ISSN 2616-5872 Vol.3, Issue 2: 40-47, DOI: 10.25236/AJEE.2021.030208

PM2.5 prediction based on CEEMD-SSA-KELM

Jiaxun Liang[1], Kexin Yan[2]

[1]School of Quality and Technical Supervision, Hebei University, Baoding, Hebei, 071002, China
[2]School of Mathematics and Statistics, Ningbo University, Ningbo, Zhejiang, 315211, China

Abstract: *At present, air pollution is still the biggest environmental health problem we are facing. Although the content of $PM_{2.5}$ is small, it has a great impact on air quality and human health. In order to accurately predict $PM_{2.5}$, this paper proposed the CEEMD-SSA-KELM model. We preprocessed the data, predicted $PM_{2.5}$, compared and evaluated the fitting effect. We decomposed the obtained non-stationary, non-linear original data series into several smoother components on different scales by CEEMD for individual prediction as data pre-processing. After preprocessing, the obtained component data were respectively used as the input variables of SSA-KELM, and the final predicted value was obtained after processing each predicted value sequence. In the evaluation section we have selected four evaluation metrics to evaluate the model and compare it with different models. From the results of the analysis and comparison, we can see that the model proposed in this paper has better fitting effect, higher accuracy of fitting, and stronger stability.*

Keywords: *complementary ensemble empirical mode decomposition, sparrow search algorithm, kernel extreme learning machine, CEEMD-SSA-KELM hybrid model, $PM_{2.5}$ prediction*

1. Introduction

$PM_{2.5}$, also known as inhalable particulate matter, refers to particulate matter with an aerodynamic equivalent diameter less than or equal to 2.5 microns. Although the content of $PM_{2.5}$ in the earth's atmosphere is very small, it is rich in a lot of toxic and harmful substances, which will seriously affect air quality and endanger human health.

Currently, there are mainly two types of models for $PM_{2.5}$ prediction: statistical models and neural network models. Statistical models often use exponential smoothing [1], gray model [2], OLS [3] (Ordinary Least Squares), etc., but these models have strong limitations in dealing with nonlinear and non-stationary time series. Aiming at the shortcomings of the traditional statistical models, the neural network methods can train the historical data and have higher prediction accuracy in terms of nonlinear time series data. At present, CNN-LSTM [4] (Convolutional neural networks- Long-short term memory), CNN-BP [5] (Convolutional neural networks-Back propagation neural network), ICNN [6] (Interpolated convolutional neural network) and other methods are widely used in prediction. They have strong adaptive capabilities and are suitable for studying $PM_{2.5}$ prediction. However, with the increase of the order of the system, the traditional feedforward network has some problems, such as slow convergence rate, difficult parameter adjustment, poor generalization ability, etc. The model proposed in this paper can solve the above defects to a certain extent.

In this article, we proposed the CEEMD-SSA-KELM hybrid model to accurately predict $PM_{2.5}$. First, we used the CEEMD (Complementary ensemble empirical mode decomposition) method to preprocess the $PM_{2.5}$ time series and decompose it into multiple IMF components and residual components with different frequencies. After preprocessing, the regularization coefficient C of KELM (Kernel extreme learning machine) and the parameters of kernel function were optimized by SSA. Then input the determined training set samples of multiple IMF components and remaining components into the SSA-KELM model to obtain the predicted value of each component. After processing each predicted value sequence, the final predicted value of $PM_{2.5}$ concentration was obtained. In order to evaluate the accuracy of the model, we compared this model with CEEMD-KELM, SSA-KELM, KELM, ELM, Elman and BP, and drew a conclusion.

ISSN 2616-5872 Vol.3, Issue 2: 40-47, DOI: 10.25236/AJEE.2021.030208

2. Proposed methodology

2.1 CEEMD

In order to overcome the problem of non-adaptability of basis functions in traditional signal processing methods, N. E. Huang [7] and others creatively proposed a new method of Empirical Mode Decomposition (EMD), which is suitable for processing of nonlinear and non-stationary signals in 1998. However, the decomposed IMF may have the problem of mode aliasing. To solve this problem, Wu and Huang [8] proposed Ensemble Empirical Mode Decomposition (EEMD). But EEMD can not completely cancel the increased white noise in the process of signal reconstruction. In addition, the number of lumped averaging is generally more than a few hundred, which is very time-consuming, and the IMF after the lumped average may no longer meet the definition of IMF.

Therefore, a new data decomposition technology, complementary ensemble empirical mode decomposition (CEEMD), proposed by Yeh [9] et al. in 2010, is the optimization of EEMD. When CEEMD is used, the number of lumped averages will be reduced from several hundred orders of EEMD to several tens orders of CEEMD, which can significantly improve the operation effect and efficiency. CEEMD adds positive and negative pairs of random Gaussian white noise to the original signal, and then uses the EMD method to offset the noise added to the signal. The specific steps are as follows:

Step1: In the original signal, we add n pairs of random white noise with the same amplitude and opposite symbols into the original signal $x(t)$, $\begin{cases} x_i^+(t) = x(t) + n_i^+(t) \\ x_i^-(t) = x(t) + n_i^-(t) \end{cases}$, $i = 1, 2, \cdots, m.$ Where m is the number of iterations, $x_i^+(t)$ is positive noise, and $x_i^-(t)$ is negative noise.

Step2: We Perform EMD decomposition on $x_i^+(t)$ and $x_i^-(t)$ to obtain two groups of IMF and residual components.

Step3: We repeat steps 1 and 2, but add different white noise each time, and finally get 2m groups of IMF and remainder.

Step4: Take the mean value of multiple sets of components to obtain the final decomposition result

2.2 SSA

Sparrow Search Algorithm (SSA) is a new swarm intelligence optimization algorithm proposed by Xue J [10] in 2020 based on the foraging and anti-predation behavior of sparrows. Compared with other algorithms, it has high search accuracy and strong robustness. The specific principle is as follows:

Suppose there are N sparrows in the d-dimensional space, $X_i = [x_{i1}, x_{i2}, \cdots, x_{id}]$ represents the position of the i -th sparrow in the d -dimensional space, and the fitness value is $f_i = f([x_{i1}, x_{i2}, \cdots, x_{id}])$. Each sparrow has three possible identities: discoverer, follower and investigator. Select the sparrows with the best position in each iteration as the discoverer and the remaining sparrows as the followers.

Update the finder's location as follows: $X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\dfrac{-i}{\alpha \cdot iter_{\max}}\right), R_2 < ST, \\ X_{i,j}^t + Q \cdot L, R_2 \geq ST. \end{cases}$

Where t represents the current number of iterations, $j = 1, 2, \cdots, d$, α is a uniform random number in $(0,1)$, $iter_{\max}$ represents the maximum number of iterations, $X_{i,j}^{t+1}$ is the position of the i -th sparrow in the j -th dimension, Q is the random number obeying the standard normal distribution, $R_2 \in [0,1]$ is the warning value, ST is the warning threshold, and the value range is $[0.5,1]$.

The follower's location update formula is as follows:

ISSN 2616-5872 Vol.3, Issue 2: 40-47, DOI: 10.25236/AJEE.2021.030208

$$X_{i,j}^{t+1}=\begin{cases} Q\cdot\exp\left(\dfrac{X_{worst}-X_{i,j}^{t}}{i^2}\right), i>\dfrac{N}{2}, \\ X_p^{t+1}+|X_{i,j}^{t}-X_p^{t+1}|\cdot A^{+}\cdot L, \textit{otherwise.} \end{cases} \tag{1}$$

Where X_{worst} represents the global worst position after the t iteration, X_p^{t+1} represents the optimal position of the discoverer in the $t+1$-th iteration, A is an $1\times d$ -dimensional matrix, whose elements are randomly assigned to 1 or -1, and $A^{+}=A^{T}(AA^{T})^{-1}$. Thus, the above expression can be simplified as:

$$X_{i,j}^{t+1}=\begin{cases} Q\cdot\exp\left(\dfrac{xw_{i,d}^{t}-x_{i,j}^{t}}{i^2}\right), i>\dfrac{N}{2}, \\ xb_{i,j}^{t}+\dfrac{1}{D}\sum_{d=1}^{D}(rand\{-1,1\})\cdot(|xb_{i,j}^{t}-x_{i,j}^{t}|), \textit{otherwise.} \end{cases} \tag{2}$$

In the formula, xw represents the worst position of the sparrow in the current population, and xb represents the best position.

The location of the scout is updated as follows:

$$X_{i,j}^{t+1}=\begin{cases} X_{best}^{t}+\beta\cdot|X_{i,j}^{t}-X_{best}^{t+1}|, f_i>f_g, \\ X_{i,j}^{t}+K\cdot\left(\dfrac{|X_{i,j}^{t}-X_{worst}^{t}|}{(f_i-f_w)+\varepsilon}\right), f_i=f_g. \end{cases} \tag{3}$$

Where β is the step length control parameter, conforming to the standard normal distribution, $K\in[-1,1]$ is a random number, f_w represents the fitness value of the sparrow in the worst position, and f_g represents the fitness value of the sparrow in the best position. ε is just a constant in case the denominator is zero.

2.3 KELM

Due to the slow learning speed of the BP neural network and the complicated parameter adjustment in different scenarios, Huang [11] et al. proposed the extreme learning machine (ELM) in 2004. ELM has faster learning speed and simpler parameter adjustment, so it is widely used in the fields of predictive regression and classification. However, the singular values are prone to appear between the sample data, resulting in the instability of the ELM. Therefore, Huang [12] and others introduced the kernel function into ELM and proposed KELM. They used kernel function mapping to replace random hidden layer feature mapping, which further increased the generalization ability and stability of ELM and improved the robustness of the model. The basic principle is as follows:

The kernel matrix is defined according to Mercer's conditions:

$$\begin{cases} \Omega_{ELM}=HH^{T}, \\ \Omega_{i,j}=h(x_i)\bullet h(x_j)^{T}=K(x_i,x_j). \end{cases} \tag{4}$$

Use the kernel matrix Ω to replace the random matrix HH^{T} in the ELM. At this time, the output weight of the ELM network is $\hat{\beta}=H^{+}T=H^{T}(\frac{I}{C}+HH^{T})^{-1}T$, where I is the diagonal matrix and C is the regularization parameter. Therefore, the KELM output function can be expressed as

$$f(x)=h(x)\hat{\beta}=h(x)H^{T}(\frac{I}{C}+HH^{T})^{-1}T=\begin{bmatrix} K(x,x_l) \\ \vdots \\ K(x,x_N) \end{bmatrix}(\frac{I}{C}+\Omega_{ELM})^{-1}T. \tag{5}$$

ISSN 2616-5872 Vol.3, Issue 2: 40-47, DOI: 10.25236/AJEE.2021.030208

In this case, the kernel function $K(\mu,\nu)$ is usually set as the RBF kernel, namely

$$K(\mu,\nu)=\exp[-(\mu-\frac{\nu^2}{\sigma})]. \tag{6}$$

2.4 CEEMD-SSA-KELM

Step 1: We use the CEEMD method to decompose the original $PM_{2.5}$ time series data into multiple IMF components and residual components with different frequencies.

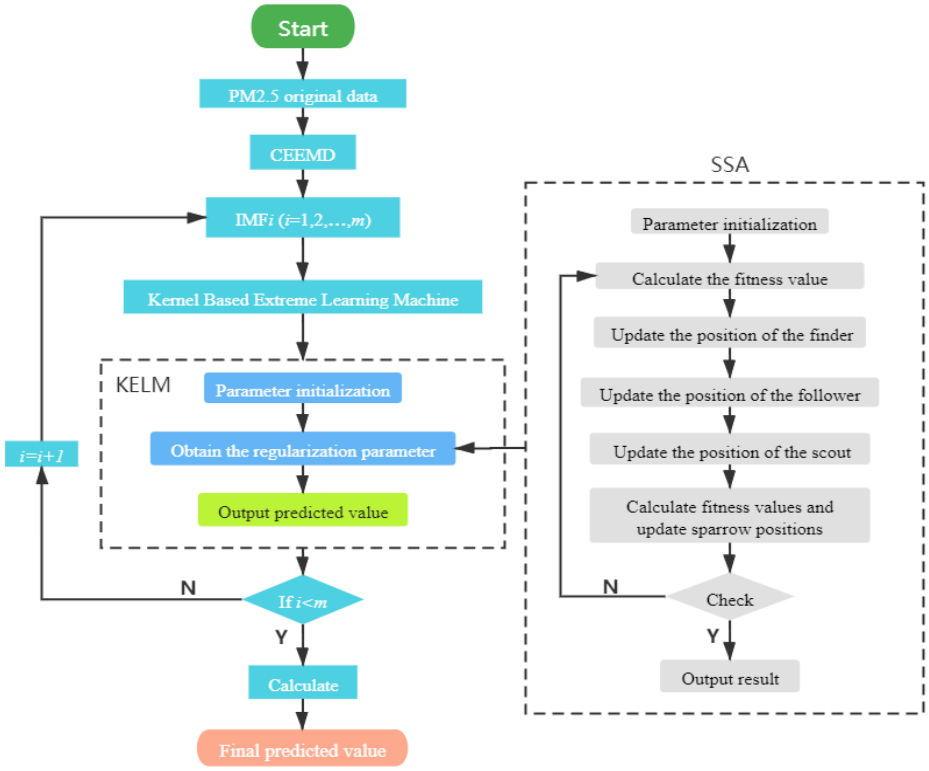

Figure 1: Flow chart of CEEMD-SSA-KELM model establishment.

Step 2: This paper initializes the relevant parameters of the SSA and KELM models, where the fitness function is chosen as the MSE of the error of the training set, and the smaller the MSE error indicates that the predicted data overlap with the original data, suggesting that the trained results are more accurate.

Step 3: Determine the training set samples and test set samples of multiple IMF components and residual components and use them as input variables of the SSA-KELM model.

Step 4: We use SSA to optimize the regularization coefficients C and kernel function parameters of KELM to obtain the predicted values of each component.

Step 5: We add the predicted value of each component to get the predicted result and compare the predicted results with the test data to obtain descriptive indicators of each data.

Step 6: We compare the prediction performance of different prediction models, analyze and summarize and draw conclusions.

3. Experiments and results analysis

1) Data sources and processing

The CEEMD-SSA-KELM model established in this paper was used to take a total of 1000 $PM_{2.5}$ data from October 4, 2018 to June 30, 2021 in a region of Hebei province as the valid data for testing, and 900 of them were taken as the training set data and 100 as the test set data. The descriptive characteristics of these data are shown in Table 1.

ISSN 2616-5872 Vol.3, Issue 2: 40-47, DOI: 10.25236/AJEE.2021.030208

Table 1: Basic statistical characteristics of sample data

Index	Number of Samples	Mean	Standard Error	Median	Standard Deviation	Variance	Kurtosis	Skewness	J-Btest
$PM_{2.5}$	1000	60.55	1.59	44	50.34	2534.13	4.79	1.96	1

From Table 1, we can see that the standard deviation of this sample data is 50.34, so the data sample has a large volatility, and the corresponding skewness and kurtosis are 1.96 and 4.79, respectively. The J-Btest test result is h=1, so this data series does not conform to the normal distribution. We can find that the $PM_{2.5}$ series is non-stationary.

The $PM_{2.5}$ data series of this place from October 4, 2018 to June 30, 2021 is shown in Fig.2, and it can be seen that this data series has a large volatility and non-stationarity. As a consequence, we decompose the non-stationary data by CEEMD method to get the IMF components and one residual component at different scales, and the decomposition results are shown in Fig. 3.

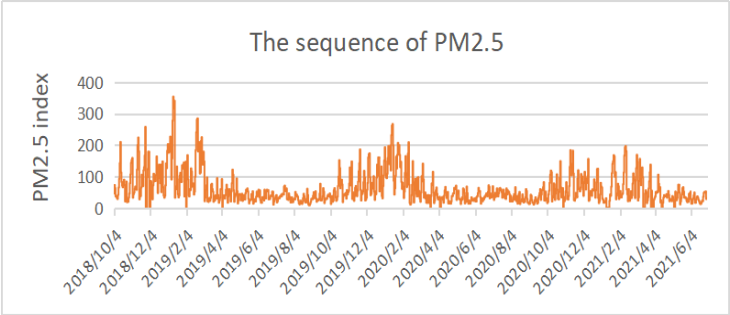

Figure 2: The sequence of PM2.5.

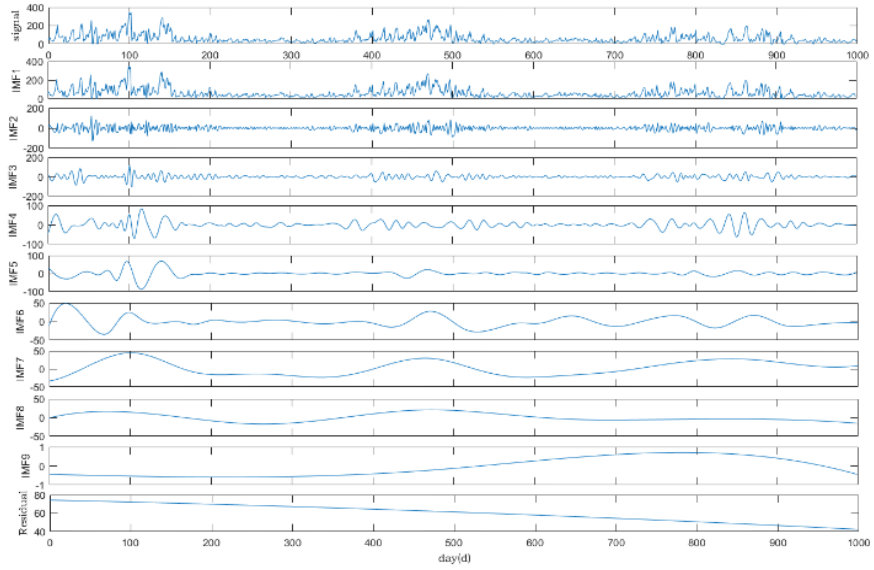

Figure 3: Graph of CEEMD decomposition results.

2) Selection of evaluation indexes

We use the following four metrics to evaluate the performance of the model in learning and testing, and we choose mean absolute error (MAE), mean squared error (MSE), mean square root error (RMSE) and coefficient of determination (R^2) as the evaluation metrics. The smaller the values of MAE, MSE and RMSE, the better the prediction effect.

3) Predicted results for each component

The prediction results of each component after the training of CEEMD-SSA-KELM modeling are shown in Fig. 4. The final sequence of $PM_{2.5}$ predicted values was compared with the actual $PM_{2.5}$ data as shown in Fig. 5. In order to verify the comparison of the fitting effect of the model proposed in this paper with other models, we selected the CEEMD-KELM model, SSA-KELM model, KELM model,

ISSN 2616-5872 Vol.3, Issue 2: 40-47, DOI: 10.25236/AJEE.2021.030208

ELM model, Elman model, and BP model for comparison. The prediction errors of each model are shown in Fig. 6, and the comparison graph of the prediction effects of each model is shown in Fig. 7.

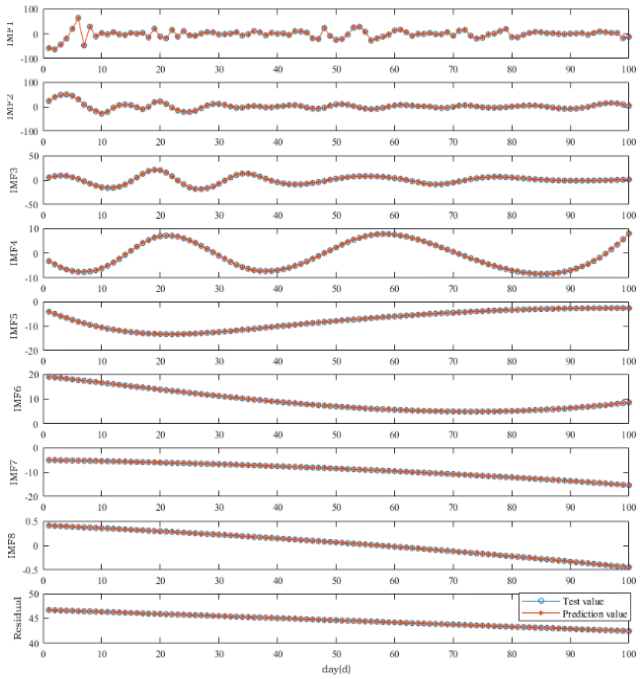

Figure 4: Prediction results of the CEEMD-SSA-KELM model.

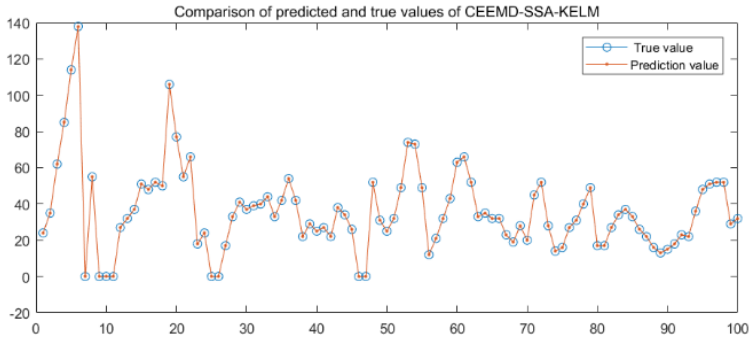

Figure 5: PM2.5 predicted value sequence.

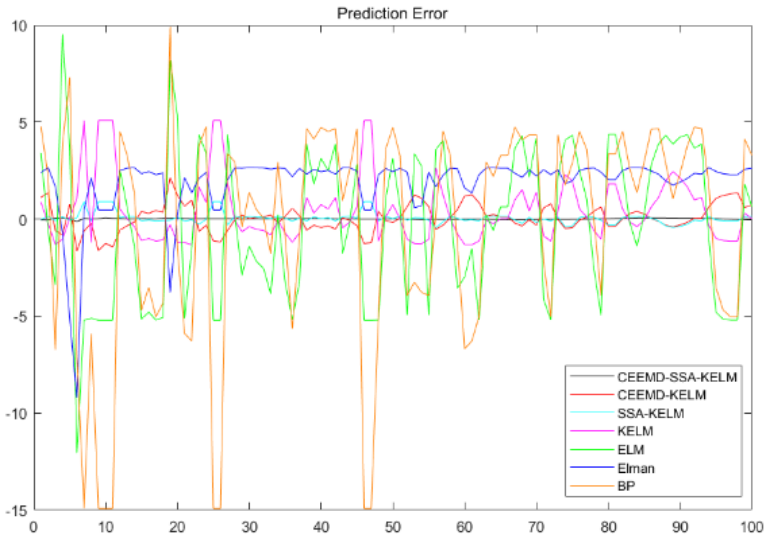

Figure 6: The prediction errors of each model.

ISSN 2616-5872 Vol.3, Issue 2: 40-47, DOI: 10.25236/AJEE.2021.030208

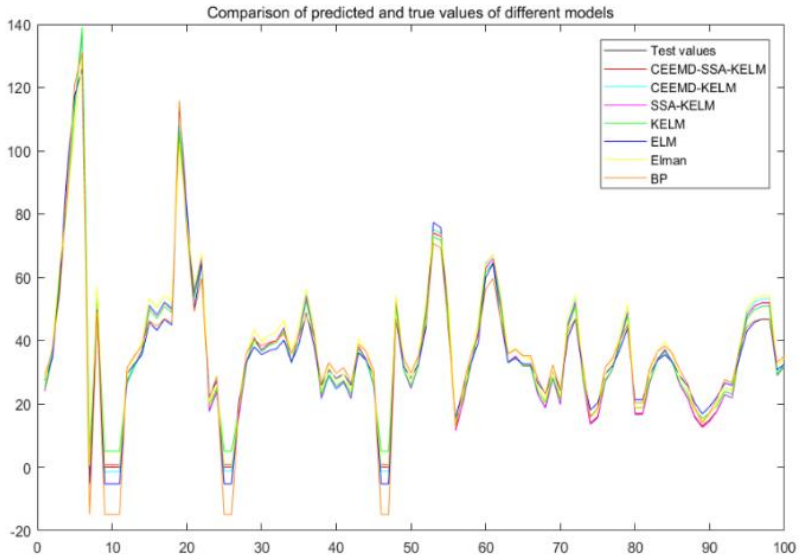

Figure 7: Comparison of predicted and ture values of different models.

4) Results Analysis

To further compare the accuracy of each model in predicting $PM_{2.5}$ series, we calculated the prediction errors of each model separately, as shown in Table 2 below.

Table 2: Comparison of error index of each model.

Model	Error index			
	MAE	MSE	RMSE	R^2
BP	4.5838	33.1279	5.7557	0.93881
Elman	2.2492	6.1087	2.4716	0.98872
ELM	3.4409	15.8713	3.9839	0.97068
KELM	1.2126	3.1055	1.7622	0.99426
SSA-KELM	0.18448	0.085569	0.29252	0.99984
CEEMD-KELM	0.53743	0.49591	0.70421	0.99908
CEEMD-SSA-KELM	0.023236	0.00088846	0.029807	0.99993

Compared with the BP neural network, the ELM learning speed is faster and the parameter adjustment is simpler, but the original ELM algorithm uses the least square method to calculate the output value, and singular values are prone to appear between sample data, which leads to unstable model performance. KELM uses kernel function mapping instead of random hidden layer feature mapping, which further increases the generalization ability and stability of ELM, and improves the robustness of the model. SSA optimizes the regularization coefficient C and kernel function parameters of KELM. It can also be seen from the data in the table that the MAE, MSE and RMSE of the CEEMD-SSA-KELM model proposed in this paper are lower than other models, and the R^2 value is higher, which indicates that the model proposed in this paper has higher fitting accuracy and stronger stability.

4. Conclusion

The accurate $PM_{2.5}$ prediction is not only helpful for the government to adjust the strategic deployment of pollution prevention and control in time, but also can improve the health level of citizens, so how to build an accurate $PM_{2.5}$ prediction model has become one of the hot spots of research. However, $PM_{2.5}$ concentration is related to many factors such as pollutants and meteorological conditions, among which there are many factors with high randomness and difficult to obtain data. To address many shortcomings in the current $PM_{2.5}$ concentration prediction models, this paper proposes a novel hybrid model of CEEMD-SSA-KELM. We use the CEEMD to decompose the obtained non-stationary original data sequence into several smoother components at different scales as individual predictions for data preprocessing. Then, the obtained component data are respectively used as the input variables of SSA-KELM, and the final predicted value is obtained after processing each predicted value sequence. And compared with the CEEMD-KELM, SSA-KELM, KELM, ELM, Elman and BP model, and conclude

ISSN 2616-5872 Vol.3, Issue 2: 40-47, DOI: 10.25236/AJEE.2021.030208

that the model proposed in this paper fits better and has a higher accuracy rate of fitting, as well as stronger stability. To a certain extent, this model has some realistic significance, and at the same time can provide a more reliable reference basis for the formulation of relevant strategic guidelines.

References

[1] Mahajan S, Chen L J, Tsai T C. Short-Term PM2.5 Forecasting Using Exponential Smoothing Method: A Comparative Analysis [J]. Sensors, 2018, 18(10).

[2] Jia Yang Li and Yang Zhao and Shen Kai Shi. The Multivariate Gray Model approach to predict the Concentration of Atmospheric Fine PM2.5 [J]. Advanced Materials Research, 2014, 3248(955-959): 2362-2365.

[3] Pohlmann, J. T., Leitner, D. W. A Comparison of Ordinary Least Squares and Logistic Regression. Ohio Journal of Science. 103(5), 118-125 (2003). 41.

[4] Huang C J, Kuo P H. A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities [J]. Sensors, 2018, 18(7): 2220.

[5] Kow P Y, Wang Y S, Zhou Y, et al. Seamless integration of convolutional and back-propagation neural networks for regional multi-step-ahead PM2.5 forecasting [J]. Journal of Cleaner Production, 2020: 121285.

[6] Chae S, Shin J, Kwon S, et al. PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network [J]. Scientific Reports.

[7] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. Proceedings of the Royal Society of London Series A, 1998, 454(1971): 903-995.

[8] WU Z H, HUANG N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method [J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41.

[9] YEH J R, SHIEH J S, HUANG N E. Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method [J]. Advances in Adaptive Data Analysis, 2010, 2(2): 135-156.

[10] Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm [J]. Systems science & control engineering, 2020, 8(1): 22-34.

[11] Huang Guangbin, Zhu Qinyu, Siew Cheekheong. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1/2 /3): 489-501.