETS carbon prices	Fang et al. [5]
Adaptive symmetric geometric mode decomposition	LightGBM + BES	Prediction error	BiLSTM	No	No, generic BES without carbon volatility customization	Carbon trading prices	Wang et al. [14]
CEEMDAN + multi-scale entropy reconstruction	No intelligent optimization	Prediction error	RF + LSTM (rolling framework)	No	No, but designed for carbon price nonlinearity and non-stationarity	Hubei and EU carbon prices	Wang & Li [13]
No decomposition (multi-source data fusion)	MCTS	Prediction error & risk evaluation	AutoEncoder + BERT	Yes (implicit in BERT)	No, but designed for multi-source heterogeneous data (market prices + policy texts)	Carbon market prices	Zhang et al. [19]
VMD optimized adaptively via IBES	IBES (chaotic initialization + dynamic inertia weight + Gaussian mutation)	Envelope entropy (decomposition quality evaluation)	CNN-BiLSTM with multi-head attention	Yes	Yes; three improvements tailored to multi-scale noise and abrupt shocks of carbon volatility	EU-ETS carbon futures volatility	This paper

er Representations from Transformers (BERT) is responsible for processing policy text information, and Monte Carlo Tree Search (MCTS) is used to optimize investment decisions in uncertain environments. The results indicate that the model has certain effectiveness [19]. A comparison of existing time-series forecasting models for carbon markets with the proposed method is provided in Table 1.

In summary, the prediction problems in the carbon market have all adopted optimized DL techniques and achieved significant results. However, existing studies have treated intelligent optimization algorithms merely as general-purpose parameter tuning tools and have not yet designed targeted optimization strategies that incorporate the intrinsic characteristics of carbon volatility, such as nonlinearity, multiscale behavior, and abrupt noise. This has resulted in a disconnect between the optimization mechanism and the temporal dynamics of the carbon market, lacking a theoretical explanation for their coupling. To address this gap, it is first necessary to clarify the typical theoretical characteristics of carbon market volatility. Carbon market volatility is driven by a combination of policy shocks, energy linkages, climate risks, and heterogeneous investor expectations, and exhibits four typical statistical characteristics: volatility clustering, long memory, leverage effects, and multiple fractal properties. Traditional GARCH-type models can only capture volatility clustering and struggle to account for both long memory and multiple fractal characteristics. Although standalone DL methods possess nonlinear fitting capabilities, they cannot explicitly distinguish multiscale volatility structures. The four typical characteristics mentioned above provide clear financial theoretical support for each module of this model: volatility clustering corresponds to CNNs extracting short-term volatility patterns through local window convolution; the long-memory characteristic relies on BiLSTM for bidirectional time-series modeling to capture long-term lag effects; the asymmetric shocks embodied by the leverage effect can be fully identified through bidirectional information flow, which distinguishes the differential transmission of positive and negative information; and the multi-fractal characteristics serve as the core driving force for introducing VMD multi-frequency decomposition, which breaks down the original sequence into intrinsic mode function (IMF) components of different scales with clear economic implications. Based on the above theoretical mapping, this study designs a triple improvement strategy to construct an improved BES (IBES) algorithm, and integrates VMD multi-scale decomposition with CNN-BiLSTM-Attention into a carbon futures volatility prediction framework. The core innovation and main contributions of this study are reflected in the following three aspects:

1. At the algorithm level, the three improvements of IBES (chaos initialization, dynamic inertia weight, Gaussian mutation) are designed for the strong nonlinearity, multi-scale coupling, and abrupt jump characteristics of carbon volatility. They belong to customized improvements for carbon finance time series, rather than simple modifications to general algorithms.
2. On the theoretical level, a one-to-one correspondence between four typical facts of carbon volatility and each module has been established, providing clear economic explanations for the model and enhancing its interpretability.
3. On the empirical level, cross market validation was conducted on three typical carbon markets, and envelope entropy was used as the optimization objective of VMD to fundamentally suppress mode mixing problems.

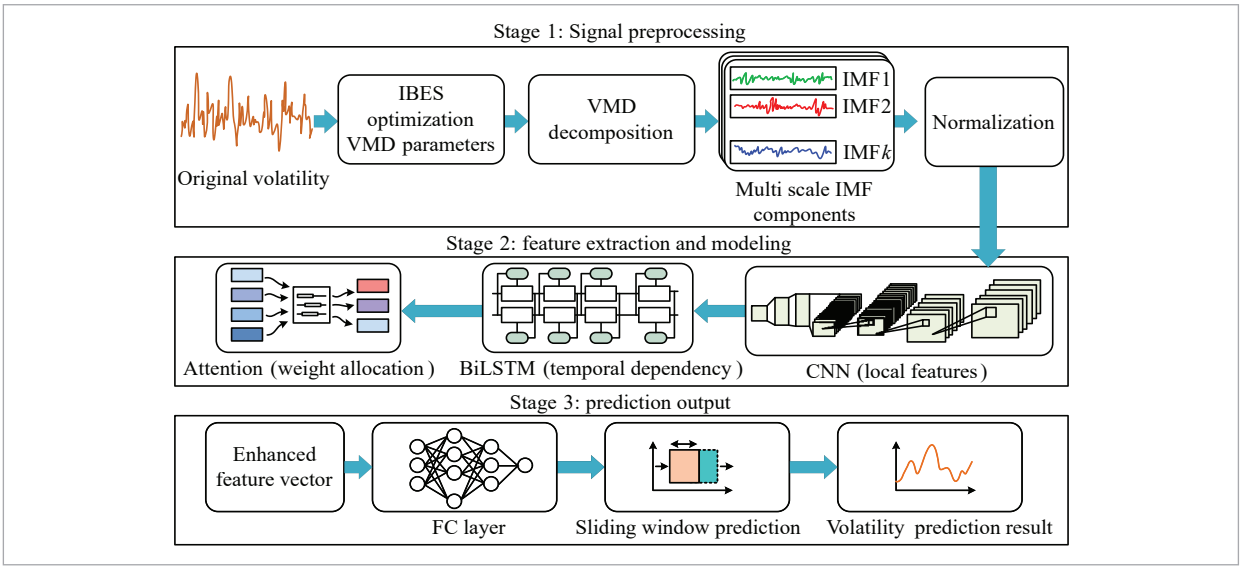
The above customized design distinguishes the framework of this article from the simple combination of existing technologies, reflecting the deep coupling between the model and the inherent characteristics of carbon volatility.

2. Methodology

2.1. Hybrid Framework for Predicting Carbon Futures Volatility

The volatility series of carbon futures exhibits significant nonlinear, non-stationary, and noise interference characteristics due to the comprehensive influence of multiple factors like macroeconomic policies, energy market price linkage, climate and environmental change, and investor expectations [4]. This kind of complex time series data contains fluctuation components of different time scales, and it is difficult for a single model to capture high-frequency noise, intermediate frequency periodic fluctuations and low-frequency trend changes at the same time.

Figure 1
Hybrid framework process for predicting carbon futures volatility.



To tackle this issue, a predictive hybrid framework was developed. This framework consists of three stages, with the specific process shown in Figure 1. Stage one in Figure 1 is signal preprocessing. After inputting the original volatility, IBES is utilized to optimize the VMD parameters, followed by VMD to obtain multi-scale IMF components. Stage two is feature extraction and modeling. The IMF component standardization is processed, then the local traits are captured through CNN and implementing temporal dependencies through BiLSTM, and attention for weight allocation is introduced. Stage three is the prediction output, which enhances the trait vectors to the fully connected (FC) layer and then uses sliding window prediction to obtain the volatility prediction result. These three stages form a progressive information processing chain: signal decomposition separates different frequency components; CNN extracts local waveform patterns; BiLSTM captures bidirectional time dependencies; Attention dynamically refines information features. This orthogonal and cascaded design ensures that each module can address different aspects of carbon fluctuation prediction without functional redundancy.

2.2. Calculation of Volatility Series

After determining the prediction mixing framework, the study first calculates the volatility series. Using the closing price of EU-ETS carbon futures as raw data, the study calculates the volatility series of carbon futures as input for subsequent modeling. Based on the daily closing price of carbon futures, this paper constructs a 22-day rolling low-frequency realized volatility series and converts the raw price data into volatility indicators. The expression for the volatility V_i on day i is displayed in Equation (1) [10].

$$V_i = \sqrt{\frac{1}{m}\sum_{q=i-m+1}^{i}\left(s_q - \overline{s}\right)^2}. \quad (1)$$

In Equation (1), s_q stands for the original logarithmic income sequence. $\overline{s}$ is the logarithmic mean of returns for m days. Referring to the slidin g window setting standards for predicting European Union Allowance (EUA) futures volatility in existing carbon market research, this study sets the rolling window length to m=22 [12]. The rolling low-frequency realized volatility defined by Equation (1) is entirely calculated based on daily logarithmic returns and does not include intraday high-frequency trading data. This estimator differs fundamentally from the classical high-frequency realized volatility constructed using intraday minute or minute level return data in terms of data foundation and information frequency. This study adopts this rolling

standard deviation estimator with the aim of using it as an effective proxy indicator for the low-frequency volatility of EUA futures.

2.3. Design of IBES Optimization VMD Method

2.3.1. VMD Process

Carbon futures volatility data exhibits nonlinear, non-stationary characteristics and is subject to noise interference. As a signal decomposition method, VMD's decomposition effect highly depends on two key hyperparameters: The total number of decomposition modes K and the penalty factor α. The core idea of VMD is to decompose the signal into several IMF with sparse characteristics, each IMF exhibiting a narrowband distribution around its center frequency. The calculation formula for the k-th IMF $q_k(l)$ at time l is displayed in Equation (2) [17].

$$q_k(l) = A_k(l)\cos\left(P_k(l)\right). \tag{2}$$

In Equation (2), k stands for the modal number. $A_k(l)$ and $P_k(l)$ represent the time-varying amplitude and phase of the k-th intrinsic mode. VMD solves all modes simultaneously through non recursive variational optimization, avoiding the error accumulation that may occur in recursive methods in carbon volatility decomposition. To minimize the bandwidth of each modality while constraining the superposition of all modalities to be equal to the original signal, the expression of the variant constraint is constructed as detailed in Equation (3).

$$\begin{cases} \min_{\{q_k\}\{\bar{\omega}_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_l \left[\left(\delta(l) + \dfrac{d}{\pi l} \right) * q_k(l) \right] e^{-d\bar{\omega}_k l} \right\|_2^2 \right\} \\ \text{s.t.} \sum_{k=1}^{K} q_k(l) = f \end{cases}. \tag{3}$$

In Equation (3), $\{q_k\}$ is the set of all mode functions to be optimized, and k=1,2,...,K. K stands for the total number of VMD modes. $\bar{\omega}_k$ stands for the set of center frequencies for all modes to be optimized. ∂_l stands for the partial derivative of time. $\delta(l)$ is the Dirac function. d is the imaginary unit, d^2=–1. $e^{-d\bar{\omega}_k l}$ stands for the complex exponential modulation term. The original input signal is the volatility series of carbon futures to be decomposed. $\|\|_2^2$ is the square of the L2 norm. The bandwidth constraint in Equation (3) forces each mode to be compactly distributed around its center frequency, thereby achieving a clear separation between long-term trends and short-term noise. By using Lagrange multiplier λ and penalty factor α, the constrained variational problem is transformed into an iteratively solvable unconstrained minimization problem. The specific expression of the augmented Lagrangian function $L(\cdot)$ is displayed in Equation (4).

$$\begin{aligned} L\left(\{q_k\},\{\bar{\omega}_k\},\lambda\right) = \alpha \sum_{k=1}^{K} \left\| \partial_l \left[\left(\delta(l) + \frac{d}{\pi l} \right) * q_k(l) \right] e^{-d\bar{\omega}_k l} \right\|_2^2 \\ + \left\| f(l) - \sum_{k=1}^{K} q_k(l) \right\|_2^2 + \left\langle \lambda(l), f(l) - \sum_{k=1}^{K} q_k(l) \right\rangle \end{aligned}. \tag{4}$$

In Equation (4), $\langle\cdot,\cdot\rangle$ is the sign of the inner product operation, corresponding to the product of the Lagrange multiplier and the constraint residual. VMD uses variational constraints to iteratively solve for the center frequencies of each narrowband mode and separates high- and low-frequency components with the goal of minimizing bandwidth, thereby mathematically suppressing modal aliasing in carbon price volatility. The modes obtained through this decomposition correspond to the long-term trend, medium-term cycle, and short-term noise in the carbon market, respectively, enabling the hierarchical separation of multiscale volatility signals.

2.3.2. IBES Algorithm Design

The volatility sequence of carbon futures has three characteristics: the coupling superposition of high and low frequencies makes K and α in VMD highly sensitive to parameter deviations; There are numerous local optima in the envelope entropy objective function under extreme market conditions; Temporal continuity can easily lead to premature convergence of traditional algorithms. The inherent defects of algorithms such as BES, PSO, GA, etc., such as the tendency of randomly initialized populations to aggregate in multimodal spaces, the weakening of fine tuning ability in the later stages of iteration due to fixed inertia weights, and the lack of disturbance mechanisms that make it difficult to escape from local optima, further exacerbate the above problems. To address this, the study adopts

IBES intelligent search for the optimal parameter combination [K,α]. IBES uses chaotic mapping initialization strategy, dynamic inertia weight adjustment strategy, and Gaussian mutation mechanism to jointly advance global search ability, convergence accuracy, and the ability to escape local optima. Below are specific introductions to these designs. The fitness function for carbon fluctuation time series is distributed extremely unevenly in the parameter space. This study employs chaotic sequences that uniformly cover the two-dimensional solution space defined by K and α, thereby preventing the initial population from clustering in local minimum regions and avoiding, at the source, the problem of decomposition failure caused by the aliasing of high- and low-frequency components in the carbon market., this method has stronger chaotic stability and a more uniform sequence distribution, making it better suited for the application scenario of multi-peak optimization in carbon finance time series [1]. The calculation process of the chaotic variable values for the u+1 th iteration of the mapping is displayed in Equation (5).

$$x(u+1)=\begin{cases} x(u)/p_1, & 0\le x(u)<p_1 \\ x(u)-p_1/0.5-p_1, & p_1\le x(u)<0.5 \\ 1-p_1-x(u)/0.5-p_1, & 0.5\le x(u)1-p_1 \\ 1-x(u)/p_1, & 1-p_1\le x(u)<1 \end{cases}. \quad (5)$$

In Equation (5), p_1 stands for the control parameter of the piecewise linear chaotic mapping. Among them, $p_1\in(0,2)$ is set as $p_1=0.5$. The initial chaos value $x(1)$ is generated by a random function to obtain $x(1)=\text{rand}$. During the VMD parameter optimization process, a broad global search is required in the early stages to traverse the range of modal numbers, followed by a refined local search in the later stages to fine-tune the penalty factors [16]. Dynamic weights decay linearly with the number of iterations, effectively balancing the need for early-stage global exploration with the need for late-stage local fine-tuning, precisely aligning with the parameter optimization logic of "broad-range preliminary selection of K and narrow-range fine-tuning of α" in carbon volatility decomposition. Therefore, IBES adopts a dynamic inertia weight adjustment strategy that balances global exploration and local development, and its calculation formula is displayed in Equation (6).

$$w=\begin{cases} w_{\max}, & u<I_{\max}/4 \\ \left(w_{\max}-w_{\min}\right)e^{\frac{I_{\max}}{2}\times\left(\frac{u}{I_{\max}}-\frac{1}{4}\right)^2}, & u\ge I_{\max}/4 \end{cases}. \quad (6)$$

In Equation (6), $w_{\max}$ and $w_{\max}$ are the max and min values of the inertia weight, respectively. $w_{\max}$=0.95 and $w_{\min}$=0.1 are set as the current iteration count. $I_{\max}$ stands for the max number of iterations. After introducing the inertia weight w, the formula for updating the position of the j-th bald eagle during the prey positioning phase in $w_{\min}$ = 0.1 space is displayed in Equation (7).

$$P_{j,\text{new}}=w\times P_j+x(j)\times\left(w\times P_j-P_{\text{mean}}\right)+y(j)\times\left(w\times P_j-P_{j+1}\right). \quad (7)$$

In Equation (7), $P_{j,\text{new}}$ is the updated new position. P_{mean} is the average position of all individuals in the current population. P_{j+1} is the current position of the j+1 th bald eagle. $x(j)$ and $y(j)$ represent a random coefficient, usually a random number in the range of [0,1], utilized to increase search randomness. Due to factors such as policy adjustments and energy price shocks, the carbon market is subject to sudden changes in volatility, resulting in a large number of local optima distributed within the corresponding envelope entropy function [21]. Applying Gaussian variation to the globally optimal individuals can continuously perturb the position of the optimal solution, thereby helping the algorithm escape the local optima traps caused by extreme market samples and ensuring that VMD decomposition effectively separates the three components of the carbon market: long-term trend fluctuations, medium-term cyclical fluctuations, and short-term noise. The global optimal value obtained during the search phase is updated through mutation operation, as detailed in Equation (8).

$$P_{\text{G,new,best}}=P_{\text{best}}+P_{\text{best}}\times\text{G}(\sigma). \quad (8)$$

In Equation (8), $P_{\text{G,new,best}}$ stands for the updated position after Gaussian mutation. P_{best} is the original iteration position. G(σ) stands for a random vector that follows a Gaussian distribution, and σ stands for

the standard deviation (SD) of the Gaussian distribution. The mean of G(σ) is μ and the variance is σ^2. The mathematical expression of the Gaussian distribution that G(σ) follows is displayed in Equation (9).

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{(x-\mu)^2}{2\sigma^2}}. \quad (9)$$

In Equation (9), f(x) stands for the probability density function of Gaussian distribution, describing the probability distribution of the value of the random variable x. exp(·) stands for the natural exponential function.

2.3.3. IBES Optimization of VMD Parameter Process

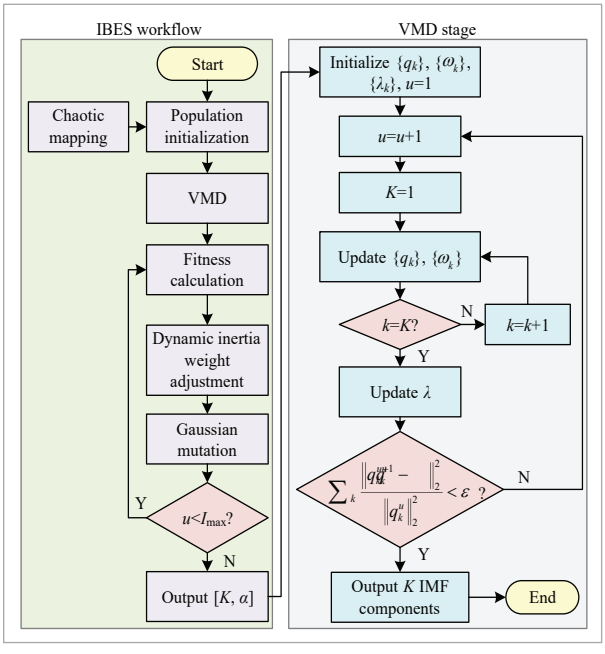
According to the designed IBES, the IBES-VMD workflow is displayed in Figure 2. In Figure 2, for the VMD stage, first the parameters of IBES are initialized, it set $\alpha\in[500,2000]$, $K\in[1,10]$, Dim=2, $I_{\max}=30$, and the population size is set to 50.

After initializing the parameters, the study selects envelope entropy as the fitness function. Envelope entropy is often utilized as a fitness function within the realm of signal processing to evaluate the quality of signal decomposition. The smaller the envelope entropy, the purer the modal components obtained through VMD, the less noise interference it receives, and the better the decomposition effect [8]. The steps of calculating envelope entropy first use Hilbert transform to extract the signal envelope, then standardize the envelope curve and calculate the entropy value V_{ent} of the envelope curve, which is calculated by Equation (10).

$$V_{\text{ent}}=-\sum\nolimits_n\left[v_n\log\left(v_n\right)\right]. \quad (10)$$

In Equation (10), n stands for the index of the envelope curve sample points, traversing all normalized envelope values. v_n denotes the Nth normalized envelope value, i.e., the envelope amplitude after standardization, with a value range of [0,1]. The goal is to minimize the sum of envelope entropies for each IMF component obtained through VMD, serving as the criterion for evaluating decomposition quality. IBES updates the eagle's position based on a nonlinear dynamic inertia weight strategy, combining a Gaussian mutation mechanism to perturb the global optimal solution, preventing the algorithm from falling into local optima. Using the current candidate parameter [K,α], the signal is decomposed via VMD. Subsequently, the envelope entropy of each IMF component is calculated individually, and the sum of all entropy values is taken as the fitness of the current parameter combination. Based on this fitness, the eagle's position (i.e., the parameter combination) in the algorithm is updated while retaining the global optimal solution. The above steps are repeated until the iteration stopping condition is met, ultimately outputting the optimal parameter combination [K,α]. Envelope entropy quantifies the purity of a modal signal: when modal aliasing is severe, the entropy value is high; when the frequency components are relatively simple, the entropy value is low. Minimizing envelope entropy means maximizing the proportion of useful information, thereby suppressing cross-frequency noise interference at its source. Compared to sample entropy, which focuses on sequence complexity, and permutation entropy, which relies on amplitude ordering, envelope entropy is constructed based on time-domain envelopes and is more sensitive to local aliasing caused by carbon market policy shocks.

Figure 2
IBES-VMD workflow diagram.



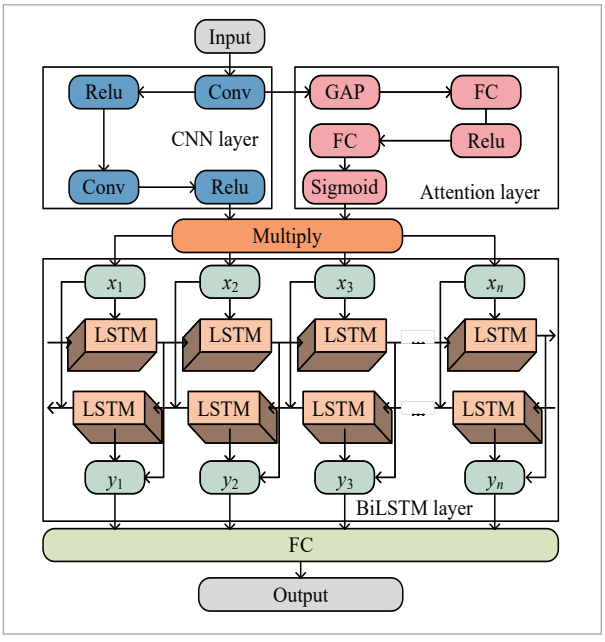
2.4. VMD data Decomposition

The VMD model with optimal parameters are utilized, the nonlinear and non-stationary volatility series V_i is decomposed and I different frequency IMF is outputted, each corresponding to a specific frequency or period of volatility characteristics. To make sure the clarity of IMF frequency separation and further obtain the spectrograms of each IMF, VMD decomposes nonlinear and non-stationary original volatility series into several IMFs with different frequency characteristics. It effectively reducing sequence complexity. It standardizes and normalizes all IMF components and raw data to eliminate dimensional differences [8]. The sample partitioning in this study adopts a global static approach, dividing all original samples into two parts in order: the training set and the reserved external samples. The first 80% of the samples are used as the training set, and the last 20% of the samples are reserved as external samples. All subsequent rolling evaluation experiments were conducted within the scope of this basic dataset.

2.5. Construction of Hybrid DL Models

A model is designed to extract temporal traits and perform predictive analysis. The three-level hybrid architecture of CNN, BiLSTM, and Attention is displayed in Figure 3. In Figure 3, each IMF component is input into a CNN layer, which extracts local spatial traits and then mines temporal dependencies through a BiLSTM layer. The Attention layer is responsible for adaptively weighting key traits, and ultimately FC outputs the predicted results.

Figure 3
CNN-BiLSTM-Attention hybrid architecture.



The CNN layer slides along the time axis through a one-dimensional convolution kernel and extracts local waveform features (spikes, abrupt changes, inflection points) of the volatility sequence in parallel. ReLU introduces nonlinearity while maintaining sparsity. The core size is 2, used to capture short-term volatility patterns within one to two trading days. The BiLSTM layer received traits extracted by CNN and captured both forward and backward temporal dependencies, comprehensively understanding the long-term dynamic patterns of the sequence. The forward LSTM processing sequence outputted the hidden states $y_1,y_2,y_3,...,y_n$ from the forward timing of $x_1 \rightarrow x_n$. The backward LSTM processing sequence outputted the hidden state $y_1,y_2,y_3,...,y_n$ from the reverse timing of $x_n \rightarrow x_1$. The BiLSTM layer selectively preserves historical information through forget gates, input gates, and output gates to capture the long-term dependence of the carbon market on policy event responses (tens of trading days). Bidirectional design can respectively depict the attenuation of historical shocks and the reverse shaping of future expectations, adapting to asymmetric information transmission under leverage effects. The Attention layer assigned higher weights to important information, allowing the model to focus on key details and further advance prediction accuracy and reliability. The global average pooling (GAP) of this layer first performed GAP on the temporal traits outputted by BiLSTM, aggregated the overall trait information of the sequence, and simplified the trait dimension. The first FC performed a linear transformation on the pooled traits and mapped them to the attention weight dimension. Afterwards, under the action of Relu and Sigmoid activation functions, the generated attention weight coefficients were multiplied bit by bit with the original output traits of BiLSTM, and finally the attention enhanced trait vector was output. It utilized the sliding window prediction method to output a volatility prediction sequence, analyzes the overall prediction performance of the model by comparing the predicted values with the true values.

3. Results and Analysis

3.1. Experimental Setup

The experiment utilized Intel i7-12700 processor, 16GB of RAM, Windows 10 operating system, and Python algorithm running environment. The model parameters set for the study are displayed in Table 2.

To evaluate the model's robustness across different time intervals, this study employed an 80%/20% time-series partitioning approach combined with a sliding-window cross-validation scheme: On top of the basic dataset divided in Section 2.4, 5 sets of sliding window subsets were further constructed for rolling prediction evaluation, with a fixed training window of 600 days and a sliding step size set at 20 days. Thus, five independent training-test subsets covering the period from 2020 to 2024. The training and test samples within each window were strictly partitioned in chronological order to prevent future information leakage. Across all windows, the SD of mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) were all below 8% of the corresponding metric means. This not only indicated that the model's predictive performance remained stable across different volatility cycles but also validated the robustness of the fixed deep learning hyperparameters in Table 2 across different time intervals. Regarding the cross-market generalizability of hyperparameters, the K and α parameters in the VMD decomposition were not arbitrarily fixed but were adaptively optimized by the IBES algorithm for each of the three carbon markets. The fixed hyperparameters for the neural network were derived from established literature, and preliminary experiments have already confirmed their robustness across various carbon market scenarios.

Table 2
Parameter settings of the model.

Model	Parameter	Set value
CNN	Number of filters (conv1)	64
	Number of filters (conv2)	128
	Convolution kernel size	2
BiLSTM	Time window size	14
	Hidden units (layer 1)	32
	Hidden units (layer 2)	64
	Dropout rate	0.2
Attention	Attention mechanism	Softmax
CNN-BiLSTM-Attention	Optimization algorithm	Adam
	Batch size	32
	Initial learning rate	0.0005
	Number of training epochs	100
	Objective function	MAE
IBES-VMD	Population size of IBES algorithm	50

3.2. Analysis of IBES-VMD Results

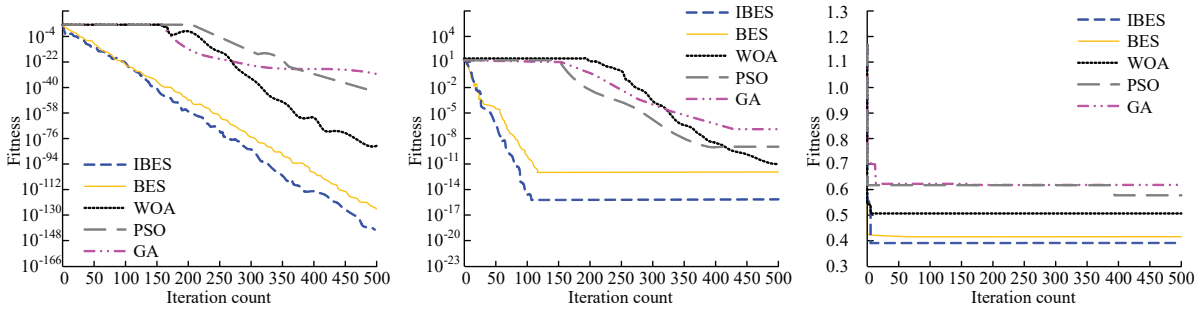
3.2.1. Performance Analysis of IBES

To make sure the effectiveness of the IBES algorithm, the fitness convergence curves of IBES and whale optimization algorithm (WOA), BES, GA, PSO intelligent optimization algorithms were compared under the same iteration times, population size, and parameter search range. The test functions were Sphere, Ackley, and Branin Function, and the convergence curves obtained are displayed in Figure 4. In Figure 4(a), IBES had the fastest convergence speed on the Sphere unimodal function, with a final fitness close to 10^{-140}, and its optimization effect was better than other algorithms. The fitness convergence curve of BES decreased slightly slower than IBES, but its convergence performance was better than WOA, GA, and PSO. In Figure 4(b), IBES converged to 10^{-15} on the Ackley multimodal function. The convergence speed of WOA increased in the later stages of iteration, but the accuracy was only 10^{-10}. In Figure 4(c), the final convergence fitness value of IBES algorithm was about 0.398, which almost completely coincided with the theoretical min value of Branin function, and was the closest to the global optimal solution among all compared algorithms. Therefore, IBES had strong advantages in convergence speed, convergence accuracy, and global optimization ability.

To further make sure the statistical significance of the performance differences between IBES and

Figure 4

Comparison of fitness convergence curves of test functions.



(a) Sphere (b) Ackley (c) Branin function

Table 3

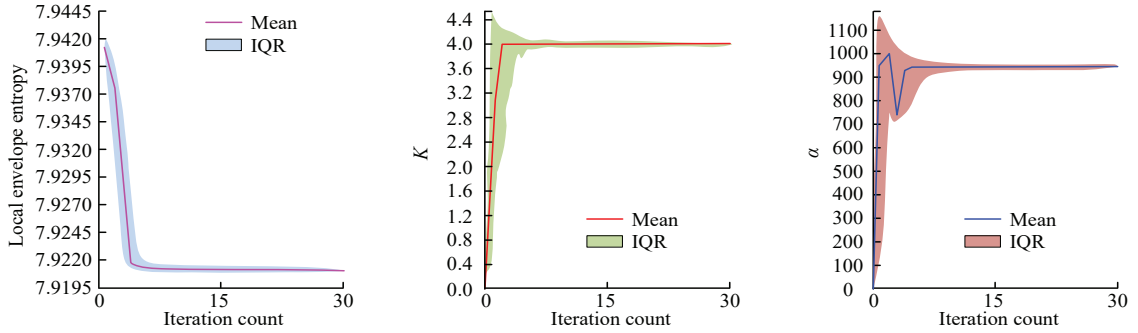
Statistical results of different algorithms on three test functions.

Test function	Algorithm	Optimal fitness mean	Optimal fitness SD	p value
Sphere	IBES	2.47×10^{-140}	1.23×10^{-139}	/
	BES	3.51×10^{-128}	2.87×10^{-97}	< 0.01
	WOA	7.82×10^{-82}	4.56×10^{-81}	< 0.01
	PSO	5.33×10^{-46}	3.21×10^{-45}	< 0.01
	GA	2.15×10^{-32}	1.84×10^{-31}	< 0.01
Ackley	IBES	8.93×10^{-16}	3.27×10^{-15}	/
	BES	4.21×10^{-12}	1.56×10^{-11}	< 0.01
	WOA	7.65×10^{-11}	2.98×10^{-10}	< 0.01
	PSO	3.44×10^{-9}	1.23×10^{-8}	< 0.01
	GA	2.18×10^{-7}	8.95×10^{-7}	< 0.01
Branin function	IBES	3.98×10^{-1}	2.34×10^{-3}	/
	BES	4.23×10^{-1}	5.67×10^{-3}	4.68×10^{-7}
	WOA	5.17×10^{-1}	8.92×10^{-3}	< 0.01
	PSO	5.89×10^{-1}	7.34×10^{-2}	1.05×10^{-8}
	GA	6.23×10^{-1}	1.04×10^{-2}	2.31×10^{-9}

other intelligent optimization algorithms, the study independently ran these 5 algorithms 30 times and recorded the optimal envelope entropy obtained from each iteration. Meanwhile, the stability of the algorithm was described using mean and SD, and Wilcoxon rank sum test was utilized to make sure the significance of the differences. The statistical results of different algorithms on three test functions are displayed in Table 3. In Table 3, IBES had the smallest SD for the optimal fitness of Sphere, Ackley, and Branin Function functions, and its optimization results were the most stable. The Wil-

Figure 5

The results of optimizing VMD hyperparameters using IBES.



(a) Envelope entropy curve (b) K optimization process (c) α optimization process

coxon rank sum test p values for IBES and the other four algorithms on Sphere and Ackley functions all less than 0.01.. On the Branin Function, the p values of IBES and various algorithms ranged from 1.05×10^{-8} to 4.68×10^{-7}. All $p<0.05$ indicate statistically significant differences.

3.2.2. VMD Effect

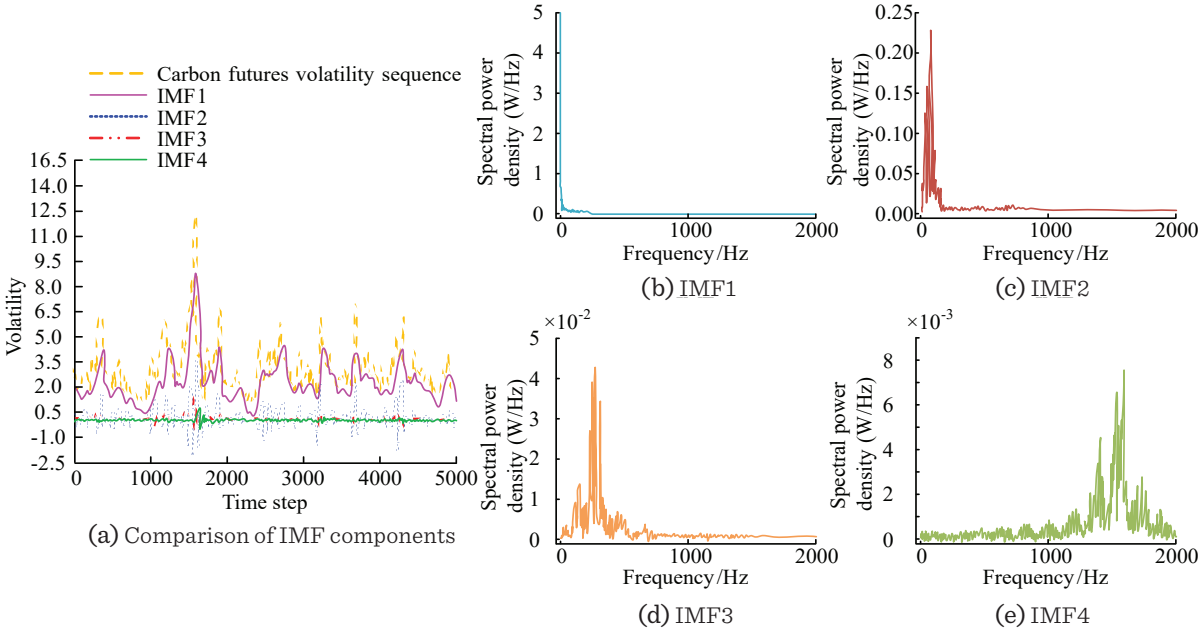
The IBES experiment to optimize VMD hyperparameters was repeated independently 30 times. Figure 5 shows the envelope entropy curve and the optimization processes for K and α. The shaded area in Figure 5 represents the interquartile range (IQR) (25th–75th percentiles). In Figure 5(a), the envelope entropy of IBES rapidly decreased in the first four iterations and reaches the global min of 7.9210 at the ninth iteration. Subsequent iterations stabilized the curve, indicating convergence to the global optimal solution, where VMD achieved the best noise suppression effect and the highest feature extraction efficiency. In Figure 5(b), the K parameter quickly converged in the initial iterations and stabilized at the optimal value of 4, making the optimal decomposition mode number 4, which was reasonable for decomposing the original volatility series into four IMF components. In Figure 5(c), the optimization curve of α oscillated rapidly and stabilized, ultimately converging to the optimal value of 942. This optimal value effectively balanced the precision and smoothness of signal decomposition. Overall, the IBES-optimized VMD optimal parameters were $K=4$ and $\alpha=942$, achieving the min envelope entropy (7.9210).

Table 4

Average envelope entropy and its corresponding validation set Mae under different parameter combinations.

Parameter combination number	K	α	Average envelope entropy	Validation Set MAE
1	6	1500	8.9102	15.2345
2	6	2000	8.8234	14.5678
3	7	1000	8.7456	13.9876
4	3	300	8.6678	13.4567
5	7	1500	8.5901	12.8765
6	3	2000	8.5123	12.3456
7	6	500	8.4345	11.7654
8	7	500	8.3567	11.2345
9	3	800	8.2789	10.6543
10	4	300	8.2012	10.1234
11	5	1500	8.1567	9.6543
12	5	500	8.1123	9.2345
13	4	2000	8.0678	8.7654
14	5	2000	8.0234	8.3456
15	5	800	7.9789	7.8765
16	6	800	7.9567	7.6543
17	4	1500	7.9345	7.4321
18	4	800	7.9234	7.1234
19	4	942	7.921	6.8512
20	3	1500	7.9567	7.5432

Figure 6
Time domain and frequency spectrum of IMF components.



(a) Comparison of IMF components (b) IMF1 (c) IMF2 (d) IMF3 (e) IMF4

To further validate the theoretical logic of envelope entropy, this study recorded the average envelope entropy values and their corresponding validation set MAE for 20 sets of different parameter combinations during the IBES-VMD optimization process; the calculation results are presented in Table 4. Furthermore, the calculated Pearson correlation coefficient between the two was r = 0.87 (p < 0.01), confirming a significant and strong positive correlation between envelope entropy and prediction error – that is, the lower the envelope entropy value, the smaller the model's prediction error. This empirical result quantitatively validated the theoretical soundness and practical effectiveness of envelope entropy as a fitness function.

To make sure the signal separation capability and noise reduction effect of the optimized VMD, the time-domain and spectral characteristics of the IMF components were analyzed. The results of these two types of diagrams are displayed in Figure 6. Figure 6(a) presents the time-domain plots of the carbon futures volatility series and the four IMF components. The original series exhibited pronounced nonlinear and non-stationary fluctuation characteristics, containing significant noise and multi-scale fluctuation components. Among them, the four IMF components obtained through VMD implied that IMF1 had the lowest frequency and the largest amplitude. IMF1 reflects the long-term trend component of carbon market volatility. The IMF2 and IMF3 components fell within the low-to-medium frequency range and correspond to medium-term cyclical fluctuations in supply and demand. The IMF4 component had a higher frequency and smaller amplitude, and primarily corresponds to short-term market noise and policy shocks. Figures 6(b)-6(e) display the spectral plots of IMF1-IMF4, respectively. The total spectral energy was calculated separately for each of the four IMF components obtained from the decomposition. The proportions of the total energy of the original carbon volatility series accounted for by each component were as follows: IMF1 accounted for 62.7%, IMF2 for 21.3%, IMF3 for 11.5%, and IMF4 for 4.5%. The spectra of these four IMF components were independent and did not overlap. Therefore, the IBES-optimized VMD could avoid trait confusion across different time scales and ensure precise frequency separation.

3.3. Ablation Experimental Results of the Model Frequency Separation

To make sure the effectiveness of the suggested model (IBES-VMD-CNN-BiLSTM-Attention), ablation experiments were conducted, and the parameter settings and experimental conditions of all models were kept consistent. The prediction results of Comparison Model 1-5 (CM1-5) and the suggested model are shown in Table 5. Each model was trained independently for 30 times, using different random seeds to initialize the network parameters, and reporting the mean and SD of each performance index. In Table 5, comparing the results of CM5 and the suggested model, it can be seen that the triple improvement of IBES, including chaos initialization, dynamic inertia weight, and Gaussian mutation, reduced MAE by 4.43% and increased R^2 by 1.25%. Compared with CM2, the MAE of CM3 decreased by 15.04%. This verified that the multi-scale decomposition of VMD could effectively eliminate the mixed noise embedded in the carbon volatilization sequence. CM4 removed the attention module while retaining the IBES-VMD backbone network. Its MAPE (4.62%) was significantly higher than the suggested model (3.56%). This indicated that the attention layer could adaptively enhance key volatile time steps and bring independent prediction gains.

The prediction curves of CM1-4 trained were compared, and the results are displayed in Figure 7. In Figure 7(a), the predicted volatility curves of CM1 and CM2 differed significantly from the actual value curves. And the performance of CM3 and CM4 was better than the first two. This was because VMD ensured minimal noise interference, thereby improving prediction accuracy. However, there was still a

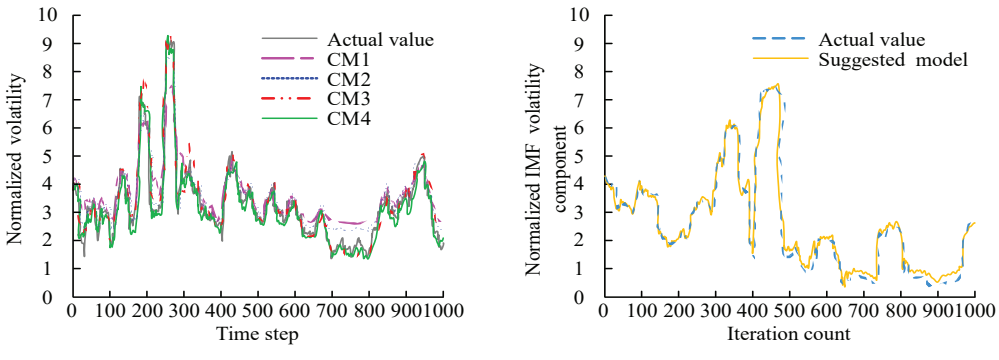
Table 5

Comparison of prediction results of different models.

Code name	Model	MAE	RMSE	MAPE/%	R2/%
CM1	CNN-BiLSTM	38.08 ± 2.14	51.05 ± 2.87	17.44 ± 0.98	76.21 ± 1.24
CM2	CNN-BiLSTM-Attention	28.70 ± 1.56	37.81 ± 2.01	10.82 ± 0.76	86.01 ± 1.12
CM3	VMD-CNN-BiLSTM-Attention	13.66 ± 0.87	18.10 ± 1.23	5.58 ± 0.45	96.83 ± 0.67
CM4	IBES-VMD-CNN-BiLSTM	13.76 ± 0.92	18.98 ± 1.31	8.80 ± 0.62	93.97 ± 0.78
CM5	BES-VMD-CNN-BiLSTM-Attention	11.28 ± 0.71	15.34 ± 1.06	4.62 ± 0.38	97.54 ± 0.52
Suggested model	IBES-VMD-CNN-BiLSTM-Attention	6.85 ± 0.54	10.50 ± 0.78	3.56 ± 0.31	98.79 ± 0.35

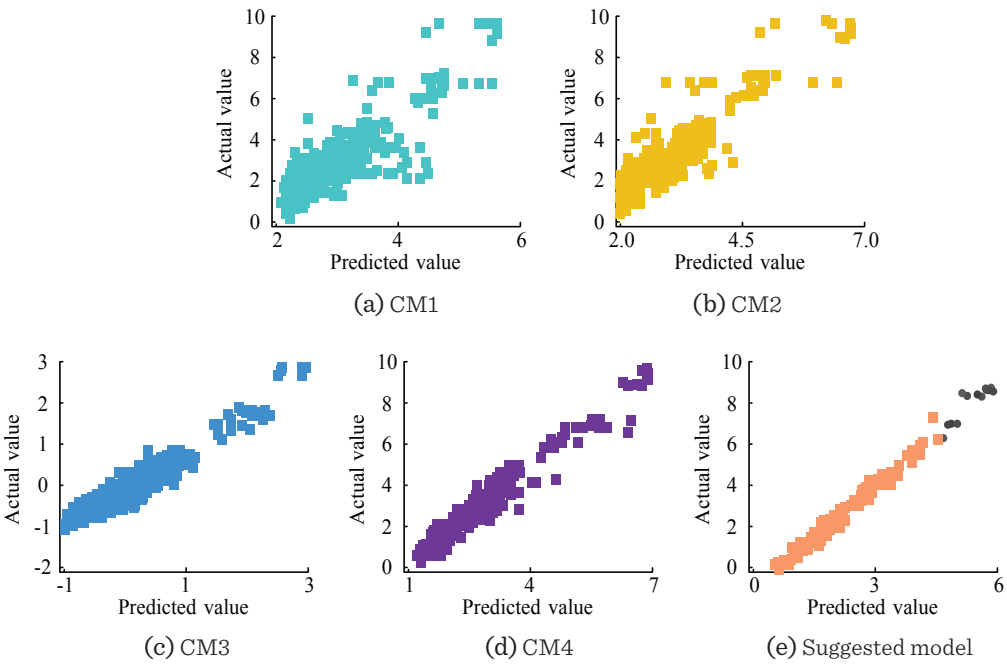
Figure 7

Changes in volatility of different models.



(a) Changes in volatility of different models (b) Changes in volatility of proposed model

Figure 8
Scatter plot of correlation between different models.



certain deviation between the curves of CM3 and CM4 and the true values in some intervals, which might be due to the impact of extreme volatility events like sudden market shocks and policy changes. Figure 7(b) implies the volatility variation of the suggested model. In contrast, the suggested model had higher prediction accuracy. This was because the synergistic effect of optimizing VMD and attention mechanism through IBES had greatly contributed to the performance of the model.

To analyze the linear correlation and fitting degree between the predicted and actual volatility values of different models, a study was conducted to compare the predictive performance of five models on carbon futures volatility. The scatter plot of the correlation between different models is displayed in Figure 8. In Figure 8(a), the scatter distribution of CM1 was relatively scattered, significantly deviating from the diagonal and with many outliers (R^2=76.21%). This indicated that the correlation between the predicted and actual values of volatility was weak, and the fitting effect was poor. In Figure 8(b), the fitting effect of CM2 was poor (R^2=86.01%), and there was still a significant error. In Figure 8(c), the coefficient of determination for CM3 was 0.9683, and its scatter distribution values ranged below 3. The overall sample coverage was insufficient, and its ability to capture some extreme values was weak. In Figure 8(d), the fitting effect of CM4 was good (R^2=93.97%), and the scatter distribution was relatively concentrated. In Figure 8(e), the linear correlation between the predicted and actual values of the suggested model was the strongest (R^2=98.79%), and the scatter height was concentrated on the diagonal with no extreme outliers. Therefore, this model performed the best.

To further explore the source of prediction error, this study conducted a driving analysis around the three major events during the sample period. The experiment centered on each event date (± 15 days), set a 30 day observation window, and calculated the MAE of suggested model and CM1. See table 4 for comparison results. The results in Table 6 showed that within the three major event windows, the MAE increase of the suggested model (8.3% -12.5%) was significantly lower than that of CM1 (29.1% -42.3%). The suggested model isolated policy shocks in the

Table 6

Comparison of MAE between the suggested model and CM1 within the three major event windows.

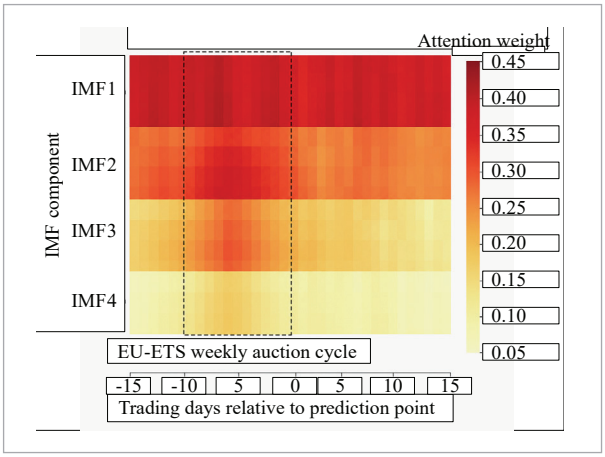
Events	Event date	Suggested model / CM1 baseline MAE (mean 30 days before the event)	Suggested model / CM1 event window MAE (± 15 days)	Suggested model / CM1 MAE increase/%
EU-ETS Phase 4 reform announcement	Feb-21	6.52 / 37.21	7.34 / 52.98	+12.5 / +42.3
Russia-Ukraine energy crisis	Mar-22	6.78 / 38.45	7.51 / 53.34	+10.8 / +38.7
EU Parliament's ratification of the Carbon Border Adjustment Mechanism	Apr-23	6.61 / 37.89	7.16 / 48.92	+8.3 / +29.1

high-frequency channel of IMF4 through VMD, avoiding their contamination of long-term trends and medium-term cycle components, and verified the source of its predictive stability under extreme events, that is, achieving effective separation of short-term noise and persistent components in the decomposition stage.

To improve the transparency of the model, this study analyzed the attention weight distribution of the suggested model on different IMF components. The attention heatmap is shown in Figure 9. According to Figure 9, the average attention weights assigned to IMF1, IMF2-IMF3, and IMF4 were 0.42, 0.38, and 0.20, respectively. This indicated that the model gave the highest attention weight to the low-frequency trend component, followed by the mid frequency periodic component, while giving the lowest weight to the high-frequency noise component. In addition, within 7 to 10 trading days before the prediction point, the attention weight was relatively high, which corresponded to the weekly auction cycle of EU-ETS, indicating that the model could effectively capture the regularity features in the market microstructure.

Figure 9

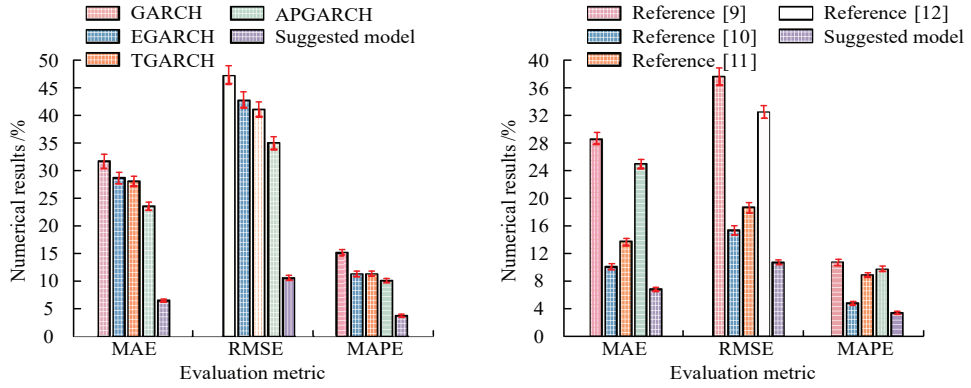
Suggested model attention heatmap on different IMF components.



3.4. Evaluation and Analysis of the Effectiveness of the Suggested Model

To further validate the superiority of the suggested model, it was compared with classical financial time series models. The compared models were GARCH, EGARCH, threshold GARCH (TGARCH), and asymmetric power GARCH (APGARCH). In addition, the suggested model would be compared and analyzed with the predictive models in references [5, 11, 13, 14]. To ensure that the experimental results were not affected by randomness, each model was independently trained for 30 times. Each time, the network parameters were initialized with different random seeds, and the mean and SD of each evaluation index were calculated. This study reproduced all the comparative models mentioned in references [5, 11, 13, 14] and evaluated them under completely consistent experimental conditions, including using the same dataset, the same 80%/20% time series partitioning method, the same prediction window length, and the same evaluation index system. The prediction accuracy of the suggested model and different types of models is displayed in Figure 10. In Figure 10(a), APGARCH performed the best in traditional financial models, with an MAE of 23.78%, RMSE of 35.16%, and MAPE of 9.84%. The suggested model had an MAE of 6.85%, RMSE of 10.50%, and MAPE of 3.56%, all of which were significantly better than traditional financial models. This proved that the strategy combining data-driven and DL had superior generalization ability and prediction accuracy

Figure 10
The prediction accuracy of the suggested model and different types of models.



(a) Comparison with traditional financial models

(b) Comparison with advanced prediction models

in processing complex nonlinear financial data. In Figure 10(b), the suggested model outperformed the comparative models from references [5, 11, 13, 14] in terms of MAE, RMSE, and MAPE evaluation metrics. This result validated the superiority of IBES optimized VMD hyperparameters and CNN-BiLSTM-Attention hybrid framework in predicting carbon futures volatility tasks. A paired t-test was used to compare the model in this paper with each of the alternative models on a pair-wise basis. The results showed that all p-values were less than 0.01, indicating that the performance improvement of the suggested model was statistically significant.

The R^2 results of different types of models were compared in the study, and the results are displayed in Table 7. In Table 7, the suggested model had an R^2 of 98.79%, ranking first among all compared models. The R^2 of traditional financial models ranged from 83.12% to 92.68%, all lower than the suggested model. Compared with the optimal traditional econometric model APGARCH (92.68%), the suggested model advanced by 6.11%. A paired t-test of the R^2 values between the suggested model and APGARCH yielded a p-value of < 0.01, indicating that the suggested model was significantly superior to the traditional econometric model. Overall, the IBES optimized VMD and CNN-BiLSTM-Attention hybrid framework could more fully explain the variability of carbon futures volatility, with stronger fitting ability and prediction accuracy.

3.5. Application Analysis of Domestic Carbon Market

To make sure the adaptability and generalization ability of the suggested model in different carbon market environments, cross market validation experiments were conducted. The study selected China carbon emission trade exchange (CCETE) and Beijing emission allowances (BEA) as supplementary empirical objects. Among them, the comparison of CCETE prediction results between different models is displayed in Figure 11. Figure 11(a) implies the results of MAE and RMSE for different models. The suggested model had the lowest error (MAE=6.99%, RMSE=10.77%), and its predictive error control ability was significantly better than all compared models. The overall errors of traditional financial models GARCH, EGARCH, TGARCH, and APGARCH were relatively high. The error level of CNN-BiLSTM was the highest among all models (MAE=39.21%, RMSE=52.35%). After adding attention mechanism, the error of CNN-BiLSTM-Attention decreased slightly (MAE=29.43%, RMSE=38.77%). In Figure 11(b), the traditional financial model had limited performance in the CCETE market, with an optimal APGARCH R^2 of only 87.65%. The suggested model advanced by about 11%, verifying that the DL hybrid framework still had significant advantages in the less mature carbon market. The suggested model optimized

VMD parameters through IBES, resulting in a nearly 38% reduction in MAPE compared to the fixed parameter VMD model. This indicated that adaptive optimization of VMD parameters played a crucial role in improving prediction accuracy. The DL model without signal decomposition performed poorly in the CCETE market, with an R^2 of only 75.12% for CNN-BiLSTM. After adding attention mechanism, it advanced to 85.43%, but there was still a significant gap from the suggested model's 98.65%. This result indicated that signal decomposition and parameter optimization were necessary steps to ensure prediction accuracy in the CCETE market with severe fluctuations and strong noise interference.

The comparison of different models in BEA prediction results is displayed in Table 8. In Table 8, the suggested model performed slightly better in the BEA market than in the CCETE market. The BEA market had a longer operating time, relatively stable trading activity, and stronger data regularity. The CCETE market was in its early stages of development, with frequent policy changes and a more complex volatility series. All traditional models performed better in the BEA market than in the CCETE market, verifying the stronger regularity of BEA market volatility. The suggested model achieved R^2 above 98.6% and MAPE below 3.7% in all three markets. BEA performed the best in the market (R^2=98.88%, MAPE=3.49%).

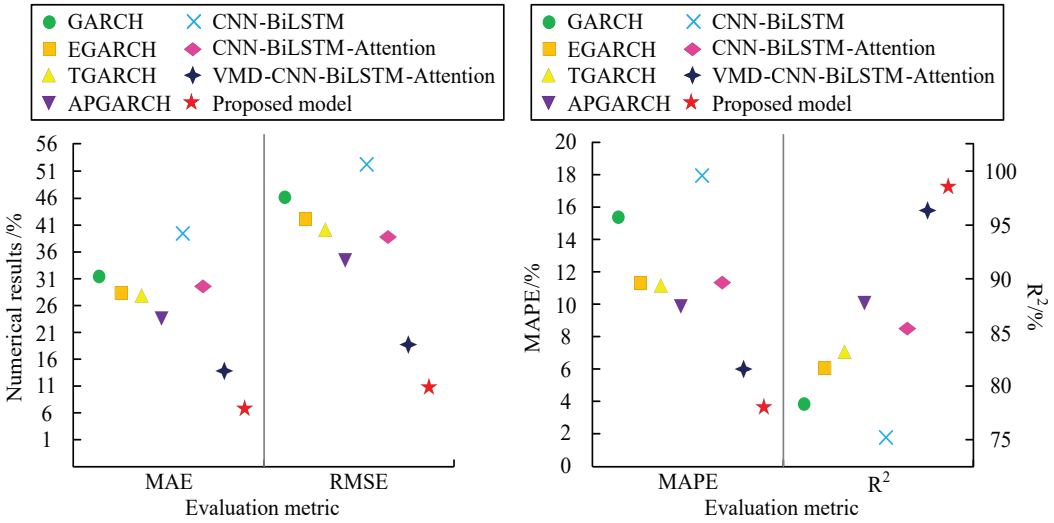
Table 7

R^2 results of different types of models.

Model	R^2/%	Model	R2/%
GARCH	83.12 ± 1.56	Reference [11]	86.01 ± 1.12
EGARCH	86.70 ± 1.34	Reference [5]	95.72 ± 0.58
TGARCH	88.18 ± 1.21	Reference [14]	93.97 ± 0.78
APGARCH	92.68 ± 0.89	Reference [13]	88.35 ± 1.08
Suggested model	98.79 ± 0.35	/	/

Figure 11

Different models predict CCETE results.



(a) Result of MAE and RMSE (b) Result of MAPE and R^2

Table 8
Different models predict BEA results.

Model	MAE	RMSE	MAPE/%	R^2/%
GARCH	29.88 ± 2.45	44.54 ± 3.56	14.32 ± 1.23	79.43 ± 1.56
EGARCH	26.75 ± 2.12	40.43 ± 3.21	10.65 ± 0.98	82.77 ± 1.34
TGARCH	26.23 ± 1.98	38.77 ± 2.98	10.54 ± 0.87	84.12 ± 1.21
APGARCH	22.12 ± 1.67	32.88 ± 2.56	9.23 ± 0.76	88.54 ± 0.89
CNN-BiLSTM	37.54 ± 2.89	50.43 ± 3.45	17.12 ± 1.23	76.43 ± 1.56
CNN-BiLSTM-Attention	28.12 ± 1.78	37.12 ± 2.34	10.77 ± 0.89	86.32 ± 1.12
VMD-CNN-BiLSTM-Attention	13.32 ± 0.92	17.99 ± 1.28	5.43 ± 0.48	96.77 ± 0.67
suggested model	6.72 ± 0.56	10.32 ± 0.82	3.49 ± 0.34	98.88 ± 0.38

4. Discussion and Interpretation

The study first utilized IBES to optimize VMD, decomposing the volatility of the carbon market into multiple different frequency components. Afterwards, components were input into CNN-BiLSTM-Attention to complete deep feature extraction and ultimately implemented prediction. The study adopted IBES to automatically find the optimal K and α in the parameter space through intelligent search, thereby improving the decomposition quality. IBES introduced chaos initialization on the basis of the original BES to enhance population diversity, dynamic inertia weight balance for global exploration and local development, and Gaussian mutation enhancement for escaping local optima. In the standard test function, the convergence accuracy and speed were better than BES, WOA, PSO, and GA. To evaluate the adaptability of the model in different market states, this study divided the market into three categories for analysis: during the stationary period, GARCH models performed stably, but the deviation gradually widened under small sustained fluctuations; During the policy shock period, the GARCH model struggled to capture asymmetric pulse shocks, resulting in a sharp increase in error; The IBES-VMD framework in this study significantly outperformed prediction stability by removing high-frequency noise and bidirectional conduction recognition. During the high volatility period at the end of the performance period, the multifractal characteristics of carbon prices became more prominent. IBES adaptive optimization could effectively split long and short-term fluctuations, which was superior to ordinary deep learning methods without optimized VMD. Compared with existing advanced prediction models, the suggested model had better performance. Reference [13] adopts a rolling decomposition prediction framework combining CEEMDAN decomposition and multi-scale entropy reconstruction. Although this method can effectively address the non-stationary characteristics of carbon price sequences, it does not adaptively optimize the decomposition parameters. Although its RF-LSTM architecture can capture high-frequency and low-frequency components, it relies on a fixed decomposition setting, which may not be applicable to all market conditions. Compared with it, the IBES-VMD framework proposed in this paper adaptively determines the optimal decomposition parameters (i.e. K and α) through intelligent search, thereby achieving better performance in decomposition quality and prediction accuracy. The R^2 values of the models in references [11] and [13] were both below 90%, indicating that DL models without optimization mechanisms or models with poor decomposition quality had limited explanatory power. The model in reference [5] adopted CEEMDAN and VMD quadratic decomposition, but the VMD parameters had not been optimized. The model in ref-

erence [14] adopted adaptive symmetric geometric mode decomposition, and the decomposition quality was greatly affected by the signal morphology. The model in reference [19] utilized POA to optimize VMD, but POA's global search capability was weaker than IBES. The introduction of IBES algorithm for optimizing VMD hyperparameters mainly solved the problems of difficult parameter selection, unstable decomposition quality, and susceptibility to local optima in the VMD process. Thereby, it advanced the data preprocessing quality, prediction accuracy, and model robustness of the carbon futures volatility prediction model.

In summary, this study provides fluctuation monitoring technology reference for the construction of domestic carbon markets, and provides quantitative basis for risk management in the transition from pilot markets in different regions to national markets.

5. Summary and Future Work

To enhance the adaptability and robustness of carbon market prediction methods, a CNN-BiLSTM-Attention framework was developed for predicting carbon futures volatility. This study solves the accuracy bottleneck of traditional econometric models and single DL models in volatility prediction, and provides a reliable technical path for volatility monitoring and risk management in the domestic carbon market. There are three main limitations to this study. At the input level, the current model only relies on historical volatility data and has not yet incorporated external information such as policy sentiment and energy prices. In the future, the study plans to integrate multimodal data and expects to reduce the prediction error during the policy window period by 10%-15%. In terms of model efficiency, the IBES-VMD optimization process took about 12 minutes. In the future, it is planned to introduce an early stop mechanism (expected to reduce the number of iterations by 30%-40%) and knowledge distillation techniques (expected to reduce inference time by 50%). At the interpretability level, the current attention mechanism can only provide explanations at the time step level. In the future, we plan to introduce SHAP values to quantify the contributions of each feature and enhance the transparency of the model.

Acknowledgement

The research is supported by: The Science & Technology Development Fund of Tianjin Education Commission for Higher Education. Research on the Transmission Mechanism and Path of ESG Performance to Enterprise Performance under the "Dual Carbon" Goals, No.: 2024SK138.

References

1. Abd Elfattah, M., Ewees, A. A., Sayed, G. I., Darwish, A., Hassanien, A. E. Carbon Price Time Series Forecasting Utilizing an Optimized ANFIS Model. Evolutionary Intelligence, 2024, 17(5), 3745-3765. https://doi.org/10.1007/s12065-024-00955-2
2. Cai, X., Yan, B. Tail Dependence of Liquidity and Volatility in Carbon Futures Market: Evidence from EU ETS. Managerial and Decision Economics, 2025, 46(6), 3538-3570. https://doi.org/10.1002/mde.4545
3. Chai, S., Zhang, Z., Zhang, Z. Carbon Price Prediction for China's ETS Pilots Using Variational Mode Decomposition and Optimized Extreme Learning Machine. Annals of Operations Research, 2025, 345(2), 809-830. https://doi.org/10.1007/s10479-021-04392-7
4. Chen, Y., Jin, M., Zhou, Z., Tian, Z. A Novel Ensemble Learning Framework Based on News Sentiment Enhancement and Multi-Objective Optimizer for Carbon Price Forecasting. Computational Economics, 2025, 66(5), 3709-3733. https://doi.org/10.1007/s10614-024-10828-6
5. Fang, M., Zhang, Y., Liang, W., Shi, S., Zhang, J. Carbon Price Prediction Research Based on CEEMDAN-VMD Secondary Decomposition and BiLSTM. Environmental Science and Pollution Research, 2025, 32(14), 8921-8942. https://doi.org/10.1007/s11356-025-36220-x
6. He, M., Jiang, W., Gu, W. TriChronoNet: Advancing Electricity Price Prediction with Multi-Module Fusion. Applied Energy, 2024, 371, 123626. https://doi.org/10.1016/j.apenergy.2024.123626

7. Hou, J., Dong, Z. Volatility-Adapted Network Model for Hedging Equity Index Options. Automatika, 2025, 66(3), 370-381. https://doi.org/10.1080/00051144.2025.2475261

8. Jiang, Z., Zhang, Z., Zhang, L. Forecasting China's Short-Term Energy Futures Price Using a Novel Secondary Decomposition-Optimized System. Computational Economics, 2025, 66(5), 4009-4043. https://doi.org/10.1007/s10614-024-10840-w

9. Liu, T., Niu, H., Yin, X. Macroeconomic Variables and Time-Varying Volatility Persistence of Carbon Market. Applied Economics, 2026, 58(6), 1021-1040. https://doi.org/10.1080/00036846.2025.2464812

10. Sadefo Kamdem, J., Miano Mukami, P., Njong, J. B. Time-Frequency Analysis and Machine Learning Models for Carbon Market Forecasting. Annals of Operations Research, 2025, 352(3), 543-562. https://doi.org/10.1007/s10479-023-05443-x

11. Sha, D., Zeng, X., Johannssen, A., Wang, R., Tran, K. P. A Two-Stage NLP-Driven Framework for Interval-Valued Carbon Price Prediction Using Sentiment Analysis and Error Correction. Journal of Forecasting, 2026, 45(2), 806-818. https://doi.org/10.1002/for.70059

12. Song, Y., Tang, X., Sun, Y. Volatility Forecasting for the European Union Carbon Emissions Allowance Trading Market Based on the VMD Method. Portuguese Economic Journal, 2026, 25(1), 57-76. https://doi.org/10.1007/s10258-025-00276-x

13. Wang, H., Li, Y. RF-LSTM Carbon Price Prediction Based on CEEMDAN Decomposition and Multiscale Entropy Reconstruction. Scientific Reports, 2026, 16(1), 1-19. https://doi.org/10.1038/s41598-026-35085-5

14. Wang, P., Zhou, X., Zeng, Z. LightGBM-BES-BiLSTM Carbon Price Prediction Based on Environmental Impact Factors. Computational Economics, 2025, 65(5), 2891-2917. https://doi.org/10.1007/s10614-024-10648-8

15. Xiao, C., Liu, Y. A Multifrequency Data Fusion Deep Learning Model for Carbon Price Prediction. Journal of Forecasting, 2025, 44(2), 436-458. https://doi.org/10.1002/for.3198

16. Xu, W., Wang, J., Zhang, Y., Li, J., Wei, L. An Optimized Decomposition Integration Framework for Carbon Price Prediction Based on Multi-Factor Two-Stage Trait Dimension Reduction. Annals of Operations Research, 2025, 345(2), 1229-1266. https://doi.org/10.1007/s10479-022-04858-2

17. Xu, Y., Che, J. A Novel Hierarchical Carbon Price Forecasting Model with Local and Overall Perspectives. Environmental and Ecological Statistics, 2024, 31(3), 749-776. https://doi.org/10.1007/s10651-024-00619-5

18. Zhang, Y., Peng, Y., Song, Y. Realized Volatility Forecasting for Stocks and Futures Indices with Rolling CEEMDAN and Machine Learning Models. Computational Economics, 2025, 66(2), 1215-1268. https://doi.org/10.1007/s10614-024-10732-z

19. Zhang, Z., Li, Q., Li, R. Leveraging Deep Learning for Carbon Market Price Forecasting and Risk Evaluation in Green Finance Under Climate Change. Journal of Organizational and End User Computing, 2025, 37(1), 1-27. https://doi.org/10.4018/JOEUC.387168

20. Zhu, B., Wan, C., Wang, P., Chevallier, J. Forecasting Carbon Market Volatility with Big Data. Annals of Operations Research, 2025, 348(1), 317-343. https://doi.org/10.1007/s10479-023-05401-7

21. Zhu, B., Wan, C., Wang, P., Chevallier, J. Forecasting Carbon Market Volatility with Big Data. Annals of Operations Research, 2025, 348(1), 317-343. https://doi.org/10.1007/s10479-023-05401-7

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).