

RESEARCH

Open Access



Noise prediction of chemical industry park based on multi-station Prophet and multivariate LSTM fitting model

Qingtian Zeng¹, Yu Liang¹, Geng Chen^{1*} , Hua Duan² and Chunguo Li³

*Correspondence:

gengchen@sdust.edu.cn

¹ College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China
Full list of author information is available at the end of the article

Abstract

With the gradual transformation of chemical industry park to digital and intelligent, various types of environmental data in the park are extremely rich. It has high application value to provide safe production environment by deeply mining environmental data law and providing data support for industrial safety and workers' health in the park through prediction means. This paper takes the noise data of the chemical industry park as the main research object, and innovatively applies the 3σ principle to the zero-value processing of the noise data, and builds an LSTM model that integrates multivariate information based on the characteristics of the wind direction classification noise data combined with the wind speed and vehicle flow information. The Prophet model integrating multi-site noise information was adopted, and the Multi-PL model was constructed by fitting the above two models to predict the noise. This paper designs and implements a comparative experiment with Kalman filter, BP neural network, Prophet, LSTM, Prophet + LSTM weighted combination prediction model. R^2 was used to evaluate the fitting effect of single model in Multi-PL, RMSE and MAE that were used to evaluate the prediction effect of Multi-PL on noise time series. The experimental results show that the RMSE and MAE of the data processed by the 3σ principle are reduced by 32.2% and 23.3% in the multi-station ordered Prophet method, respectively. Compared with the above comparison models, the Multi-PL model prediction method is more stable and accurate. Therefore, the Multi-PL method proposed in this paper can provide a new idea for noise prediction in digital chemical parks.

Keywords: Chemical industry park, Noise prediction, Prophet, LSTM, Multi-PL

1 Methods/experimental

In order to improve the accuracy of environmental noise prediction in chemical industry park, this paper proposes a multivariate and multi-station neural network model (Multi-PL) based on LSTM and Prophet. According to the periodicity of environmental data in the park, it is divided into multivariate data and multi-station data. Secondly, the structure and implementation of the model are introduced and explained in detail. Finally, the prediction accuracy under different proportions of training sets is compared through experiments, and different data sets and different models are used for experiments. Experiments with and without the use of the 3σ criterion were conducted to compare

the single model in different data sets, and to compare the model with other models. The results of experiment 4.2 show that the application of 3σ criterion and multivariate and multi-station data can improve the prediction performance of the single model. In addition, experiment 4.3 proves that Multi-PL is better than single model, traditional prediction method and LSTM + Prophet linear combination model.

2 Introduction

With the spread of 5G high-speed transmission technology, chemical industrial complexes are also entering the Era of Internet of Things (IoT) through sensors [1]. As the chemical park brings good economic benefits through the gathering of factories, pollution problems are gradually exposed. Exhaust gas and wastewater can be recycled and reused through Ecological Industrial Park, and noise, as a threat that is often overlooked, continues to affect human mental and hearing health. Factory noise may cause mild or moderate noise deafness [2]; noise can also cause headaches, insomnia, unresponsiveness, hearing loss and other symptoms [3–6]. The chemical park is surrounded by farmland and villages, and noise will have a negative impact on villagers' lives, animal breeding and natural ecology [7]. How to use effective methods to predict noise and dig out noise rules to reduce the impact on life and physical health is a problem that needs to be considered and solved.

IoT data contain a lot of useful information, such as satellite Industrial Internet of Things (IIoT) data can be used to solve service quality problems [8]. Noise prediction is restricted by many conditions. With the development and change of artificial intelligence technology, existing technologies can solve learning trends, big data classification and trend prediction problems by introducing environmental factors [9]. Information transmission in the IIoT is also limited by spectrum resources, so data loss is a common situation [10]. It is an extremely important research topic to dig out the laws of noise and predict the future noise level to be able to mitigate noise hazards [11]. Noise prediction research has received increasing attention. For example, the literature [11] proposed a gradient boosting model to predict noise, which combines multiple characteristics to analyze areas with severe noise exposure, and performs well under specific frequency sensors. [12] proposed a two-layer long short-term memory (LSTM) network to predict environmental noise under a large amount of data, which can reflect the change of noise level within a day, but only the time regularity of noise is considered. [13] proves that the LSTM model is better than the traditional ARIMA time series forecasting model. In literature [14], LSTM model is used for airport noise prediction, and metadata of aircraft type, trajectory information and weather data are also integrated into the model, resulting in higher prediction accuracy, but lack of consideration of spatio-temporal characteristics of noise. [15] proposed an integrated model of airport noise prediction based on space fitting and BP neural network, which integrates time and space characteristics to improve the accuracy and fault tolerance of prediction. However, the application area of this model is limited and not flexible enough. [16] established a feature-weighted support vector regression model FWSVR based on the time series similarity, which has generalization ability. [17] simulates the noise of a typical road network based on the existing traffic flow model. The above two methods are limited to univariate prediction

and lack information integrity. [18] uses the improved Federal Highway Administration (FHWA) model to predict the noise level. This method integrates multivariate information, but the information is not perfect in practical application. Environmental noise prediction still faces the following challenges: Noise has superposition and mutability, how to capture the noise law of the park? How to reduce the influence of sparse zero and outliers caused by sensor faults on the prediction without affecting the noise law? In addition to noise prediction, Prophet, Stackelberg model and extended Kalman filter have also been used by some researchers to achieve good results [19–21]. However, a single forecasting method cannot capture the distribution of complex time series patterns. More and more researchers are capturing complex time series distribution patterns based on hybrid forecasting models in order to obtain better forecast accuracy and performance [22]. There are three types of hybrid models for time series prediction. Hybrid model based on ARMA and machine learning [23, 24]: Literature [23] combined ARMA, PSO-SVM and clustering method for wind power generation prediction, and [24] uses the combined EMD-GM-ARMA model for coal mine safety production situation prediction. Hybrid model based on ARIMA and machine learning [25–27]: In literature [25], the mixed SSA-ARIMA-ANN model was used to predict daily rainfall, in [26], the combined ARIMA and ANN model was used to predict daily radiation and in [27], the mixed ARIMA and SVM model was used to predict corn futures price. Hybrid model based on machine learning [28–30]: Literature [28] uses CNN and AI-tuned SVM for power consumption prediction, literature [29] uses CNN-LSTM hybrid model for price sequence prediction, and literature [30] uses LSTM-RNN combined model for low-traffic flow forecast. The prediction accuracy obtained by applying the mixed model in the above literature is better than that of the single model, so the mixed model will be the key method to solve the problem of time series prediction of park noise. The above-mentioned literature focuses on noise pollution mainly on road traffic, airport, and urban environmental noise, ignoring the harm of noise in chemical parks. Motivated by the studies mentioned above, this paper studies the noise prediction of chemical industry park from the perspective of mixed model, which fills in the blank of the research direction of noise prediction in chemical industry park.

Based on the existing sensor distribution and traffic data in the chemical park, this paper builds a scene model suitable for the distribution characteristics of the park, constructs a noise multivariate data set and a multi-station data set according to the scene, and introduces the 3σ criterion to deal with the zero value of noise in order to improve the prediction accuracy. A Multi-PL model based on LSTM and Prophet models is proposed. Multivariate data set features such as wind speed, vehicle flow, and noise data based on wind direction classification are used in the multivariate LSTM model to improve the prediction accuracy. The multi-station noise data set is used as an additional regression variable for the Prophet model. Fitting the above model forms Multi-PL prediction model with higher accuracy.

The rest of this article is structured as follows. The second part introduces the research background, data set and preprocessing. The third part introduces the principle and construction of Multi-PL model. In the fourth part, the experimental results of the training model are given and evaluated in detail. The last part is summary and prospect.

3 System model and data set

3.1 System model

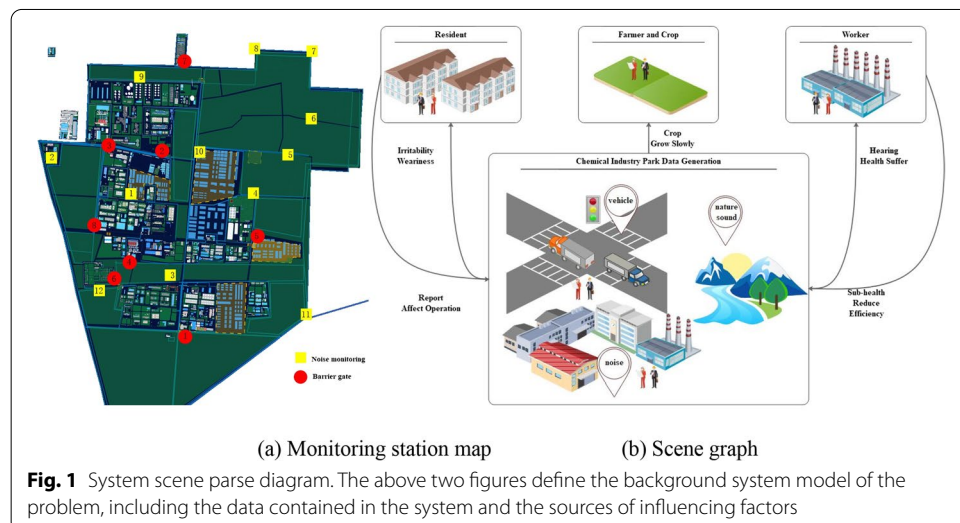
The research scene of this paper is an engineering plastics industrial park in Shandong Province, China. Based on the original smart chemical industry park, noise monitoring data are obtained through sensors. The collected data are accurate and effective, which provides an effective data basis for noise prediction.

The Park covers an area of 8.97 km² and is equipped with 12 air monitoring stations (no data at Station 11 due to failure) and 8 vehicle gate monitoring stations. At the mark in Fig. 1a, this paper takes the data of no.10 monitoring station and gate for analysis. There are three main sources of noise in the park:

1. There are a large number of vehicles in the park for the transportation, loading and unloading of chemical raw materials. The volume of vehicles will affect the noise level.
2. Chemical plants generally operate 24 h a day, and the impact of noise is not only periodic but also persistent.
3. Natural sounds, such as wind, also affect the overall noise level. Different wind directions will bring different regional sound effects.

According to Fig. 1b, noise affects the hearing health of workers in the park, reduces the growth rate of crops, and causes residents to be irritable and tired. Conversely, hearing loss leads to decreased work efficiency, and residents' behaviors affect the operation of the park.

In the face of many problems in the scene, noise prediction and risk identification can assist the park in planning the operation cycle and reduce the operation of noise source equipment during periods of high noise to avoid the occurrence of the above situations.



3.2 Data set construction

Based on the system model, we constructed the park noise prediction data set as shown in Fig. 2. Part A represents the noise data and natural environment information monitored by the air monitoring station, and part B represents the vehicle information recorded by the gates. The information is uploaded to the gateway and stored in the park database server.

We carry out preprocessing by reading the data in the server. In this paper, all data are constructed into two sub-data sets according to requirements: multivariate data set and multi-station data set.

3.2.1 Data set preprocessing

The data sets used in this paper are from the scenarios in Sect. 2.1 and span from 14:00 on August 22, 2020 to 01:00 on February 2, 2021. As shown in Fig. 2, data pre-processing mainly includes the following three tasks:

Step 1: Data cleaning. Noise has mutability, and the irregular 0 dB value of the data has a great influence on the prediction accuracy. The 3σ criterion is introduced to deal with outlier zero value. Sparse missing data are completed by KNN adjacent interpolation.

Step 2: Data screening. The original noise data interval is 30 s, and a noise sensor has 470,760 pieces of data. The data are too dense. The training process can be accelerated by resampling experimental data according to 10-min intervals.

Step 3: Traffic data parsing: All the vehicle information in the park is classified with the vehicle entry and exit status as tags, and statistics are made at 10-min intervals.

After the data set preprocessing is completed, we construct sub-data set and verify the correlation between noise data and different variables, laying a foundation for the subsequent prediction work.

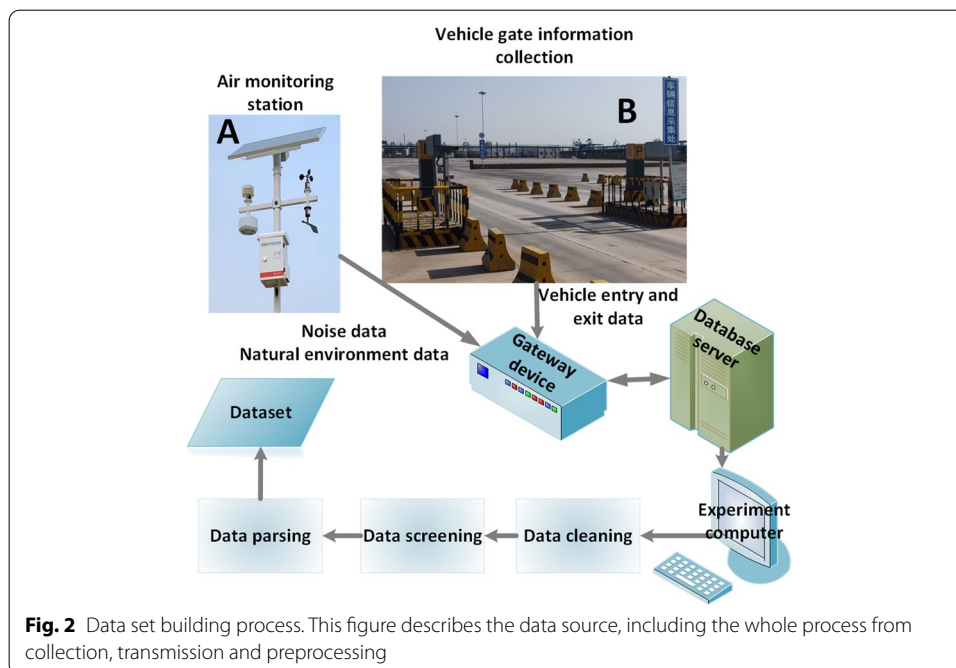


Fig. 2 Data set building process. This figure describes the data source, including the whole process from collection, transmission and preprocessing

3.2.2 Multivariate data set and multi-site data set

Multivariate data sets include vehicle flow, noise characteristics of adjacent stations based on wind direction classification (the construction method is located in Sect. 3), wind speed and noise. The multi-station dataset contains noise data from 11 monitoring stations.

The noise data and natural environment data are derived from part A of Fig. 2, including information such as temperature, wind speed, wind direction, light, noise, and PM2.5. The traffic flow data come from part B of Fig. 2. In Fig. 3b, c, the X-axis represents the time interval index (in days), and (b) the Y-axis represents the noise decibel value and wind speed. The blue and red curves represent the noise value and wind speed, respectively, (c) the Y-axis represents the noise decibel value and the number of traffic flows. The blue, red and green curves indicate the number of vehicles entering and leaving the park and the decibel level, respectively. In order to analyze the correlation of representative data in air monitoring stations, Pearson correlation coefficient ρ is introduced as follows:

$$\begin{cases} \rho_{NW} = \frac{\text{cov}(N^{(T,Y)}, W^{(T,Y)})}{\sigma_{N^{(T,Y)}}\sigma_{W^{(T,Y)}}} \\ \rho_{NN} = \frac{\text{cov}(N^{(T,Y)}, N^{(T,Y')})}{\sigma_{N^{(T,Y)}}\sigma_{N^{(T,Y')}}} \end{cases} \quad (1)$$

where $\text{cov}(\cdot)$ refers to the covariance operator, σ is the standard deviation, ρ_{NW} means in the same station at the same time the correlation coefficient of wind speed, $N^{(T,Y)}$ and $W^{(T,Y)}$, respectively, represent the noise and wind speed values of Y station at time T . ρ_{NN} represents the correlation coefficient between the noise values of different stations

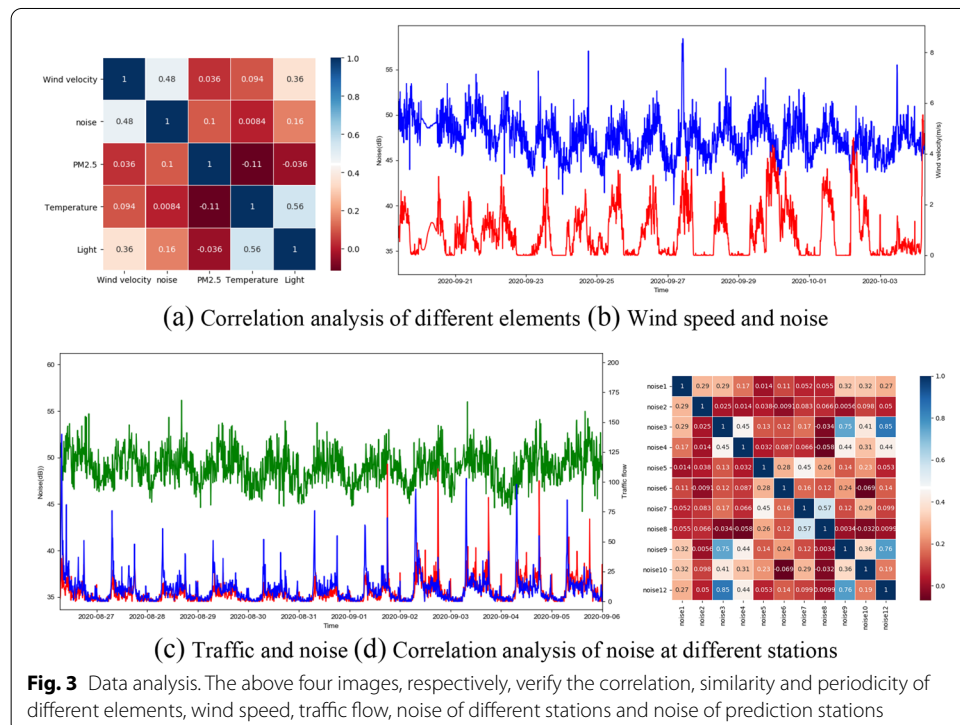


Fig. 3 Data analysis. The above four images, respectively, verify the correlation, similarity and periodicity of different elements, wind speed, traffic flow, noise of different stations and noise of prediction stations

at the same time, $N^{(T,Y)}$ and $N^{(T,Y')}$ represent the noise of Y and Y' at time T , respectively. Draw conclusions based on the information in Fig. 3:

1. Correlation of data. According to (a), the correlation coefficient between noise and wind speed is 0.48, which is the main influencing factor in the existing information. According to (d), the noise data of different stations are correlated.
2. Similarity of data. According to (a), the fluctuation trend of noise and wind speed is similar. It is necessary to correlate wind speed information to predict noise more accurately.
3. Periodicity of data: Traffic flow and noise level have similar periodicity. Among them, at zero o'clock, the peak of vehicle entry and exit is reached, and the second peak of traffic in the park is reached around 12 noon.

The multivariate data set contains the influence of traffic flow, wind speed and wind direction on noise change, and the multi-station data set contains the correlation between the noise of neighboring stations and the stations to be measured. The Multi-PL noise prediction method is proposed according to the unique data attribute of park.

4 Multi-PL model based on Prophet and LSTM

4.1 Multi-element LSTM model

LSTM (long short-term memory) network model is an improvement of RNN (recurrent neural network). The infrastructure of LSTM contains a part that controls the storage state, which can solve the problem of gradient disappearance encountered by RNN [31]. In this paper, the method of supervised learning is adopted, which does not require artificial construction of time series features. The time series curve can be fitted through deep learning network, and the long-term dependence of time sequence relationship can be captured for feature learning and prediction. The principle of LSTM is shown in Fig. 4.

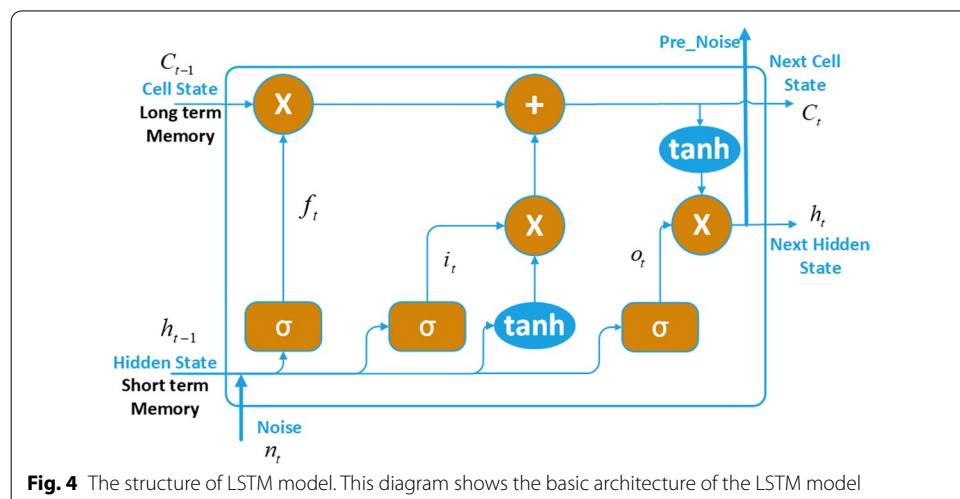


Fig. 4 The structure of LSTM model. This diagram shows the basic architecture of the LSTM model

When $f_t = 1$, it means that the short-term memory is completely retained. After the noise data are input, whether it can be stored in the cell depends on the input gate, and the output of the input gate is C_t as in the formula (3).

$$f_t = \sigma(W_f[h_{t-1}, n_t]) + b_f, f_t \in [0, 1] \quad (2)$$

$$C_t = f_t * C_{t-1} + \sigma(W_i[h_{t-1}, n_t] + b_i) * \tanh(W_C[h_{t-1}, n_t] + b_C) \quad (3)$$

n_t represents the input noise of the current layer. h_{t-1} is the output noise of the previous layer and the hidden state of the current layer. The above formula represents the state of the new cell after discarded useless information and retained some new information, where $i_t = \sigma(W_i[h_{t-1}, n_t] + b_i)$ and it represents the probability of new information being retained, and the prediction noise of output depends on the output gate:

$$o_t = \sigma(W_o[h_{t-1}, n_t] + b_o) \quad (4)$$

$$Y_t = h_t = o_t * \tanh(C_t) \quad (5)$$

o_t is the output probability. Multiplying o_t and hyperbolic tangent function $\tanh(C_t)$ can achieve the purpose of controlling the cell state filtering, and the output Y_t is the hidden state of the next layer. In the above expression, W_f, W_i, W_C, W_o are the function parameter weight vectors and b_f, b_i, b_C, b_o are the bias vectors.

The essence of realizing multivariate is to form a sample with multiple dimensions of multiple information and transform it into a supervised learning problem, so as to achieve the purpose of multiple inputs and single output. There are 32 neurons in the first hidden layer, 1 neuron in the output layer is used to predict noise, and the input variables are four-dimensional information including wind speed, noise of neighboring station based on wind direction, traffic flow information and noise of prediction station. The output is prediction noise of prediction station with 2 prediction steps and time interval of 10 min. The model was trained 100 times with a batch size of 128, tracking training and test losses during training by setting the validation_data parameter in the fit () function.

Multi-factor features were extracted based on LSTM model for noise prediction. The prediction error was large during the abrupt change period: In January, the noise plunged about 4.5 dB, and the high error of the prediction result was about 2 dB. Therefore, the Prophet model was introduced to fuse multi-station information to improve the prediction accuracy.

4.2 Prophet model based on spatial multi-station regression

The Prophet prediction model has great advantages in processing periodic data with abnormal values and trend changes, and the noise of chemical parks has strong micro-abruptness and macro-regularity. Therefore, Prophet model is introduced for noise prediction in this paper. Prophet model decomposes the time series according to the following formula:

$$P(t) = g(t) + s(t) + h(t) + \varepsilon(t) \quad (6)$$

In formula (6), $g(t)$ represents the noise trend term, which is mainly used to fit aperiodic changes in the time series. We use a trend term model based on piecewise linear functions:

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (7)$$

In formula (7), m is the offset, k represents the growth rate, and δ represents the change in the growth rate. The indicator function is: $a(t) = (a_1(t), \dots, a_S(t))^T$. $a(t) \in \{0, 1\}^S$, $a_j(t) = \begin{cases} 1, & \text{if } t \geq S_j \\ 0, & \text{otherwise} \end{cases}$, $\gamma = (\gamma_1, \dots, \gamma_S)^T$, $\gamma_j = -S_j \delta_j$, where S represents the number of mutation points.

$s(t)$ is a periodic term modeled by Fourier series:

$$s(t) = \sum_{n=1}^N \left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] \quad (8)$$

In Formula (8), t represents a fixed period, $2n$ represents the number of periods expected to be used in the model, P represents the period of the time series, and $P=7$ represents a period of weeks.

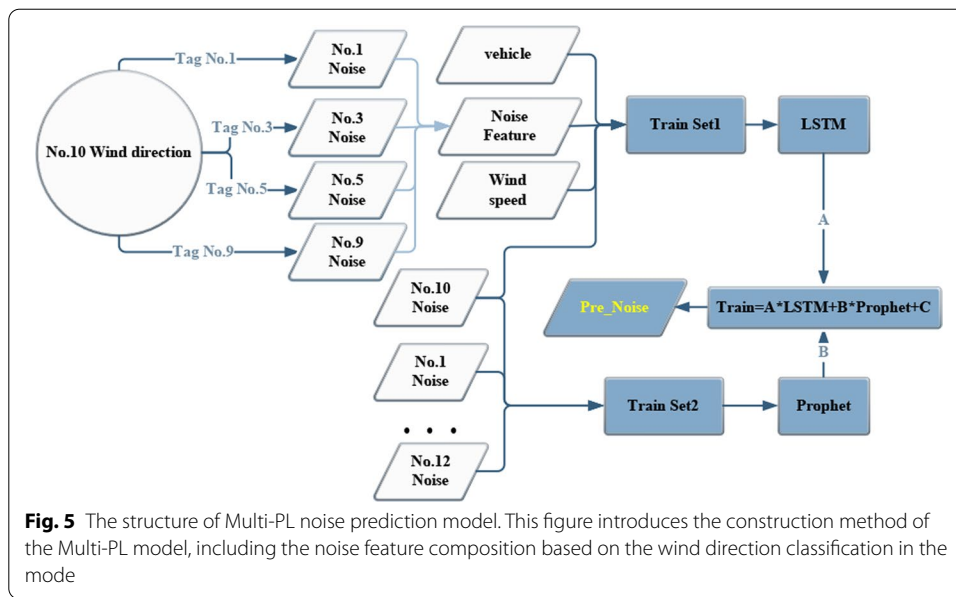
$h(t)$ is a holiday item that regards the influence of each holiday at different times as an independent model. $\varepsilon(t)$ represents the error term or interference term, which represents random and unpredictable fluctuations. Prophet algorithm can add up trend terms, season terms and so on to be the predicted value of time series.

In this paper, the method `add_regressor()` was used to add data from multiple stations as regression variables for fitting. First, the noise time series data of other sites were added to Prophet in turn for prediction. Then, the sites were sorted according to the RMSE size of the prediction results, and the ranking results were added to Prophet model in turn to improve the prediction accuracy. Although the Prophet model is flexible, it cannot consider the influence of the characteristics of multi-dimensional factors. Therefore, achieving accurate prediction requires a more complete prediction scheme.

4.3 Multi-PL model based on Prophet and LSTM combination

Based on the characteristics of the Prophet and LSTM models, we propose the Multi-PL model to make up for the limitations of a single model, and can effectively use the park information and the advantages of the two models to achieve higher-precision noise prediction.

Firstly, the noise feature sequence of adjacent stations based on wind direction was constructed, and the wind direction was classified as direction labels with time series features. Extract the noise value of the corresponding site during the time according to the tag, stitch the extracted noise value into a new time series feature, which is the noise feature in Fig. 5, and construct a multi-element LSTM model by combining the time series features of traffic flow and wind speed. The above work is based on the multivariate data set Train Set 1. The data of each site in the multi-station dataset Train Set 2 were, respectively, used in the Prophet model, sorted according to the size of RMSE of different sites, and added to the Prophet model in the order of RMSE from small to large.



Use the cftool (Curve Fitting Tool) curve fitting toolbox in MATLAB to fit the two model prediction results and the real noise value in the training set, and obtain the formula (9) between the actual noise value and the model prediction value:

$$\text{Train_true} = A * \text{LSTM} + B * \text{Prophet} + C \tag{9}$$

The method of obtaining the relationship between the actual value and the predicted value by fitting method is closer to the true value than the linear weighting method of the predicted value of the two models, and has the property of constant compensation, which prevents the training result of a certain model from being too high or too low leading to deviations in forecast results.

5 Experiment and result analysis

Firstly, the proportion and evaluation indexes used in the training set are described, and then, the 3σ criterion and multivariate multi-station prediction results analysis are introduced. Finally, the Multi-PL proposed in this paper is compared with Prophet + LSTM linear weighted combination model, LSTM, Prophet, BP neural network model, traditional Kalman filter prediction model and other prediction models, to verify that the proposed method has better accuracy and prediction ability.

5.1 Train set proportion and evaluation index

The proportion of training set and test set in multivariate data set and multi-station data set is determined by experimental comparison. The LSTM deep neural network is prone to overfitting, and the Prophet model has good stability. Taking single-site prediction as an example, the difference in RMSE between different data set ratio experiments does not exceed 0.5. Therefore, the LSTM model is used as the basis for data set division to ensure, however, the best is selected on the basis of fitting. According to Table 1, 72% of the training set is finally determined, and the rest is the test set.

Table 1 LSTM prediction results

| Train set weight (%) | RMSE\dB | MAE\dB |
|----------------------|--------------|------------------------|
| 50 | 2.742 | 2.588 |
| 60 | 1.300 | 0.976 |
| 70 | 1.262 | 0.986 |
| 72 | 1.162 | 0.935 |
| 80 | 1.280 | 1.076 |
| 90 | 1.287 | 0.816 (Overfit) |

In order to verify the validity of Multi-PL prediction model, this paper uses three evaluation indexes: root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2). The calculation formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (x - \tilde{x})^2} \quad (10)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |x - \tilde{x}| \quad (11)$$

$$R^2 = 1 - \frac{\sum_{t=1}^n (x - \tilde{x})^2}{\sum_{t=1}^n (x - \bar{x})^2} \quad (12)$$

\bar{x} is the mean value of the true value of noise, $x = (x_1, x_2, \dots, x_n)$, $x_i \in R^n$ is the true value of noise, $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$, $\tilde{x}_i \in R^n$ is the predicted value of noise in Eqs. (10) and (11), expressed as the fitted value of the predicted values of the two models in Eq. (12), and n is the number of time series values. The smaller the number of values, RMSE and MAE, the better the predictive ability of the model. The closer R^2 is to 1, the better the predictive effect of the fitted model.

5.2 Analysis of forecast results

The 3σ criterion assumes that a set of data contain only random errors, and the noise value noise $\in (u - 3\sigma, u + 3\sigma)$ interval accounts for about 99.74%. It is believed that any error exceeding this interval is not a random error but a gross error. The data containing this error should be removed or replaced, u represents the mean value of noise, σ is the noise standard deviation, and *noise* is the noise value.

This article replaces the noise range at $0 \leq \text{noise} < u - 3\sigma$ (dB) with the mean value. Take the noise data of Station 10 in Fig. 6 as an example, part A is the original noise value containing the zero value of the sparse mutation point, and the unbiased standard deviation of the sample is 4.45. Part B represents the noise value after the above 3σ treatment, and the unbiased standard deviation of the sample is 4.28.

According to Table 2, RMSE decreases by at least 0.1 dB and MAE also decreases for both single-station and multi-station predictions using the 3σ criterion; compared

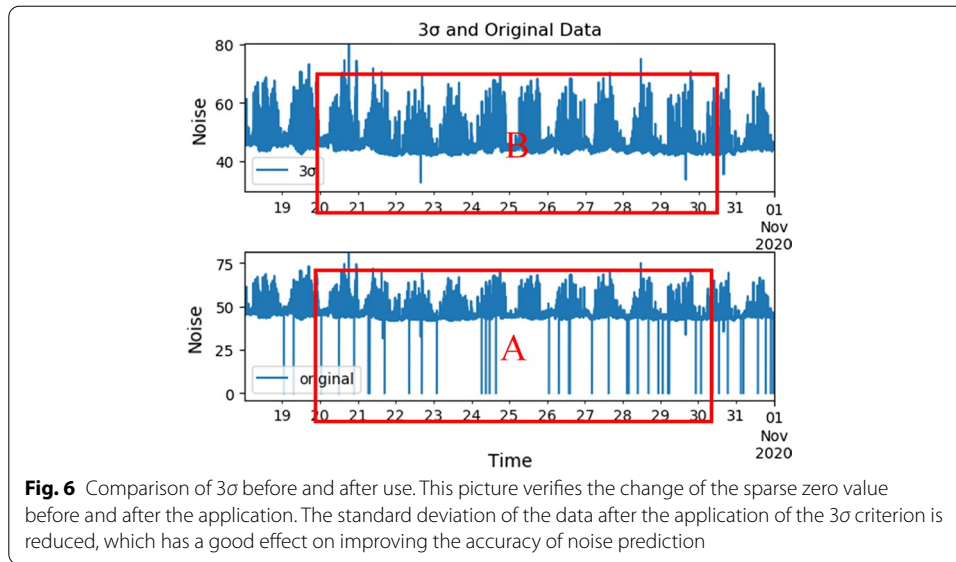


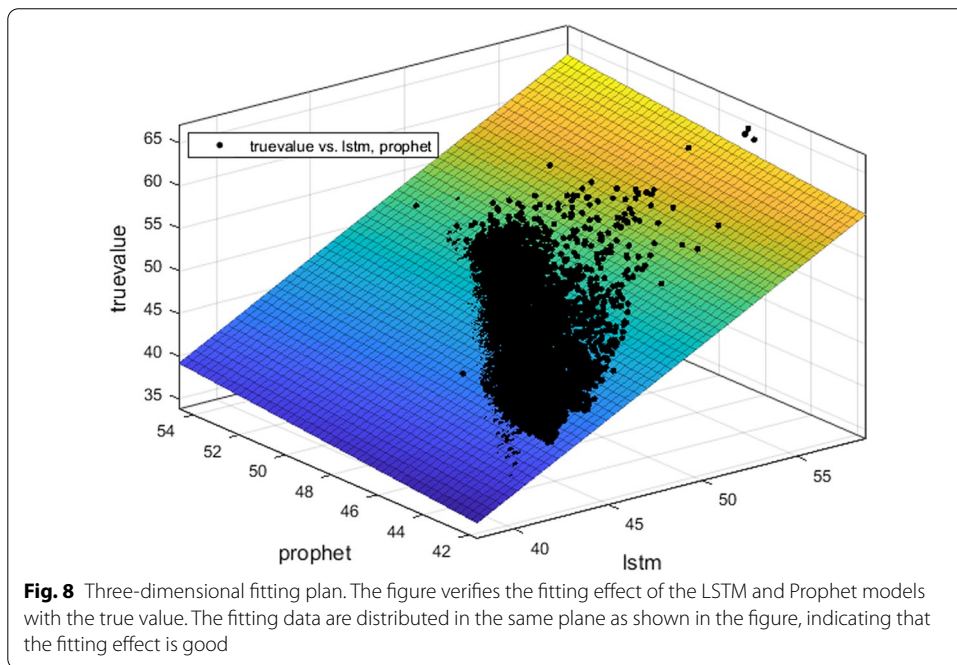
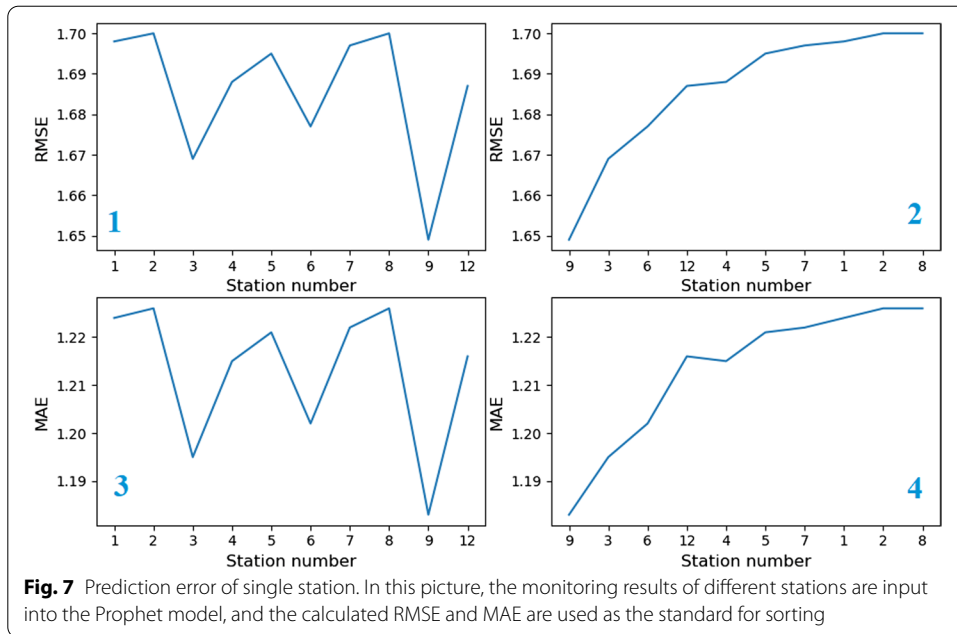
Fig. 6 Comparison of 3σ before and after use. This picture verifies the change of the sparse zero value before and after the application. The standard deviation of the data after the application of the 3σ criterion is reduced, which has a good effect on improving the accuracy of noise prediction

Table 2 Comparative experimental results

| Deal with zero | Model | Data set | RMSE | MAE |
|----------------------------|---------|-----------------------|-------|------|
| Use 3σ criterion | Prophet | Single station | 1.70 | 1.23 |
| | | Multi-stop (Disorder) | 1.61 | 1.14 |
| | | Multi-stop (Orderly) | 1.186 | 0.92 |
| | LSTM | Single station | 1.29 | 1.13 |
| | | Multi-information | 1.17 | 0.95 |
| Unused 3σ criterion | Prophet | Single station | 1.814 | 1.28 |
| | | Multi-stop | 1.75 | 1.20 |
| | LSTM | Single station | 1.46 | 1.27 |
| | | Multi-information | 1.27 | 0.92 |

with single-station data, the predicted RMSE and MAE of multi-station data set used in Prophet model are reduced by 5.3% and 7.3%, respectively. Each station is used for noise prediction of station 10. The RMSE and MAE of each station are shown in sub-pictures 1 and 3 in Fig. 7. After the stations are sorted according to RMSE, they are shown in sub-pictures 2 and 4. According to the order, the multi-station data are added to the Prophet model as regression variables, and the RMSE and MAE of the disorderly prediction are reduced by 26.3% and 22.8%, respectively. In the multi-site ordered Prophet method, the RMSE and MAE of the data processed by the 3σ principle are reduced by 32.2% and 23.3%; compared with single-site data, the RMSE and MAE predicted by using the multivariate data set in the LSTM model are reduced by 9.3% and 15.9% dB, respectively.

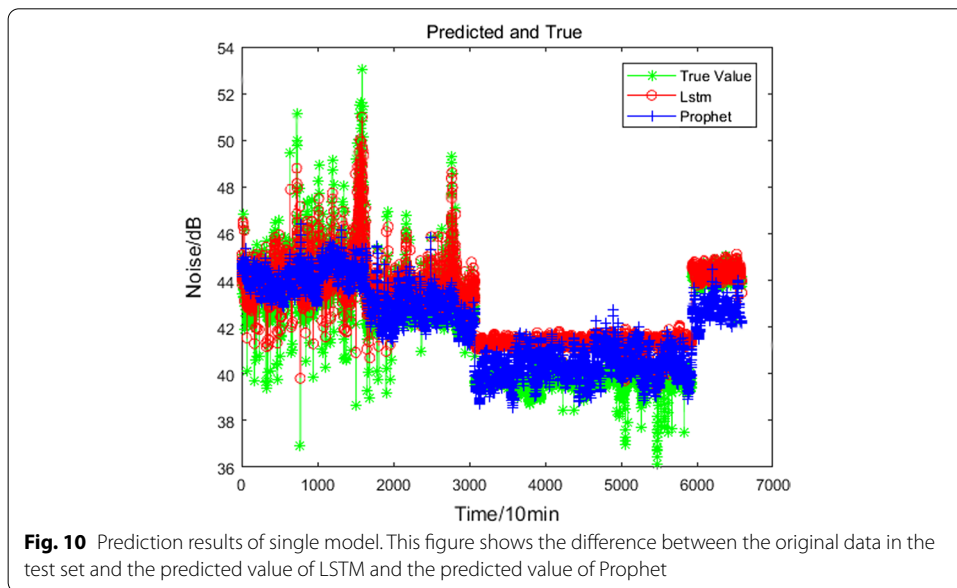
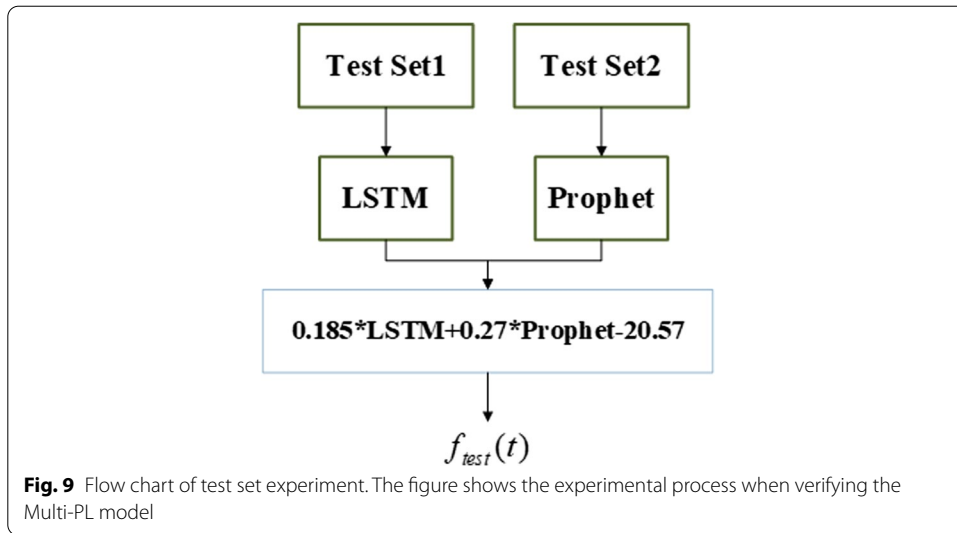
It can be seen from Table 2 that the prediction results with the application of 3σ criterion have higher accuracy. The Prophet model uses multi-station ordered data with the highest accuracy, and the LSTM model uses multivariate data sets with higher accuracy than the original station. On this basis, the data predicted by LSTM and Prophet training set were fitted, and the relationship between the real noise value of the training set and the predicted value of LSTM and Prophet training set was obtained as shown in



Eq. (13), where $L(t), P(t)$ are the predicted results of LSTM and Prophet training set, respectively. $f(t)$ is the fitting predicted value.

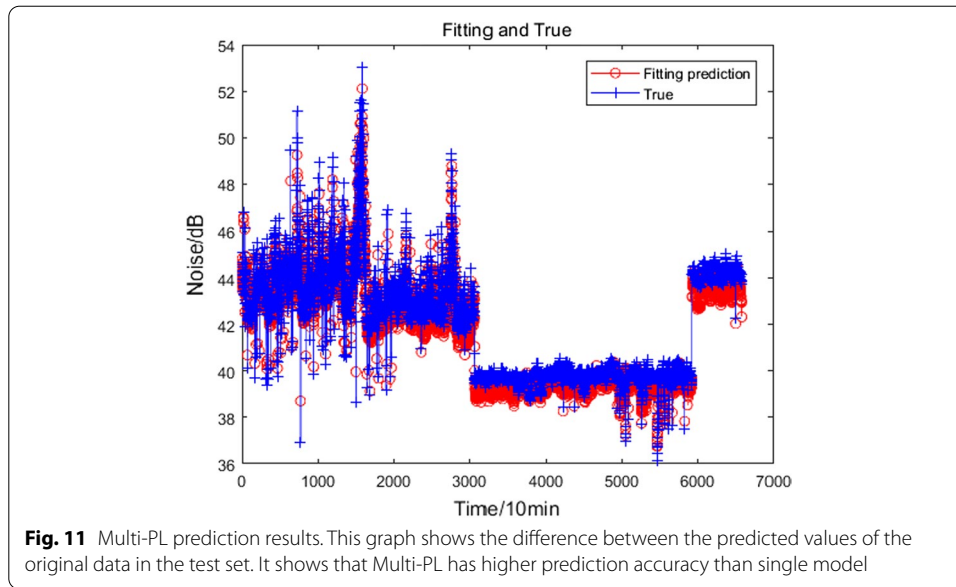
$$f(t) = -20.57 + 1.185 * L(t) + 0.27P(t) \tag{13}$$

The RMSE of $f(t)$ obtained by fitting and the true value is 0.54. The data points in Fig. 8 are basically fitted to the same plane and the coefficient of determination $R^2 = 0.962$. The fitting effect is good. After the test set was fed into LSTM and



Prophet, the predicted value was put into the verification Eq. (13), and the prediction result $f_{test}(t)$ of the Multi-PL model was obtained as shown in Fig. 9.

Among them, Test Set 1 is from multivariate data set and Test Set 2 is from multi-station data set. Figure 10 shows the true value, LSTM and Prophet noise predicted value. Compared with the predicted value fitted by the Multi-PL model in Fig. 11, Multi-PL makes up for the prediction deviation of the two models and improves the prediction accuracy of outliers contained in the noise. The RMSE and MAE of $f_{test}(t)$ and the true value were 0.53 and 0.46 dB, respectively. The prediction result of Multi-PL model is obviously better than that of single LSTM and Prophet model.

**Table 3** Analysis of results of different prediction models

| Model | | | RMSE\dB | MAE\dB |
|----------------|-----------------|--------------------|--------------|--------------|
| Kalman | | | 1.91 | 1.16 |
| BP | | | 1.34 | 1.25 |
| Prophet | | | 1.18 | 0.92 |
| LSTM | | | 1.16 | 0.95 |
| Prophet + LSTM | ω_{LSTM} | $\omega_{Prophet}$ | RMSE\dB | MAE\dB |
| | 0.0 | 1.0 | 1.186 | 0.919 |
| | 0.1 | 0.9 | 1.12 | 0.875 |
| | 0.2 | 0.8 | 1.065 | 0.843 |
| | 0.3 | 0.7 | 1.024 | 0.826 |
| | 0.4 | 0.6 | 0.996 | 0.82 |
| | 0.5 | 0.5 | 0.985 | 0.824 |
| | 0.6 | 0.4 | 0.989 | 0.832 |
| | 0.7 | 0.3 | 1.01 | 0.846 |
| | 0.8 | 0.2 | 1.046 | 0.86 |
| | 0.9 | 0.1 | 1.095 | 0.899 |
| | 1.0 | 0.0 | 1.157 | 0.954 |
| Multi-PL | | | 0.53 | 0.46 |

5.3 Comparison results of different prediction models

In order to verify the prediction performance of Multi-PL model, two evaluation indexes, RMSE and MAE in Sect. 4.1, are used to evaluate Kalman filter prediction, BP neural network, LSTM, Prophet, and Prophet + LSTM linear weighted model (optimal weight: $\omega_{LSTM}=0.5$, $\omega_{Prophet}=0.5$). According to Table 3, the prediction results of Multi-PL are better than other prediction methods, and the accuracy of RMSE and MAE is improved by 45.9% and 25.9%, respectively, compared with the linear weighted model. Multi-PL model can be used as an effective prediction model for chemical industry parks.

6 Conclusions

It is very important to analyze the noise law and influencing factors in chemical industry park and improve the prediction accuracy of noise, which is of great significance to guide the working time planning and workers' hearing health protection. Based on the appearance law of time series data such as noise, traffic flow, wind direction and wind speed in a chemical park, this paper uses the 3σ criterion to replace the zero value of noise, and proposes a Multi-PL model based on multivariate information and multi-station information. Design and implement the comparative experiment with Prophet + LSTM weighted model, single model, Kalman filter prediction model and traditional BP neural network model under each weight coefficient. The results show that the time series data of park noise processed by the 3σ criterion have better performance in the prediction model, and the prediction error of multi-station Prophet and multivariate LSTM neural network model is lower than the traditional Kalman filter prediction model and BP neural network model. Moreover, Prophet + LSTM linear weighted combination model has a slightly higher prediction accuracy than the above models, and Multi-PL model which can effectively use park data and has constant compensation property has the best effect. Compared with linear weighted combination model, RMSE and MAE errors are reduced by 0.45 dB and 0.36 dB, respectively. Multi-PL can be used as an effective noise prediction model in chemical industry park. On the basis of the wide application of intelligent parks, this study can provide a new idea for noise prediction in parks.

This paper only constructs the prediction model fitted by two multi-factor models. In the future, the traditional prediction model based on statistical method can be introduced to make up the disadvantage of neural network and get more accurate noise prediction results. In addition, transfer learning or reinforcement learning can be used to predict the overall noise level of the park.

Abbreviations

3σ criterion: A method of outlier discrimination based on normal distribution; RNN: Recurrent neural network; LSTM: Long short-term memory; Prophet: An open source time series prediction algorithm provided by Facebook; Multi-PL: Noise prediction model based on LSTM and Prophet; Kalman: An algorithm for optimal estimation of system state by using linear system state equation; BP: Back propagation; RMSE: Root mean-squared error; MAE: Mean absolute error; R^2 : R squared.

Acknowledgements

Not applicable.

Authors' contributions

QTZ, YL, GC and CGL conceived and designed the experiments. QTZ and YL performed the experiments. GC, HD analyzed the data. YL, GC and QTZ wrote the paper. All authors have contributed to this research work and read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China under Grant No. 61701284, U1931207 and 61702306, the Innovative Research Foundation of Qingdao under Grant No.19-6-2-1-cg, the Application Research Project for Postdoctoral Researchers of Qingdao, the Sci. & Tech. Development Fund of Shandong Province of China under Grant No. ZR202102230289, ZR202102250695 and ZR2019LZH001, the Humanities and Social Science Research Project of the Ministry of Education under Grant No.18YJAZH017, the Taishan Scholar Program of Shandong Province, the Shandong Chongqing Science and technology cooperation project under Grant No. cstc2020jscx-lyjsAX0008, the Sci. & Tech. Development Fund of Qingdao under Grant No. 21-1-5-zlyj-1-zc, SDUST Research Fund under Grant No. 2015TDJH102, and the Science and Technology Support Plan of Youth Innovation Team of Shandong higher School under Grant No. 2019KJN024.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China. ²College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao 266590, China. ³School of Information Science and Engineering, Southeast University, Nanjing 210096, China.

Received: 2 September 2021 Accepted: 20 October 2021

Published online: 29 October 2021

References

1. X. Liu, X. Zhang, Rate and energy efficiency improvements for 5G-based IoT with simultaneous transfer. *IEEE Internet Things J.* **6**(4), 5971–5980 (2019)
2. J. Wu, H. Miao, J. Liu, Hearing status of workers in automobile industry and the correlation of influencing factors of noise deafness. *Clin. Res. China* **30**(5), 713–716 (2017)
3. Z.U.R. Farooqi, Assessment of noise pollution and its effects on human health in industrial hub of Pakistan. *Environ. Sci. Pollut. Res. Int.* **27**(3), 2819–2828 (2020)
4. G.R. Taffere, M. Bonsa, M. Assefa, Magnitude of occupational exposure to noise, heat and associated factors among sugarcane factory workers in Ethiopia, 2017. *J. Public Health (Berl.)* **28**, 517–523 (2020)
5. I.P. Nyarubeli, A.M. Tungu, Variability and determinants of occupational noise exposure among iron and steel factory workers in Tanzania. *Ann. Work Expos. Health* **62**(9), 1109–1122 (2018)
6. K. Ayda, G. Perihan, "Noise factory": a qualitative study exploring healthcare providers' perceptions of noise in the intensive care unit. *Intensive Crit. Care Nurs* **63**, 102975 (2020)
7. M. Susan, Life: Industrial noise disrupts plants: effects on animals alter dispersal of seeds, pollen. *Sci. News* **181**(8), 15 (2012)
8. X. Liu, X.B. Zhai, W. Lu, C. Wu, QoS-guarantee resource allocation for multibeam satellite industrial internet of things with NOMA. *IEEE Trans. Ind. Inform.* **17**(3), 2052–2061 (2021)
9. H. Zhang, Intelligent 5G: when cellular networks meet artificial intelligence. *IEEE Wirel. Commun.* **24**(5), 175–183 (2017)
10. X. Liu, X. Zhang, NOMA-based resource allocation for cluster-based cognitive industrial internet of things. *IEEE Trans. Ind. Inf.* **16**(8), 5379–5388 (2020)
11. W. Po-Jiun, H. Chihpin, Noise prediction using machine learning with measurements analysis. *Appl. Sci.* **10**(18), 6619 (2020)
12. X. Zhang, M. Zhao, R. Dong, Time-series prediction of environmental noise for urban IoT based on long short-term memory recurrent neural network. *Appl. Sci.* **10**(3), 1144 (2020)
13. S. Jaffry, Cellular traffic prediction with recurrent neural network//2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT). (IEEE, 2020)
14. A.E. Vela, Y. Oleyaei-Motlagh, Ground level aviation noise prediction: a sequence to sequence modeling approach using LSTM recurrent neural networks//2020 IEEE/AIAA 39th Digital Avionics Systems Conference (DASC). (IEEE, 2020)
15. X. Tao, S. Han, Y. Guoqing, Integrated model of airport noise prediction based on spatial fitting and neural network. *China Environ. Sci.* **36**(04), 1250–1257 (2016)
16. C. Liu, *Research on Interactive Prediction of Airport Noise Monitoring Points Based on Time Series Similarity Measure* (Nanjing University of Aeronautics and Astronautics, Nanjing, 2018)
17. B. Sun, L. Chen, Univariate traffic noise prediction considering traffic flow state of road network. *Noise Vib. Control* **41**(02), 190–195 (2021)
18. S. Sameer, K. Satish, Assessment and prediction of environmental noise generated by road traffic in Nagpur City, India. *Environ. Pollut.* **77**, 167–180 (2018)
19. K. Mashael, L. Kaouther, A. Nada, Time series Facebook Prophet model and python for COVID-19 outbreak prediction. *Comput. Mater. Continua* **67**(3), 3781–3793 (2021)
20. J. Song, H. Xie, B. Gao, Maximum likelihood-based extended Kalman filter for COVID-19 Prediction. *Chaos Solitons Fractals* **146**, 110922 (2021)
21. F. Li, K. Lam, X. Liu, J. Wang, K. Zhao, L. Wang, Joint pricing and power allocation for multibeam satellite systems with dynamic game model. *IEEE Trans. Veh. Technol.* **67**(3), 2398–2408 (2018)
22. Z. Liu, Z. Zhu, J. Gao, C. Xu, Forecast methods for time series data: a survey. *IEEE Access* **9**, 91896–91912 (2021)
23. Y. Wang, D. Wang, Y. Tang, Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods. *IEEE Access* **8**, 17071–17079 (2020)
24. M. Wu, Y. Ye, N. Hu, Q. Wang, 'EMD-GM-ARMA model for mining safety production situation prediction.' *Complexity* **2020**, 1–14 (2020)

25. P. Unnikrishnan, V. Jothiprakash, Hybrid SSA-ARIMA-ANN model for forecasting daily rainfall. *Water Resour. Manag.* **34**(11), 3609–3623 (2020)
26. C.G. Ozoegwu, Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number. *J. Cleaner Prod.* **216**, 1–13 (2019)
27. S. Wu, F. Shao, R. Sun, "Corn futures price forecast based on ARIMA time series and support vector machine. in Proceedings of 4th International Conference on System Computing and Big Data, vol. 5, pp. 41–49 (2019)
28. S. Chan, I. Oktavianti, V. Puspita, A deep learning CNN and AI-tuned SVM for electricity consumption forecasting: Multivariate time series data, IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 488–494 (2019)
29. I.E. Livieris, E. Pintelas, P. Pintelas, A CNN-LSTM model for gold price time-series forecasting. *Neural Comput. Appl.* **32**, 17351–17360 (2020)
30. B.B. Sahoo, R. Jha, A. Singh et al., Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys.* **67**, 1471–1481 (2019)
31. S. Hochreiter, J. Schmidhuber, Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Qingtian Zeng received the B.S. degree and the M.S. degree in computer science from Shandong University of Science and Technology, Taian, China, in 1998 and 2001, respectively, and the Ph.D. degree in computer software and theory from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. He is currently a Professor with Shandong University of Science and Technology, Qingdao, China. His research interests are in the areas of Petri nets, process mining, and knowledge management.

Yu Liang received the B.S. degree in Communication Engineering from Shandong University of Science and Technology, Qingdao, China, in 2019. She is currently pursuing the master degree in the College of Electronic and Information Engineering of Shandong University of Science and Technology. Her current research interests include the Internet of Things and noise prediction in chemical parks.

Geng Chen received the B.S. degree in electronic information engineering and the M.S. degree in communication and information system from Shandong University of Science and Technology, Qingdao, China, in 2007 and 2010, respectively, and the Ph.D. degree in information and communications engineering from Southeast University, Nanjing, China, in 2015. He is currently an Associate Professor with the College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, China. His current research interests are in the areas of heterogeneous networks, ubiquitous networks and software defined mobile networks, with emphasis on wireless resource management and optimization algorithms, and precoding algorithms in large scale MIMO.

Hua Duan received the B.S. and M.S. degrees in applied mathematics from the Shandong University of Science and Technology, Tai'an, China, in 1999 and 2002, and the Ph.D. degree in applied mathematics from Shanghai Jiao Tong University, in 2008. She is currently a Professor with the Shandong University of Science and Technology. Her research interests include process mining and machine learning.

Chunguo Li received the bachelor's degree in wireless communications from Shandong University in 2005, and the Ph.D. degree in wireless communications from Southeast University in 2010. In July 2010, he joined the Faculty of Southeast University, Nanjing, where he became Associate Professor in 2012 and Full Professor in 2017. He also has been the Supervisor of Ph.D. candidate since 2016. From June 2012 to June 2013, he was the Post-Doctoral researcher with Concordia University, Montreal, Canada. From July 2013 to August 2014, he was with the DSL laboratory as the visiting associate professor, supervised by Prof. J. M. Cioffi. His research interests are in the 5G cellular transmission, underwater communications, machine learning for video signal processing, and next generation of WiFi. He is currently the Associate Editor for IET Communications, Area Editor of the Elsevier AEU-International Journal of Electronics and Communication, Editor of Telecommunications Systems, Associate Editor of Circuits, Systems and Signal Processing, Editor of the KSII Transactions on Internet and Information Systems.