Risk control of metal raw materials based on deep learning

Cheng Zhao¹, Yexin Wang², Yuefeng Cen³, Lebin Wu¹ and Jie Zhou¹

Abstract
Stabilizing the purchase cost of metal raw materials is of great significance to the metal manufacturing industry. Most enterprises use futures hedging strategies to cope with the risks arising from fluctuations in the prices of metal raw materials. However, the difference between spot and futures prices of metals makes it impossible to fully control the risk. In order to further improve the efficiency of hedging and controlling corporate risks, it is necessary to accurately predict futures prices. However, the decomposition algorithms in traditional mixed models are prone to modal aliasing and have limited ability to extract nonlinear features from futures prices. Therefore, this paper proposes a variational modal decomposition-sample entropy-Cascaded Long Short-Term Memory Neural Network Model (VMD-SE-CLSTM). This paper proposes SE combined with VMD algorithm to determine the decomposition number to suppress the aliasing phenomenon of subsequence patterns, and introduces CLSTM network to improve the extraction ability of nonlinear features in futures data. The experiments are compared with 10 mainstream model methods and the method proposed in this paper. The experimental results show that the model reduces the prediction error and improves the prediction accuracy of the model, which is of great significance for enterprises to improve hedging efficiency, reduce operating risks and control production costs.

Keywords
Cascaded long-short term memory network, hedging strategies, risk control, sample entropy, variational mode decomposition

Introduction
Metal is vital for industrial production. With the rapid development of the modern industry, the enterprises’ demand for metal materials is gradually rising, denoting that metal materials have an important role in social development.

Changes in metal futures prices are influenced by a variety of factors, such as monetary policy, environmental protection policy, resource endowment, supply, demand, etc. It is a typical non-stationary time series. For metal processing enterprises, the purchase cost of raw materials can account for more than 60% of enterprise costs, and price fluctuations can have a significant impact on enterprise production. Hence, enterprises wish to establish a profit and loss hedging mechanism between the metal raw material futures market and the spot market through hedging strategies to reduce the risks associated with futures price fluctuations. Among them, hedging strategy refers to a company’s trading in the spot market and futures market in equal but opposite directions, expecting to buy or sell futures at some point in time to compensate for the price risk caused by price fluctuations in the spot market. The error between spot and futures is called the forward margin. The premise of this strategy is that the volatility of the forward margin is highly coordinated. However, the volatility of the forward margin is obvious, and when the difference is too large, it is difficult to hedge risks using conventional financial means. Therefore, if we can achieve accurate forecasting of metal futures prices in the futures market, we can guide...
With the continuous development of the futures market, two types of futures price forecasting models have been formed: traditional single forecasting models and hybrid forecasting models. A traditional single forecasting model is one in which data are simply processed and fed directly into the model for forecasting. Dooley and Lenihan compared the results of a lagged futures price model and an autoregressive integrated moving average model (ARIMA) for forecasting lead and zinc price series and found that ARIMA achieved better forecasting results. Although statistical models have improved forecasting performance to some extent, they all have inherent limitations and have a smaller range of applications. Statistical models can capture the linearity of the data, however, there are many nonlinearities in the price series of metal raw materials. As a result, statistical models cannot achieve good prediction accuracy. In recent years, the use of artificial intelligence models for time series forecasting has become more and more widely used. Liu et al. used decision trees to predict the short- and long-term changes in copper prices and achieved robust prediction results with better prediction accuracy than traditional regression models, but traditional artificial intelligence models dealing with time series tend to fall into local optima. Among them, the recurrent neural network-based improved long- and short-term memory network (LSTM) has obvious advantages in time series modeling, improving the long-term dependency problem in Recurrent Neural Networks (RNN), and is suitable for modeling time series. Compared with other time series prediction models, LSTM has a more complete network structure with the design of forget and memory gates and has better prediction ability for highly nonlinear time series data, which can improve the long-term dependency performance of the model. Muzaffar and Afshari used LSTM networks to predict short-term electric load changes and achieved better prediction accuracy than statistical models and traditional neural networks. However, Futures price series data are highly complex and the traditional single model has many limitations in forecasting.

Accordingly, a hybrid model was developed. The basic idea is to address the shortcomings of a single model and generate a synergistic effect in forecasting. Hybrid models are generally divided into two types: time-series hybrid prediction methods based on residuals and time-series hybrid prediction methods based on decomposition-aggregation. Among them, the hybrid model based on the decomposition-aggregation principle decomposes the time series data into multiple subseries, reducing the volatility of the series and capturing the trend of the series more easily. The model can effectively improve the forecasting accuracy and is suitable for futures price series forecasting. Empirical mode decomposition (EMD) is an adaptive signal decomposition algorithm proposed by Huang. It decomposes sequence data into smaller frequency components, which can be referred to as intrinsic modal components (IMFs). The decomposed IMF has lower volatility, which is convenient for model analysis and processing. EMD uses a recursive approach to decompose the data. Decomposition usually over-decomposes the series, which leads to the gradual accumulation of decomposition errors. Especially when there are jump changes in the data, the decomposition is likely to produce modal confounding, and the series values will fluctuate up and down several times and with large amplitude in a very short period of time, affecting the accuracy of prediction. Xian et al. combined ensemble EMD and independent component analysis to predict the price of gold, extract deep features in the sequence, and obtain better prediction results. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) added a limited adaptive white noise to EEMD to solve the incompleteness and reconstruction error of EEMD with the addition of white noise. Lin et al. used a hybrid model combining LSTM and CEEMDAN to forecast stock price changes, comparing a single model with a hybrid model, and achieved the best prediction effect. Both EEMD and CEEMDAN algorithms solve the modal aliasing phenomenon by adding white noise to the original sequence. None of the above decomposition algorithms can separate components with similar frequencies, which limits the decomposition effect of the sequence. Decomposition methods such as EEMD will decompose the sequence too much, and residual white noise will cause information interference. The non-original data added in the decomposition method changes the original information of the data and affects the accuracy of prediction.

We propose a VMD-SE-CLSTM network futures price forecasting hybrid method, which can be applied to hedging strategies, which can effectively enhance the ability of metal processing companies to respond to fluctuations in raw material prices and reduce the risks caused by market price fluctuations. VMD has a solid theoretical foundation and can decompose a time series into sub-sequences of a specific bandwidth and a specified number. We use the SE algorithm to calculate the SE of the sub-sequence and determine the decomposition number K to prevent the original sequence from being over-decomposed, which obviously suppresses the modal aliasing phenomenon, and does not change the original data. The SE algorithm is usually used to characterize the irregularity and complexity of the signal, and to quantify the regularity and unpredictability of the time series. Compared with approximate entropy, the calculation of SE does not depend on the length of the data, has better consistency, and is suitable for measuring the complexity of financial time series. At the same time, to better extract the nonlinear features in the futures sequence data, we introduced the
cascaded long-term short-term memory network, which has more nonlinear structures and deeper network levels and can better extract the features in the sequence data. Enterprises can predict the changing trend of the closing price of the futures market through the proposed model and implement the hedging trading strategy in the futures market, thereby reducing the risks caused by market price fluctuations and weakening the risks caused by raw material market price fluctuations.

This study has the following contributions:

1. In order to better suppress the modal aliasing phenomenon in the decomposition process, we propose SE combined with the VM algorithm to clarify the decomposition number K in the variational modal decomposition process, measure the complexity of the subsequence, and prevent the original sequence from being overly decomposed.

2. To better extract nonlinear features in futures sequence data, we introduce a cascaded long-term short-term memory network. Based on this, a decomposed hybrid model VMD-SE-CLSTM combined with the futures market hedging strategy is proposed to reduce the risk of market price fluctuations.

3. By comparing with the single model, the advantages of the mixed model in futures price forecasting are proved. At the same time, the proposed hybrid model is compared with 10 mainstream models, and the best prediction accuracy is obtained.

This paper is arranged as follows: Section 2 introduces the framework of the hybrid model and algorithms in the models. Section 3 presents the futures data and evaluation index and model parameters in the study. Model comparison experiments and back-testing experiments are presented in section 4. Section 5 summarizes the full text.

**Methods**

**Model construction for metal futures hedging**

Considering the non-stationarity, autocorrelation, and asymmetry of the spot price series and futures price series. In this paper, we introduce the GJR (1,1) model with negative information correction term to fit the expected volatility and expected volatility of the time series of asset returns. The GJR (1,1) model is constructed as:

\[
\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha e_{t-1}^2 + \gamma s_{t-1}^2
\]

\[
r_t = \mu_t \sigma_{t-1} + e_t
\]

where \(r_t\) is the rate of return on assets at time \(t\), \(\mu_t\) is the drift rate of the asset return at time \(t\), \(\sigma_t^2\) is the variance of the return at moment \(t\), \(\omega\) is a constant term representing the level of stable variance over time for the entire sample time series, \(\alpha\) and \(\beta\) are two parameters of the GJR model, reflecting the degree of influence of volatility in the previous period on the current estimate and the rate of decay of the influence, \(s_{t-1}^2\) is the stochastic tail between the estimated and actual volatility in the GJR model. \(s_{t-1}^2\) is a virtual signal function reflecting positive news and negative information. \(\gamma\) is the leverage coefficient, and the larger \(\gamma\) indicates the greater the impact of the information on the continuous shock to the price.

On day \(t\), the closing price of the futures is \(F_t\), and the rate of return is \(r_t^f\); Spot \(S_t\), the rate of return is \(r_t^s\), for the purpose of hedging, which constructs a futures and spot investment portfolio. At time \(t\), it contains the long spot position \(L_s\) unit and the short futures position \(L_f\) unit. At time \(t\), the hedging ratio is \(h_t\), then:

\[
h_t = \frac{L_f F_t}{L_s S_t}
\]

In the portfolio, the amount of spot planned to be hedged is constant, and the hedge ratio is a variable that changes dynamically over time. Therefore, the expected return rate \(r_t^p\) of a hedging combination on day \(t\) is:

\[
r_t^p = \frac{L_s S_{t-1} r_{t-1}^s - L_f F_{t-1} r_{t-1}^f}{L_s S_{t-1}} = r_t^s - h_t r_t^f
\]

This paper reduces the risk of market price fluctuations by predicting the price changes of raw material futures. The hedging strategy is the most commonly used risk hedging scheme in the futures market. Therefore, the GJR model is introduced here for illustration. The core parameter in the GJR model is \(rt\), the predicted value of the asset return at time \(t\), which can be obtained from the mixed model in Section 2.2. Therefore, the predictive effect of the model is directly related to the effect of the hedging strategy. Decision makers can execute corresponding trading strategies by predicting price changes.

**The proposed hybrid method**

This study aims to propose a hybrid method for futures price prediction, while identifying the appropriate decomposition number, performing VMD decomposition to decrease volatility, and extracting features via the CLSTM model. The model frame structure is exhibited in Figure 1. Initially, the original futures data are imported and the original data and other preprocessing operations are normalized to integrate the data into the decomposition algorithm. After which we use the VMD decomposition algorithm to decompose the original sequence, calculate the SE of the subsequence, and select the most appropriate number of decomposition
K. Decompose the original sequence into K sub-sequences and input them into the CLSTM network. Thereafter, the appropriate decomposition number k is determined according to SE and the sub-sequence data are incorporated into the CLSTM network. Finally, by synthesizing the prediction results of each sub-sequence, the final prediction result is obtained.

VMD

VMD is a completely non-recursive mode decomposition and signal processing method proposed by Dragomiretskiy and Zosso. Its decomposition process involves the construction and elucidation of variational problems. Notably, a relatively stable sub-sequence with multiple frequency scales can be decomposed by iteratively searching for the optimal solution of the variational model so that the center frequency and bandwidth of each decomposed component can be determined. VMD can decompose data into specific frequency sub-sequences, such that the bandwidth of each sub-signal in the frequency domain has a specific sparse nature. Thus, the VMD decomposition algorithm can be applied to the volatile futures price data to reduce the complexity of the sequence, suppress the appearance of modal aliasing, and obtain more sequence features.

VMD is used to decompose the original input signal $f(T)$ into k finite bandwidth modal components $u_k$ with the center frequency $\omega_k$, and the sum of the estimated bandwidth of each modal is kept to the minimum. The smaller the bandwidth of the signal, the more stable the decomposed sub-sequence. The steps for obtaining modal components are as follows: (1) first, the Hilbert transform is applied to each mode $u_k$ to obtain a single side spectrum. (2) Utilizing exponential tuning, the center frequency of each modal component is modulated to the fundamental frequency region. (3) The bandwidth of each modal component is computed using the Gaussian smoothness (i.e. square of the L2 norm) of the demodulated signal.

The VMD construction constrained variational problem is as follows:

$$\min_{\{u_k\}, \{w_k\}} \left\{ \sum_{k=1}^{K} \left\| \frac{\partial}{\partial t} \left[ \delta(t) + \frac{j}{\pi f} * u_k(t) \right] e^{-j\omega_k t} \right\|^2_2 \right\}$$

subject to:

$$\sum_{k=1}^{K} u_k(t) = f(t)$$

where $f(T)$ is the signal to be decomposed, $\delta(t)$ is the Dirac distribution, $u_k$ refers to each modal component, $\omega_k$ is the central frequency corresponding to the mode, and $*$ is the convolution operation. The above constrained variational problem is transformed into an unconstrained optimization problem through the addition of the quadratic penalty term $\alpha$ and Lagrange multiplier $\lambda(t)$. Furthermore, the augmented Lagrange expression is attained as follows:

$$L(\{u_k\}, \{w_k\}, \lambda) = \alpha \sum_{k=1}^{K}$$

$$\left\| \frac{\partial}{\partial t} \left[ \delta(t) + \frac{j}{\pi f} * u_k(t) \right] e^{-j\omega_k t} \right\|^2_2$$

$$+ \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|^2_2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle$$

(6)
Using the alternating direction method of multipliers, the saddle point of the augmented Lagrange expression is updated iteratively, and optimal solutions are obtained. The iterative update method is as follows:

$$u_k^{t+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\delta(\omega)}{1 + 2\alpha(\omega - \omega_k)^2}}{\hat{u}_k(\omega)}$$ (7)

$$\omega_k^{t+1} = \frac{\int_0^\infty \omega|\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}$$ (8)

**SE**

SE is used to measure the complexity of time-series signals based on the approximate entropy method. It uses a nonnegative number to represent the complexity of time series and reflects the richness of information contained within. Compared to approximate entropy, SE has two advantages: the calculation of SE does not rely on the length of the data; the change of parameters $m$ and $r$ has the same effect on SE, and has good consistency, anti-interference, and anti-noise ability. The lower the value of SE, the higher the self-similarity of the sample sequence. The greater the SE, the more intricate the sample sequence. After the VMD decomposition of the original sequence, SE was employed to judge the complexity of the sub-sequence. We determined the appropriate decomposition number and whether the sequence has been completely decomposed. The algorithm is described as follows:

1. For a time-series consisting of $N$ data \(\{x(n)\} = x(1), x(2), \ldots, x(N)\).
2. Reconstruct a set of vector sequences of dimension $m$, $X_m(1), \ldots, X_m(N - m + 1)$, among them $X_m(i) = \{X(i), X(i+1), \ldots, X(i+m-1)\}, 1 \leq i \leq N - m + 1$.
3. We define the distance $d(X_m(i), X_m(j))$ between vectors $X_m(i)$ and $X_m(j)$ as the absolute value of the largest difference between the two corresponding elements:

$$d(X_m(i), X_m(j)) = \max_{k = 0, \ldots, m-1} |x(i + k) - x(j + k)|$$ (9)

4. For a given $X_m(i)$, the number of times the distance between $X_m(i)$ and $X_m(j)$ ($1 \leq j \leq N - m, j \neq i$) is less than or equal to $r$, for $1 \leq j \leq N - m$, is denoted as $B_i$:

$$B_i^k(r) = \frac{1}{N - m - 1} B_i$$ (10)

5. Define $B^{(m)}(r)$ as:

$$B^{(m)}(r) = \frac{1}{N - m - 1} \sum_{i=1}^{N-m} B_i^k(r)$$ (11)

6. By increasing the dimension to $m + 1$, the number of distances between $X_{m+1}(i)$ and $X_{m+1}(j)$ ($1 \leq j \leq N - m, j \neq i$) less than or equal to $r$ are calculated and denoted as $A_i$. $A^{(m)}(r)$ is defined as:

$$A^{(m)}(r) = \frac{1}{N - m - 1} A_i$$ (12)

7. $A^{(m)}(r)$ is defined as follows:

$$A^{(m)}(r) = \frac{1}{N - m - 1} \sum_{i=1}^{N-m} A_i$$ (13)

$B^{(m)}(r)$ is the probability that two sequences match $m$ points under a similarity tolerance $r$. SE is presented as:

$$SampEn(m, r) = \lim_{N \to \infty} \left\{ -\ln \frac{A^{(m)}(r)}{B^{(m)}(r)} \right\}$$ (14)

When $N$ is a finite value, it can be estimated using the following formula:

$$SampEn(m, r) = -\ln \frac{A^{(m)}(r)}{B^{(m)}(r)}$$ (15)

**CLSTM**

LSTM is an improved recurrent neural network derived from RNN, which is used to solve the problem of gradient disappearance or explosion caused by training long-distance dependent data. Hence, it is suitable for time series, natural language processing, and other fields. The LSTM network mainly controls the input door and the forgotten door in the network structure to determine the information remembered and forgotten about the network. Moreover, the output door is also regulated to determine the output value.

The difference between the CLSTM network emphasized in this study and the traditional LSTM network is that the former can extract the main features of the time series to better fit the trend of futures prices. Primarily, the cascading structure of the model can increase the depth of the network and further extract the main features of the sequence. Moreover, adding nonlinear units can enhance the characteristics of time series, and the improved gating mechanism can better memorize the
characteristics of the input sequence. The network structure diagram of CLSTM constitutes the following three parts (Figure 2):

Part I:

The first part is to adjust on the basis of LSTM; the difference is that the current state \( C_t \) is contingent on the previous state and is controlled by the forgetting gate \( f_t \), the input gate \( i_t \), and the input control gate \( g_t \).

\[
f_t = \sigma(W_f * [H_{t-1}, X_t, C_{t-1}] + B_f) \quad (16)
\]

\[
i_t = \sigma(W_i * [X_t, H_{t-1}, C_{t-1}] + B_i) \quad (17)
\]

\[
g_t = \tanh (W_c * [H_{t-1}, X_t, C_{t-1}] + B_c) \quad (18)
\]

where \( W_f \) and \( B_f \) are the weight and deviation of the forgetting gate, \( W_i \) and \( B_i \) are the weight and deviation of the input gate, \( W_c \) and \( B_c \) are the weight and deviation of the current state, respectively. \( \tanh \) function was utilized to generate a new current cell state. Also, \( X_t \) is the input, \( H_{t-1} \) is the output of the previous time, \( C_{t-1} \) is the unit state of the previous time, \( \sigma \) is the sigmoid function, square brackets are the connections of tensors, and round brackets are the equations.

Lastly, the forgetting gate is multiplied by the last time state, and the input gate is multiplied by the control gate to acquire the updated unit state \( C_t \), where \( \odot \) refers to the element multiplication operation.

\[
C_t = f_t \odot C_{t-1} + i_t \odot g_t \quad (19)
\]

Part II:

The network structure of the second part is similar, but the nonlinear operation is completed before the \( M_{t-1} \) input at the last time. \( f_t', i_t', \) and \( g_t' \) are forgetting gate, input gate, and control gate of the second part, respectively.

The process of the second part is similar to that of the first part. The gate is determined by the input \( X_t \), the cell state \( M_{t-1} \) of the second part at the previous moment, and the cell state \( C_t \) of the first LSTM. The forgotten door of Part 2 \( f_t' \) is denoted by:

\[
f_t' = \sigma(W_f' * [X_t, C_t, M_{t-1}] + B_f') \quad (20)
\]

\[
i_t' = \sigma(W_i' * [X_t, C_t, M_{t-1}] + B_i') \quad (21)
\]

\[
g_t' = \tanh (W_c' * [X_t, C_t, M_{t-1}] + B_c') \quad (22)
\]

where \( W_f' \) and \( B_f' \) are the weight and deviation of the forgetting gate, \( W_i' \) and \( B_i' \) are the weight and deviation of the input gate, \( W_c' \) and \( B_c' \) are the weight and deviation of the current state, respectively, and \( M_{t-1}' \) is the cell state of the second part of the last time.

Finally, a nonlinear operation is performed on \( M_{t-1}' \) to increase the nonlinear degree of the network, and \( \odot \) represents the multiplication operation to obtain the unit state \( M_t \) of the second part.

\[
M_t = f_t' \odot \tanh (W_M * M_{t-1}') + i_t' \odot g_t' \quad (23)
\]

Part III:

\[
o_t = \tanh (W_o * [X_t, C_t, M_t]) \quad (24)
\]

\[
H_t = o_t \odot \tanh (W_H * [C_t, M_t]) \quad (25)
\]

\( W_o \) is the weight of the candidate state, and the candidate state \( o_t \) is determined by the cell state \( C_t \), the cell state \( M_t \) of the second part, and the input \( X_t \). The final output \( H_t \) is determined by \( C_t \) and \( M_t \), and \( W_H \) is the
weight of the convolution operation of the two cell states, where the function represented by $\sigma$ is expressed as follows:

$$\sigma(x) = \text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (26)$$

The $\text{tanh}$ function is as follows:

$$\text{tanh} (x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (27)$$

### Experimental data and index design

In this section, we introduce the datasets used in this study, the statistical metrics of the experimental results, and the setting of relevant parameters in the algorithm. These futures returns were traded from January 2010 to December 2019, lasting for approximately 10 years. The futures transaction data of each trading day are considered as the raw data, including opening price, highest price, lowest price, closing price, trading volume, and so on.

### Dataset

The futures data used in this study come from four raw material futures markets in Shanghai (i.e. RB, AL, CU, ZN), comprising futures contracts with the longest trading time and the largest trading volume in the Shanghai futures market. RB represents the raw material of rebar, AL aluminum, CU copper, and ZN zinc. The data are obtained from https://www.joinquant.com/data (accessed on June 24, 2021), which can be downloaded in the sub-category of “Futures Data” under the grouping of “Data Dictionary.” The data can also be downloaded for free through github (accessed on June 24, 2021). After downloading the original data, it is normalized and then divided for each of the futures. The first 65% of the data were employed as the training set, 15% were used as the validation set, and the last 20% were utilized as the test set. Consequently, the training, validation, and test datasets of four futures with labeled data can be obtained.

### Evaluation criteria

The loss error is usually used to evaluate the prediction effect and better evaluate the prediction ability of the model. Notably, this research uses three evaluation indexes. The smaller the value of the evaluation index, the better the prediction effect, including the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Meanwhile, the goodness of fit ($R^2$) was also used to measure the degree of fit between the predicted and true values; the closer $R^2$ is to 1, the better the fit. All the evaluation metrics are outlined in Table 1.

In Table 1, $y_{\text{predict}}$ specifies the predicted value, $y_{\text{actual}}$ represents the true value, $\bar{y}$ signifies the mean of the series, and RMSE is used to measure the deviation between the observed value and the true value. It is used to measure the average size of the error and is more sensitive to the performance of outliers.

MAE is the average value of the absolute error and reflects the error of the predicted value, which can better reflect the actual situation of the predicted value error and is a more general form of the average value of the error. MAPE is expressed as a percentage and has nothing to do with the ratio but can be used to compare the forecasts of different ratios. The smaller the three indicators of RMSE, MAE, and MAPE, the lower the prediction error of the model and the better the prediction effect. $R^2$ was utilized in the regression analysis to test the density of the sample data points clustered around the regression line. Furthermore, it is suitable for the prediction of general regression series to verify the fitting degree of the model. The closer the goodness of fit is to 1, the better the fit of the predicted value of the model to the true value. These indicators can explain the prediction accuracy of the model as well as the robustness and stability of the model prediction from different aspects.

### Parameter settings

**Model parameter settings.** Here, we explain the setting of the CLSTM network model parameters, as demonstrated in Table 2. The model parameter table is shown in Tables 3 and 4.
In the SE algorithm, the SE value is related to the values of \( m \) and \( r \). According to Pincus,\(^4\) \( m = 1 \) or \( m = 2 \), \( r = 0.1 \sim 0.25 \times SD \), where SD is the standard deviation of the sub-sequence data and the obtained SE value has reasonable statistical characteristics. This study sets \( m = 2 \), \( r = 0.2 \times SD \). We predict the following day’s price using 10 days of data; that is, the time window \( t \) is 10. \( p \) stands for the number of time lags of the autoregressive model. \( d \) stands for the degree of differencing. \( q \) stands for the order of the moving-average model.

The parameter selection in the above parameter table is found through grid search, and the fairness of the comparison between models is ensured by selecting the optimal parameters in the model. Epoch is the number of training iterations for which the model is most suitable by early stopping.

**Determination of the decomposition number \( K \).** This study uses a variational model to decompose the input futures price series. Among them, the key parameter decomposition number \( K \) had a greater impact on VMD decomposition. If the chosen decomposition number \( K \) happens to be too small, the sequence data will not be fully decomposed, and the volatility of the sub-sequence will still be large. Conversely, if the decomposition number is too large, the sequence data will be over-decomposed, but the difference in the sub-sequences will be small, the decomposition accuracy shall be reduced, and the modal aliasing phenomenon will also occur.

This research employs SE to measure the complexity of a sequence signal. The larger the SE, the higher the complexity of the sequence. According to the change in SE of the sub-sequence, it is determined whether the sequence is fully decomposed to ascertain the appropriate decomposition number \( K \).

The change trend of SE after the decomposition of different futures price series are displayed in Table 1. Interestingly, the trend changes first increase and then decrease. When the value of \( K \) selected by VMD decomposition is small, the original futures price sequence is not fully decomposed. After the sequence is decomposed, the value of the SE increases instead. When an appropriate decomposition number \( K \) is selected, the SE is at a turning point, implying that the complexity of the sequence is reduced after decomposition. Still, the sequence will be over-decomposed when we choose a very large decomposition number \( K \), and modal aliasing shall occur easily.

According to decomposition Table 5 of the rebar price sequence, when the number of sub-sequences is less decomposed, the intricacy of the sequence is lower, and the phenomenon of sub-sequence spectrum aliasing can be effectively suppressed, but the SE value gradually increases. When \( K \) is 5, the SE value reaches its maximum threshold, and the entropy value of the subsequence decomposition sample decreases, indicating that when \( K = 5 \), the sequence is fully decomposed. We choose the \( K \) value at the turning point where the SE

| Name | Input neurons | Hidden neurons | Activation function | Hidden layers | Output neurons |
|------|--------------|---------------|---------------------|--------------|---------------|
| First hidden layer | 64 | 50 | Tanh | 1 | 32 |
| Second hidden layer | 32 | 50 | Tanh | 1 | 1 |

| Futures | p | d | q |
|---------|---|---|---|
| RB      | 1 | 0 | 2 |
| AL      | 1 | 0 | 0 |
| CU      | 2 | 0 | 1 |
| ZN      | 1 | 0 | 0 |

Table 4. Related parameters for training the 10 common models.

| Models          | Related parameters                      |
|-----------------|-----------------------------------------|
| RNN             | input size = 64; hidden size = 32; output size = 1; layer num = 1; activation function = Tanh; optimization function = Adam with learning rate = 0.009, betas = (0.9, 0.999); loss function = MSE; stop training epoch = 120 |
| GRU             | stop training epoch = 80 |
| LSTM            | stop training epoch = 100 |
| CLSTM           | stop training epoch = 40 |
| EMD-CLSTM       | decompose number = 8; stop training epoch = 30 |
| EEMD-CLSTM      | decompose number = 9; stop training epoch = 30 |
| CEEMDAN-CLSTM   | decompose number = 9; stop training epoch = 30 |
| VMD-RNN         | stop training epoch = 70 |
| VMD-GRU         | stop training epoch = 50 |
| VMD-LSTM        | stop training epoch = 50 |
tends to be smooth to prevent excessive decomposition of the original price series. The black bold part in the table is the K value of the selected decomposition number of the futures price selected by the model. Thus, 5 is chosen as the decomposition number for the rebar futures price series. Likewise, we can obtain the decomposition number of AL futures price as 4, the decomposition number of CU futures price as 6, and the decomposition number of ZN futures price as 6.

**Experiments**

In this section, the SE algorithm is used to determine the complexity of the sub-sequence and choose the appropriate decomposition number for the futures price. Then, we compare the prediction results of the single model and the mixed model with regard to the futures data. EMD, EEMD, CEEMDAN, and VMD are employed to decompose the futures price sequence and then compare the forecasts. We then compare the methods to determine the effect of the VMD decomposition number K on the model prediction effect. Additionally, we use the VMD decomposition method and the same decomposition number for the model predictions, and compare the results to the conventional regression algorithm to visualize the prediction effect of our model.

**Comparison of mixed model and single model**

Figure 3 highlights a comparison of evaluation indicators concerning five types of neural network models for different futures (i.e. RB, AL, CU, ZN) price series predictions.

Compared with the traditional statistical model ARIMA, the neural network model has higher prediction accuracy for financial futures time series. For example, it can be seen from the RMSE, MAPE, and MAE in Figure 3 that the prediction error of the mixed model VMD-CLSTM is less than that of a single model. Furthermore, the mixed model has the highest degree of fit for futures prices from the perspective of the goodness of fit. Compared to the single-model prediction method, the hybrid model combined with the decomposition algorithm has a good prediction effect, improves the prediction performance of the model, and minimizes the prediction error. Moreover, in the comparison between prediction and evaluation indicators of a single model, the prediction effect of the CLSTM network was better than that of the conventional time series model. When we adjusted the number of hidden layers of the model for experimental comparison, we found that increasing the depth of the model failed to improve its accuracy.

**Evaluating the effectiveness of the VMD algorithm**

Different decomposition algorithms are used to decompose the original sequence, including EMD, EEMD, CEEMDAN, and VMD. Subsequently, the CLSTM network is utilized to predict the sub-sequence. The outcomes obtained for the different futures price forecasts are depicted in Figure 4.

VMD has better prediction accuracy than different futures price series decomposition methods based on EMD, EEMD, and CEEMDAN (refer to Figure 4), signifying that VMD has a better futures price sequence decomposition effect. Algorithms such as EMD adopt the self-use decomposition method, which leads to excessive decomposition of the sequence and reduction of the prediction accuracy of the model. The evaluation indexes RMSE, MAPE, and MAE obtained from the VMD decomposition prediction in the experiment were also lower, and the prediction results were more accurate. This finding confirms that using the VMD decomposition method to decompose the original data can provide more sequence information and reveal the characteristics of each component more completely. The sub-sequence obtained through VMD decomposition is more stable, so the prediction of the sub-sequence after decomposition can provide a prediction result with a smaller error and better fitting degree.

**Evaluate the validity of sample entropy to determine K value**

In this study, we used the SE algorithm to determine the K value of the VMD decomposition algorithm. The
The EMD method is typically used for the data first, and the K value determined by the EMD method is used as the K value of the VMD algorithm.33

The experiment compares the VMD decomposition number K determined by two different decomposition methods and proves that using SE to find the decomposition number can improve the prediction ability of the model. The futures price sequence is decomposed by VMD using different decomposition numbers. Thereafter, the decomposed sub-sequences are incorporated into the CLSTM network for prediction, and the final result is obtained. The results of predicting a variety of futures prices are listed in Table 6.

The SE algorithm selects the appropriate decomposition number by judging the change trend of the entropy value. It can be seen from the table that the decomposition number K determined by SE can achieve better prediction accuracy and smaller prediction errors than the general K value determination method. From the goodness of fit index, the decomposition number determined by SE also has a better decomposition effect on the original data.

**Evaluation of the proposed hybrid model**

VMD is employed to decompose the original price series and use different networks for predictions. The proposed VMD-CLSTM model can obtain more accurate prediction results, outlining that a combination of VMD and CLSTM networks can better capture the nonlinear and deep features of the original sequence (refer to Table 7). For the four raw material futures datasets, the prediction effect of the proposed VMD-CLSTM prediction model is the most accurate, validating that the proposed model has good prediction accuracy and robustness for various types of futures.

The VMD-CLSTM model shows good performance in raw material futures price forecasting and achieves the best accuracy in the above-mentioned comparison model. The MAPE, RMSE, and R side of VMD-CLSTM in the rebar futures sequence prediction were 0.87, 35.58, and 0.97, respectively. Meanwhile, the mixed model threaded steel futures forecast error is the smallest (refer to Figure 6(a)). Compared with other benchmark models, the MAPE and RMSE of VMD-CLSTM are smaller, relaying that the prediction effect of the model is better than that of the benchmark model, with a better effect in futures price prediction. From Figure 5(a), we can see that the rebar futures price prediction curve can better fit the price trend and has the highest goodness of fit. We use VMD-CLSTM to predict on other futures prices and obtain better forecast accuracy and smaller forecast errors, as shown in Figures 5 and 6. Compared with other models, this
model has smaller prediction errors for out-of-sample futures prices, indicating that the model is more robust.

As can be seen from Figure 5(d), due to the impact of the global epidemic from 2019 to 2020, the unilateral market situation of zinc futures has led to persistent deviations in all models, and the statistical characteristics of the market have been weakened. A good forecast is achieved, but overall, the model can make the best forecast for most futures prices.

**Evaluation of simulated trading**

Using four mixed models of LSTM, GRU, RNN, and CLSTM transactions are carried out according to the predicted change trend of rebar futures prices. In this simple trading strategy, we buy the futures if the next day’s opening price is predicted to be above the current day’s opening price, accounting for the transaction costs.

As illustrated in Figure 7, the test data time range was 476 days in total (i.e. approximately 2 years). The average return rate of the VMD-RNN model is 33.21%, the average rate of return of the VMD-GRU model is 50.62%, the average rate of return of the VMD-LSTM model is 62.43%, the average rate of return of the VMD-CLSTM model is 103.26%, and the results show that under the same initial capital premise, the four futures simulated by the VMD-CLSTM

---

**Table 6. Influence of different decomposition numbers of VMD on the prediction.**

| Futures contracts | Decomposition number | RMSE | MAPE | MAE | R_Square |
|-------------------|----------------------|------|------|-----|----------|
| RB                | 8(EMD)               | 86.02| 2.14 | 80.31 | 0.88     |
|                   | 5(SE)                | 35.58| 0.87 | 32.25 | 0.97     |
| AL                | 8(EMD)               | 135.78| 0.88 | 122.71 | 0.98     |
|                   | 4(SE)                | 61.22| 0.33 | 46.89  | 0.97     |
| CU                | 9(EMD)               | 516.12| 0.89 | 444.81 | 0.92     |
|                   | 6(SE)                | 392.52| 0.72 | 358.91 | 0.96     |
| ZN                | 9(EMD)               | 800.34| 3.59 | 734.93 | 0.87     |
|                   | 6(SE)                | 517.69| 1.75 | 390.28 | 0.95     |

---

**Figure 4.** (a) sequence diagram of the comparison of RMSE of futures data forecasts with different decomposition methods, (b) sequence diagram of the comparison of MAE of futures data forecasts with different decomposition methods, (c) sequence diagram of the comparison of MAPE of futures data forecasts with different decomposition methods, and (d) sequence diagram of the comparison of $R^2$ of futures data forecasts with various decomposition methods.
model have the highest average yield, followed by VMD-GRU, VMD-LSTM, and finally VMD-RNN.

Figure 7 shows excess returns that can be obtained by all prediction models. Higher excess returns can be achieved because VMD-CLSTM has a better prediction effect. For enterprises, futures can be purchased or sold through a hedging strategy to reduce the risk in the current market.

**Conclusions**

This paper proposes a hybrid model, VMD-CLSTM, to predict changes in daily raw material metal prices. First, we decompose the original data based on the VMD decomposition algorithm, and decompose the original data fully according to the decomposition number K value determined by the SE value. Afterward, the decomposed sub-sequence is trained by the CLSTM network. The four data sets of the Shanghai Futures Exchange were then evaluated using three error indicators (RMSE, MAPE, and MAE) and the goodness of fit ($R^2$) to evaluate the performance of the prediction model. The hybrid VMD-CLSTM model achieved the best prediction effect. The findings confirm the effectiveness of the CLSTM network and also show that using the VMD decomposition algorithm and the SE value to determine the decomposition number K value can improve the predictive ability of the model. Besides, the mixed model prediction results of three decomposition algorithms are compared with four benchmark models to verify the effectiveness of the proposed method. The proposed VMD-CLSTM model proved to be superior to the compared model for all evaluation indicators, reflecting the effectiveness of the model for forecasting futures price data. Simultaneously, smaller forecast error can reduce the enterprise’s risk in the market. In hedging strategy, enterprises can choose the time of futures trading by the change of futures price predicted by the model, which can effectively reduce the enterprise’s risk in production and operation. When the futures market is impacted by sentiment, the statistical properties of the market will weaken, and uncertainty will increase. We hope that we can design a comprehensive model to deal with this relatively special market sit-
Figure 6. (a) Four hybrid models predicting deviation from target values of rebar futures, (b) four hybrid models predicting deviation from target values of aluminum futures, (c) four hybrid models predicting deviation from target values of copper futures, and (d) four hybrid models predicting deviation from target values of zinc futures.
uation, combining the news data in the market to predict futures prices. This is also our future research direction.

Acknowledgements
We would like to thank Editage (www.editage.cn) for English language editing.

Author contributions
C.Z.: Resources, supervision, conceptualization, methodology and project administration. Y.W.: Programming, validation of the results, analyses, writing & editing, and data curation. Y.C.: Resources, supervision and project administration. All authors have read and agreed to the published version of the manuscript.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China under Grant 61902349.

ORCID ID
Yexin Wang https://orcid.org/0000-0003-0808-677X

Data availability statement
The data are from https://www.joinquant.com/data (accessed on 24 June 2021), under the “Futures Data” sub-category of the “Data Dictionary” category. The data are also available for free through https://github.com/ssxzsqy/Futures-data-mtal (accessed on 24 June 2021).

References
1. Shao Y and Wang S. Productivity growth and environmental efficiency of the nonferrous metals industry: an empirical study of China. J Clean Prod 2016; 137: 1663–1671.
2. Gargano A and Timmermann A. Forecasting commodity price indexes using macroeconomic and financial predictors. Int J Forecast 2014; 30: 825–843.
3. Gu Q, Chang Y, Xiong N, et al. Forecasting Nickel futures price based on the empirical wavelet transform and gradient boosting decision trees. Appl Soft Comput 2021; 109: 107472.
4. Wang B and Wang J. Energy futures and spot prices forecasting by hybrid SW-GRU with EMD and error evaluation. Energy Econ 2020; 90: 104827.
5. Sharma R, Saxena A and Vagrecha K. Supply chain optimization of zinc industry: opportunities, strategies and challenges. Glob J Enterp Inf Syst 7(3): 62–70.
6. Erb CB and Harvey CR. The strategic and tactical value of Commodity Futures. Financ Anal J 2006; 62: 69–97.
7. Li J, Huang L and Li P. Are Chinese crude oil futures good hedging tools? Finance Res Lett 2021; 38: 101514.
8. Herrera AM, Hu L and Pastor D. Forecasting crude oil price volatility. Int J Forecast 2018; 34: 622–635.

| Futures contracts | Model  | RMSE  | MAPE  | MAE   | R_Square |
|-------------------|--------|-------|-------|-------|----------|
| RB                | VMD-RNN| 62.21 | 1.33  | 49.22 | 0.94     |
|                   | VMD-GRU| 53.58 | 1.09  | 40.77 | 0.95     |
|                   | VMD-LSTM| 81.26 | 1.65  | 61.05 | 0.90     |
|                   | VMD-CSTM| 35.88 | 0.87  | 32.35 | 0.97     |
| AL                | VMD-RNN| 114.6 | 0.63  | 89.26 | 0.91     |
|                   | VMD-GRU| 125.07| 0.51  | 72.38 | 0.94     |
|                   | VMD-LSTM| 139.59| 0.68  | 93.94 | 0.9      |
|                   | VMD-CLSTM| 61.22 | 0.33  | 46.89 | 0.97     |
| CU                | VMD-RNN| 474.8 | 0.82  | 400.94| 0.94     |
|                   | VMD-GRU| 454.84| 0.76  | 373.44| 0.95     |
|                   | VMD-LSTM| 498.36| 0.81  | 398.84| 0.94     |
|                   | VMD-CLSTM| 392.52| 0.72  | 358.91| 0.96     |
| ZN                | VMD-RNN| 1773.08| 6.56 | 1480.86| 0.39    |
|                   | VMD-GRU| 1288.99| 4.98 | 1114.37| 0.68    |
|                   | VMD-LSTM| 486.24| 0.78  | 382.73| 0.94     |
|                   | VMD-CLSTM| 454.78| 1.9   | 399.66| 0.96     |

Figure 7. Four hybrid models simulating the futures price of rebar.
Liu Y, Yang C, Huang K, et al. Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network. *Knowl Based Syst* 2020; 188: 105006.

Dooley G and Lenihan H. An assessment of time series methods in metal price forecasting. *Resour Policy* 2005; 30: 208–217.

Liu W, Wang C, Li Y, et al. Ensemble forecasting for product futures prices using variational mode decomposition and artificial neural networks. *Chaos Solitons Fractals* 2021; 146: 110822.

Gürsoy and Engin SN. A wavelet neural network approach to predict daily river discharge using meteorological data. *Meas Control* 2019; 52: 599–607.

Li C, Hu Z, Li Y, et al. Forecasting copper prices by decision tree learning. *Resour Policy* 2017; 52: 427–434.

Wang B and Wang J. Energy futures price prediction and evaluation model with deep bidirectional gated recurrent unit neural network and RIF-based algorithm. *Energy* 2021; 216: 119299.

Hochreiter S and Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9: 1735–1780.

Chen M-R, Zeng G-Q, Lu KD, et al. A Two-Layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM. *IEEE Internet Things J* 2019; 6: 6997–7010.

Zhao F, Zeng G-Q and Lu KD. EnLSTM-WPEO: short-term Traffic Flow Prediction by Ensemble LSTM, NCNT weight integration, and population extremal optimization. *IEEE Trans Vehicular Technol* 2020; 69: 101–113.

Muzaffar S and Afshari A. Short-term load forecasts using LSTM Networks. *Energy Proc* 2019; 158: 2922–2927.

Zhang T, Tang Z, Wu J, et al. Multi-step-ahead crude oil price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm. *Energy* 2021; 229: 120797.

Xiong T, Li C, Bao Y, et al. A combination method for interval forecasting of agricultural commodity futures prices. *Knowl Based Syst* 2015; 77: 92–102.

Niu H, Xu K and Liu C. A decomposition-ensemble model with regrouping method and attention-based gated recurrent unit network for energy price prediction. *Energy* 2021; 231: 120941.

Tang L-H, Bai Y-L, Yang J, et al. A hybrid prediction method based on empirical mode decomposition and multiple model fusion for chaotic time series. *Chaos Solitons Fractals* 2020; 141: 110366.

Zhang Y, Yan B and Aasma M. A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM. *Expert Syst Appl* 2020; 159: 113609.

Ding J, Chen G, Huang Y, et al. Short-term wind speed prediction based on CEEMDAN-SE-improved PIO-GRNN model. *Meas Control* 2021; 54: 73–87.

Wang R, Wang J and Xu Y. A novel combined model based on hybrid optimization algorithm for electrical load forecasting. *Appl Soft Comput* 2019; 82: 105548.