PREDICTION OF DUST CONCENTRATION BASED ON THE BOLTZMANN AND MULTIVARIATE LINEAR MODEL

by

Ying ZHANG**, *Luqi JIN, and Qiangqiang ZHAO

School of Energy and Environment, Zhong Yuan University of Technology, Zhengzhou, China

Original scientific paper
https://doi.org/10.2298/TSCI2503951Z

In order to reduce the dust concentration in the port and ensure that the dust concentration meets the requirements of environmental protection, a combination of Boltzmann model and multiple linear regression model is proposed to predict the continuous changes in dust concentration during the loading and unloading process of imported and exported ores in the port. The quantitative relationship between port dust concentration and falling mass of mineral powder, wind speed, and moisture content of mineral powder is simulated using the Boltzmann model and the multivariate linear regression model. This quantitative relationship is an effective means of describing the change in dust concentration and reducing the port dust concentration. Moreover, the findings have the potential to enhance the quality of life for those residing in the vicinity of the port.

Key words: *multivariate linear regression model, Boltzmann coefficient, port, dust concentration*

Introduction

Since the reform and opening up, China's GDP has grown rapidly, and the steel market has developed synchronously and rapidly. However, China's mineral resources are limited, and the domestic supply of mineral resources is insufficient. This has led to an increase in the total amount of imported ore in China. The continuous improvement of port ore throughput has also led to an increase in environmental problems. A large amount of dust is generated during the ore loading and unloading process, causing significant damage to the environment near the port and affecting the health of residents in the area. Given the importance of environmental protection and public health, China has continued to expand research on dust pollution control methods in recent years. Accurately predicting the continuous changes in dust concentration in ports is of great significance for solving the problem of mining dust pollution in ports.

The primary application of machine learning in research is the prediction of dust concentration. This is achieved through the use of various techniques, including decision trees [1, 2], support vector machines [3], linear regression [4-6], neural networks [7, 8], *etc.* Xu [9] developed a non-linear autoregressive model by integrating time series and neural networks to assess the efficacy of the model in forecasting mine dust concentration. Although the aforementioned models predict changes in dust concentration, the application scope of those mod-

* Corresponding author, e-mail: zhangying@zut.edu.cn

els is limited as the effect of time on dust concentration is considered. Therefore, the predicted results are too simplistic and lack representativeness.

In order to consider the effects of factors such as mineral powder quality, wind speed, and ore moisture content on port dust concentration, a combination of linear and non-linear prediction models is proposed. This combination employs Boltzmann and multiple linear regression to flexibly predict the continuous change of port dust concentration in ports.

Theoretical basis of the model

Theoretical basis of Boltzmann model

The Boltzmann model is a fundamental model of collision dynamics theory [10]. It is employed to describe the motion of sparse gases and to elucidate the temporal evolution of non-equilibrium states in complex systems comprising a multitude of particles [11]. The model's curve evolution aligns with the temporal trend of port dust concentration, and it is capable of accurately capturing the continuous change in port dust concentration over time. This enables the precise determination of the four coefficients A_1, A_2, x_0, and dx of the Boltzmann equation.

The Boltzmann equation has the form:

$$y = \frac{A_1 - A_2}{1 + \mathrm{e}^{(x - x_0)/\mathrm{d}x}} + A_2 \tag{1}$$

where y is the dependent variable (predictor), x_0 and dx – the constant coefficients, A_1 and A_2 – the constant variables, and x – the independent variable.

Theoretical basis of multiple linear regression model

The multiple linear regression model considers the relationship between multiple independent variables and dependent variables. The main steps are establishing a multiple linear regression equation, predicting the target value, and solving the coefficients of the multiple linear regression model. Duan *et al.* [12] employed a multiple linear regression equation to predict the fluctuation of PM2.5 concentration. The multiple linear regression model was used to simulate the linear relationship between the four coefficients of the Boltzmann equation and the quality of mineral powder, wind speed, and water content. Furthermore, the quantitative relationship between dust concentration and mineral powder quality, wind speed, and water content was also obtained.

The general form of multiple linear regression analysis is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon \tag{2}$$

where Y is the dependent variable (predictor), β_0 – the constant coefficient, β_1, β_2, …, β_p – the regression coefficients of the independent variables X_1, X_2, …, X_p, and ε – the residual error (the difference between the observed and predicted values).

Prediction of dust concentration based on Boltzmann model

Acquisition of original data

The dataset comprises 12 experimental conditions for bauxite, iron ore powder, and Australian iron ore powder. The experiments generated concentration data for total suspended particulate matter (TSP), particulate matter with a diameter of 10 microns (PM10), and particulate matter with a diameter of 2.5 microns (PM2.5) at specific measurement points under different conditions, which include weight, ore moisture content, wind speed and distance. The experimental conditions are shown in tab. 1.

Table 1. Experimental conditions

Experimental condition No.	Grab material mass [kg]	Experimental wind speed [ms^{-1}]	Water content [%]
1	25	1.5	3%
2	25	3	7%
3	25	4.5	5%
4	20	1.5	7%
5	20	3	3%
6	20	6	5%
7	15	1.5	5%
8	15	4.5	3%
9	15	6	7%
10	10	3	5%
11	10	4.5	7%
12	10	6	3%

The experiment employs a small grab device to take a quantity of materials. The ores of 10 kg, 15 kg, 20 kg, and 25 kg are weighed by the mass-weighing device and placed in the small grab device. The grab is then moved to the top of the material recovery device, and a dust tester is placed at a release height of 0.5 m to record the amount of dust emitted from the source of the dust. According to the specified test conditions, the feeding rate of the ore material is evenly controlled, and the falling time is recorded. The dust concentrations are monitored at points 1.5 m, 3.0 m, 4.5 m, and 6.0 m away from the dust source, which are designated as A, B, C and D.

Fifteen data points representing significant fluctuations in TSP concentration at the dust source were selected from the ore material to align with the observed changes in dust concentration. The background data were excluded from further analysis. The dust concentration control model was obtained at measuring points A, B, C, and D in each experimental condition. These conditions were selected for analysis based on their relevance to bauxite experimental condition 1-11. The TSP concentration at the dust source under different experimental conditions of bauxite was employed as the original data source. The TSP concentration changes of measuring points A, B, C, and D (dust source) under experimental condition 1 for bauxite are shown.

As the ore is unloaded at the unloading point, the dust concentration at the dust source begins to change. From the trend of TSP dust concentration changes at measuring points A, B, C, and D illustrated in figs. 1-4, it can be seen that the TSP dust concentration changes over time with a certain development trend, and the relationship between its concentration change trend and time variable is not a simple linear relationship. Using the Boltzmann model to fit the change trend of dust concentration, a corresponding mathematical correlation can be obtained to represent the change in dust concentration at the dust source.

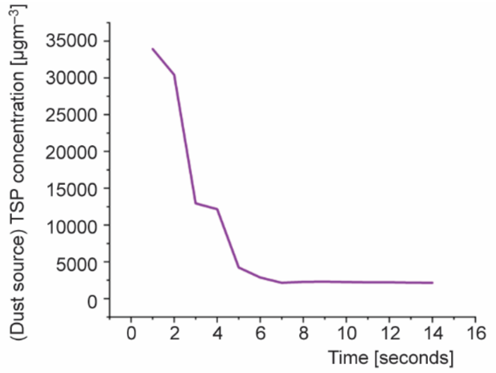

Figure 1. The TSP concentration at Point A

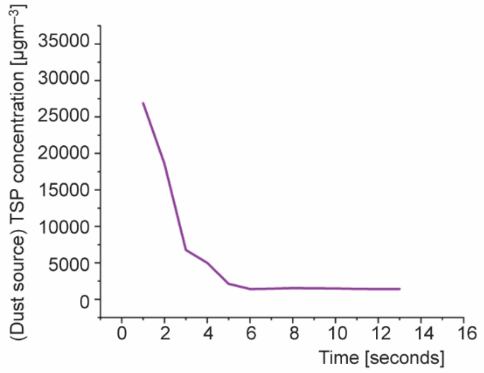

Figure 2. The TSP concentration at Point B

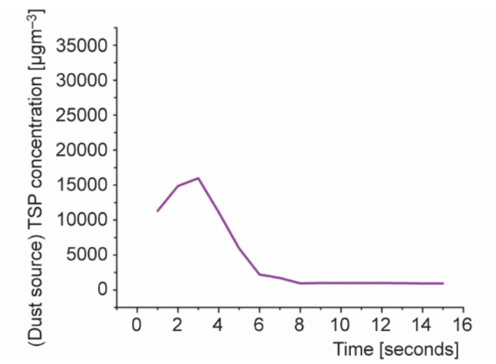

Figure 3. The TSP concentration at Point C

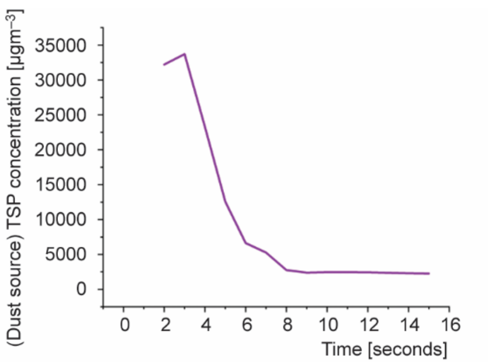

Figure 4. The TSP concentration at Point D

Establishment of Boltzmann model

The Boltzmann model was employed to assess the change in TSP concentration at the dust source, resulting in eqs. (3)-(6) and fitting figs. 5- 8. According to the Boltzmann eq. (1), the Boltzmann equation for the TSP concentration at measuring points A, B, C, and D under Condition 1 for bauxite can be obtained. Furthermore, the Boltzmann equation for measuring points A, B, C, and D under other conditions of bauxite can be derived by analogy.

The Boltzmann equation of TSP concentration at dust source at measuring point A under Condition 1 for bauxite is:

$$y = \frac{37274}{1+\mathrm{e}^{(x-0.57)/1.58}} + 2239 \tag{3}$$

The Boltzmann equation of TSP concentration at dust source at measuring point B under Condition 1 for bauxite is:

$$y = \frac{51328}{1+\mathrm{e}^{(x-0.49)/1.69}} + 1419 \tag{4}$$

The Boltzmann equation of TSP concentration at dust source at measuring point C under Condition 1 for bauxite is:

$$y = \frac{12924}{1 + \mathrm{e}^{(x-0.56)/0.58}} + 1133 \tag{5}$$

The Boltzmann equation of TSP concentration at dust source at measuring point D under Condition 1 for bauxite is:

$$y = \frac{33834}{1 + \mathrm{e}^{(x-0.47)/1.01}} + 2756 \tag{6}$$

The data presented in figs. 5-8 of the TSP concentration at measuring points A, B, C, and D under experimental Condition 1 for bauxite were derived from data fitting out by Boltzmann and Boltzmann eqs. (3)-(6).

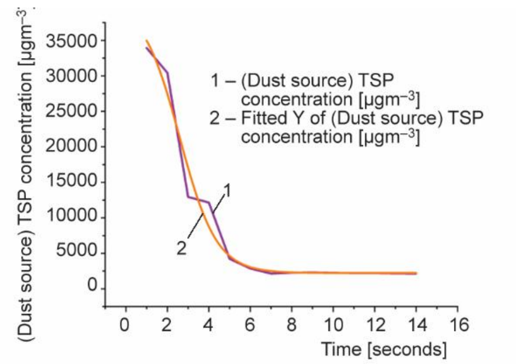

Figure 5. Fitted TSP concentration at Point A

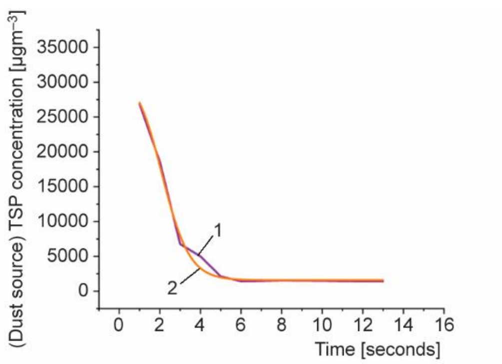

Figure 6. Fitted TSP concentration at Point B

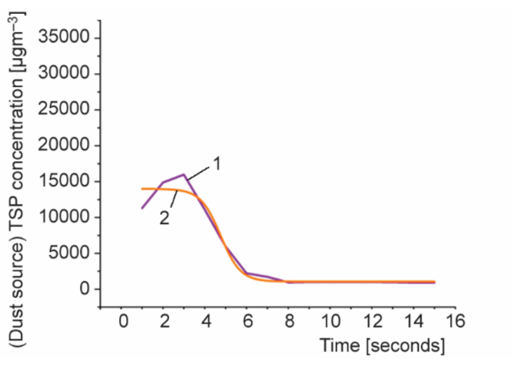

Figure 7. Fitted TSP concentration at Point C

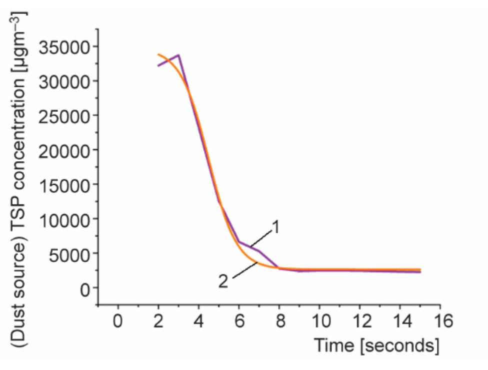

Figure 8. Fitted TSP concentration at Point D

The data presented in figs. 5-8 indicates that the Boltzmann model can effectively fit the TSP concentration at measuring point A, B, C, and D under experimental Condition 1 for bauxite.

Model test

The $X = \{x_i, I = 1, 2, \ldots, n\}$ represents the historical time-by-time sequence of a factor, and $Y = \{y_i, i = 1, 2, \ldots, n\}$ represents the historical time-by-time sequence of the other factor. Let $X_1 = \{x_i, i = 1, 2, \ldots, n\text{-}t\}$, $Y_1 = \{x_i, i = 1, 2, \ldots, n\text{-}t\}$, where $0 \le t \le n-1$. Then the correlation coefficient of R_{X1Y1} and X_1, Y_1 can be obtained:

$$R_{X_1Y_1} = \frac{Cov(X_1, Y_1)}{\delta_{X_1}\delta_{Y_1}} \tag{7}$$

The R_{X1Y1} represents the correlation coefficient between time factor X and current time factor Y before time $t(0 \le \mathrm{t} \le n-1)$. Factors X and Y can be the same, which represents an autocorrelation analysis between the same factor at the current time and the previous time t.

The feasibility of Boltzmann equation of TSP concentration at measuring point A, B, C and D under experimental Condition 1 for bauxite was tested by correlation coefficient. Among them, the correlation coefficients of Boltzmann equation of TSP concentration are 0.937, 0.999, 0.961, and 0.988. The feasibility of Boltzmann eqs. (3)-(6) can be verified.

Prediction of dust concentration based on multiple linear regression model

Original data processing

According to the Boltzmann model, the four coefficients A_1, A_2, x_0, and dx of the Boltzmann equation were obtained under different experimental conditions. Using data obtained from measurement point A under various conditions as an example, a total of 12 sets of data were obtained. A multiple linear regression model was used to study the relationship between each coefficient and the weight of mineral powder, wind speed, and the moisture content. Then, multiple linear regression equations were obtained for A_1, A_2, x_0, and dx with respect to the weight of mineral powder, x_1, wind speed, x_2, and moisture content, x_3.

In order to achieve better fitting, improve prediction accuracy, and eliminate the possible impact of different numerical magnitudes on each data point, the original data is normalized, which are shown in tab. 2. The following methods can be used to normalize data. The $\{m(q);\ q \in N\}$ represents the original input data of the neural network model and $\{n(q);\ q \in N\}$ indicates the normalized processing data of $\{m(q)\}$.

Table 2 The four coefficients, A_1, A_2, x_0, and dx under different experimental conditions

Condition No.	1	2	3	4	5	6	7	8	9	10	11	12
x_1	1.00	1.00	1.00	0.67	0.67	0.67	0.33	0.33	0.33	0.00	0.00	0.00
x_2	0.00	0.33	0.67	0.00	0.33	1.00	0.00	0.67	1.00	0.33	0.67	1.00
x_3	0.00	1.00	0.50	1.00	0.00	0.50	0.50	0.00	1.00	0.50	1.00	0.00
A_1	0.03	0.00	0.03	0.00	0.01	0.02	0.04	0.02	0.00	0.03	0.00	0.00
A_2	0.44	0.01	0.35	0.22	0.12	0.22	0.57	0.93	0.00	0.91	0.01	0.18
X_0	0.71	0.38	0.33	0.58	0.08	0.00	1.00	1.00	0.00	0.67	0.25	0.79
dx	0.11	0.84	0.17	0.28	0.56	0.75	0.17	0.05	0.32	0.80	0.14	0.18

The data is normalized to the interval [0, 1]:

$$n(q)=\frac{m(q)-\min\{m(q)\}}{\max\{m(q)\}-\min\{m(q)\}} \tag{8}$$

The corresponding normalized reduction formula is:

$$m(q)=\left(\max\{m(q)\}-\min\{m(q)\}\right)\times n(q)+\min\{m(q)\} \tag{9}$$

Establishment of multiple linear regression equation

According to the data in original data, multiple linear regression models were used to establish multiple linear regression eqs. (10)-(13) for coefficients A_1, A_2, x_0, and dx and their corresponding mineral powder mass x_1, wind speed x_2 and moisture content x_3.

$$A_1=0.0285+0.0014x_1-0.0134x_2-0.0150x_3 \tag{10}$$

$$A_2=0.8145-0.2917x_1-0.3199x_2-0.3575x_3 \tag{11}$$

$$X_0=1.1227-0.4016x_1-0.5363x_2-0.3425x_3 \tag{12}$$

$$\mathrm{d}x=0.1847+0.1272x_1+0.0616x_2+0.0616x_3 \tag{13}$$

Model validation

The feasibility of the multivariate linear regression eqs. (10)-(13) is tested through the analysis of the residual plot. The residual plot is defined as a tool for estimating whether the observed or predicted error (residual) is consistent with the random error. The following figure presents a quantitative residual diagram of the coefficients A_1, A_2, x_0, and dx, the ore powder mass x_1, wind speed x_2, and water content x_3, respectively. The specific residuals are shown in figs. 9-12.

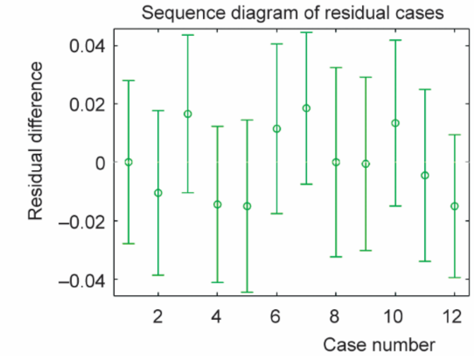

Figure 9. Resiudal of coefficient A_1

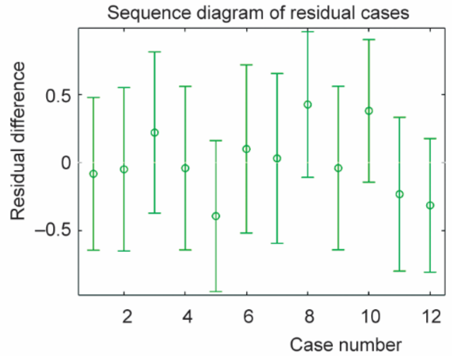

Figure 10. Resiudal of coefficient A_2

By analyzing the residual plot of the quantitative relationship between the coefficient A_1, A_2, x_0, and dx, and the mineral powder mass x_1, wind speed x_2 and moisture content x_3, it was found that the residual was consistent with the random error. The residuals are randomly distributed around 0 with no discernible trends or funnel-shaped patterns, indicating the validity of the multiple linear regression model. As demonstrated in figs. 9-12, the residual plots exhibit randomness, absence of systematic trends, and conformity to normality. The feasibility of multiple linear regression eqs. (10)-(13) can be verified.

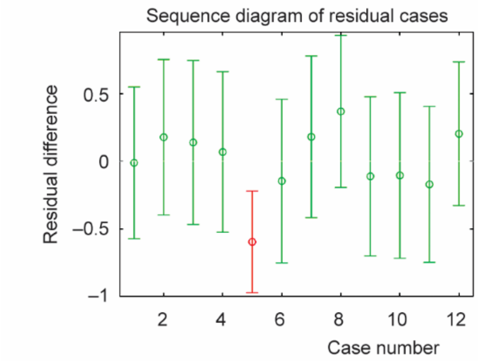

Figure 11. Resiudal of coefficient x_0

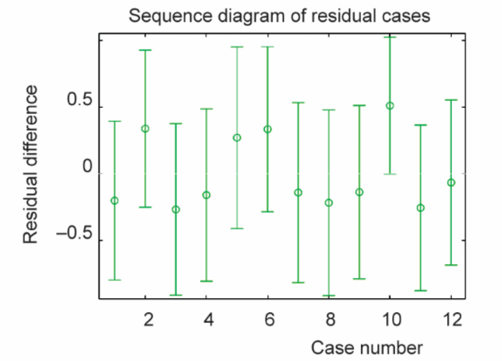

Figure 12. Resiudal of coefficient dx

Conclusions

- This study employed the Boltzmann equation to fit the change of dust concentration and time. On this basis, multiple linear regression models were used to establish multiple linear regression equations for coefficients A_1, A_2, x_0, and dx, and their corresponding mineral powder mass x_1, wind speed x_2 and moisture content x_3.
- This study employs the example of measuring point A under various experimental conditions to derive the corresponding multiple linear regression equation. The conclusions drawn from the data obtained at measuring point A under various experimental conditions can be extrapolated to other measuring points.
- The quantitative relationship between port dust concentration and falling mass of mineral powder, wind speed, and moisture content of mineral powder is an effective means of describing the change in dust concentration and reducing the port dust concentration.

References

[1] Lin, M. *et al.*, Nonlinear Relationship between Urban Form and Street-Level PM2.5 and CO Based on Mobile Measurements and Gradient Boosting Decision Tree Models, *Building and Environment, 205* (2021), 108265

[2] Yang, T., *et al.*, Anomaly Detection of Dust Removal System through Gradient Boosting Decision Tree Algorithm, *Proceedings*, International Conference on Communications, Information System and Computer Engineering (CISCE), Beijing, China, pp. 685-688, 2021

[3] Wang, W., *et al.*, Online Prediction Model Based on Support Vector Machine, *Neurocomputing, 71* (2008), 4-6, pp. 550-558

[4] Nickovic, S., *et al.*, A Model for Prediction of Desert Dust Cycle in the Atmosphere, *Journal of Geophysical Research: Atmospheres, 106* (2001), D16, pp. 18113-18129

[5] Wang, M., *et al.*, Prediction of Road Dust Concentration in Open-Pit Coal Mines Based on Multivariate Mixed Model, *Plos one, 18* (2023), 4, e0284815

[6] Park, S. Y., *et al.*, Proposal of a Methodology for Prediction of Indoor PM2. 5 Concentration Using Sensor-Based Residential Environments Monitoring Data and Time-Divided Multiple Linear Regression Model, *Toxics, 11* (2023), 6, 526

[7] Li, X., *et al.*, Long Short-Term Memory Neural Network for Air Pollutant Concentration Predictions: Method Development and Evaluation. *Environmental pollution, 231* (2017), Part 1, pp. 997-1004

[8] Cabaneros, S. M., *et al.*, A Review of Artificial Neural Network Models for Ambient Air Pollution Prediction, *Environmental Modelling & Software, 119* (2019), Sept., pp. 285-304

[9] Xu, Z., *et al.,* Mine Dust Concentration Prediction Based on Nonlinear Autoregressive Model (in Chinese), *Journal of Hebei United University (Natural Science Edition), 043* (2021), 004, pp. 127-133
[10] Goldstein, S., Lebowitz, J. L., On the (Boltzmann) Entropy of Non-Equilibrium Systems, *Physical D: Nonlinear Phenomena, 193* (2004), 1-4, pp. 53-66
[11] Dubois, F., Equivalent Partial Differential Equations of a Lattice Boltzmann Scheme, *Computers & Mathematics with Applications, 55* (2008), 7, pp. 1441-1449
[12] Duan, J., *Research on PM2.5 Concentration Prediction Based on Three Prediction Models* (in Chinese), Xi'an University of Electronic Science and Technology, Xi'an, China, 2023

Paper submitted: May 21, 2024
Paper revised: July 12, 2024
Paper accepted: July 12, 2024