

## Research Article

# A Novel Hybrid Deep Learning Model for Sugar Price Forecasting Based on Time Series Decomposition

Jinlai Zhang,<sup>1</sup> Yanmei Meng ,<sup>1</sup> Jin Wei,<sup>1</sup> Jie Chen,<sup>1</sup> and Johnny Qin<sup>2</sup>

<sup>1</sup>College of Mechanical Engineering, Guangxi University, Nanning 530004, China

<sup>2</sup>Energy, Commonwealth Scientific and Industrial Research Organisation, 1 Technology Court, Pullenvale QLD4069, Australia

Correspondence should be addressed to Yanmei Meng; [gxu\\_mengyun@163.com](mailto:gxu_mengyun@163.com)

Received 27 June 2021; Revised 21 July 2021; Accepted 28 July 2021; Published 3 August 2021

Academic Editor: Francesco Lolli

Copyright © 2021 Jinlai Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sugar price forecasting has attracted extensive attention from policymakers due to its significant impact on people's daily lives and markets. In this paper, we present a novel hybrid deep learning model that utilizes the merit of a time series decomposition technology empirical mode decomposition (EMD) and a hyperparameter optimization algorithm Tree of Parzen Estimators (TPEs) for sugar price forecasting. The effectiveness of the proposed model was implemented in a case study with the price of London Sugar Futures. Two experiments are conducted to verify the superiority of the EMD and TPE. Moreover, the specific effects of EMD and TPE are analyzed by the DM test and improvement percentage. Finally, empirical results demonstrate that the proposed hybrid model outperforms other models.

## 1. Introduction

Sugar is an important food commodity around the world, and the fluctuations of food prices have a huge impact on people's daily lives due to its impact on overall inflation dynamics of many countries [1, 2]. Therefore, it is essential to forecast the sugar price accurately.

Forecasting of sugar prices has attracted a lot of interest of researchers for several decades, and it can be divided into statistical methods and machine learning methods. The statistical methods have the advantages of low complexity and fast calculation speed [3]. In 1975, Meyer and Kim [4] applied the autoregressive integrated moving average (ARIMA) method in sugar price forecasting. However, the ARIMA requires the time series data to be stable or stable after being differentiated, which might limit the application of this method. In 2009, Xu et al. [5] used a neural network with multiple fully connected layers for sugar price forecasting using a Chinese database. In 2011, Ribeiro and Oliveira [6] introduced a hybrid model built upon artificial neural networks (ANNs) and Kalman filter. In 2019, Silva et al. [7] investigated ANNs, extreme learning machines (ELMs), and echo state networks (ESNs) for sugar price

forecasting. However, one limitation of abovementioned three methods is they do not optimize the hyperparameter of neural networks. Hyperparameter optimization is a commonly used strategy in machine learning area [8] especially in time series forecasting [3] to improve the performance of machine learning models. This is largely due to the explosion in the field of machine learning in recent years [9] and makes some very common technologies appeared in recent years in the field of machine learning, such as SGD [10], and Adam [11] are proposed after the year of 2014, while most sugar price forecasting literatures are published before 2014. On the other hand, the nonstationarity and nonlinearity of sugar prices [12] make it harder to accurately predict the future sugar price. In order to handle nonstationarity features, the inputs for machine learning models need to be properly preprocessed [13]. Therefore, some multiresolution analysis techniques are widely used in many forecasting problems [14, 15]. Conventionally, the discrete wavelet transformation (DWT) was used [16, 17]. Hajiabotorabi et al. [18] improved the recurrent neural network (RNN) with the multiresolution based on B-spline wavelet produced by an efficient DWT. Yong and Awang [19] used DWT for improving the forecast accuracy. However, the DWT generally requires

a lengthy trial and error process [20]. Moreover, the empirical mode decomposition [21] (EMD) multiresolution technique is introduced to time series, which provides self-adaptability. EMD extracts the salient features via the temporal local decomposition method and isolates these significant features into subseries that represent the physical structure of the time series [22]. The EMD-based machine learning models have been adopted in time series forecasting. Ali and Prasad [20] predicted the significant wave height by ELM and improved complete ensemble EMD. Bedi and Toshniwal [23] adopted the EMD for electricity demand forecasting. This broadly adapted technique can boost forecasting performance.

To this end, in this paper, to investigate the power of hyperparameter optimization and multiresolution analysis in sugar price forecasting, we propose a hybrid deep learning model for sugar price forecasting. The model uses Tree of Parzen Estimators (TPEs) [24] to optimize long short-term memory (LSTM) networks [25]. A time series decomposition technique named empirical mode decomposition (EMD) is used to decompose the sugar price to extract the salient features. The effectiveness of the proposed approach is tested at the daily sugar price of London Sugar Futures. To fairly compare with the mainstream methods for sugar price forecasting, we build the deep neuron networks (DNNs) with multiple fully connected layers which is equal to models in [5–7] in the machine learning field and the ARIMA compared with [4]. The results are compared against other machine learning algorithms such as the support vector regression (SVR) machine [15, 26–29], the DNN, and traditional time series model ARIMA.

The rest of this paper is organized as follows: Section 2 describes the theoretical background, such as the LSTM, EMD, and TPE. Section 3 describes the proposed hybrid model in detail. Section 4 provides details of experiments and evaluations. Section 5 shows the discussion of experimental results, and Section 6 concludes the paper and points out possible future work.

## 2. Theoretical Background

**2.1. Long Short-Term Memory (LSTM).** The LSTM neural network is heavily used as a basic building block in the modern deep learning-based time series forecasting model [30], which is an improved version of the recurrent neural network (RNN) and mainly solves the problem of gradient vanishing by its internal memory unit and gate mechanism. It can make the network memorize for a longer time and make the network more reliable. It was proposed by Meyer and Kim [4] in 1997. It solves problems that RNN cannot learn the long-term dependence of time series data. It has been widely used in the fields of sentiment analysis [31], speech recognition [32], early crop classification [33], and so on and has achieved satisfactory results. The key mathematical equation of the LSTM model is as follows:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f), \quad (1)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i), \quad (2)$$

$$g_t = \tanh(w_g \cdot [h_{t-1}, x_t] + b_g), \quad (3)$$

$$c_t = f_t * c_{t-1} + i_t * g_t, \quad (4)$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o), \quad (5)$$

$$h_t = o_t * \tanh(c_t), \quad (6)$$

where  $f_t$ ,  $g_t$ ,  $i_t$ , and  $o_t$  are the output value of the forget gate, update gate, output gate, and input gate, respectively. Moreover,  $*$  and  $w$  denote the product operation and the network parameters;  $b_{f,i,g,o}$  are the bias vectors;  $\sigma$  is the sigmoid activation function; and  $f_t$  is the memory cell. The former LSTM output value  $h_{t-1}$  and the input data  $x_t$  are the inputs of the four gates.

**2.2. Empirical Mode Decomposition (EMD).** Empirical mode decomposition (EMD) [21] is a time series decomposition technique, and it was proved to be effective in time series forecasting [20]. Considering the nonlinearity and complexity of sugar price sequences, accurately capturing sugar price characteristics will be a difficult task. Thus, the time series decomposition strategy EMD is adopted to conduct a decomposition in terms of the original sugar price sequences. The procedure of EMD technology is described as follows.

- (1) Identify all local minima ( $l_{\min}$ ) and local maxima ( $l_{\max}$ ) in sugar price sequences  $x_{(t)}$ ,  $t = 1, 2, 3, \dots, T$
- (2) Connect all  $l_{\min}$  and  $l_{\max}$  to form upper envelopes ( $x_{\text{up}(t)}$ ) and lower envelopes ( $x_{\text{low}(t)}$ )
- (3) Compute the average  $m_t = (x_{\text{up}(t)} + x_{\text{low}(t)})/2$
- (4) Extract the intrinsic mode functions  $\text{IMF} = x_t - m_t$
- (5) Iterate on the residual  $m_t$

**2.3. Tree of Parzen Estimators (TPEs).** As stated by James et al. [24], TPE is a global optimization algorithm based on a sequence model. The algorithm uses a probabilistic model to model the loss function and make informed guesses about the specified number of iterations to find the best hyperparameters. When optimizing multiple hyperparameters, this algorithm has shown performance over grid search and random search, especially for deep learning models that usually have more hyperparameters than traditional machine learning models [34].

TPE uses the Bayes rule, and the probabilistic model  $p(y|x) = p(x|y)p(y)/p(x)$ , and  $p(y|x)$  is broken down into  $l(x)$  and  $g(x)$ , such that

$$p(x|y) := \begin{cases} l(x)y < y^*, \\ g(x)y \geq y^*, \end{cases} \quad (7)$$

where  $l(x)$  means that one distributions for the hyperparameter where objective function value is less than the threshold and  $g(x)$  means that another one distribution for the hyperparameter where objective function value is larger than the threshold.

The expected improvement (EI) metric is used to identify which hyperparameters to be chosen based on the probabilistic model. Given some set of hyperparameters and a threshold value for the objective function,  $y^*$ , the EI is given by

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} dy. \quad (8)$$

When EI is positive, this means that the hyperparameter set  $x$  is expected to obtain an improvement over the threshold  $y^*$ . Therefore, the working principle of TPE is to extract sample hyperparameters from  $l(x)$ , evaluate them according to  $l(x)/g(x)$ , and then return the set  $x$  that gives the best EI value.

### 3. Materials and the Proposed Hybrid Model

**3.1. The Proposed LSTM Model.** In this paper, for a fair comparison with the mainstream sugar price forecasting model, we proposed two-layer LSTM model which is illustrated in Figure 1, and the DNN model used in the three sugar price forecasting literatures will be compared with the proposed LSTM model in the same network structure. The performance of TPE and EMD will validate experimentally via comparing LSTM and DNN.

**3.2. Hyperparameter Optimization.** For the deep learning models, the hyperparameters are model parameters that are defined in advance before training [35]. There are several hyperparameters used in this paper:

$l_{i,n}$  means the number of neurons of  $i_{th}$  hidden layer. As shown in Figure 1, the deep learning models usually contain multiple hidden layers and have several neurons in each hidden layer.

$l_{i,a}$  means the activation function of  $i_{th}$  hidden layer. The activation function is a function that runs on the neurons of the neural network and is responsible for mapping the input of the neuron to the output [36]. Each neuron contains an input, output, weight, and processing unit. The output signal of the neuron is obtained after processing by the activation function. tanh, ReLU, and LeakyReLU activation functions [37] are used in this paper.

**Dropout Rate.** Dropout is a very useful and successful technique that can effectively control the overfitting problem [38]. Generally speaking, dropout will randomly delete some neurons with the probability of dropout rate to train different neural network architectures on different batches. The dropout rate is a real number between 0 and 1.

**Optimizer.** In deep learning, the loss defines the performance of the model. The loss is used to train the network to make it perform better. Essentially, a lower loss means that the model will perform better. A process of minimizing (or maximizing) any mathematical expression is called optimization, and the optimizer is used to minimize loss. The rmsprop [39], Kingma and Jimmy [11], and sgd [40] optimizers are used in this paper.

**Batch Size.** The deep learning model updates its parameters for each minibatch of training datasets with a batch size [41]. For the same number of epochs, the number of batches required for a large batch size is reduced, so training time can be reduced. However, within a certain range, increasing batch size helps the stability of convergence; however, as the batch size increases, the performance of the model will decrease [10]. Therefore, batch size is an important hyperparameter.

**Learning Rate.** Choosing the optimal learning rate is important because it determines whether the neural network can converge to the global minimum. Choosing a higher learning rate may bring undesirable consequences on the loss, so that the neural network may never reach the global minimum, because the neural network is likely to skip it. Choosing a smaller learning rate will help the neural network converge to the global minimum, but it will take a lot of time because it only makes very few adjustments to the weight of the network and more time need to be spent to train the neural network. A smaller learning rate is also more likely to trap the neural network in a local minimum because the smaller learning rate is relatively hard to jump out of the local minimum. Therefore, we must be very careful when setting the learning rate.

**3.3. Proposed Hybrid Forecasting Model.** Figure 2 illustrates the whole process of the proposed hybrid deep learning LSTM neural network model. The model is based on EMD technique and TPE algorithm. The modeling includes three steps as follows.

*Step 1.* Sugar prices sequence is decomposed by EMD and forms a set of IMFs.

*Step 2.* The lagged values of sugar price sequences and IMFs are used as input of the deep learning model. The TPE algorithm is applied to optimize the hyperparameters of the LSTM model, then the LSTM model with the best hyperparameter combination is obtained, and then its forecasting performance is tested on the test set.

*Step 3.* The trained model can be deployed to real world sugar price forecasting for police makers.

### 4. Experiments and Evaluations

In this section, sugar price sequences from the market are used to evaluate the performance of the proposed hybrid deep learning model via performing two distinct tests. All

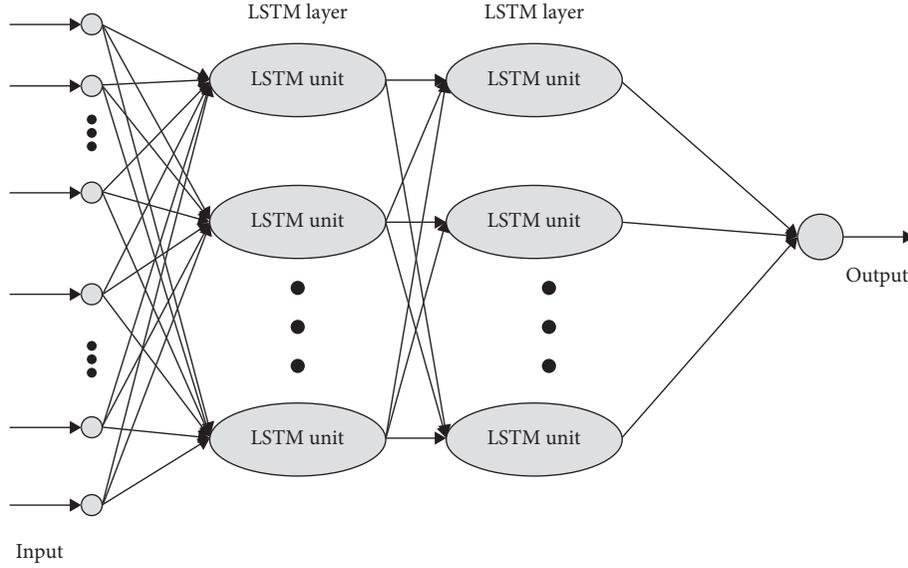


FIGURE 1: The proposed LSTM model.

the tests are run on Ubuntu 18.04 operation system with Intel Xeon (R) E5-2650 v4 CPU, GeForce GTX 1080 GPU, and 32 GB RAM. Furthermore, to avoid influence of random factor, each method is run 20 times, and the averaging value is used as final results.

**4.1. Data Set Description.** The dataset used in this paper was the price of London Sugar Futures from April 2010 to May 2020. Data are fetched from the investing.com. Table 1 shows the statistic analysis of this dataset.

To implement the different experiments, we divide the data into three sets: training set, validation set, and testing set. The training set is used to train different models. The validation set is used to select the optimal hyperparameters, and the testing set is used to compare the models. The details of those three sets are described in Table 2.

**4.2. Performance Evaluation.** To assess the superiority of the proposed hybrid deep learning model, the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) are employed for forecasting one day ahead sugar price. The formulas of MAE, MAPE, and RMSE are as follows:

$$\text{MAE} = \frac{\sum_{i=1}^m |y_i - \hat{y}_i|}{m}, \quad (9)$$

$$\text{MAPE} = \frac{\sum_{i=1}^m |y_i - \hat{y}_i| / y_i}{m} \times 100\%, \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}}, \quad (11)$$

where  $m$  is the number of prediction datasets,  $y_i$  is the real value of feed valve opening, and  $\hat{y}_i$  is the prediction value.

**4.3. Experiment I.** In this section, the dataset from London Sugar Futures is used to verify the superiority of time series decomposition technique EMD and the hyperparameters optimization algorithm TPE, respectively. Three models such as the model used EMD and TPE combining LSTM (TPE-EMD-LSTM), the model used EMD combining LSTM (EMD-LSTM), and the single LSTM model are put together for comparison. In order to have a fair comparison, those models are using the same structure which is shown in Figure 3.

For the LSTM model, we set 100 neurons in each LSTM layer, and with ReLU activation function, the dropout rate is 0.5 of dropout layer. The EMD-LSTM model has the same hyperparameter with the LSTM model. After optimization, the obtained optimal hyperparameters of TPE-EMD-LSTM are summarized in Table 3. Moreover, the RMSE, MAE, and MAPE are used to test the performance of the TPE-EMD-LSTM and other compared models. The results are shown in Table 4. The TPE-EMD-LSTM model attained 2.415 of MAE, 0.682 of MAPE, and 2.969 of RMSE. These values are the smallest among the three models, indicating the TPE-EMD-LSTM model surpasses the LSTM and EMD-LSTM model. Figure 4 shows the IMFs after decomposition. The performance of the EMD-LSTM is better than that of the LSTM in terms of achieving smaller MAE, MAPE, and RMSE, further ascertaining the performance of EMD to handle non-stationarity features in sugar price series.

**4.4. Experiment II.** In order to further explore the effectiveness of TPE-EMD-LSTM, experiment II compares the hybrid deep learning model with two categories of forecasting models. For the first category, comparisons are made across some basic single models, including the traditional statistical model ARIMA, the well-known support vector regression (SVR) model, and some deep learning models, such as the DNN model, and the selected single model LSTM, and the

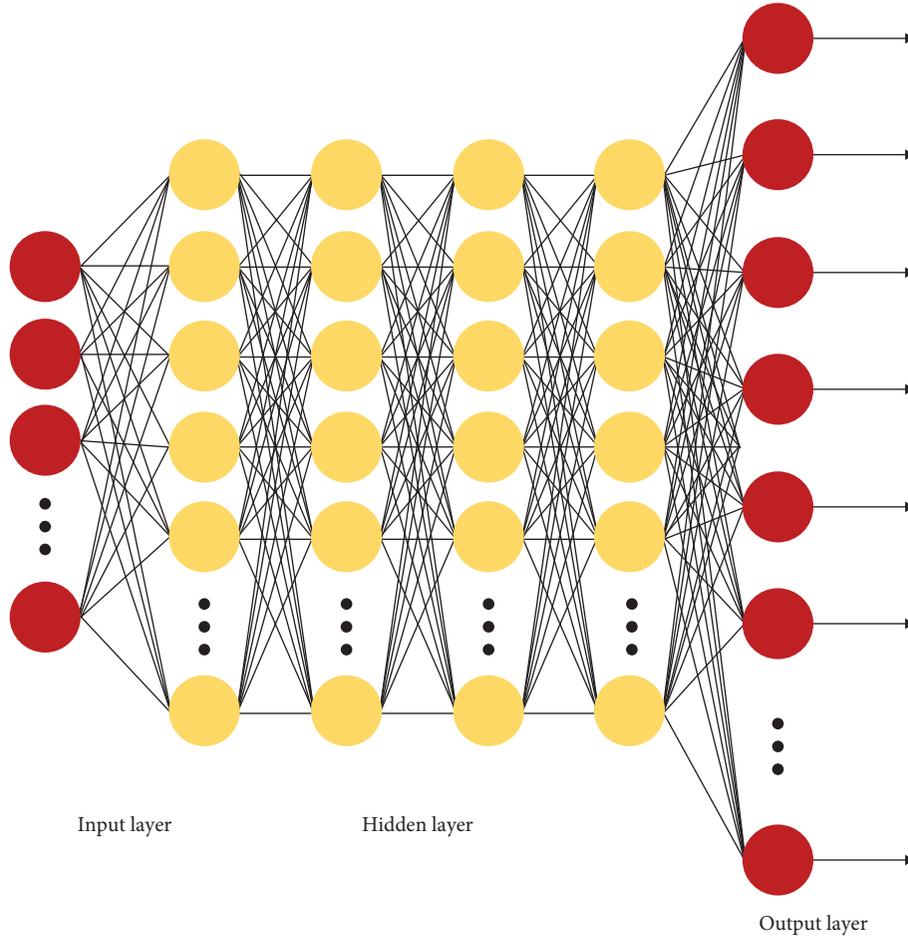


FIGURE 2: The general structure of deep neural networks.

TABLE 1: Statistical analysis of sugar price.

Minimum	Maximum	Mean	Standard deviation	Count
294.00	876.00	473.95	125.77	2561

TABLE 2: Details of the sets.

Sets	Time span	Numbers
Training	19/04/2010–29/05/2018	2063
Validation	30/05/2018–22/05/2019	249
Testing	23/05/2019–15/05/2020	249

proposed hybrid model. The DNN model has the same structure of the LSTM model, which is 100 neurons in each DNN layer and with ReLU activation function, and the dropout rate is 0.5 of dropout layer. For the second category, in order to fully test the hybrid model, two other prediction models are applied, including TPE-EMD-DNN and EMD-DNN. After optimization, the obtained optimal hyperparameters of TPE-EMD-DNN are summarized in Table 5.

The performance evaluation metrics for this test are listed in Table 6. The traditional time series model ARIMA has achieved considerable results with MAE value of 4.135, MAPE value of 1.189, and RMSE value of 5.747, which outperform the DNN and EMD-DNN model by a large

margin. However, after optimization, the TPE-EMD-DNN model performs the best in this section with MAE value of 2.648, MAPE value of 0.748, and RMSE value of 3.193. It demonstrated the effectiveness of the TPE algorithm. Moreover, the TPE-EMD-DNN model shows a significant improvement over the EMD-DNN; this not only shows the universality of TPE (effective for both LSTM and DNNs) but also shows its effectiveness.

## 5. Discussion

In order to evaluate the performance of the proposed hybrid deep learning model more comprehensively and find methods to improve forecasting capabilities and accuracy, followed by [15, 28, 42, 43], we perform the Diebold–Mariano (DM) test [44] and improvement percentage.

*5.1. DM Test.* The DM test is a method for making a comparison between the forecasting models and determines whether forecasts are significantly different. The DM test is described as follows:

$$DM = \frac{\bar{d}}{\sqrt{2\pi\hat{f}_d/T}} \text{Normal}(0, 1), \quad (12)$$

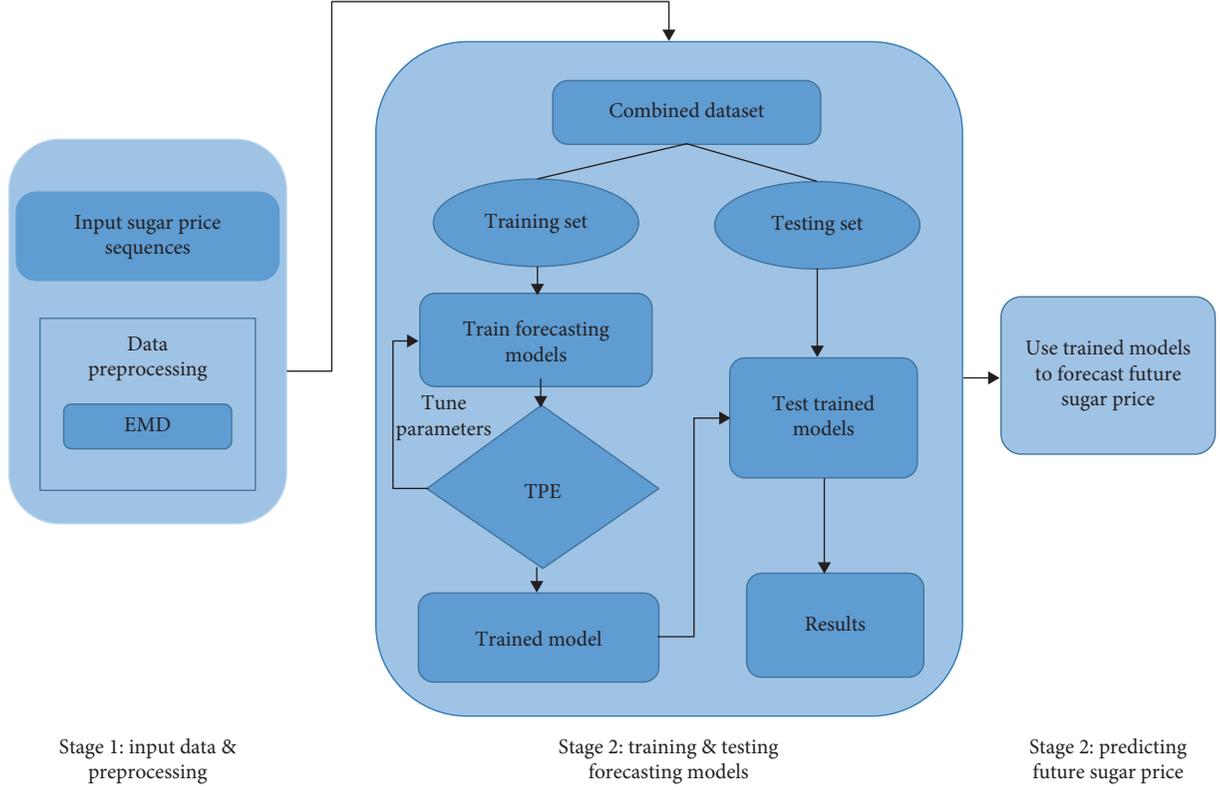


FIGURE 3: Decomposition results of sugar price.

TABLE 3: Optimal hyperparameters for the TPE-EMD-LSTM model.

Hyperparameter	Value
$l_{1,n}$	142
$l_{1,a}$	LeakyReLU
$l_{2,n}$	150
$l_{2,a}$	LeakyReLU
Dropout rate	0.2
Optimizer	Rmsprop
Batch size	32
Learning rate	0.01

TABLE 4: Comparison of the predictive accuracy of the various models.

Models	MAE	MAPE	RMSE
LSTM	6.208	1.796	8.187
EMD-LSTM	4.795	1.352	5.125
TPE-EMD-LSTM	<b>2.415</b>	<b>0.682</b>	<b>2.969</b>

where  $\hat{f}_d$  is the consistent estimate of spectral density of loss-differential,  $\bar{d}$  is the mean of the loss-differential between two forecasts, and  $T$  is the length of forecasting time series. The results of the DM test between our proposed TPE-EMD-LSTM and other models are shown in Table 7.

According to Table 7, the DM values range from 4.2 to 30.4, all far above 2.58, the upper boundary of the 1% significance level. That is to say, there is a significant difference between our proposed TPE-EMD-LSTM model and other models. This means that using the proposed TPE-

EMD-LSTM model will ultimately achieve the most important forecasting.

**5.2. Improvement Percentage.** As discussed in the DM test section, the TPE-EMD-LSTM model does significantly better than all other models. However, the details of forecasting performance improvement are not clear. Therefore, in this section, we apply three evaluation matrices ( $IP_{MAE}$ ,  $IP_{RMSE}$ ,  $IP_{MAPE}$ ) to discuss the superiority of TPE and EMD more specifically.  $IP_{MAE}$ ,  $IP_{RMSE}$ , and  $IP_{MAPE}$  represent the improvement percentages of MAE, RMSE, and MAPE, respectively. They are defined as follows:

$$IP_{MAE} = \frac{MAE_{m1} - MAE_{m2}}{MAE_{m1}} \times 100\%, \quad (13)$$

$$IP_{RMSE} = \frac{RMSE_{m1} - RMSE_{m2}}{RMSE_{m1}} \times 100\%, \quad (14)$$

$$IP_{MAPE} = \frac{MAPE_{m1} - MAPE_{m2}}{MAPE_{m1}} \times 100\%, \quad (15)$$

where  $MAE_{m1}$ ,  $RMSE_{m1}$ , and  $MAPE_{m1}$  are the MAE, RMSE, and MAPE values of the compared model and  $MAE_{m2}$ ,  $RMSE_{m2}$ , and  $MAPE_{m2}$  are the MAE, RMSE, and MAPE values of the model being compared. For example, for EMD-DNN vs. DNN,  $IP_{MAE}$  shows the improvement percentages of EMD-DNN compared with DNN; if the value of  $IP_{MAE}$  is positive, it means that the EMD-DNN surpasses DNN in terms of MAE, and vice versa.

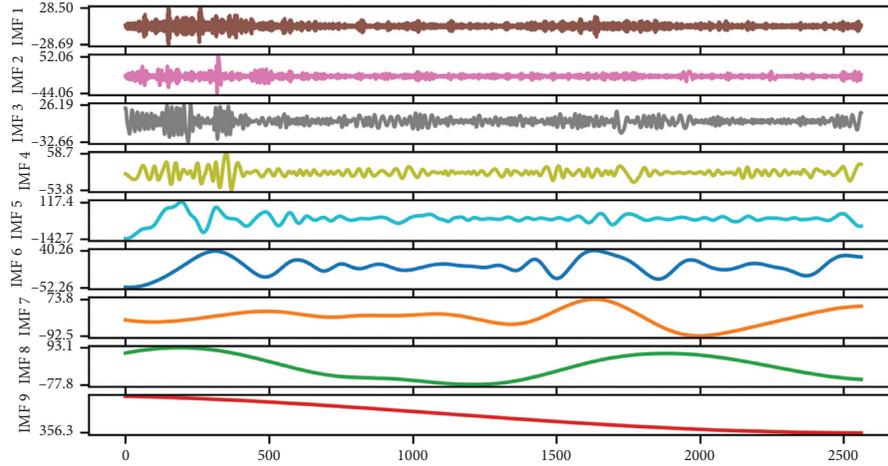


FIGURE 4: The process of sugar price forecasting based on the proposed model.

TABLE 5: Optimal hyperparameters for the TPE-EMD-DNN model.

Hyperparameter	Value
$l_{1,n}$	39
$l_{1,a}$	LeakyReLU
$l_{2,n}$	186
$l_{2,a}$	LeakyReLU
Dropout rate	0.2
Optimizer	Rmsprop
Batch size	32
Learning rate	0.01

TABLE 6: Comparison of the predictive accuracy of the various models.

Models	MAE	MAPE	RMSE
ARIMA	4.135	1.189	5.747
SVR	30.877	9.251	33.884
DNN	12.679	3.675	15.915
EMD-DNN	14.083	3.968	14.756
TPE-EMD-DNN	<b>2.648</b>	<b>0.748</b>	<b>3.193</b>

TABLE 7: DM test results between our proposed model and other models.

Model	Value
ARIMA	13.525
SVR	7.177
LSTM	13.039
EMD-LSTM	30.444
DNN	13.387
EMD-DNN	27.912
TPE-EMD-DNN	4.221

Two comparisons are conducted. Firstly, EMD-DNN and EMD-LSTM are compared with DNN and LSTM, respectively. Secondly, TPE-EMD-DNN and TPE-EMD-LSTM are compared with EMD-DNN and EMD-LSTM, respectively. Moreover, we found that the traditional time series forecasting model ARIMA still has strong performance on sugar price forecasting. Therefore, ARIMA is compared with LSTM and DNN to discuss the traditional time series model and deep learning based model. Table 8

TABLE 8: Results of improvement percentages.

Compared models	$IP_{MAE}$	$IP_{MAPE}$	$IP_{RMSE}$
EMD-DNN vs. DNN	-11.073	-7.973	7.282
EMD-LSTM vs. LSTM	22.761	24.722	37.401
TPE-EMD-DNN vs. EMD-DNN	81.597	81.149	78.361
TPE-EMD-LSTM vs. EMD-LSTM	49.635	49.556	42.068
TPE-EMD-LSTM vs. TPE-EMD-DNN	8.799	8.824	7.015
ARIMA vs. LSTM	33.392	33.797	29.803
ARIMA vs. DNN	67.387	67.646	63.889

summarizes the improvement percentage results of different models.

First of all, compared with DNN, it should be noted that EMD is more effective for LSTM network. EMD-LSTM vs. LSTM shows 22.761, 24.722, and 37.401 of  $IP_{MAE}$ ,  $IP_{MAPE}$ , and  $IP_{RMSE}$ , respectively. However, the EMD-DNN vs. DNN only shows improvement in  $IP_{RMSE}$ , and it indicates that the DNN is not as good as the LSTM network in capture EMD features for time series prediction. For TPE optimization, it shows a huge improvement for EMD-DNN and EMD-LSTM. The TPE optimization enhances the forecasting precision to a great extent as the improvement percentages relative to the contrasted models are prominent. From the comparison of TPE-EMD-LSTM with TPE-EMD-DNN, it can be seen that even if TPE optimization brings a huge improvement to the DNN, the LSTM network is still better than the DNN. Finally, ARIMA still outperforms DNN and LSTM, which shows the effectiveness of the traditional time series model and can be found in many time series applications [30]. However, it also should be noticed that after hyperparameter optimization, the deep learning-based model shows significant improvement in forecasting accuracy and surpassing the ARIMA model.

In the next step, ensemble forecasting will be investigated as it was surged recent years and shows state of the art performance on many time series forecasting task [33]. Also, in order to verify its generalizability and robustness, the proposed model needs to be applied to predict other food commodities.

## 6. Conclusions

Sugar price forecasting plays a vital role in policy making of sugar industries. In order to accurately predict sugar price, a hybrid deep learning model that utilizes the merit of time series decomposition technology and a hyperparameter optimization algorithm is proposed. This enhances forecasting performance of the proposed model compared with all other models. A large number of experiments have been conducted to prove the effectiveness of TPE and EMD. Moreover, DM tests are conducted to find improvement percentage and to reveal their specific effects. The DM values from Diebold–Mariano tests between our proposed TPE-EMD-LSTM model and other models range from 4.2 to 30.4, all far above 2.58, the upper boundary of the 1% significance level, indicating huge significance between the compared models. The improvement percentage of the proposed TPE-EMD-LSTM model to the other three models (EMD-DNN, EMD-LSTM, and TPE-EMD-DNN) in the values of IPRMSE is 78, 42, and 7, indicating a significant improvement. This enhanced forecasting performance will contribute to the sugar factory's recruitment plan for new employees and the sugarcane farmers' plan for planting sugarcane. After finishing this sugar price forecasting work, our future work is to develop ensemble forecasting as it is surged recent years and shows state of the art performance on many time series forecasting task [33]. Also, in order to verify its generalizability and robustness, the proposed model needs to be applied to predict other food commodities.

## Data Availability

The dataset used in this paper was the price of London Sugar Futures from April 2010 to May 2020. Data are fetched from the investing.com.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The project was supported by National Natural Science Foundation of China (nos. 51465003 and 61763001) and Innovation Project of Guangxi Graduate Education (nos. YCBZ2020012 and YCBZ2021019).

## References

- [1] I. Abdul-Aziz and I. Paul Alagidede, "Monetary policy and food inflation in South Africa: a quantile regression analysis," *Food Policy*, vol. 91, Article ID 101816, 2020.
- [2] R. Trostle, *Global Agricultural Supply and Demand: Factors Contributing to the Recent Increase in Food Commodity Prices*, Diane Publishing, Collingdale, PA, USA, 2010.
- [3] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A review and evaluation of the state-of-the-art in pv solar power forecasting: techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, Article ID 109792, 2020.
- [4] J. E. Meyer and Y. Kim, "An autoregressive forecast of the world sugar future option market," *Journal of Financial and Quantitative Analysis*, vol. 10, no. 5, pp. 821–835, 1975.
- [5] Y. Xu, S. Shen, and Z. Chen, "Research on forecast of sugar price based on improved neural network," in *Proceedings of the 2009 Second International Symposium on Intelligent Information Technology and Security Informatics*, pp. 12–15, Moscow, Russia, January 2009.
- [6] O. Ribeiro and S. M. Oliveira, "A hybrid commodity price-forecasting model applied to the sugar–alcohol sector," *Australian Journal of Agricultural and Resource Economics*, vol. 55, no. 2, pp. 180–198, 2011.
- [7] N. Silva, I. Siqueira, S. Okida, S. L. Stevan, and H. Siqueira, "Neural networks for predicting prices of sugarcane derivatives," *Sugar Tech*, vol. 21, no. 3, pp. 514–523, 2019.
- [8] M. Pan, C. Li, R. Gao et al., "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization," *Journal of Cleaner Production*, vol. 277, Article ID 123948, 2020.
- [9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [10] P. Goyal, P. Dollár, R. Girshick et al., *Accurate, Large Minibatch Sgd: Training Imagenet in 1 Hour*, <https://arxiv.org/abs/1706.02677>, 2017.
- [11] D. P. Kingma and B. Jimmy, "Adam: a method for stochastic optimization," 2014, <https://arxiv.org/abs/1412.6980v9>.
- [12] S. J. Taylor and B. G. Kingsman, "Non-stationarity in sugar prices," *Journal of the Operational Research Society*, vol. 29, no. 10, pp. 971–980, 1978.
- [13] D. Wang, S. Wei, H. Luo, C. Yue, and O. Grunder, "A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine," *Science of The Total Environment*, vol. 580, pp. 719–733, 2017.
- [14] S. Zhu, X. Lian, L. Wei et al., "PM2.5 forecasting using SVR with PSO-GSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors," *Atmospheric Environment*, vol. 183, pp. 20–32, 2018.
- [15] S. D. Harsh, D. Deb, and M. Josep, "Hybrid machine intelligent svr variants for wind forecasting and ramp events," *Renewable and Sustainable Energy Reviews*, vol. 108, pp. 369–379, 2019.
- [16] R. Prasad, R. C. Deo, Y. Li, and T. Maraseni, "Input selection and performance optimization of ann-based streamflow forecasts in the drought-prone murray darling basin region using iis and modwt algorithm," *Atmospheric Research*, vol. 197, pp. 42–63, 2017.
- [17] Y. Jiang, G. Huang, X. Peng, Y. Li, and Q. Yang, "A novel wind speed prediction method: hybrid of correlation-aided dwt, lssvm and garch," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 174, pp. 28–38, 2018.
- [18] Z. Hajiabotorabi, A. Kazemi, and F. Faramarz, "Improving dwt-rnn model via b-spline wavelet multiresolution to forecast a high-frequency time series," *Expert Systems with Applications*, vol. 138, Article ID 112842, 2019.
- [19] N. K. Yong and N. Awang, "Wavelet-based time series model to improve the forecast accuracy of PM10 concentrations in Peninsular Malaysia," *Environmental Monitoring and Assessment*, vol. 191, no. 2, p. 64, 2019.
- [20] M. Ali and R. Prasad, "Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition," *Renewable and Sustainable Energy Reviews*, vol. 104, pp. 281–295, 2019.

- [21] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.
- [22] Z. Wu, E. Norden, J. M. W. Huang, V. Brian, and X. Chen, "On the time-varying trend in global-mean surface temperature," *Climate Dynamics*, vol. 37, no. 3, pp. 759–773, 2011.
- [23] J. Bedi and D. Toshniwal, "Energy load time-series forecast using decomposition and autoencoder integrated memory network," *Applied Soft Computing*, vol. 93, Article ID 106390, 2020.
- [24] B. James, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," *Advances in Neural Information Processing Systems*, vol. 24, 2011.
- [25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [26] Y. Meng, S. Yu, J. Zhang et al., "Hybrid modeling based on mechanistic and data-driven approaches for cane sugar crystallization," *Journal of Food Engineering*, vol. 257, pp. 44–55, 2019.
- [27] Y. Meng, J. Zhang, J. Qin et al., "Research on the adaptive control in sugar evaporative crystallization using lssvm and sade-elm," *International Journal of Food Engineering*, vol. 15, pp. 5–6, 2019.
- [28] Y. H. Chen, W.-C. Hong, W. Shen, and N. N. Huang, "Electric load forecasting based on a least squares support vector machine with fuzzy time series and global harmony search algorithm," *Energies*, vol. 9, no. 2, 70 pages, 2016.
- [29] Y. Meng, S. Yu, Z. Qiu et al., "Modeling and optimization of sugarcane juice clarification process," *Journal of Food Engineering*, vol. 291, Article ID 110223, 2021.
- [30] B. S. Omer, U. G. Mehmet, and M. O. Ahmet, "Financial time series forecasting with deep learning: a systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, Article ID 106181, 2020.
- [31] C. Baziotis, N. Pelekis, and C. Doukeridis, "Datastories at semeval-2017 task 4: deep lstm with attention for message-level and topic-based sentiment analysis," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 747–754, Vancouver, Canada, August 2017.
- [32] M. W. Y. Lam, X. Chen, S. Hu et al., "Gaussian process lstm recurrent neural network language models for speech recognition," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7235–7239, Brighton, UK, May 2019.
- [33] M. Pawlikowski and A. Chorowska, "Weighted ensemble of statistical models," *International Journal of Forecasting*, vol. 36, no. 1, pp. 93–97, 2020.
- [34] K. Swanson, S. Trivedi, J. Lequieu, K. Swanson, and R. Kondor, "Deep learning for automated classification and characterization of amorphous materials," *Soft Matter*, vol. 16, no. 2, pp. 435–446, 2020.
- [35] J. Lago, F. De Ridder, and B. De Schutter, "Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms," *Applied Energy*, vol. 221, pp. 386–405, 2018.
- [36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, USA, 2016.
- [37] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, *Activation Functions: Comparison of Trends in Practice and Research for Deep Learning*, <https://arxiv.org/abs/1811.03378v1>, 2018.
- [38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [39] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude," *Coursera: Neural Networks for Machine Learning*, vol. 4, no. 2, pp. 26–31, 2012.
- [40] S. Ruder, *An Overview of Gradient Descent Optimization Algorithms*, <https://arxiv.org/abs/1609.04747v2>, 2016.
- [41] T. Nakama, "Theoretical analysis of batch and on-line training for gradient descent learning in neural networks," *Neuro-computing*, vol. 73, no. 1–3, pp. 151–159, 2009.
- [42] G.-F. Fan, S. Qing, H. Wang, W.-C. Hong, and H.-J. Li, "Support vector regression model based on empirical mode decomposition and auto regression for electric load forecasting," *Energies*, vol. 6, no. 4, pp. 1887–1901, 2013.
- [43] R. Karthik and S. Abhishek, *Machine Learning Using R: With Time Series and Industry-Based Use Cases in R*, Apress, New York, NY, USA, 2019.
- [44] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 134–144, 2002.