Front. Agr. Sci. Eng. 2025, 12(2): 208–230 https://doi.org/10.15302/J-FASE-2024599

RESEARCH ARTICLE

An improved deep learning model for soybean future price prediction with hybrid data preprocessing strategy

Dingya CHEN[1], Hui LIU (✉)[1], Yanfei LI[2], Zhu DUAN[1]

1 Institute of Artificial Intelligence & Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China.
2 School of Mechatronic Engineering, Hunan Agricultural University, Changsha 410128, China.

KEYWORDS

Deep extreme learning machine, hybrid data preprocessing, optimization algorithm, soybean future price prediction

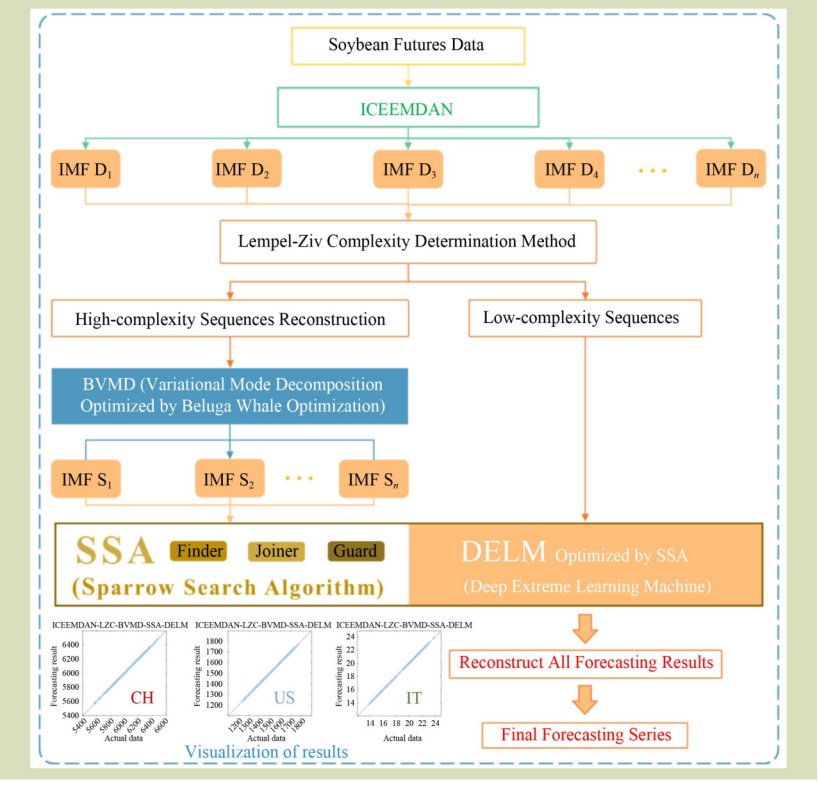
HIGHLIGHTS

- A new hybrid forecasting model for soybean futures closing prices is proposed.
- Variational mode decomposition parameters optimized by the beluga whale method is proposed as the secondary decomposition algorithm.
- The predictor of the secondary decomposition subsequences is optimized by sparrow search algorithm.

Received July 8, 2024;
Accepted September 23, 2024.

Correspondence: csuliuhui@csu.edu.cn

GRAPHICAL ABSTRACT



ABSTRACT

The futures trading market is an important part of the financial markets and soybeans are one of the most strategically important crops in the world. How to predict soybean future price is a challenging topic being studied by many researchers. This paper proposes a novel hybrid soybean future price prediction model which includes two stages of data preprocessing and deep learning prediction. In the data preprocessing stage, futures price series are decomposed into subsequences using the ICEEMDAN (improved complete ensemble empirical mode decomposition with adaptive noise) method. The

Lempel-Ziv complexity determination method was then used to identify and reconstruct high-frequency subsequences. Finally, the high frequency component is decomposed secondarily using variational mode decomposition optimized by beluga whale optimization algorithm. In the deep learning prediction stage, a deep extreme learning machine optimized by the sparrow search algorithm was used to obtain the prediction results of all subseries and reconstructs them to obtain the final soybean future price prediction results. Based on the experimental results of soybean future price markets in China, Italy, and the United States, it was found that the hybrid method proposed provides superior performance in terms of prediction accuracy and robustness.

1 Introduction

As one of the most important crops in the world, soybean prices are of concern to governments, investors and farmers. However, soybean prices are affected by a combination of many factors and have great uncertainty[1]. Governments find it difficult to monitor and predict the future trend of soybeans, and are unable to formulate appropriate policies in response. Investors lack scientific decision-making methods to guide them and can only trade soybeans based on inadequately informed speculations, which exacerbates the instability of soybean prices. Farmers also need scientific and practical price forecasts to help maximize profits and provide confidence in their forward planning[2]. Global commodity market trends, investor perceptions and political events, and currency factors have influenced soybean future price, and the price series are highly nonlinear and volatile. Therefore, it is of great significance to study how to develop a general and effective soybean future price forecasting model[3].

Many scholars have conducted relevant studies on the forecasting of agricultural prices with good results. The earliest invented method was statistically-based. The most used statistical method is the autoregressive integrated moving average (ARIMA) model. Darekar & Reddy[4] and Panasa et al.[5] used the ARIMA model to forecast monthly soybean prices and maize prices in India to provide a reasonably predictive analysis. Bhardwaj et al.[6] first used the Box-Jenkins autoregressive integrated moving average model to forecast agricultural prices in New Delhi and found that the ability to capture the volatility of the data was not satisfactory, and then used the generalized autoregressive conditional heteroscedastic model for forecasting and the results obtained were better than ARIMA in all indicators, proving that the model has better forecasting performance. In addition, the accuracy of forecasting can also be further improved by combining models from different statistical methods. Şahinli[7] proposed a Holt-Winters model combined with ARIMA to improve the accuracy of forecasting, and the performance of the proposed model was more reliable compared with ARIMA alone.

In recent years, machine learning has become increasingly popular and intelligent models are widely used in endeavors such as image processing[8], price prediction[9], and fault diagnosis[10]. Mahto et al.[11] and Jaiswal et al.[12] used artificial neural networks and deep long-term short-term memory-based models on predicting agricultural price series and compared them with ARIMA, respectively. Zong & Zhu[13] used radial basis function (RBF) and back propagation neural networks to forecast Chinese agricultural prices. Xu[14] conducted univariate modeling of spot prices and bivariate modeling of spot and futures prices with neural networks to forecast the price series of the last 500 agricultural commodities in the USA. Xu & Zhang[15] evaluated the impact of different model settings of neural networks, such as the number of hidden neurons and the proportion of training and validation sets, on the prediction performance of the canola and soybean oil price forecasting models.

It is evident that the performance of single models is still not robust for high-complexity and noisy time series, so most of the research at this stage integrates preprocessing[16], optimization algorithms[17] and postprocessing[18] to improve the performance of hybrid models. Zhang & Na[19] introduced a hybrid model that includes fuzzy information granulation and uses the mind evolutionary algorithm to optimize the hybrid model with the support vector machine and concluded that the prediction of the UN Food and Agriculture Organization issued price. Li et al.[20] proposed a model combining wavelet transformation and exponential smoothing and experimentally demonstrated superiority over long short-term memory (LSTM) and support vector regression.

Wang et al.[21] proposed a combined forecasting method based on a global optimization method, using three decomposition algorithms and five forecasting models to forecast soybean and maize futures prices, demonstrating the superiority of the combined model. Liang & Jia[22] introduced a hybrid Gray Wolf Optimizer-convolutional neural network-LSTM model for forecasting prices of maize, soybean and others, and introduced the Baidu index, Google trends and computed transfer entropy to improve real-time forecasting. Zhang et al.[23] proposed a quantile regression-RBF neural network model using a hybrid algorithm, combining gradient descent and genetic algorithms, to achieve global and local search. In particular, the introduction of this algorithm provided enhanced optimization capabilities for the quantile regression-RBF model, thus further improving the performance and stability of the model.

At this stage, price forecasting of agricultural products has achieved satisfactory forecasting accuracy, but in other price forecasting research fields, secondary decomposition and other optimization methods have been introduced to optimize the forecasting accuracy of time-series data. Liu & Long[24] used a novel framework for forecasting stock closing prices with higher prediction results compared with established models. The deep hybrid framework consists of a data processing component, a deep learning predictor component and a predictor optimization method. The empirical wavelet transform is used for data preprocessing. The LSTM is the main part of the hybrid framework and is optimized by a combination of dropout strategies and particle swarm optimization algorithms. Sun & Huang[25] proposed a secondary decomposition model based on empirical mode decomposition (EMD) and variational mode decomposition (VMD), and used LSTM to forecast the entire carbon market in China. Considering price dynamics that include cyclical growth, seasonal variations and irregular fluctuations, Zhu et al.[26] proposed a hybrid model combining a loess-based seasonal-trend decomposition procedure, support vector regression and autoregressive moving average to forecast hog prices for the next farming cycle. Liu et al.[27] proposed a hybrid model consisting of secondary decomposition, ensemble method and error correction. The decomposition is performed by wavelet decomposition to obtain the wind speed subseries. The SampEn algorithm is used to estimate the unpredictability of the subsequences. The most unpredictable subsequences are decomposed again by the VMD. The subsequences are obtained as predicted subsequences by the ensemble method neuron network, and the modified predicted sequence is reconstructed to obtain the final predicted sequence. Liu & Zhang[28] proposed a hybrid AQI time series forecasting model based on secondary decomposition, imperialist competitive algorithm, feature selection and echo state network. This hybrid model has broad application prospects and research value in the field of AQI forecasting.

In summary, although scholars have achieved good results in soybean future price forecasting, there is still scope for exploration of hybrid methods. The most mainstream hybrid learning methods are currently based on standard machine learning or deep learning with signal preprocessing methods or optimization algorithms. This study uses a secondary decomposition-based signal preprocessing method that combines a highly adaptive optimization algorithm and a neural network model with powerful information mining capabilities. Based on this, a new forecasting method is proposed and applied to the field of agricultural futures forecasting for the first time.

In this paper, an improved deep learning model for soybean future prices prediction is proposed, which includes ICEEMDAN (improved complete ensemble empirical mode decomposition with adaptive noise), Lempel-Ziv complexity (LZC) determination method, variational mode decomposition optimized by beluga whale optimization (BWO) algorithm (BVMD), sparrow search algorithm (SSA), and deep extreme learning machine (DELM), and is designated as ICEEMDAN-LZC-BVMD-SSA-DELM. The main contributions of this study can be summarized as follows.

(1) We propose a new hybrid deep learning model for high-accuracy forecasting of international soybean future closing price. The proposed deep learning forecasting model framework consists of ICEEMDAN decomposition of soybean futures closing price series, reconstruction of high-complexity subsequences based on LZC method, VMD secondary decomposition based on the BWO algorithm, the DELM forecasting algorithm and hyperparameter optimization based on the SSA. The experimental results on three soybean future data sets from China, Italy, and the United States reflect the ability of the proposed hybrid model to forecast with high accuracy. In addition, this forecasting method has good generalizability and can be extended to other work.

(2) A hybrid data preprocessing strategy of soybean future price was applied by integrating the LZC evaluation method. The ICEEMDAN algorithm was first used to decompose the original soybean future price. Then the LZC evaluation method was used to evaluate the resulting high-frequency subsequences, which was reconstructed and decomposed to further reduce the complexity of the series and help improve the accuracy of the forecast.

(3) The VMD algorithm optimized by the BWO algorithm is proposed as the secondary decomposition algorithm. Adopting the beluga whale optimization algorithm to optimize the parameters of the decomposition level K and penalty factor α of VMD can better preserve the signal characteristics and reduce the modal mixing problem so that the subsequent predicted soybean future price subsequence is smoother.

(4) The deep learning predictor for the secondary decomposition subsequences is optimized by the SSA algorithm, which consists of multiple extreme learning machine-auto encoder (ELM-AE) stacks with randomly generated input layer weights and thresholds. By optimizing the predictor using the SSA algorithm, we significantly improved the accuracy of soybean future price prediction.

2 Methodology

2.1 Framework of the proposed hybrid model

The framework of the hybrid model ICEEMDAN-LZC-BVMD-SSA-DELM proposed in this paper is shown in Fig. 1. The process of the hybrid model is summarized as follows.

Figure 1(a): The soybean future price series from China, Italy, and the United States were decomposed using the ICEEMDAN algorithm to obtain multiple primary decomposition subsequences.

Figure 1(b): The complexity of each subsequence of soybean future price decomposed by ICEEMDAN is evaluated by the

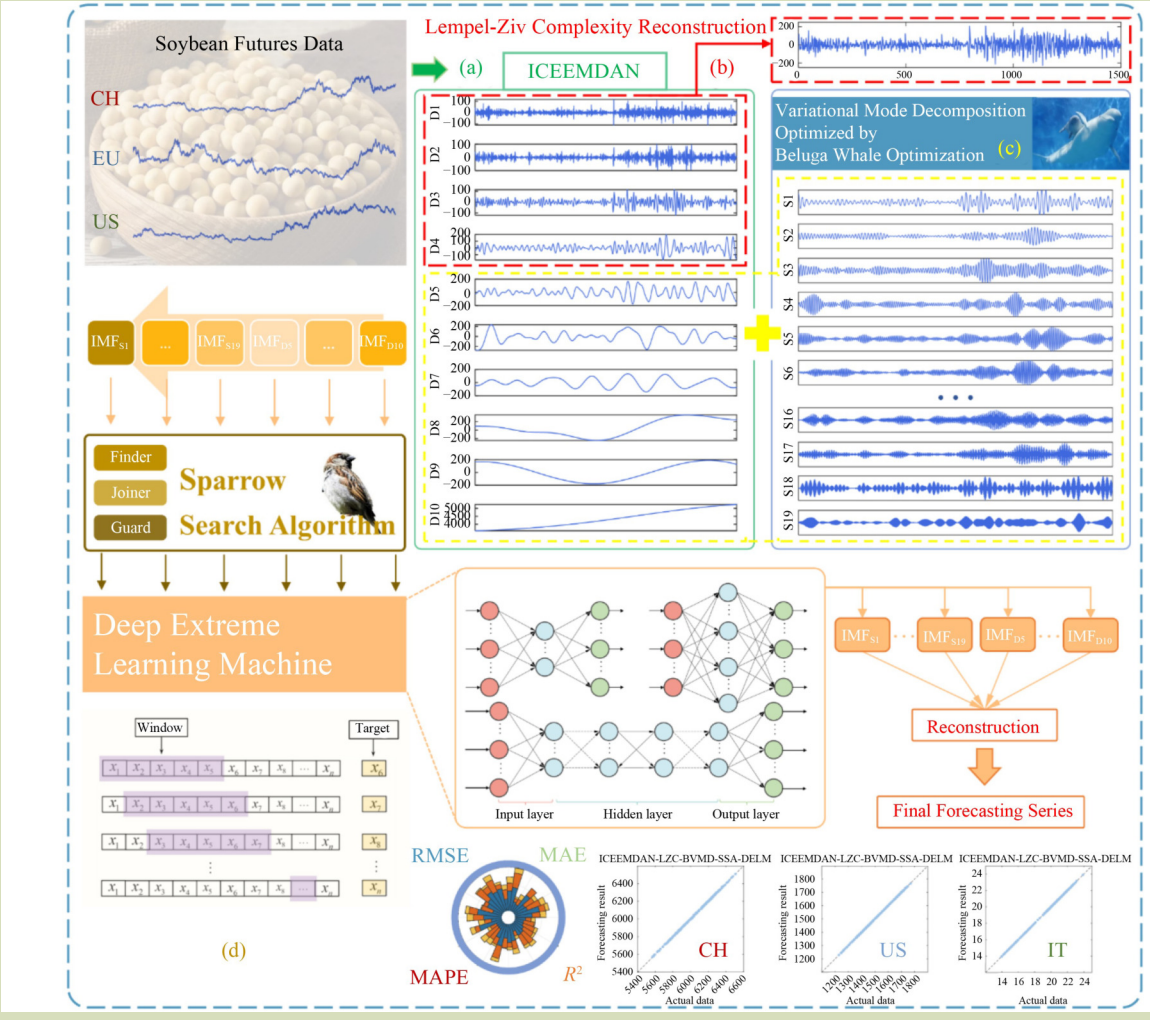


Fig. 1 Framework of the proposed hybrid model.

LZC determination method, and the selected subsequences with high complexity are reconstructed.

Figure 1(c): The reconstructed high-complexity sequences are decomposed again using VMD optimized by BWO algorithm to further decompose them into lower-complexity price subsequences. The price subsequences of low complexity derived from the LZC determination method is assembled with the price subsequences obtained from the BVMD secondary decomposition to obtain the combined soybean future price subsequences.

Figure 1(d): The DELM optimized by SSA is used to forecast all the subsequences obtained from the secondary decomposition, and then the forecasting results of all the subsequences are reconstructed to get the final soybean future price forecasts, and the weights and thresholds of DELM are optimized by the SSA algorithm.

2.2 Improved complete ensemble empirical mode decomposition with adaptive noise

To effectively suppress problems such as the modal blending phenomenon between the modes of the EMD algorithm and further improve the accuracy and stability of the decomposition, the ICEEMDAN algorithm includes adaptive noise suppression techniques[29]. The ICEEMDAN decomposition proceeds as follows.

Step 1: Introduce the operator E_1 in the EMD algorithm.

Step 2: Construct the noise signal $y^i(t) = y(t) + \beta_0 E_1(\eta^i(t))$ and use EMD to calculate the first-order residual $r_1(t) = M(y^i(t))$, which is added in the initial stage and is used to remove the noise. i is the number of times the noise is added.

Step 3: When j is 1, the first-order Intrinsic Mode Function (IMF) component is found, which is $\mathrm{IMF}_1(t) = y(t) - r_1(t)$.

Step 4: Using the average of the local means, the second-order residuals and the second-order IMF components are found as:

$$\mathrm{IMF}_2(t) = r_1(t) - \left\langle M\left(r_1(t) + \beta_1 E_2\left(\eta^i(t)\right)\right)\right\rangle \quad (1)$$

Step 5: Derive j-th order mode as:

$$r_j(t) = \left\langle M\left(r_{j-1}(t) + \beta_{j-1} E_j\left(\eta^i(t)\right)\right)\right\rangle \quad (2)$$

$$\mathrm{IMF}_j(t) = r_{j-1}(t) - r_j(t) \quad (3)$$

Step 6: Repeat Step 5 for all residuals and IMF components.

2.3 Complexity evaluation and data reconstruction

2.3.1 Lempel-Ziv complexity evaluation

The LZC algorithm was proposed by Lempel and Ziv to evaluate the randomness and chaos of sequences of a particular length[30]. The greater the complexity of a sequence, the more it converges to a random state and the richer the frequency component contained in the sequence[31]. It is widely used in applications such as electrocardiography signals, gene sequences and spoken texts.

2.3.2 Data reconstruction

Based on the complexity results of the subsequence and the Lempel-Ziv complexity principle, the process of dividing the high and low frequencies of soybean future price into the three markets and the reconstruction of the features are shown below.

Step 1: Compute the Lempel-Ziv complexity C_i for each subsequence S_i, $i = 1, 2, \ldots, m$.

Step 2: Set the critical value as $\lambda_0 = 0.8$ and find out the first k subsequences satisfying the following formula.

$$\eta = \frac{\sum_{i=1}^{k} C_i}{\sum_{i=1}^{m} C_i} \geqslant \lambda_0 \ (k \in m) \quad (4)$$

Step 3: Identify the sequence between subsequences S_1 to S_k as high-frequency sequences and the sequence between subsequences S_{k+1} to S_m as low-frequency sequences.

Step 4: The sequences between S_1 and S_k are summed and reconstructed as the final high-frequency reconstructed sequence S_{High}, and the data between S_{k+1} and S_m are summed and reconstructed as the final low-frequency reconstructed sequence S_{Low}.

Therefore, the determination of k in the subsequence reconstruction process based on the Lempel-Ziv complexity is crucial.

2.4 Improved variational model decomposition

2.4.1 Beluga whale optimization algorithm

The BWO algorithm is a meta-heuristic algorithm inspired by the observation and simulation of beluga whales during swimming, feeding and dying, which correspond to the three phases of exploration, exploitation and whale fall. In addition, a Lévy flight strategy is introduced in the development phase to

keep it from being limited to local optimality in this phase further[32].

2.4.2 Variational mode decomposition

VMD is a non-recursive signal processing algorithm[33], which decomposes the raw signal into a series of patterns with a specific spectral domain bandwidth, which is excellent for processing non-smooth, nonlinear and noisy signals.

For VMD, to obtain the best decomposition results, it is necessary to find the most suitable number of modes K and penalty factor α. K is the parameter that controls the number of modes obtained from the decomposition. When the value of K is too large, although more modes can be decomposed and the local characteristics of the signal can be better preserved, it also tends to over-fit the noise and local fluctuations, and even the phenomenon of modal confusion can occur. However, when the value of K is too small, the signal is decomposed into fewer modes, which is helpful in preserving the global features of the signal, but some detailed information is also lost, and a complete decomposition cannot be achieved.

Secondly, α is the parameter that controls the bandwidth of each mode. When the value of α is too large, it allows the bandwidth of the modes to be wider and better preserve the global features, but there will be excessive smoothing of the signal, which will cause some loss of details and local features. When the value of α is too small, it narrows the bandwidth of each mode, allowing better preservation of details and local features, but may also spread the signal too much[27].

2.4.3 Parameter optimization for variational mode decomposition

To investigate how to choose the appropriate K and α to achieve an optimal balance between signal feature preservation and noise removal, BVMD is proposed, with the pseudocode shown in **Algorithm 1**.

In this study, the fitness function for VMD optimized by the BWO algorithm is the envelope entropy[34]. The smaller the envelope entropy, the more signal features are retained, the more noise signals are removed, and the VMD effect is more thorough.

The optimal K and α of the VMD are determined by comparing the corresponding fitness function values at different update positions. The formula of envelope entropy is:

$$\begin{cases} p_j = a(j) \Big/ \sum_{i=1}^{N} a(j) \\ E_p = -\sum_{i=1}^{N} p_j \lg p_j \end{cases} \tag{5}$$

where, $a(j)$ is the envelope signal, p_j is the normalized form of $a(j)$, and E_p is envelope entropy.

Although several advantages of VMD and its improved models have been reported in the literature, the application of hybrid forecasting models based on VMD is not yet widespread in the field of agricultural price forecasting. Therefore, there is a need for researchers to conduct more relevant studies in the future to explore its potential for application in agricultural price forecasting and to further improve its performance.

2.5 SSA-optimized DELM

2.5.1 Sparrow search algorithm

The SSA algorithm was derived from the foraging and anti-predatory behavior of sparrow populations[35]. In the SSA algorithm, the magnitude of the fitness value indicates the strength of the finder to search for food, and finder position is updated during iteration as follows[35].

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(-\dfrac{i}{\alpha \cdot i_{\mathrm{ter}_{\max}}}\right) & R_2 < ST \\ X_{i,j}^{t} + Q \cdot L & R_2 \geqslant ST \end{cases} \tag{6}$$

where, t is the number of current iterations; j is the number of dimensions, $i_{\mathrm{ter}_{\max}}$ is the maximum number of iterations, α is a random number in (0,1], ST is the safety value and R_2 is the warning value. $R_2 < ST$ means the population is in a safe area and the finder can forage randomly, whereas $R_2 \geqslant ST$ means there is a predator around the population and it needs to be moved immediately to bring safe area for foraging.

If joiners in the population perceive that a finder has foraged for better food, they will immediately compete for it. If the joiners are successful, they are given access to finder food, otherwise they will continue to watch the finder for food. Update the location of the joiners using this formula[35]:

$$E_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\dfrac{X_{\mathrm{worst}}^{t} - X_{i,j}^{t}}{i^2}\right) & i > n/2 \\ X_P^{t+1} + \left|X_{i,j}^{t} - X_P^{t+1}\right| \cdot A^{+} \cdot L & \text{otherwise} \end{cases} \tag{7}$$

where, A is a $1 \times d$ matrix with each element randomly assigned 1 or −1. $i > n/2$ means that the i-th accession with a low fitness value is not getting food and needs to go elsewhere to feed. In addition, 10%–20% of sparrows in the entire population are randomly generated, named guards, and they randomly generate initial positions, as Li et al.[35]:

Algorithm 1 VMD optimized by BWO algorithm

Input

Reconstructed soybean future price series after ICEEMDAN decomposition y
Maximum number of iterations $T_{\max}$

Output

Number of decomposition modes K
Penalty factor α

Algorithm

1: Randomly initialize the initial position of each beluga whale in the search range $[K,\alpha]$

2: **for** $T = 1 : T_{\max}$ **do**

3: Decompose the soybean future price series y according to each beluga whale position $[K,\alpha]$ using VMD

4: Calculate the fitness value for each beluga whale using the envelope entropy of the decomposed sequence as the fitness function

5: Calculate the balance factor B_f and the whale fall probability W_f using the following equations:

$$B_f = B_o(1 - T/2T_{\max}$$

$$W_f = 0.1 - 0.05T/T_{\max}$$

6: **for** each X_i **do**

7: **if** $B_f > 0.5$ then

8: The beluga whale can enter the exploration phase and then update its position using the following equation:

$$\begin{cases} X_{i,j}^{T+1} = X_{i,Pj}^{T} + (X_{r,P1}^{T} - X_{i,Pj}^{T})(1 + r_1)\sin(2\pi r_2), \ldots\ldots\ldots j = \text{even} \\ X_{i,j}^{T+1} = X_{i,Pj}^{T} + (X_{r,P1}^{T} - X_{i,Pj}^{T})(1 + r_1)\cos(2\pi r_2), \ldots\ldots\ldots j = \text{odd} \end{cases}$$

9: **else** if $B_f \leqslant 0.5$ **then**

10: The beluga whale enters the exploitation phase and updates its position using the following formula:

$$X_i^{T+1} = r_3 X_{\text{best}}^{T} - r_4 X_i^{T} + C_1 L_f (X_r^{T} - X_i^{T})$$

11: **end if**

12: Calculate and rank the fitness values for the new position

13: **end for**

14: **for** each X_i **do**

15: **if** $B_f \leqslant W_f$ **then**

16: Update $X_{\text{step}} = (u_b - l_b)\exp(-C_2 T/T_{\max})$, $C_2 = 2W_f \times n$.

17: The beluga whale enters the whale fall phase and updates its position and calculates the new fitness value, and the position is updated by the following formula:

$$X_i^{T+1} = r_5 X_i^{T} - r_6 X_r^{T} + r_7 X_{\text{step}}$$

18: **end if**

19: **end for**

20: Find the optimal fitness value, i.e. the minimum value of the envelope entropy and record the corresponding beluga whale position parameter $[K,\alpha]$

21: **end for**

22: Output the optimal $[K,\alpha]$ for the VMD

$$C_{i,j}^{t+1} = \begin{cases} X_{\text{best}}^{t} + \beta \cdot \left| X_{i,j}^{t} - X_{\text{best}}^{t} \right| & f_i > f_g \\ X_{i,j}^{t} + P \cdot \left(\dfrac{\left| X_{i,j}^{t} - X_{\text{worst}}^{t} \right|}{(f_i - f_w) + \varepsilon} \right) & f_i = f_g \end{cases} \tag{8}$$

where, β is the step control parameter; X_{best}^{t} is the position of the safest sparrow; f_i is the adaptation value of the current individual sparrow; f_g and f_w denote the current best and worst adaptation values, respectively, and when $f_i > f_g$, it means that the sparrow is in the edge position at this time and is relatively threatened; $f_i = f_g$ means that the middle sparrow is aware of predator arrival at this time and thus tries to approach the nearby sparrow; ε is the minimum constant.

2.5.2 Deep extreme learning machine

The ELM has the advantages of strong generalization ability and fast learning speed[36]. For high-dimensional feature expressions, the output weight β of the hidden layer is expressed as:

$$\beta = \left(HH^T + \frac{1}{C}\right)H^T X \tag{9}$$

where, $\beta = \left[\begin{array}{cccc} \beta^1 & \beta^2 & \cdots & \beta^n \end{array}\right]$, $\beta^i\ (i = 1, 2, \cdots, n)$ is the connection weight of the hidden node to the output node, C is the regularization parameter and X is the input data.

For the equal dimensional feature expression, the output weight β of the hidden layer is expressed as:

$$\beta = TH^{-1} \tag{10}$$

The DELM is a stack of multiple ELM-AEs. The DELM enables the mapping of data features to improve both the forecasting accuracy and the generalization capability of the model[37]. The structure of DELM is shown in Fig. 2.

In the training process of the DELM, the training time series is generally used as the output of the first ELM-AE layer to find the output weight β^1. The output of the hidden layer of the DELM-AE model is then used as the input data of the second ELM-AE layer[37]. This method is repeated for each layer of the ELM-AE.

2.5.3 The deep extreme learning machine optimized by the sparrow search algorithm

The SSA is used to perform parameter optimization of the input layer weight and bias of the DELM. The optimization process flow is shown in Fig. 3.

3 Experiments and results

3.1 Soybean future data

This paper uses experimental time-series data from three typical soybean markets, China, Italy, and the United States, to validate the model. The three market price studies were developed based on soybean future price, which are widely used in the study of soybean prices as indicators of uniform markets, complete data and timely updates.

The data resolution in the experiment was a single day, and the specific price movements and detailed characteristics are shown in Figs. 4–6 and Table 1, respectively. Specifically, soybean prices for each market included 1500 data values, divided into two parts: a training set and a test set. The training set was data values 1–1200 and the test set was data values 1201–1500. The Chinese soybean future price data set was chosen from the closing prices of the main contract of the

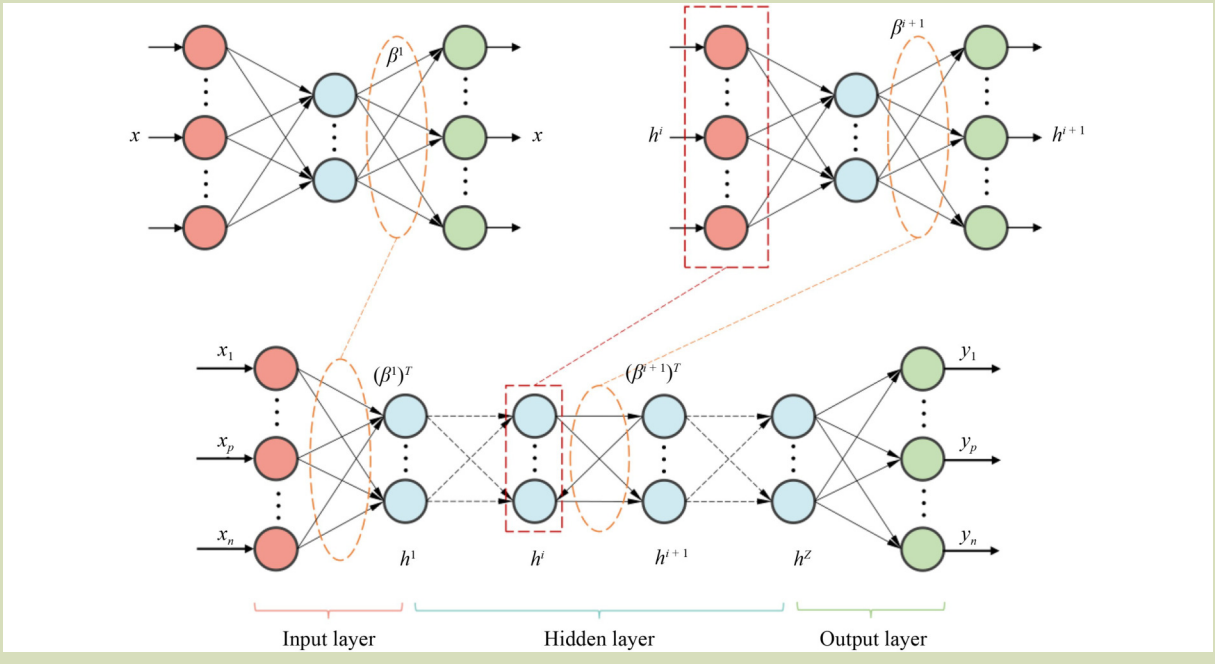


Fig. 2 Structure of the deep extreme learning machine.

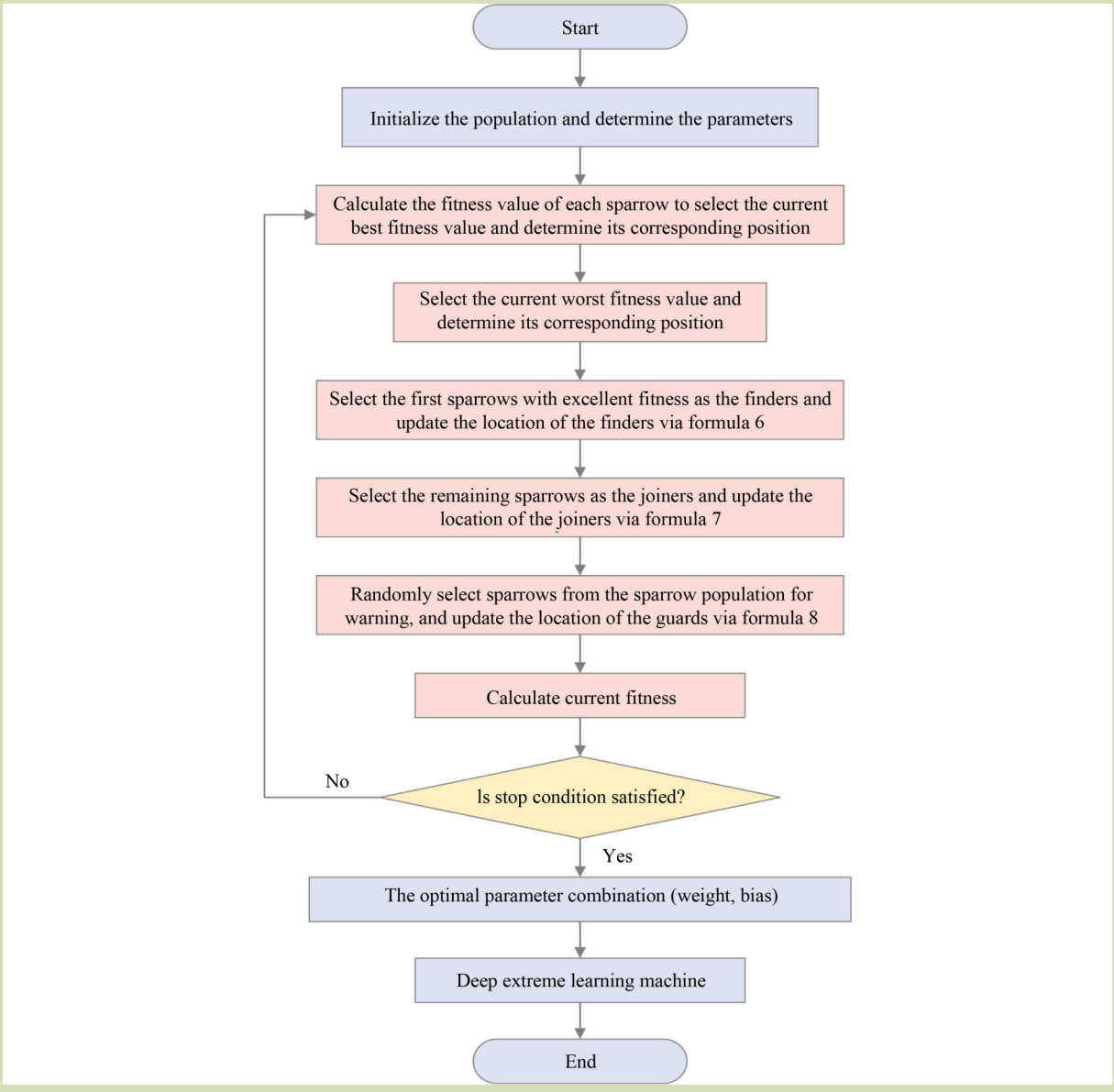


Fig. 3 The flow chart of the deep extreme learning machine optimized by the sparrow search algorithm.

Yellow Soybean 1 futures on the Dalian Exchange in China from 29 August 2016 to 2 November 2022. The United States soybean future price data set was selected from the closing prices of soybean future on the Chicago Board of Trade for the period 30 March 2017 to 27 January 2023. Additionally, to demonstrate model generalizability across financial instruments, the Italy soybean ETF (exchange traded fund) price data set was introduced. ETFs typically track specific indices or commodities, such as soybeans, which are directly linked to futures contracts. However, unlike futures contracts that are traded in the futures market, ETFs are listed and traded on exchanges like stocks, allowing for real-time buying and selling. On the WisdomTree Soybeans ETF, traded on the Italian Stock Exchange (Borsa Italiana), for the period from 19 June 2015 to 29 October 2021. The European Soybean ETF (exchange-traded fund) tracks the price of soybeans, and the unit of measurement is the 'share,' which represents a fraction of the total value of the ETF, linked to soybean future. All experiments were conducted on the MATLAB R2021b platform, which was run on a computer with Intel Core i5-12600K 3.70 GHz, 32G RAM, RTX 3070ti and Windows 10 Professional Edition operating system.

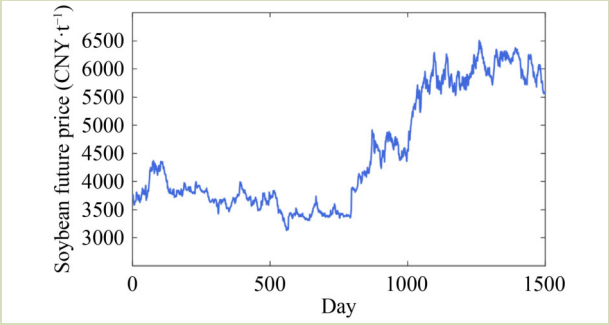


Fig. 4 Soybean future price in China.

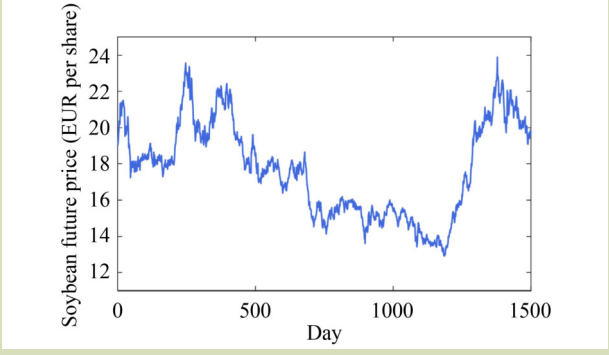


Fig. 5 Soybean future price in Italy.

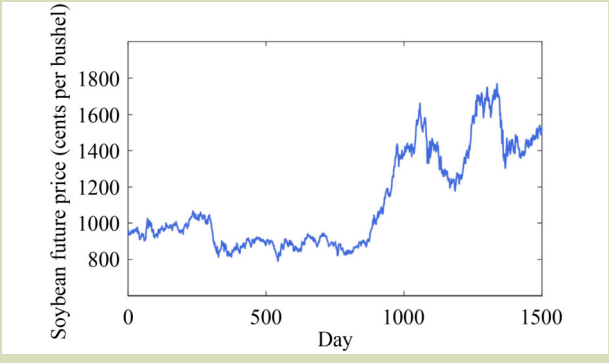


Fig. 6 Soybean future price in United States.

3.2 Experimental evaluation metrics

3.2.1 Performance evaluation indexes

Four main evaluation metrics were used to evaluate the predictions obtained in this study, including mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE) and R^2 calculated as:

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(t)-\widehat{X}(t))^2} \tag{11}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|X(t)-\widehat{X}(t)| \tag{12}$$

$$\mathrm{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{X(t)-\widehat{X}(t)}{X(t)}\right|\times 100 \tag{13}$$

$$R^2=\frac{\sum_{i=1}^{N}(X(t)-\overline{X}(t))(Y(t)-\overline{Y}(t))}{\sqrt{\sum_{i=1}^{N}(X(t)-\overline{X}(t))^2}\sqrt{\sum_{i=1}^{N}(Y(t)-\overline{Y}(t))^2}} \tag{14}$$

where, N is the number of samples, $X(t)$ is the actual closing price and $\widehat{X}(t)$ is the predicted closing price. RMSE, MAE and MAPE are inversely proportional to prediction accuracy, and R^2 is positively proportional to prediction accuracy. The optimal values of MAE, MAPE, RMSE and R^2 are 0, 0%, 0 and 1, respectively.

3.2.2 Performance improvement indexes

To compare the performance of the models, the percentage improvement of the evaluation indicators is used in this study to further evaluate the experimental results.

The percentage improvement of each evaluation indicator was calculated according to the formulae:

$$\begin{cases} P_{\mathrm{MAE}}=(\mathrm{MAE}_1-\mathrm{MAE}_2)/\mathrm{MAE}_1 \\ P_{\mathrm{MAPE}}=(\mathrm{MAPE}_1-\mathrm{MAPE}_2)/\mathrm{MAPE}_1 \\ P_{R^2}=\left(R_2^2-R_1^2\right)/R_2^2 \\ P_{\mathrm{RMSE}}=(\mathrm{RMSE}_1-\mathrm{RMSE}_2)/\mathrm{RMSE}_1 \end{cases} \tag{15}$$

A metric with a subscript of 1 is the performance of the baseline model, and a metric with a subscript of 2 is the performance of the proposed model.

Table 1 Descriptive statistics of the data set

Region	Max	Min	Median	Mean	Std.	Unit
China	6508	3129	3978	4510.3	1047.2	CNY·t^{-1}
Italy	23.9	12.9	17.8	17.7	2.6	EUR per share
United States	1769	791	983.5	1116.9	269.3	Cents per bushel

3.3 Model analysis

To demonstrate the superiority of the models proposed in this paper, 14 models were selected for comparison, including six single machine learning prediction models and eight hybrid prediction models. The single prediction models include ELM, RBF, deep belief network (DBN), LSTM, gated recurrent unit (GRU) and DELM. The hybrid prediction models include SSA-DELM, ICEEMDAN-DELM, ICEEMDAN-SSA-DELM, VMD-DELM, BVMD-DELM, BVMD-SSA-DELM, ICEEMDAN-LZC-BVMD-DELM, and the proposed model. For the simplicity of illustration and comparison, in some figures, S1 to S6 is used to denote the six single forecasting models above, and D1 to D8 to denote the eight hybrid forecasting models.

The results of the soybean future price error assessment for the single forecasting models and the hybrid forecasting models are shown in Table 2.

3.3.1 Comparative analysis of single models

To verify that DELM can accurately predict soybean future prices, experiments were conducted comparing DELM with other single models.

The ELM is a fast feed-forward neural network that reduces training time by randomly generating weights for the hidden layers. The RBF is a generating function that maps the input space to a high-dimensional space and is used to handle nonlinear classification and regression problems. The DBN is a multilayer neural network model consisting of multiple stacked restricted Boltzmann machines. The DBN has a self-encoder feature that The LSTM and GRU are variants of recurrent neural networks for processing time-series data. The LSTM has a memory capability to store historical information, and the GRU is a simplified version of the LSTM.

To visualize the forecasting performance of all single models, prediction error plots are given in Fig. 7. The prediction results of a single model for the Chinese soybean future price series are given in Fig. 8. Table 2 shows the accurate results of all the single models.

After the experiments, this paper found that for soybean futures price forecasting in the three markets, compared with the other five neural network models, the DELM had better evaluation performance, with MAPE values of 0.941, 1.23 and 1.32 for soybean future price forecasting in the three markets China, Italy, and the United States, respectively, which were the lowest values. In the error evaluation plot, DELM has the smallest variation, indicating that it outperforms the other models, which proves that DELM has the best stability and forecasting performance among the six single models proposed.

Table 2 Performance evaluation of the models

Model		China				Italy				United States			
		MAE	MAPE	RMSE	R^2	MAE	MAPE	RMSE	R^2	MAE	MAPE	RMSE	R^2
S1	ELM	63.3	1.05	77.5	0.843	0.237	1.22	0.312	0.983	25.0	1.65	32.4	0.933
S2	RBF	61.2	1.00	79.8	0.811	0.243	1.25	0.324	0.981	22.0	1.43	28.6	0.945
S3	DBN	62.5	1.04	77.1	0.880	0.249	1.27	0.322	0.982	23.1	1.55	29.8	0.958
S4	LSTM	68.8	1.14	84.8	0.812	0.278	1.42	0.362	0.977	24.8	1.64	31.9	0.939
S5	GRU	66.7	1.10	83.6	0.794	0.241	1.23	0.318	0.982	25.4	1.68	32.4	0.930
S6	DELM	57.3	0.941	72.4	0.883	0.243	1.23	0.318	0.982	19.9	1.32	26.2	0.962
D1	SSA-DELM	47.3	0.787	64.5	0.907	0.202	1.02	0.276	0.986	17.8	1.18	23.5	0.969
D2	ICEEMDAN-DELM	36.2	0.599	45.9	0.954	0.138	0.709	0.178	0.994	13.8	0.921	18.1	0.982
D3	VMD-DELM	42.5	0.707	53.2	0.932	0.139	0.714	0.184	0.994	14.4	0.955	18.7	0.980
D4	BVMD-DELM	19.5	0.323	23.4	0.988	0.0936	0.491	0.112	0.998	8.8	0.588	11.0	0.993
D5	ICEEMDAN-SSA-DELM	19.8	0.328	26.0	0.985	0.0763	0.388	0.105	0.998	7.35	0.487	9.69	0.995
D6	BVMD-SSA-DELM	7.82	0.129	9.94	0.9978	0.0361	0.184	0.047	0.9996	3.23	0.217	4.24	0.999
D7	ICEEMDAN-Lz-BVMD-DELM	6.36	0.106	7.79	0.9986	0.0262	0.133	0.033	0.9998	2.28	0.151	2.87	0.999
D8	ICEEMDAN-Lz-BVMD-SSA-DELM	2.42	0.042	3.01	0.9998	0.0115	0.058	0.015	0.9999	0.987	0.065	1.28	0.999

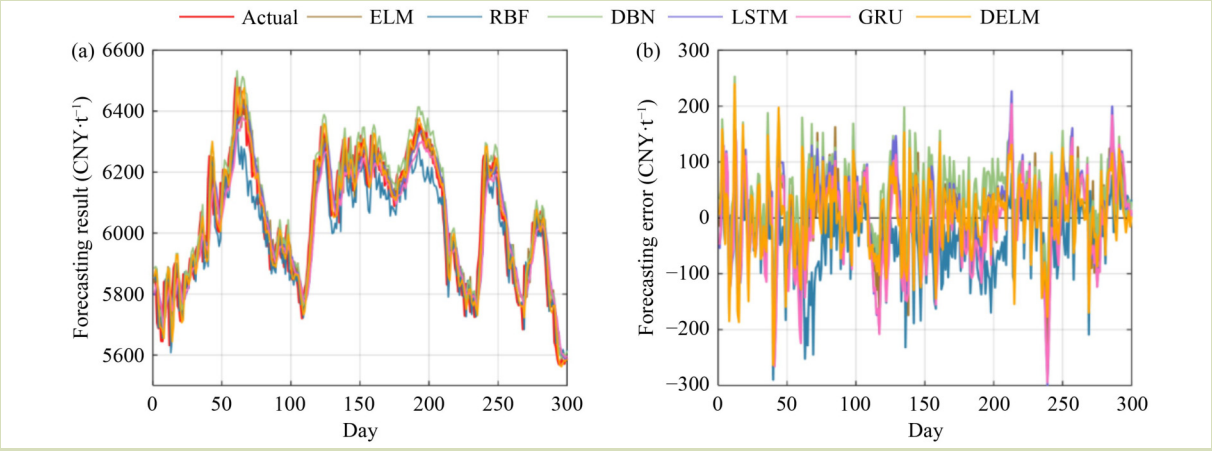


Fig. 7 Overview of soybean future price prediction in China. (a) Overall performance of prediction results; (b) prediction errors from various models, including ELM: Extreme Learning Machine, RBF: Radial Basis Function, DBN: Deep Belief Network, LSTM: Long Short-Term Memory, GRU: Gated Recurrent Unit, and DELM: Deep Extreme Learning Machine.

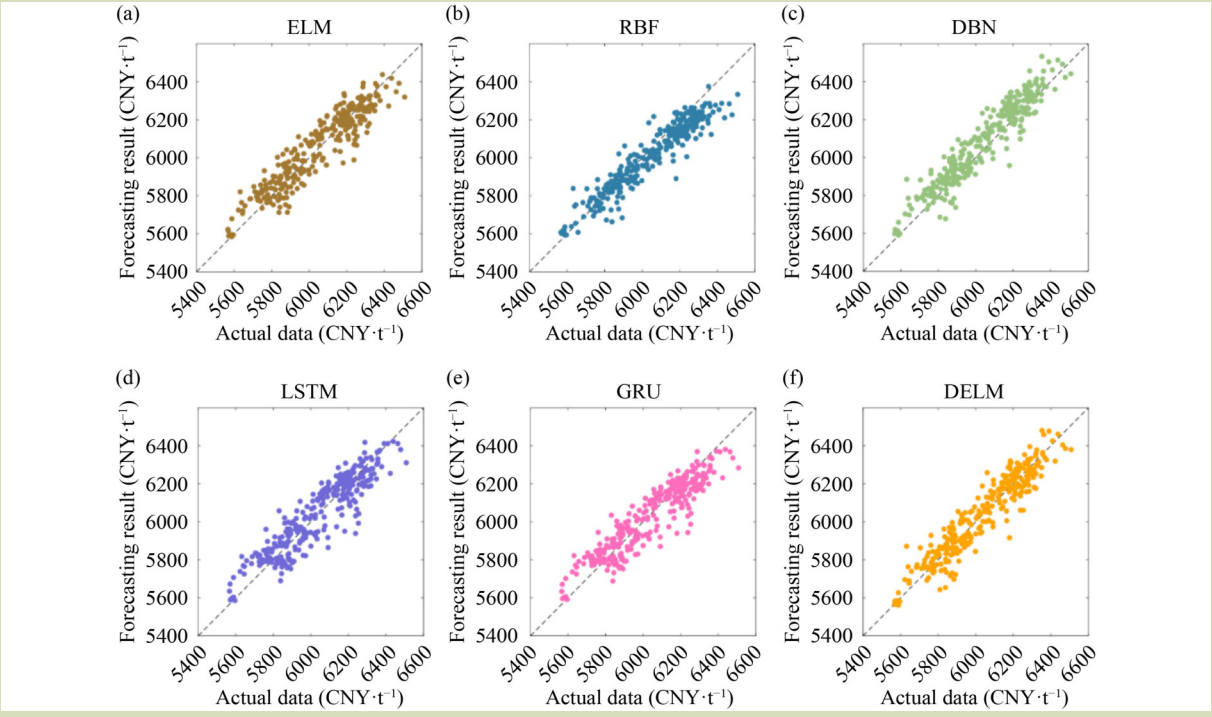


Fig. 8 Actual vs. predicted scatter plots of soybean future price series in China. (a–f) Results from six different models, with brief abbreviations for each model as indicated in Fig. 7.

3.3.2 Comparative analysis of decomposition methods

In this study, we first calculated the Lempel-Ziv complexity of soybean future data for three markets based on the principles of the LZC algorithm (Table 3).

According to the calculation results of the complexity of data subsequences in each market in Table 3, where "Res" is presented as the residuals—indicating the differences between the observed and predicted values. The calculation results of η

Table 3 Lempel-Ziv complexity of subsequences after ICEEMDAN in markets

Market	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	Res
China	0.928	0.689	0.513	0.337	0.232	0.119	0.091	0.028	0.028	0.021
Italy	0.914	0.717	0.555	0.330	0.196	0.119	0.077	0.042	0.035	0.021
United States	0.970	0.724	0.520	0.344	0.239	0.084	0.063	0.056	0.028	0.021

with the change of k value in the three markets are shown in Table 4.

Subsequences were divided into high and low-frequency sequences based on complexity similarity[38]. When k is 4, the value of η will be greater than the set threshold $\lambda_0(0.8)$. This shows that the first four subsequences obtained by ICEEMDAN decomposition are all high-frequency sequences, which can be reconstructed into a new subsequence. At the same time, the remaining ones represent low-frequency sequences and residual sequences and are retained. The process of the secondary decomposition for Chinese market is shown in Fig. 9. In the next step, the decomposed subsequences are used in forecasting models to assess their impact on prediction accuracy, as analyzed below.

Figure 9(a): Decomposition algorithms can help deep learning models to better handle large-scale data and can also enhance the generalization ability of neural networks, thus reducing prediction accuracy. The soybean future price is full of uncertainty and randomness, so decomposition is imperative. In this study, the experimental results of soybean future price prediction using ICEEMDAN and using the VMD algorithm in combination with DELM are compared with the prediction results of DELM alone, with the specific percentage improvements shown in Table 5. It is concluded that the use of the decomposition algorithm reduces the degree of nonlinearity in the raw data and improves the forecasting accuracy of soybean future price. Using MAPE as the evaluation metric, the ICEEMDAN-DELM for the three markets improved the forecasting accuracy by 36.3% compared with DELM, and the VMD-DELM improved the forecasting accuracy by 24.9% compared with DELM, which represents the effectiveness of the data decomposition algorithm.

Figure 9(b): The secondary decomposition method can decompose high-complexity time-series data into multiple stable regular subsequences efficiently. In contrast, ICEEMDAN is an adaptive decomposition algorithm, VMD requires empirical manual adjustment of the decomposition level K and penalty factor α, which increases the operational difficulty of this hybrid model and the risk of experimental failure due to blind adjustment. Therefore, BVMD was introduced to decompose the decomposed ICEEMDAN high-frequency subsequences. To fully evaluate the optimized decomposition method proposed in the soybean future price forecasting model, the BVMD-DELM was compared with the VMD-DELM in experiments (Table 5). It is evident that compared with VMD, BVMD optimized by BWO algorithm can better decompose the original series to reduce the difficulty faced by the forecasting network. For example, in comparison with VMD-DELM, the MAE, MAPE and RMSE of BVMD-DELM in China are reduced by 54.1%, 54.3% and 56.0%, respectively. While the R^2 value improved by 5.7%, demonstrating that the forecasting accuracy of the model will be significantly improved after optimization by the BWO algorithm.

Figure 9(c): Based on the high-frequency reconstructed series from the three markets, the randomness of the time series and the components included remain very high, which requires a secondary decomposition to continue the research. The data for the secondary decomposition of soybean future price in the three markets compared with the primary decomposition are shown in Table 5. Comparing the prediction results of ICEEMDAN-LZC-BVMD-DELM for the three markets with

Table 4 Variation of the Lempel-Ziv complexity ratio

Market	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	Res
China	0.311	0.541	0.713	0.826	0.904	0.944	0.974	0.984	0.993	1.000
Italy	0.304	0.542	0.727	0.836	0.902	0.942	0.967	0.981	0.993	1.000
United States	0.318	0.555	0.726	0.839	0.917	0.945	0.965	0.984	0.993	1.000

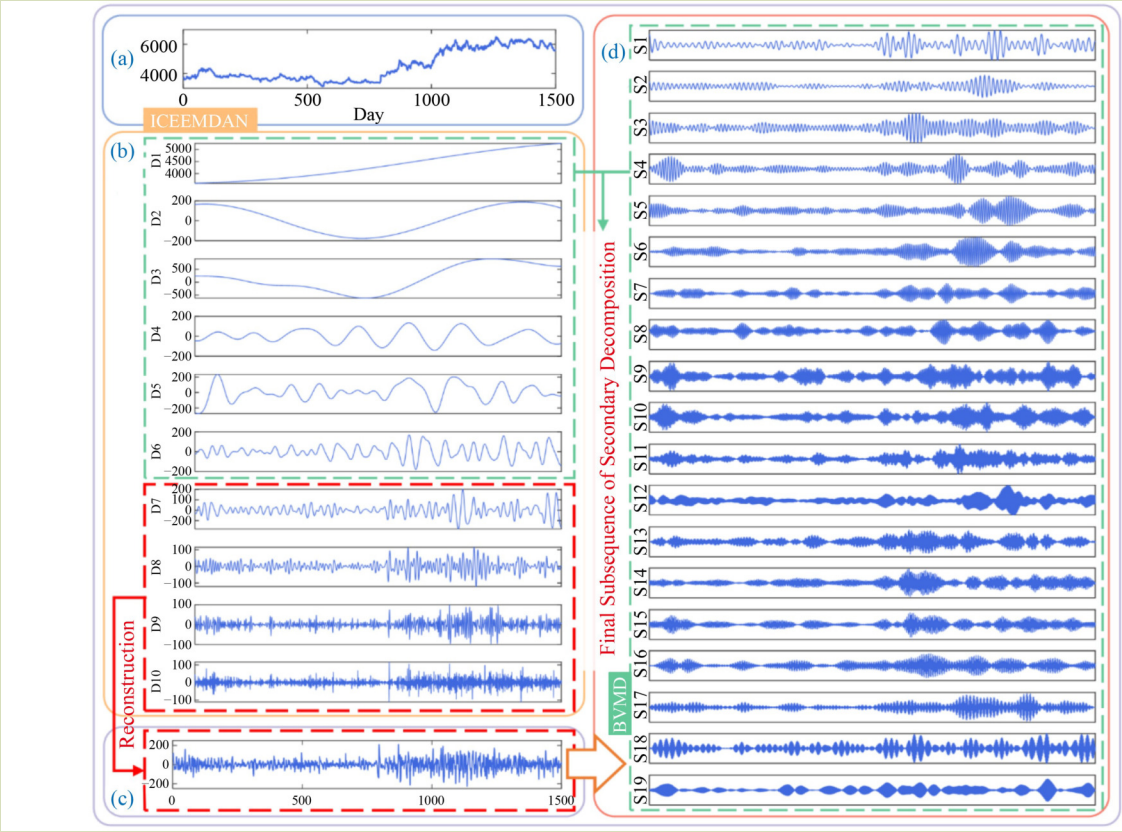


Fig. 9 Process of the secondary decomposition in the Chinese soybean market. (a) The raw soybean future data for the Chinese market; (b) the high-complexity and low-complexity subsequences, highlighted in red and green boxes, respectively, obtained from the primary decomposition; (c) the high-complexity subsequence from the red box is reconstructed into a new sequence; (d) the reconstructed high-complexity sequences are decomposed again into lower-complexity price subsequences.

ICEEMDAN-DELM and BVMD-DELM, respectively, using MAE as the error evaluation index, there was an improvement of 82.4% and 67.4% for China. This demonstrates that the secondary decomposition preprocessing based on ICEEMDAN-LZC-BVMD can provide better data identification and information extraction for the original time series compared with the conventional primary decomposition preprocessing and is a completely necessary data preprocessing method. The ICEEMDAN-LZC-BVMD-DELM shows an increase in MAE values of close to 90% in the three markets compared with the DELM. In the case of the original soybean futures price series with high complexity, containing considerable information and redundancy, which makes identification and prediction difficult, the ICEEMDAN-LZC-BVMD method can reduce the nonlinearity of the original data and can effectively improve the prediction accuracy. The ICEEMDAN and BVMD are used for secondary decomposition to obtain multiple subsequences, which can greatly reduce the difficulty of sequence identification, and the prediction effect of DELM is greatly improved.

3.3.3 Forecasting model optimization method enhancement analysis

To be able to thoroughly evaluate the performance of the SSA optimization algorithm, four comparison experiments were conducted for the three markets, with the specific parameters shown in the table below.

The results of the error assessment of the parametric optimization model are given in Table 2. The percentage improvement for the parametric optimization method is given in Table 6. The prediction results of soybean price series of the hybrid models for the three markets are shown in Figs. 10–12. The prediction errors of the hybrid models for the three markets are shown in Figs. 13–15. The scatter plot of hybrid

Table 5 Percentage improvements with decomposition methods

Model	Metric	China	Italy	United States
S6/D2	MAE	36.8%	43.2%	30.4%
	MAPE	36.3%	42.4%	30.4%
	RMSE	36.5%	44.0%	30.9%
	R^2	7.46%	1.25%	2.05%
S6/D3	MAE	25.8%	42.8%	27.2%
	MAPE	24.9%	42.0%	27.8%
	RMSE	26.4%	42.1%	28.6%
	R^2	2.61%	0.745%	1.08%
D3/D4	MAE	54.1%	32.7%	39.1%
	MAPE	54.3%	31.2%	38.4%
	RMSE	56.0%	39.1%	41.5%
	R^2	5.72%	0.391%	1.39%
D2/D7	MAE	82.4%	81.0%	83.5%
	MAPE	82.3%	81.2%	83.6%
	RMSE	83.0%	81.5%	84.2%
	R^2	4.44%	0.550%	1.77%
D4/D7	MAE	67.4%	72.0%	74.1%
	MAPE	67.2%	72.9%	74.3%
	RMSE	66.8%	70.5%	73.8%
	R^2	1.05%	0.210%	0.610%
S6/D7	MAE	88.9%	89.2%	88.5%
	MAPE	88.7%	89.2%	88.6%
	RMSE	89.2%	89.6%	89.1%
	R^2	11.6%	1.790%	3.78%

Note: The representative meanings of the models are shown in Table 2.

Table 6 Percentage improvement with optimization methods

Model	Metric	China	Italy	United States
S6/D1	MAE	17.5%	16.9%	10.3%
	MAPE	16.4%	17.1%	10.7%
	RMSE	10.8%	13.2%	10.2%
	R^2	2.67%	0.456%	0.753%
D2/D5	MAE	45.5%	44.7%	46.8%
	MAPE	45.2%	45.3%	47.1%
	RMSE	43.5%	41.0%	46.6%
	R^2	3.08%	0.371%	1.29%
D4/D6	MAE	59.9%	61.4%	63.3%
	MAPE	60.1%	62.5%	63.1%
	RMSE	57.6%	58.0%	61.3%
	R^2	0.972%	0.190%	0.561%

(*Continued*)

Model	Metric	China	Italy	United States
D7/D8	MAE	62.0%	56.1%	56.6%
	MAPE	60.4%	56.2%	57.0%
	RMSE	61.4%	54.5%	55.5%
	R^2	0.120%	0.010%	0.040%

Note: The representative meanings of the models are shown in Table 2.

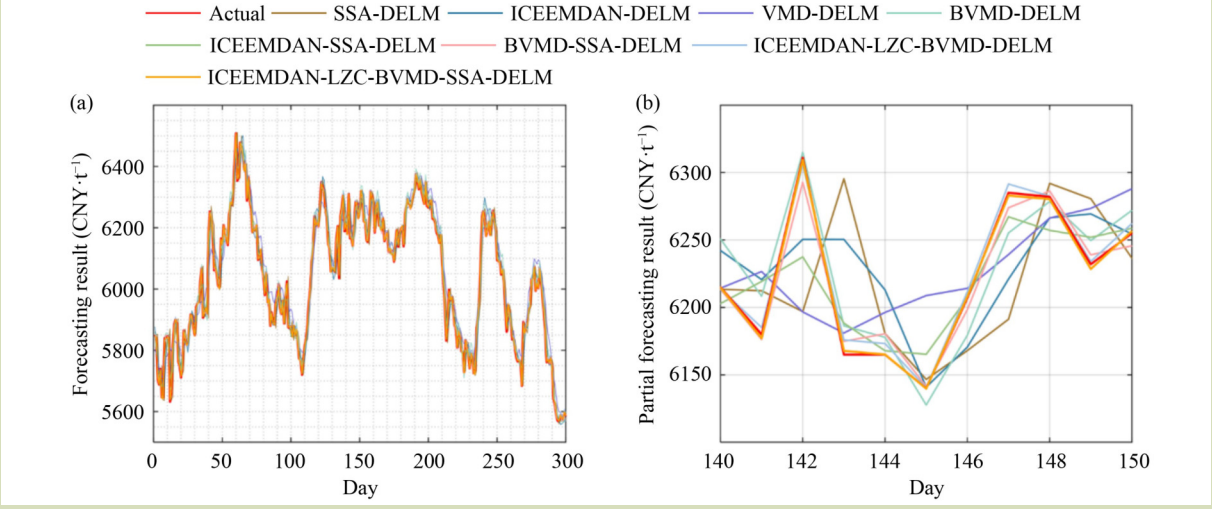


Fig. 10 Prediction of soybean price series using hybrid models in China. (a) Overall performance of the prediction results; (b) partial performance of the prediction results. The models include SSA-DELM: Sparrow Search Algorithm-Based Deep Extreme Learning Machine, ICEEMDAN-DELM: Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise-Based Deep Extreme Learning Machine, VMD-DELM: Variational Mode Decomposition-Based Deep Extreme Learning Machine, BVMD-DELM: Beluga Whale Optimization Algorithm-Based Variational Mode Decomposition and Deep Extreme Learning Machine, ICEEMDAN-SSA-DELM: Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Sparrow Search Algorithm-Based Deep Extreme Learning Machine, and ICEEMDAN-Lz-BVMD-DELM: Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, Lempel-Ziv Complexity, and Beluga Whale Optimization Algorithm-Based Variational Mode Decomposition and Deep Extreme Learning Machine.

models for three markets are shown in Figs. 16–18.

Figs. 10–18 and Table 6 reveal that the forecasting results of the model with the parametric optimization method outperformed the model without the parametric optimization method in all cases. This demonstrates that the use of the parameter optimization algorithm results in a more robust data analysis capability. The SSA parameter optimization method proposed in this paper can analyze the characteristics of each subsequence, select the optimal parameters and obtain better prediction results, effectively improving the accuracy and generalization of the prediction model. Therefore, the SSA method has good potential for improving the accuracy of soybean future price forecasting systems.

3.3.4 Comparison of the evaluated models

To evaluate the effectiveness of the ICEEMDAN-LZC-BVMD-SSA-DELM, single network models (ELM, RBF, DBN, LSTM, GRU and DELM) and hybrid models (SSA-DELM, ICEEMDAN-DELM, VMD-DELM, BVMD-DELM, ICEEMDAN- SSA-DELM, BVMD-SSA-DELM and ICEEMDAN-LZC-BVMD-DELM) were compared with the proposed models. The evaluation of these models is shown in Figs. 19–21.

All hybrid models gave better predictions than the single models. The hybrid models integrated multiple signal processing methods, which dramatically improved the performance of the models. As concluded above, the secondary

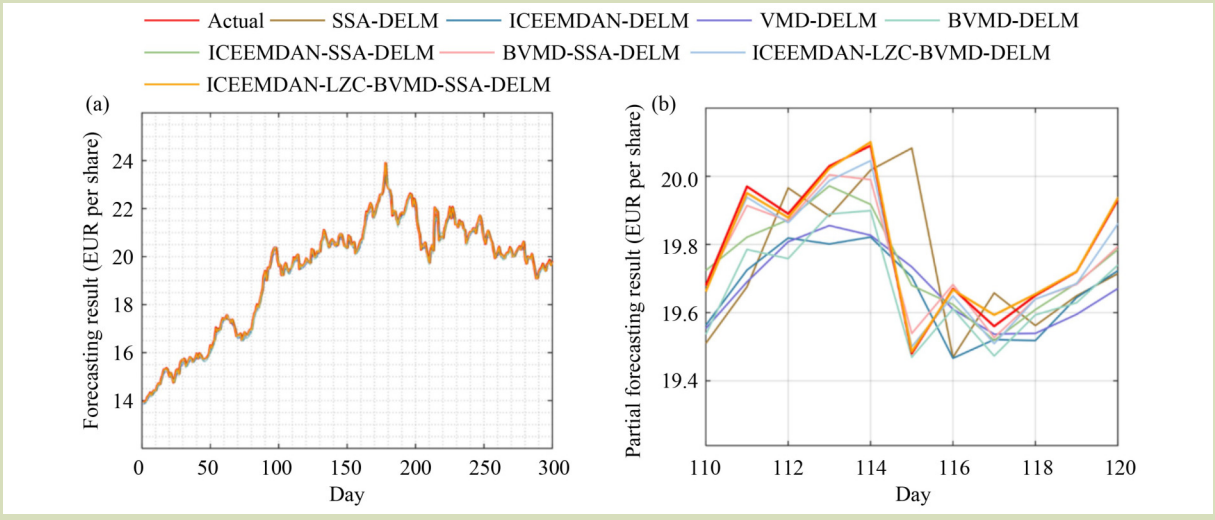


Fig. 11 Prediction of soybean price series using hybrid models in Italy. (a) Overall performance of the prediction results; (b) partial performance of the prediction results.

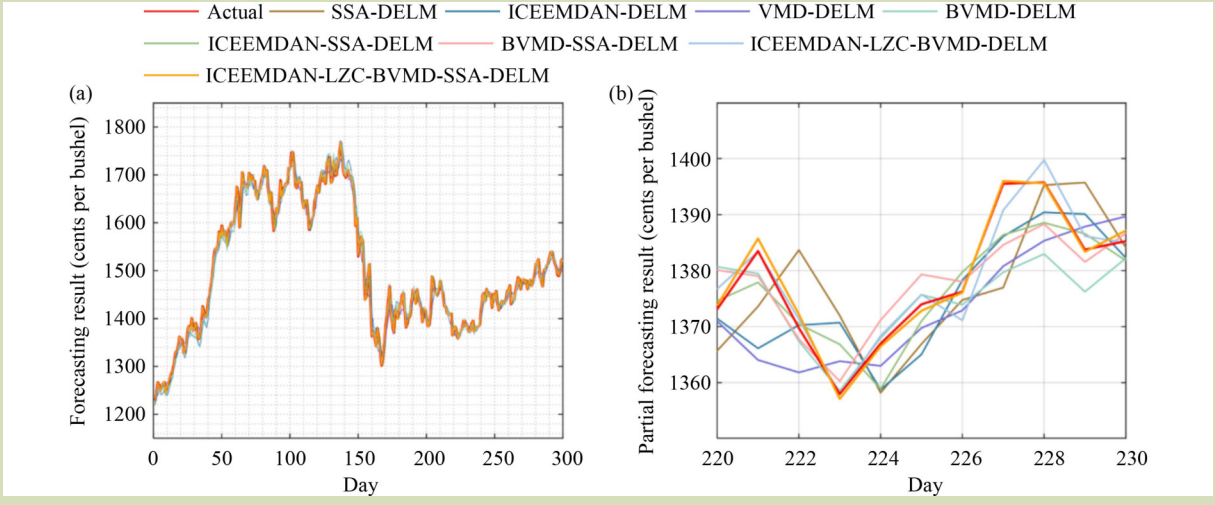


Fig. 12 Prediction of soybean price series using hybrid models in United States. (a) Overall performance of the prediction results; (b) partial performance of the prediction results.

decomposition method can effectively decompose highly complex time-series data into multiple subsequences with stable patterns. ICEEMDAN is an adaptive decomposition algorithm, but VMD requires empirical manual adjustment of the decomposition level K and penalty factor α, which increases the operational difficulty and poses the risk of experimental failure. Therefore, BVMD is introduced to decompose the reconstructed sequences of ICEEMDAN high-frequency subsequences. The secondary decomposition can boost the forecasting accuracy of the neural network by noise reduction and nonlinearity removal and also improve the generalization ability of the neural network by reducing the non-smoothness of the time-series data.

DELM was used to predict the soybean future price, and it performed the best of the six single models. However, the random generation of input layer weights and thresholds in DELM led to unstable predictions, negatively impacting prediction accuracy. To address this, the SSA algorithm was used for parameter optimization.

In a comparative analysis of the Chinese and the United States

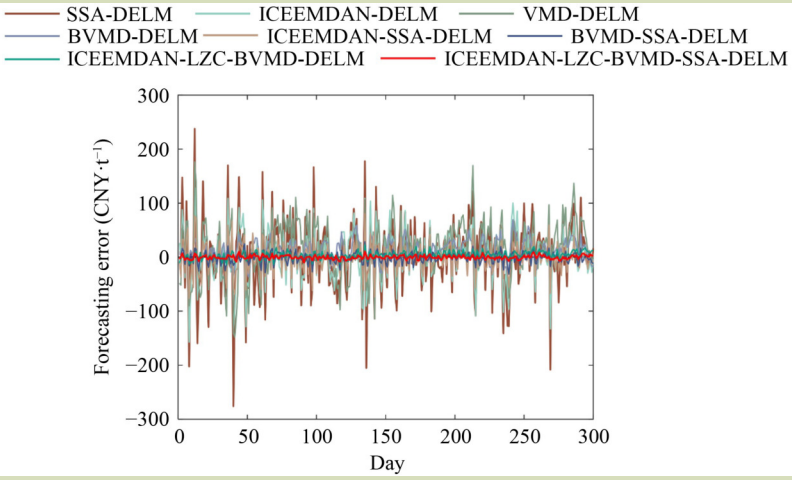


Fig. 13 Prediction errors of soybean price series of the hybrid models in China.

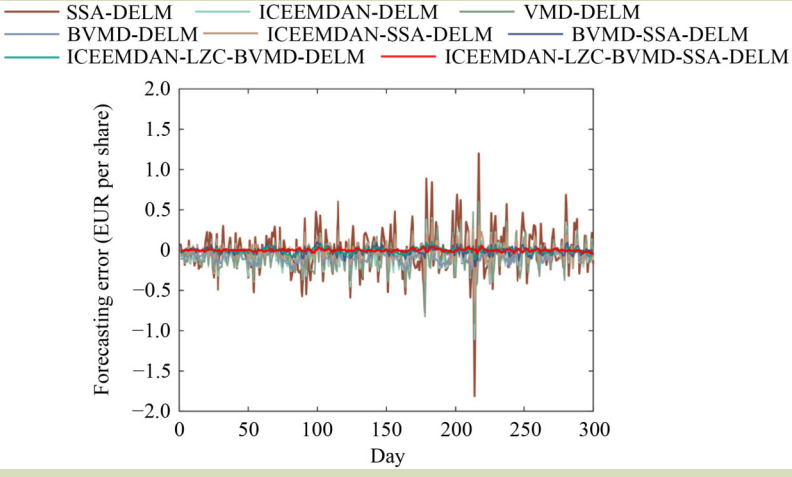


Fig. 14 Prediction errors of soybean price series with hybrid models in Italy.

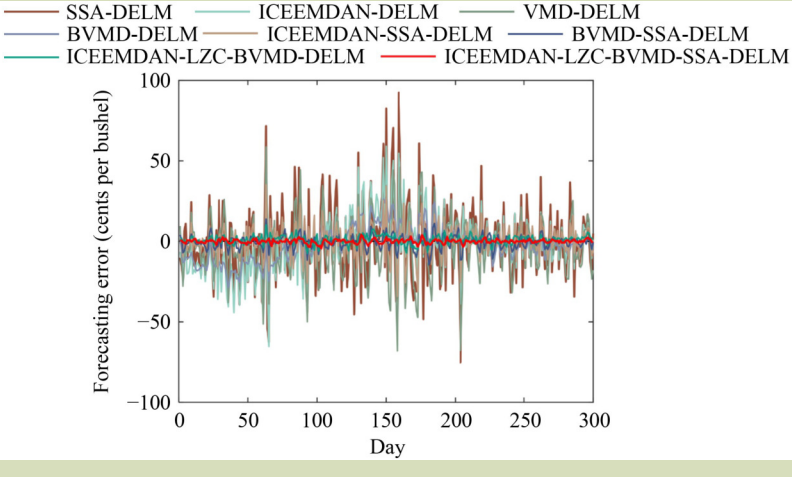


Fig. 15 Prediction errors of soybean price series with hybrid models in United States.

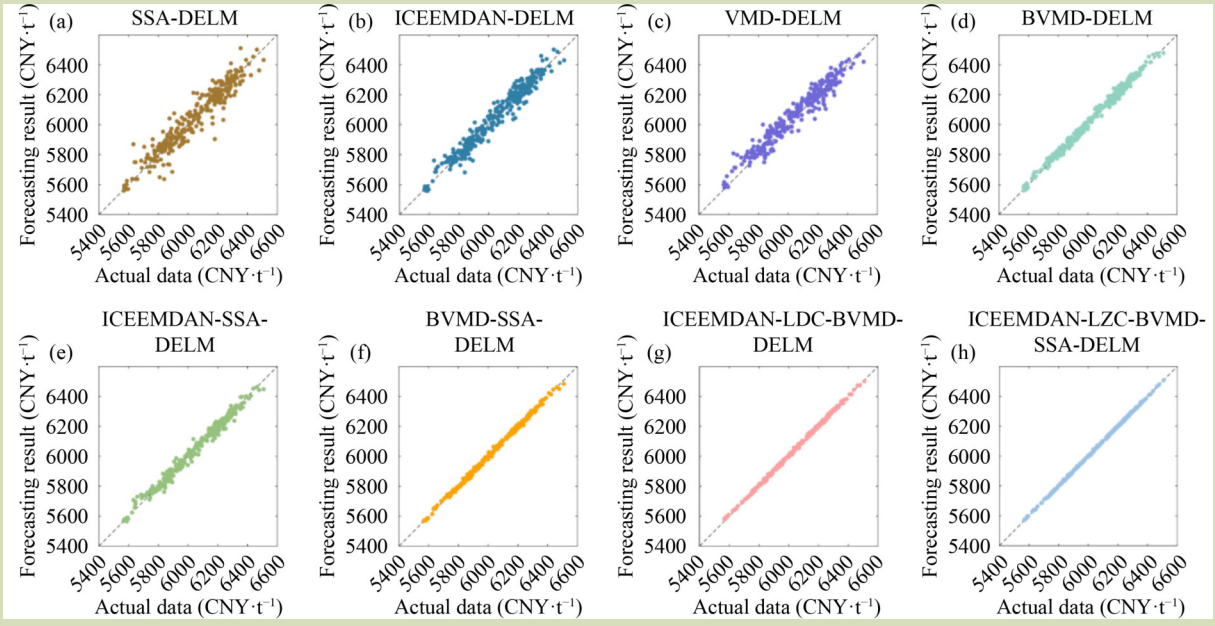


Fig. 16 Actual vs. predicted scatter plots of soybean future price series in China. (a–h) Results from eight different models, with brief abbreviations for each model as indicated in Fig. 10.

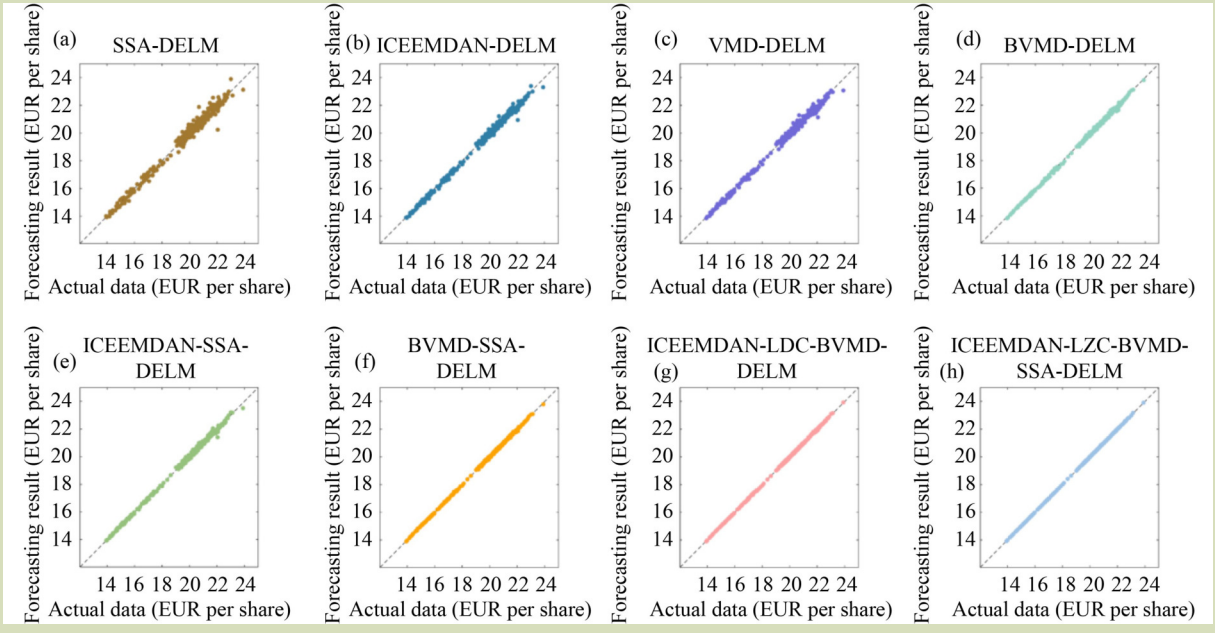


Fig. 17 Actual vs. predicted scatter plots of soybean future price series in Italy. (a–h) Results from eight different models, with brief abbreviations for each model as indicated in Fig. 10.

soybean future price data sets against the Italy soybean ETF price data set, it is found that the MAPE results of the single network models and hybrid models for the Italy data set were improved more substantively compared with the other two data sets. This finding indicates that the proposed model enhances overall forecasting performance and excels in specific markets (e.g., extremes or outliers), as illustrated in Fig. 22, which shows the MAPE comparison of representative models

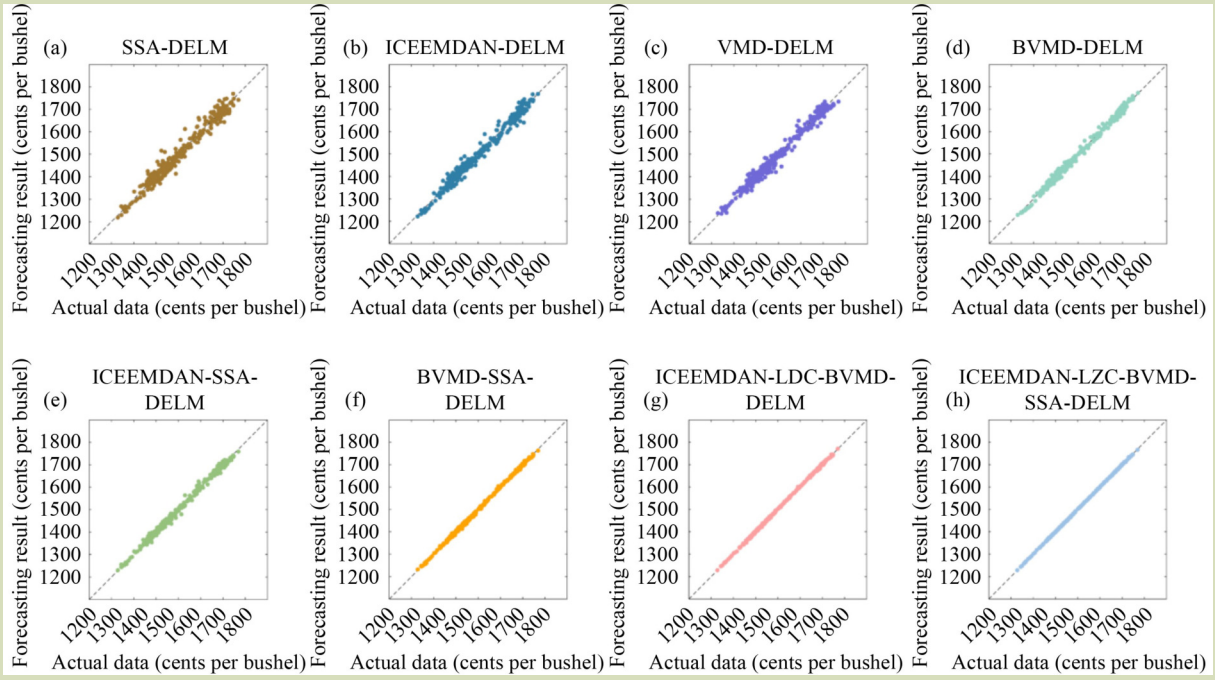


Fig. 18 Actual vs. predicted scatter plots of soybean future price series in United States. (a–h) Results from eight different models, with brief abbreviations for each model as indicated in Fig. 10.

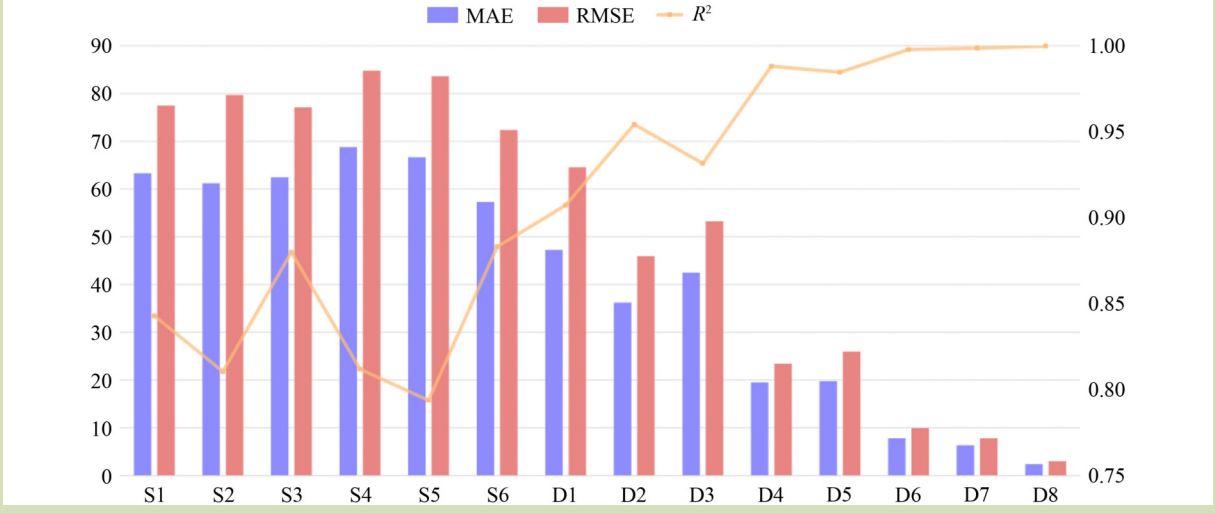


Fig. 19 Evaluation of all models in China.

across the markets. While the improvement in R^2 was not statistically significant, as shown in Fig. 23, this may be attributed to the already high R^2 value of the Italy data set. Nonetheless, values close to 1 indicate that the proposed model effectively fits the data and the proposed model performs quite well on the Italy soybean ETF price data set.

Of the models compared, the proposed model achieved the highest prediction accuracy in the three markets and for all evaluation metrics. It fully illustrates that the ICEEMDAN-LZC-BVMD-SSA-DELM model can effectively predict soybean futures price series and has strong prospects for application and further research in the field of agricultural futures price forecasting.

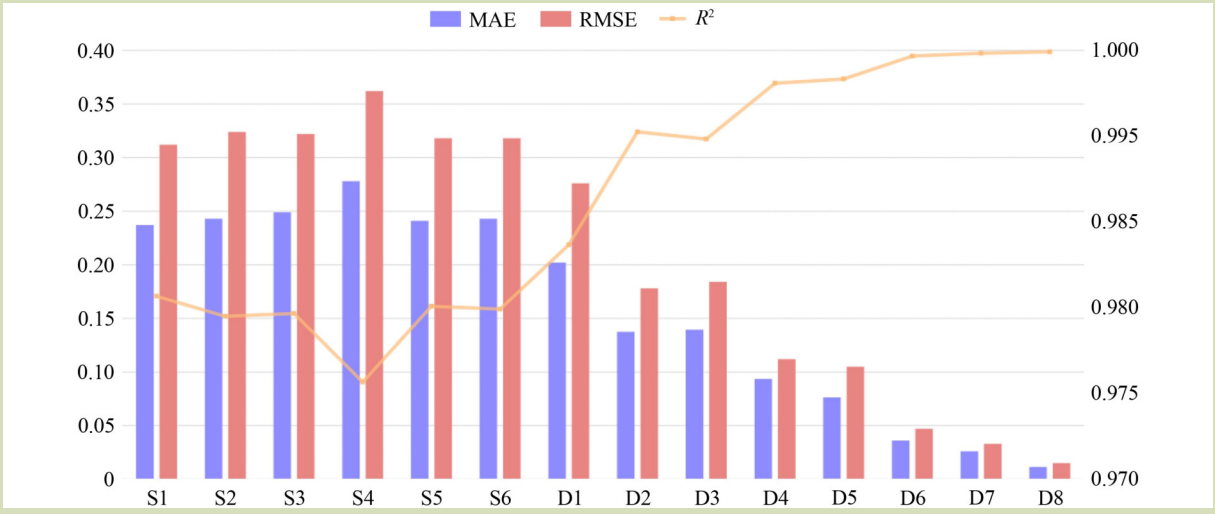


Fig. 20 Evaluation of all models in Italy.

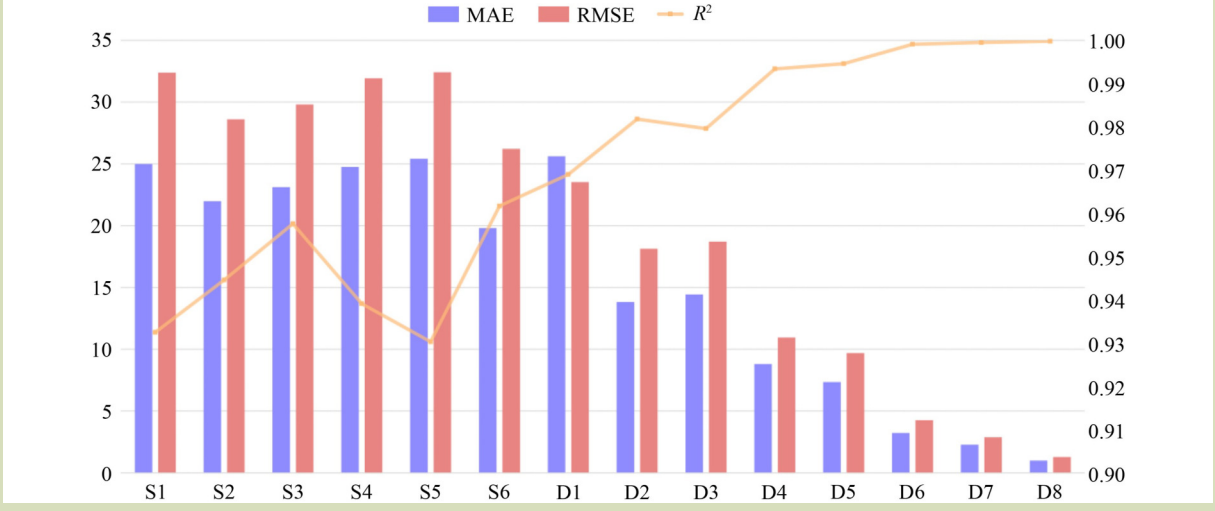


Fig. 21 Evaluation of all models in United States.

4 Conclusions

To achieve high-precision forecasting of soybean future price series, an improved deep learning model for soybean future prices prediction with hybrid data preprocessing strategy is proposed.

The main conclusions from the analysis of the comparative experimental results are as follows. (1) Of the six single models compared, the DELM model performs the best. (2) The secondary decomposition method has greatly improved the forecasting effect of the proposed model. In contrast to the non-decomposition and primary decomposition methods, the subsequences obtained by the secondary decomposition proposed are more stable and orderly, which greatly reduces the difficulty of time-series identification and is an excellent data preprocessing method. (3) The BWO algorithm used in this study can significantly enhance the decomposition effect of VMD, and the SSA algorithm can effectively increase the prediction accuracy of DELM network. (4) The proposed hybrid model achieves sound experimental results for soybean futures closing price time series in the three markets evaluated. Notably, the proposed model also performs well in predicting

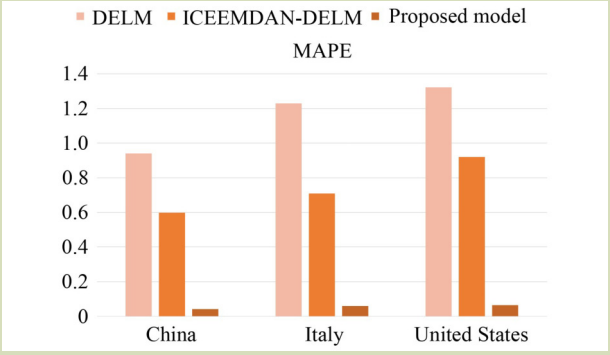


Fig. 22 Mean absolute percentage error comparison of models across three markets.

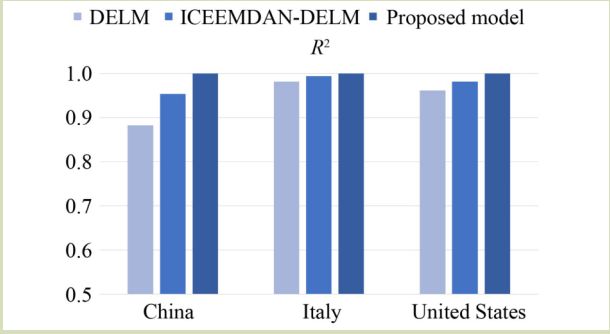


Fig. 23 R^2 comparison of models across the three markets.

the typical more flexible Italy soybean ETF price data set, demonstrating its strong potential for generalization.

The hybrid forecasting framework proposed in this paper provides new ideas for agricultural price forecasting research and also has clear potential for generalization and application to the forecasting of other financial time series.

Acknowledgements

The study is fully supported by the National Natural Science Foundation of China (52072412).

Availability of data and materials

Data for this study were obtained from two major commodity exchanges: Dalian Commodity Exchange (DCE) and Chicago Mercantile Exchange (CME Group). Additionally, financial information regarding soybean ETFs was retrieved from the Bloomberg platform.

Compliance with ethics guidelines

Dingya Chen, Hui Liu, Yanfei Li, and Zhu Duan declare that they have no conflicts of interest or financial conflicts to disclose. All applicable institutional and national guidelines for the care and use of animals were followed.

REFERENCES

1. Isengildina O, Irwin S H, Good D L. Evaluation of USDA interval forecasts of corn and soybean prices. *American Journal of Agricultural Economics*, 2004, **86**(4): 990–1004
2. Yang Q H, Du X Q, Wang Z H, Meng Z C, Ma Z H, Zhang Q. A review of core agricultural robot technologies for crop productions. *Computers and Electronics in Agriculture*, 2023, **206**(13): 107701
3. Kenyon D E. Producer ability to forecast harvest corn and soybean prices. *Applied Economic Perspectives and Policy*, 2001, **23**(1): 151–162
4. Darekar A, Reddy A A. Predicting market price of soybean in major india studies through ARIMA model. *Indian Journal of Pulses Research*, 2017, **30**(2): 73–76
5. Panasa V, Kumari R V, Ramakrishna G, Kaviraju S K. Maize price forecasting using auto regressive integrated moving average (ARIMA) model. *International Journal of Current Microbiology and Applied Sciences*, 2017, **6**(8): 2887–2895
6. Bhardwaj S P, Paul R K, Singh D R, Singh K N. An empirical investigation of Arima and Garch models in agricultural price forecasting. *Economic Affairs*, 2014, **59**(3): 415–428
7. Şahinli M A. Potato price forecasting with Holt-Winters and ARIMA methods: a case study. *American Journal of Potato Research*, 2020, **97**(4): 336–346
8. Gao Z M, Khot L R, Naidu R A, Zhang Q. Early detection of grapevine leafroll disease in a red-berried wine grape cultivar using hyperspectral imaging. *Computers and Electronics in Agriculture*, 2020, **179**: 105807
9. Yin H L, Jin D, Gu Y H, Park C J, Han S K, Yoo S J. STL-ATTLSTM: vegetable price forecasting using STL and attention mechanism-based LSTM. *Agriculture*, 2020, **10**(12): 612
10. Craessaerts G, De Baerdemaeker J, Saeys W. Fault diagnostic systems for agricultural machinery. *Biosystems Engineering*,

2010, **106**(1): 26–36
11. Mahto A K, Alam M A, Biswas R, Ahmed J, Alam S I. Short-term forecasting of agriculture commodities in context of indian market for sustainable agriculture by using the artificial neural network. *Journal of Food Quality*, 2021, **2021**: e9939906
12. Jaiswal R, Jha G K, Kumar R R, Choudhary K. Deep long short-term memory based model for agricultural price forecasting. *Neural Computing & Applications*, 2022, **34**(6): 4661–4676
13. Zong J J, Zhu Q Y. Price forecasting for agricultural products based on BP and RBF Neural Network. In: 2012 IEEE International Conference on Computer Science and Automation Engineering. *IEEE*, 2012, 607–610
14. Xu X J. Corn cash price forecasting. *American Journal of Agricultural Economics*, 2020, **102**(4): 1297–1320
15. Xu X J, Zhang Y. Canola and soybean oil price forecasts via neural networks. *Advances in Computational Intelligence*, 2022, **2**(5): 32
16. Liu H, Yin S, Chen C, Duan Z. Data multi-scale decomposition strategies for air pollution forecasting: a comprehensive review. *Journal of Cleaner Production*, 2020, **277**: 124023
17. Liu H, Wu H P, Li Y F. Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction. *Energy Conversion and Management*, 2018, **161**: 266–283
18. Yin S, Liu H, Duan Z. Hourly PM2.5 concentration multi-step forecasting method based on extreme learning machine, boosting algorithm and error correction model. *Digital Signal Processing*, 2021, **118**: 103221
19. Zhang Y L, Na S G. A novel agricultural commodity price forecasting model based on fuzzy information granulation and MEA-SVM model. *Mathematical Problems in Engineering*, 2018, **2018**: e2540681
20. Li G Q, Chen W, Li D H, Wang D J, Xu S W. Comparative study of short-term forecasting methods for soybean oil futures based on LSTM, SVR, ES and wavelet transformation. *Journal of Physics: Conference Series*, 2020, **1682**(1): 012007
21. Wang J, Wang Z, Li X, Zhou H. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *International Journal of Forecasting*, 2022, **38**(1): 21–34
22. Liang J Y, Jia G Z. China futures price forecasting based on online search and information transfer. *Data Science and Management*, 2022, **5**(4): 187–198
23. Zhang D Q, Zang G M, Li J, Ma K P, Liu H. Prediction of soybean price in China using QR-RBF neural network model. *Computers and Electronics in Agriculture*, 2018, **154**: 10–17
24. Liu H, Long Z H. An improved deep learning model for predicting stock market price time series. *Digital Signal Processing*, 2020, **102**: 102741
25. Sun W, Huang C C. A novel carbon price prediction model combines the secondary decomposition algorithm and the long short-term memory network. *Energy*, 2020, **207**: 118294
26. Zhu H M, Xu R, Deng H Y. A novel STL-based hybrid model for forecasting hog price in China. *Computers and Electronics in Agriculture*, 2022, **198**: 107068
27. Liu H, Duan Z, Han F Z, Li Y F. Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm. *Energy Conversion and Management*, 2018, **156**: 525–541
28. Liu H, Zhang X Y. AQI time series prediction based on a hybrid data decomposition and echo state networks. *Environmental Science and Pollution Research International*, 2021, **28**(37): 51160–51182
29. Colominas M A, Schlotthauer G, Torres M E. Improved complete ensemble EMD: a suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control*, 2014, **14**: 19–29
30. Aboy M, Hornero R, Abasolo D, Alvarez D. Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Transactions on Biomedical Engineering*, 2006, **53**(11): 2282–2288
31. Kurtin D L, Scott G, Hebron H, Skeldon A C, Violante I R. Task-based differences in brain state dynamics and their relation to cognitive ability. *NeuroImage*, 2023, **271**: 119945
32. Zhong C T, Li G, Meng Z. Beluga whale optimization: a novel nature-inspired metaheuristic algorithm. *Knowledge-Based Systems*, 2022, **251**: 109215
33. Dragomiretskiy K, Zosso D. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 2014, **62**(3): 531–544
34. Liu Q, Zhao R Z, Yang B Z. Research of fault recognition method of rolling bearings based on K-VMD envelope entropy and SVM. *Noise and Vibration Control*, 2022, **42**(3): 92
35. Li X H, Guo M M, Zhang R R, Chen G M. A data-driven prediction model for maximum pitting corrosion depth of subsea oil pipelines using SSA-LSTM approach. *Ocean Engineering*, 2022, **261**: 112062
36. Mi X W, Liu H, Li Y F. Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. *Energy Conversion and Management*, 2017, **151**: 709–722
37. Tissera M D, McDonnell M D. Deep extreme learning machines: supervised autoencoding architecture for classification. *Neurocomputing*, 2016, **174**(Part A): 42–49
38. Zhu B Z, Ye S X, Wang P, He K J, Zhang T, Wei Y M. A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. *Energy Economics*, 2018, **70**: 143–157