Transactions on Computational and Scientific Methods | Vo. 1, No. 1, 2024
https://pspress.org/index.php/tcsm
Pinnacle Science Press

A Predictive Model for NOx Emissions in Coal-Fired Boilers Based on PCA-MFOA-LSSVM Integration Method

Lisa White[1], Christopher Harris[2]

[1]University at Buffalo, Lisalll99@gmail.com

[2]Drexel University, christopherh879@gmail.com

Abstract: The long prediction cycle and low accuracy of NOx emissions forecasts from power plant boilers significantly hinder effective pollutant control and emission reduction efforts. To address this, a predictive method integrating Principal Component Analysis (PCA), Modified Fruit Fly Optimization Algorithm (MFOA), and Least Square Support Vector Machine (LSSVM) is proposed. Initially, the high-dimensional sample matrix undergoes preprocessing. The sample space is then segmented based on the normalized levels of NOx emissions. PCA is employed to extract principal components from each subspace, thereby reducing dimensionality. The FOA is adapted into MFOA by incorporating an adaptive step size and modifying the odor determination value. Subsequently, LSSVM is utilized to develop prediction models for each subspace, with the kernel parameter and penalty factor optimized globally using MFOA. Finally, these sub-models are integrated through segmentation fitting to generate the overall model output. Simulation results demonstrate that the PCA-MFOA-LSSVM integrated method enhances prediction accuracy and reduces prediction time compared to other prediction models.

Keywords: Principal Components Analysis; Fruit fly Optimization Algorithm; Least Square SupportVector Machine; NOx emissions.

1. Introduction

As the enhancing management of pollutant emissions generated by coal-fired units in state, the accurate and effective prediction models for pollutant emissions are of great significance for the further control of pollutant emissions reduction. Nitrogen Oxides (NOx) is one of the main components of thermal power generation pollutants, which is often of multivariate and strong coupling result in the difficulty of describing the process mathematically using simple traditional models (Song et al., 2018).

At present, the common algorithm used for NO_x emissions prediction modeling are the machine learning algorithm (Gu et al., 2015; Li et al., 2018; Zhen et al., 2019), such as Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Least Square Support Vector Machine (LSSVM). Wang et al. (2016) used Genetic Algorithm (GA) to solve the problem of super-parameter optimization for LSSVM and constructed prediction model of NO_x emissions for coal-fired boilers which improved the precision and reduced the training time. Liu et al. (2019) proposed a boiler NO_x emissions prediction modeling method based on the Whale Optimization Algorithm-Least Squares Support Vector Machine (WOA-LSSVM) which enhanced the simulation performance. However, the following defects existed in this kind of predictive modeling method. The original input information could not be retained more completely when facing high-dimensional sample sets by using mechanism analysis of NO_x formation, variable screening and other methods, which may result in worse accuracy improvement of prediction. The parameter optimization process of LSSVM is too long to meet the requirements of short- term prediction in the case of slightly larger sample

sets, and takes general effect. In order to reduce the dimensionality of high-dimensional sample sets for LSSVM model, Zhen et al. (2019) proposed a modeling method for NOx emissions of boiler flue gas based on multi-model clustering integration, which improved the accuracy of modeling. Principal Component Analysis (PCA) is a method which can effectively preserve the original input information, reduce the dimensionality, simplify the system structure and overcome the correlation between variables (Zhong et al., 2015; Peng et al., 2016). The data segmentation fitting can perform better block calculation, lessen the calculation amount and increase the calculation speed (Lv et al., 2012) which can solve the problem that the optimization time is too long and the effect is not remarkable when the sample set is large. Then Fruit fly Optimization Algorithm (FOA) as a global optimization algorithm because of its simple algorithm, few parameters and high precision, it is often combined with LSSVM in various fields for predictive modeling. But it is easy to be limited to local optimum (Zhang et al., 2016; Xiao et al., 2016).

An integrated prediction algorithm combined PCA and MFOA-LSSVM for NO_x emissions of power boilers is proposed based on the above problems and referring to the documents of predecessors. According to the level of NO_x value, the sample space is divided into two subspaces and the subspaces are reduced by PCA. Then the principal components whose cumulative contribution rate higher than 90% are regarded as the input of LSSVM. Next, the parameter kernel function width and penalty factor of LSSVM are optimized by Modified Fruit fly Optimization Algorithm (MFOA). Then MFOA-LSSVM is used to establish the prediction model of each subspace, and the overlapping sample points are processed finally by segmentation fitting to get final output of the model. The simulation results show that the proposed PCA-MFOA-LSSVM integration model shortens the prediction time and achieves higher precision and better generalization ability.

2. Fundament of PCA, LSSVM and MFOA

2.1. Principal Component Analysis (PCA)

PCA (Abdi et al., 2010), a multivariate statistical technique, is used to mine internal correlations between multidimensional data. The method reduce dimension by using orthogonal transformation to transform a variable that may have correlation in the initial sample set into a new set of linear uncorrelated variables. The orthogonal transformation of PCA has different expressions in algebra and geometry. In algebra, it appears to transform from the covariance matrix or the correlation coefficients matrix of the initial vector into a diagonal matrix. In geometry, it appears to transform from the original coordinate system to a new orthogonal coordinate system. Therefore, such a transformation result in the variable points to p orthogonal directions in which the data points are distributed the most and the new variable retains the information of the initial variable to a large extent at the same time.

An initial sample $X=(X_1, X_2, ..., X_p)$ is assumed, here, p is the number of variables, $xi=(x_1,x_2,...,x_n)^T$, then the initial sample matrix can be represented the orthogonal matrix a by PCA in the form:

$$yi= a_{1i}X_1+a_{2i}X_2+...+a_{2i}X_n=a_iX \quad (1)$$

Orthogonal matrix a can be obtained by formula:

$$| R-\lambda I |=0 \quad (2)$$

2.2. Modified Fruit fly Optimization Algorithm (MFOA)

Large errors of final prediction modeling often result from directly using LSSVM, because the value of the kernel function σ and the penalty factor c will importantly affect the accuracy of the model prediction. Unreasonable parameter settings will result in poor reliability of modeling results (Si et al., 2017). As an interactive evolutionary optimization algorithm, Fruit fly Optimization Algorithm (FOA) achieves global optimization by simulating the foraging process of fruit fly groups. It is easy

to make the whole population become local optimal for FOA, but it may lead to slow convergence and weak convergence. Therefore, MFOA is used to optimize σ and c to improve prediction accuracy. The following improvements are based on FOA:

(1) In order to prevent precocity of FOA, an adaptive step size is introduced to change the search step size with the progress of iteration.

(2) It is supposed that i is the current number of iterations, and δ(i) is the search step size at the i-th iteration. That is, the search step size which S*(i-1) reaches S*(i) required. And S*(i) is the optimal odor concentration value at the i-th time iteration, and θ=|S*(i)-S*(i-1)| is the absolute value of the odor error.

$$\begin{cases} \delta(i) = 3 or 15, i = 1 \\ \delta(i+1) = \delta(i)\left[1 - \frac{S^*(i)}{S^*(i-1)}\right], i \geq 2, S^* < S^*(i-1) \\ \delta(i+1) = \delta(i)\left[1 + \frac{S^*(i-1)}{S^*(i)}\right], i \geq 2, S^*(i) \geq S^*(i-1) \end{cases} \tag{3}$$

The initial search step δ(1) is set to 3 or 15, and the initial optimal odor concentration value is S*(1). It indicates that the optimal odor concentration value of the i-th generation is better than that of the previous generation when S*(i)< S*(i-1), and then the search step size should be narrowed to improve the optimization precision. At the same time, it means that it is close to the optimal value when the value of θ is small and the search step should be sped up to quickly converge. The search step size should be increased to expand the search range to improve global optimization ability of the algorithm when S*(i) ≥ S*(i-1). The optimization effect at this time is not good when the value of θ is small, and the search step should be increased to find new spaces to continue searching.

(2) In order to overcome the premature convergence of the algorithm and improve the accuracy, the odor determination value is modified as follows:

$$W(i)=1/D(i)+\eta \quad i=1, 2, ..., \eta \tag{4}$$

$$\eta=\rho \cdot D(i) \tag{5}$$

Where, ρ is uniformly distributed and K is a constant.

At the same time, the fruit fly group is randomly divided into two parts because the FOA algorithm is easy to fall into local optimum which results in precocity of the algorithm. Part of the population start searching for food in a small range at the beginning, and the other part begins to search for food in a large range, which prevents the occurrence of local optimality.

3. Integration Model PCA-MFOA-LSSVM

3.1. Data Preparations

3.1.1 Selection of input variables The research object is a sub-critical 600MW double tangential boiler of a power plant. There are 48 pulverized coal burners distributed on the front and rear walls in six different heights, and there are five separate SOFA injectors on the top of the main bellows. The boiler system is a direct combustion system that uses six medium speed mills to produce pulverized coal and deliver it to six burners. The 3233 experimental data group of the boiler intercepted from the start-up to the steady-state in operating sampling cycle as initial sample set, which can better reflect the power system lifting load process. Figure 1 shows the load change of the initial sample set.

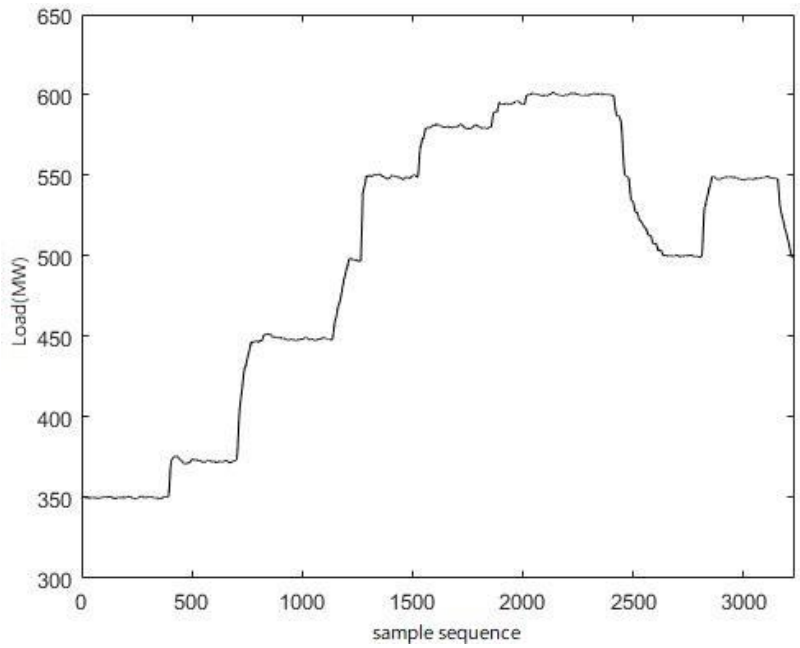


Fig 1. Load change of initial sample set

The characteristics of sample set include unit load, damper opening, working fluid flow and other index parameters, of which there are 37 input, and the output is NO_x emissions. The meaning of the specific operating parameters of the initial sample set and its variation range are given in Table 1. The type of coal quality will have a huge impact on it for the NO_x emissionsof coal-fired boilers, but the coal quality is not replaced during operation for the experimental data, the impact on the results is not considered.

3.1.2 Data processing In order to effectively prevent the occurrence of overfitting phenomenon, the initial sample set was processed by pauta criterion and the abnormal sample points were removed by using moving average filtering. The first 3000 sets of the initial sample set were used for training by the hold-out method to preserve the fidelity of the results, and the remaining samples were used for testing. Figure 2 shows the output portion of the initial sample set and how it is divided.

Table 1. Variable description of initial sample set

Variable meaning/(unit)	variation range
Unit load/(MW)	349.259~601.780
Total air volume/(t•h^{-1})	1144.100~1704.700
Total coal/(t•h^{-1})	182.065~317.144
Air flow on sides A and B/(t•h^{-1})	505.970~962.439
Furnace SOFA1-5 layer #1 corner burning secondary damper opening/ (%)	-0.001~98.966
Coal mill A-F inlet primary damper opening/ (%)	0.952~99.331
Coal mill A-F outlet temperature/(°C)	43.763~83.080
Secondary air flow at the inlet of the coal mill/(t•h-1)	365.745~523.516
Coal feeder A-F instantaneous flow/(t•h^{-1})	0.217~65.168
Air preheater A, B outlet secondary air pressure/(kPa)	0.385~1.012
Air preheater A, B outlet secondary air temperature/(°C)	273.066~314.182
Furnace outlet flue gas temperature 1, 2/(°C)	613.926~808.890
Main steam pressure A, B/(kPa)	7.029~15.196
NOx emissions/(mg/m-3)	246.712~345.750

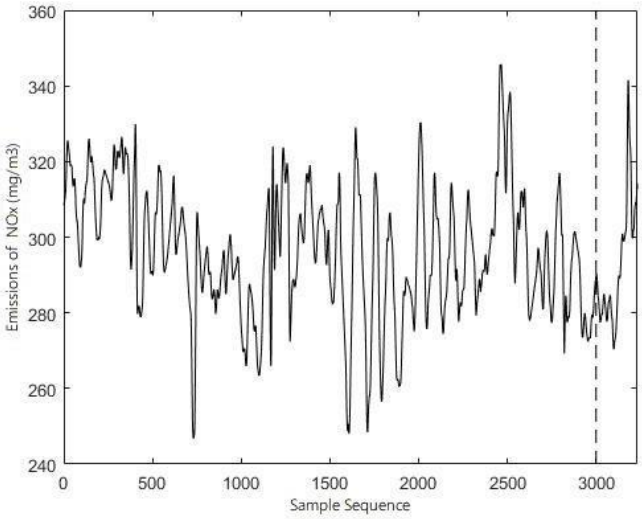


Fig 2. Output section of initial data sample set

There are often different dimensions and orders of magnitude, because each type of sample has different meanings and characteristics. For the comprehensive consideration and validity of the modeling results, all variables in the sample set were first normalized and defined as:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{6}$$

According to the output NO_x value, the sample data set was initially divided into two data subspaces D_l and D_h. The overlapping parts of the two subspaces can better reflect the dynamic characteristics of the system. The specific division rules are defined as:

$$\begin{cases} D\_l = \{X | 0 \le y_i < 0.55\} \\ D\_h = \{X | 0.45 \le y_i \le 1\} \end{cases} \tag{7}$$

3.2. Results of Model Simulation

Sample sub-spaces were normalized, and the NO_x emissions impact factor of each subspace was reduced according to the PCA. The main components and their contribution rates are shown in Table 2 and Table 3. It can be seen from Table 2 and Table 3 that the accumulative contribution rate of the first four principal components is 92.3142% for the sample space D_l and the cumulative contribution rate of the first three principal components is 90.5680% for the sample space D_h, which are all exceeding 90%. The PCA requirements can be extracted when the cumulative contribution rate is higher than 85%, so they are used to replace the initial input variables on the respective sample spaces.

Table 2. PCA results (partial) of space D_l

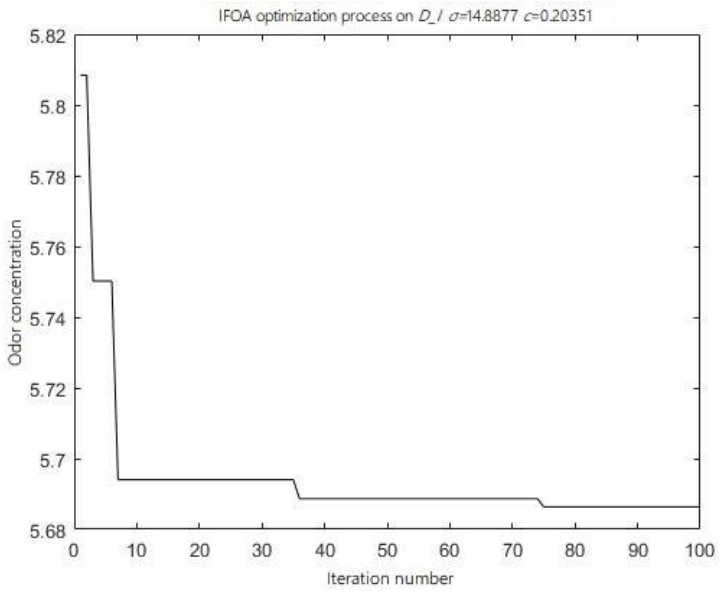
Principal component number	Eigenvalues	Contribution rate (%)	Cumulative contribution rate (%)
1	1.7132	58.5533	58.5533
2	0.4988	17.0470	75.6003
3	0.3794	12.9676	88.5679
4	0.1079	3.6868	92.2547
5	0.0603	2.0595	94.3142
6	0.0520	1.7762	96.0904
7	0.0298	1.0188	97.1092

8	0.0182	0.6209	97.7301

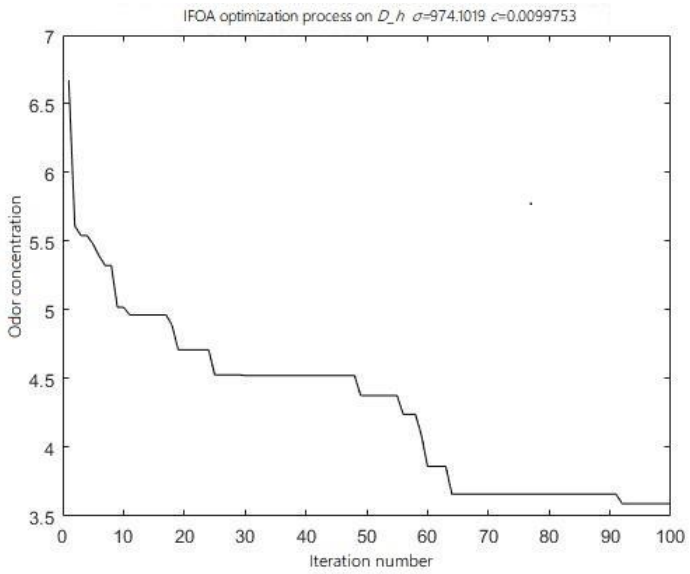
Table 3. PCA results (partial) of space D_h

Principal component number	Eigenvalues	Contribution rate (%)	Cumulative contribution rate (%)
1	2.2849	70.8227	70.8227
2	0.3457	10.7143	81.5370
3	0.2914	9.0310	90.5680
4	0.0798	2.4744	93.0424
5	0.0657	2.0357	95.0781
6	0.0412	1.2770	96.3552
7	0.0334	1.0344	97.3895
8	0.0196	0.6080	97.9976

The kernel function width σ and penalty factor c of LSSVM were super-parametric optimized by MFOA. The maximum number of iterations was n=100, the population size was m=20, and the population position coordinates were randomly initialized. Figure 3 (a) and (b) show the iterative process on subspaces D_l and D_h. And then the σ and c were obtained after parameteroptimization.



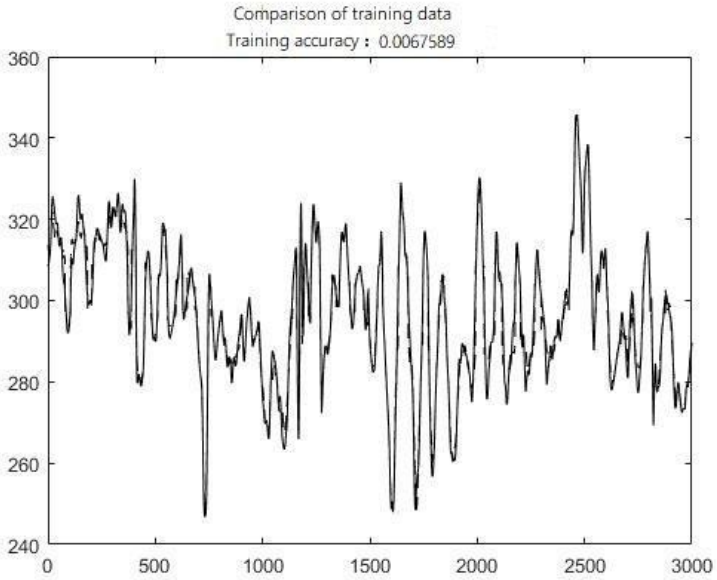
(a) Optimization process results of D_l space



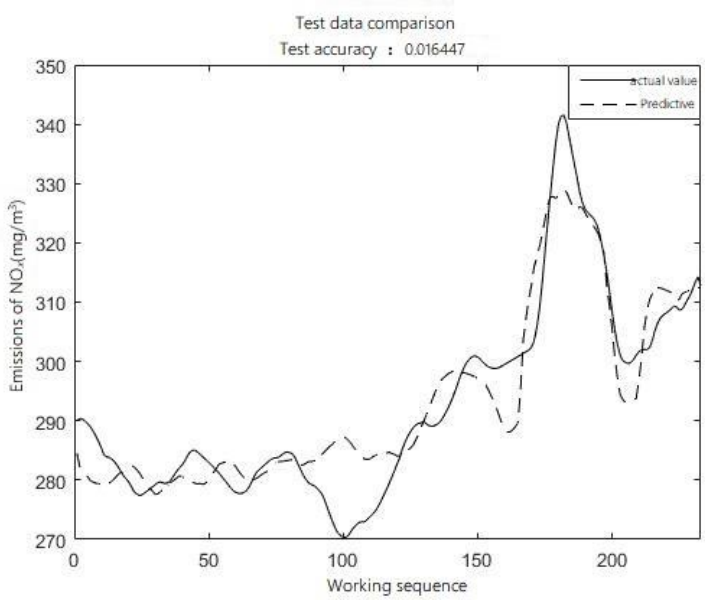
(b)Optimization process results of D_h space

Fig 3. Iterative process of MFOA

Subspace parameters obtained by MFOA optimization were substituted into LSSVM to establish a model of each subspace, and then the segmentation method was used to weight the overlapping sample points of the subspace that obtained the output of the model for the training sample and the test sample which was shown in Figure 4 (a) and (b) respectively.



(a) Results of training



(b) Results of test

Fig 4. Comparison of measured and predicted values

3.3. Comparison of the Prediction Performance Among the Selected Models

In order to verify that the PCA-MFOA-LSSVM integration model has better performance than other models, the genetic parameters and penalty factors of LSSVM were optimized by Genetic optimization Algorithm (GA), Particle Swarm Optimization (PSO) and Fruit fly Optimization Algorithms (FOA) respectively. At the same time, the models based on LSSVM and PCA-LSSVM was used as a comparative study. The output results for test set of different models are comparing given in Figure 5 and Table 4. Among them, LSSVM adopted grid parameter optimization. PSO, GA and FOA selects RMSE of LSSVM result as optimization function and odor concentration judgment function respectively.

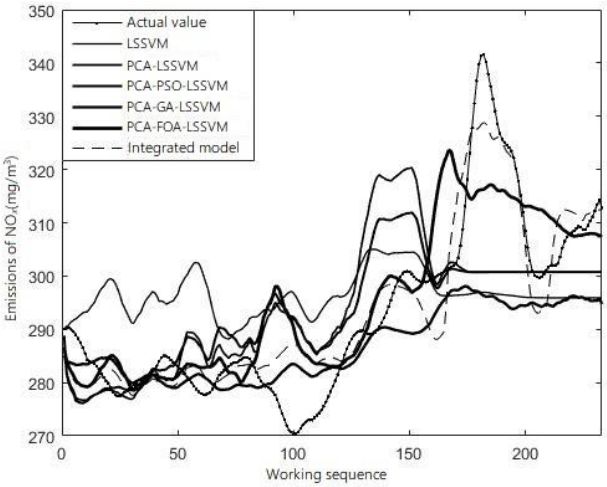


Fig 5. Comparison of measured and predicted values

Table 4. Performance comparison of different prediction models

Model	MAE/ (%)	MRE/ (%)	t/(s)
LSSVM	14.0698	4.78	3615.114
PCA-LSSVM	10.5114	3.52	3490.850
PCA-PSO-	9.6713	3.24	6144.510

LSSVM			
PCA-GA-LSSVM	8.5763	2.80	1202.339
PCA-FOA-LSSVM	6.9494	2.36	1909.733
Proposed model	2.1894	1.64	1052.034

It can be seen from Figure 5 and Table 4 that the integration model PCA-MFOA-LSSVM has a strong ability to track the change trend of the sample, and can significantly improve the prediction accuracy when compared to other selected models because the choice of kernel parameters and penalty factors will have a huge impact on the results for LSSVM, and general optimization methods are prone to fall into local optimums and precocity may occur. Therefore, the MFOA algorithm can be used to overcome these defects, which can further improve the accuracy of the model.

In the meanwhile, the integration model PCA-MFOA-LSSVM could obtain prediction results in a shorter time than other models and shorten the simulation time, because the calculation amount of kernel parameter matrix increases a lot when the LSSVM processes a sample set with a large amount of data which is time consuming. Therefore, the amount of computation is reduced by dividing the data space and partition modeling, so the operation time is shorten.

4. Conclusion

A PCA-MFOA-LSSVM integration method is proposed to predict NO_x emissions of power plant boilers and other similar prediction models are selected to serve as comparative study. The influence of correlation between inputs and output can be effectively eliminated after using PCA. And the prediction accuracy is obviously promoted and the prediction cycle is sharply shortened by means of applying MFOA and segmentation fitting when comparing with other selected prediction models. While the optimization performance of FOA has been enhance. In summary, the PCA-MFOA-LSSVM integration model can accurately predict the NOx emissions, effectively solve the problems of low modeling accuracy, low generalization, long prediction time, etc.

Literature Cited

[1] Gu L.-J., Y.-H. Li, L. Li; "Hybrid Model Prediction of Utility Boiler Combustion Optimization," Proceedings of the CSEE, 35, 2231-2237 (2015).

[2] H E Abdi and Williams L J; "Principal component analysis," Wiley Interdisciplinary Reviews: Computational Statistics, 2, 433- 459 (2010).

[3] Li C.B., S.-K. Li, Y.-Q. Liu; "Power load forecasting by wavelet least squares support vector machine with improved fruit fly optimization algorithm," Applied Intelligence, 33, 1122-1143 (2017).

[4] Li G.-Q., B. Chen, K.C.C. Chan, et al.; "Modeling Thermal Efficiency of a 300 MW Coal-Fired Boiler by Online Least Square Fast Learning Network," Journal of Chemical Engineering of Japan, 51, 100-106 (2018).

[5] Li H.-Z., S. Guo, H.-R. Zhao, C.-B. Su and B. Wang; "Annual electric load forecasting by a least squares support vector machine with a fruit fly optimization algorithm," Journal of Energies, 5, 4430-4445 (2012).

[6] Liu H.-Y, C.-G. Zhen; "Prediction Model of Boiler NOx Emissions Based on WOA-LSSVM," Journal of North China Electric Power University (Natural Science Edition), 46, 84-91 (2019).

[7] Lv Y., J.-Z. Liu, W.-J. Zhao, T.-T. Yang; "Steady-state detecting method based on piecewise curve fitting," Chinese Journal of Scientific Instrument, 33, 194-200 (2012).

[8] Peng T., Z. Cheng, X. Ji, et al.; "NOx emissions model for coal-fired boilers using principle component analysis and support vector regression," Journal of Chemical Engineering of Japan, 49, 211-216 (2016).

[9] Si G.-Q., S.-W. Li, J.-Q. Shi, et al.; "Least Squares Support Vector Machine Parameters Optimization Based on Improved Fruit Fly Optimization Algorithm with Applications," Journal of Xi'an Jiaotong University, 51, 14-19 (2017).

[10] Song Q.-K.and Y.-J. Hou; "Modeling Optimization for Boiler based on Modified Fruit fly Algorithm," Computer Integrated Manufacturing Systems, 35, 98-102+120 (2018).

[11] Wang G.-L., M. Lv, W.-J. Zhao; "LS-SVM Modeling for NOx Emissions of Power Plant Boiler Based onGenetic Algorithm," Automation & Instrumentation, 70-72 (2016).

[12] Xiao F. and G.-C. Chen; "Wind Power Short-Term Prediction Based on SVM Trained by Improved FOA," Journal of East China University of Science and Technology, 42, 420-426 (2016).

[13] Zhang W.-G., Y. Zhang, Y.-Z. Sun, et al.; "Combustion optimization for CFB boiler based on least square support vector machine and modified fruit fly optimization algorithm," Thermal Power Generation, 45, 44-49 (2016).

[14] Zhen C.-G., H.-Y. Liu; "Model for predicting NOx emissions from boilers based on MWOA-LSSVM integration," Journal of Chemical Engineering of Japan, 52, 702-709 (2019).

[15] Zhen C.-G., H.-Y. Liu; "Prediction model of NOx emissions from coal-fired boiler based on multi-model clustering ensemble," Thermal Power Generation, 48, 33-40 (2019).

[16] Zhong Y.-L., H.-S. Li, F.-S. Liu, et al.; "PCA-SVR model based NOx emissions prediction for coal-fired boilers," Thermal Power Generation, 44, 87-90 (2015).