Urban PM$_{2.5}$ Diffusion Analysis Based on the Improved Gaussian Smoke Plume Model and Support Vector Machine

Peng He[1], Bohong Zheng[1], Jian Zheng[2*]

[1] *School of Architecture and Art, Central South University, Changsha 410075, China*
[2] *School of Architecture, South China University of Technology, Guangzhou 510640, China*

ABSTRACT

With the acceleration of urbanization in China, haze has become a growing threat to human health. However, comprehensive research on the diffusion and evolution of PM$_{2.5}$ is still lacking. Therefore, this study proposed an improved Gaussian smoke plume model that considered the influence of multiple factors, such as rain wash, gravity sedimentation, and surface rebound, on PM$_{2.5}$. Additionally, the evolution of PM$_{2.5}$ was predicted by selecting 9 factors with a large influence. In the prediction, a support vector machine (SVM) and radial basis function kernel were adopted to construct classifiers and obtain the maximum distinction degree, respectively. Finally, the diffusion simulation and experimental evolution prediction were verified using data obtained from nine PM$_{2.5}$ monitoring stations in Wuhan. The experimental results showed that the algorithm could obtain considerably accurate simulation results of the PM$_{2.5}$ diffusion with low error for measured values. Therefore, this model may be useful in government plans for formulating strategies that control and reduce environmental pollution.

Keywords: PM$_{2.5}$; Diffusion simulation; Evolution prediction; Gaussian smoke plume model; SVM.

INTRODUCTION

Industrialization in China has remained dynamically innovative over the past 30 years; however, as a consequence, China is now facing significant environmental problems (Bell *et al*., 2014). Common environmental problems include air pollution, water pollution, and soil pollution. Among them, air pollution, especially PM$_{2.5}$ particles in the air, threatens human health directly (Chafe *et al*., 2015). PM$_{2.5}$ refers to fine particulate matter in air that are less than or equal to 2.5 µm in diameter. These particles can absorb bacteria, viruses, and harmful pollutants (Atkinson *et al*., 2014). The concentration of PM$_{2.5}$ is closely related to haze weather, which directly influences human health and causes various kinds of diseases, such as cardiovascular diseases (Kloog *et al*., 2015). Therefore, governance on haze weather and PM$_{2.5}$ has become the focus of the government, environmental protection departments, and the general public (Kioumourtzoglou *et al*., 2016).

Thus far, numerous scholars have studied the causes, related factors, and the harmful effects of PM$_{2.5}$ to the environment and human health. Huang *et al*. (2017) studied

PM$_{2.5}$ diffusion using NOAA, NAQFC, and other indexes; Kim *et al*. (2017) conducted long-term prediction of PM$_{2.5}$ index diffusion using the history prediction model; Hu *et al*. (2015) explored PM$_{2.5}$ diffusion using U.S. MODIS data with a three-phase model; Gan (2012) completed diffusion prediction through a point source Gaussian smoke plume model for the first time. Gao *et al*. (2017) predicted winter PM$_{2.5}$ concentration in the north of China, and also conducted diffusion research using the Gaussian plume model; Cao *et al*. (2016) conducted PM$_{2.5}$ diffusion research using the multi-scale adaptive KLMS kernel function; Zhou *et al*. (2003) comprehensively analyzed the influencing factors of atmospheric pollution through weather forecasting and a statistical method, and identified the influencing factor associated with PM$_{2.5}$; Di *et al*. (2016) conducted PM$_{2.5}$ diffusion research using the chemical transmission model, and predicted PM$_{2.5}$ using linear regression; Zhang *et al*. (2013) predicted the evolution of PM$_{2.5}$ through multiple linear regressions, and achieved good prediction. You *et al*. (2016) predicted the evolution of PM$_{2.5}$ using the geometric weighting regression model; Zhan *et al*. (2017) completed PM$_{2.5}$ evolution prediction of China's fog areas using the airspace dominant machine learning method; Huang *et al*. (2014) achieved accurate prediction of PM$_{2.5}$ with non-point source Gaussian smoke plume model, considering the influence of various factors on PM$_{2.5}$ to ensure more accurate prediction results. On this basis, Liu and Ding (2015) determined the influencing factors of

---

[*] Corresponding author.
 Tel.: +8617773188808; Fax: 008673184623968
 E-mail address: hpjzyysxy@csu.edu.cn

PM$_{2.5}$, identified key factors with the largest impact on PM$_{2.5}$ through a principal component analysis model, and established the theoretical basis for the evolution prediction of PM$_{2.5}$. Li (2014) built the Gaussian plume model by considering factors such as wind speed, simulated the diffusion of PM$_{2.5}$, and obtained ideal results.

According to existing literature, the diffusion simulation and evolution prediction of PM$_{2.5}$ still require further study. In this study, an improved Gaussian smoke plume model was built on the basis of the existing Gaussian smoke plume model. The new model took rain wash, gravity sedimentation, and surface rebound into full consideration. In the evolution prediction of PM$_{2.5}$, the key influencing factors were extracted from the literature through principal component analysis. Then, they were transformed into feature vectors, and input into support vector machine (SVM) for training. Next, the trained SVM was used to complete the PM$_{2.5}$ evolution prediction of the follow-up time. Finally, a corresponding simulation experiment was conducted using data provided by nine PM$_{2.5}$ monitoring sites in Wuhan, and the improved algorithm proposed in the paper was verified.

IMPROVED GAUSSIAN SMOKE PLUME MODEL

Gaussian Smoke Plume Model

The Gaussian smoke plume model is a pollutant diffusion model of PM$_{2.5}$ in the atmosphere established on the assumption of normal distribution. Among numerous diffusion prediction models, Gaussian smoke plume models based on turbulence statistics can provide more accurate predictions for simulating actual conditions (Ristic *et al.*, 2015). Therefore, research on diffusion models of PM$_{2.5}$ generally adopt Gaussian smoke plume models. The distribution of the Gaussian function can help simulate all kinds of random processes. The current Gaussian smoke plume model includes a point diffusion model, closed diffusion model, and surface diffusion model, which are applied to the diffusion simulation of different pollutants. The point diffusion model is usually constructed in the diffusion modeling of PM$_{2.5}$.

Gaussian smoke plume models can simulate the concentration distribution of PM$_{2.5}$, and areas of dangerously high concentrations, as well as indicate the concentration diffusion radius, and also reflect the concentration peak of a certain concentration diffusion point with changes in time. Fig. 1 shows the classic Gaussian point source smoke plume diffusion model.

In the model, $t = 0$ signifies the origin coordinates of the PM$_{2.5}$ point pollution sources; after t, the contamination concentration of the point (x, y, z) in 3D space is marked with $C(x, y, z)$. According to the calculation equation of surface flux, the flux Q per unit area per unit time could be obtained:

$$Q = -\sigma grad C(x, y, z) \tag{1}$$

σ denotes diffusion coefficient; $grad C(x, y, z)$ denotes pollution concentration gradient because the pollutants diffuse from high to low concentration areas. The transformation equation of pollutants could be obtained from turbulent diffusion and atmospheric molecular diffusion theory:

$$\frac{\partial C}{\partial t} = \sigma \nabla (grad C) = \sigma_x \frac{\partial^2 C}{\partial x^2} + \sigma_y \frac{\partial^2 C}{\partial y^2} + \sigma_z \frac{\partial^2 C}{\partial z^2},$$
$$t > 0, x, y, z \in [-\infty, +\infty] \tag{2}$$

The X axis signifies the downwind diffusion direction axis of the pollution source; the Y axis is the horizontal direction axis; the Z axis is the vertical upward direction axis; and σ_x, σ_y, and σ_z are the diffusion coefficients of the respective corresponding directions, and the coefficients are closely related to the atmospheric stability and downwind distance. ∇ signifies the dissolution degree.

Eq. (2) is the parabolic partial differential equation in infinite space. The source function of coordinate origin was recorded as initial conditions:

$$C(x, y, z, 0) = Q\delta(x, y, z) \tag{3}$$

Q denotes the total amount of pollutants; $\delta(x, y, z)$ is the intensity function of the point source. From the limit of the
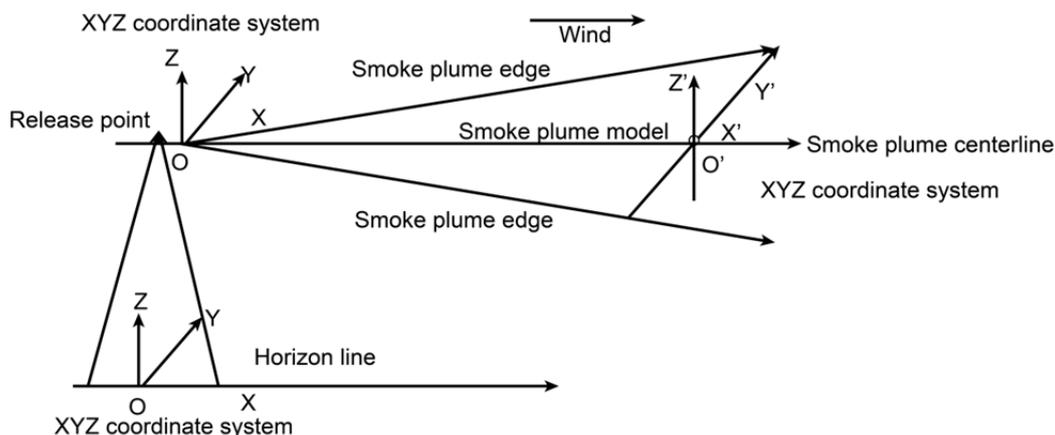

Fig. 1. Classic Gaussian point source smoke plume diffusion model.

initial conditions and after the introduction of the Gaussian function, the Gaussian smoke plume diffusion model can be defined as follows:

$$C(x, y, z, H(m)) = \frac{Q}{2\pi k \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right)$$
$$\left[\exp\left(-\frac{(z - H(m))^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z + H(m))^2}{2\sigma_z^2}\right)\right] \quad (4)$$

$H(m)$ signifies the effective emission height of $PM_{2.5}$; k denotes the average wind speed. The model assumes that the wind speed is horizontal in the X-axis direction.

Through the definition of Gaussian smoke plume model $C(x, y, z, H(m))$, the model could simulate the diffusion of $PM_{2.5}$ concentration at various points through normal distribution in the three-dimensional space. In fact, because the model was based on turbulent statistics, the following assumptions were made during the process of modeling:
(1) Pollutants presented Gaussian distribution in space;
(2) The space wind speed was uniform, and the direction of wind did not change;
(3) The distribution of pollutants was continuous in space;
(4) The process of diffusion followed the law of conservation of energy, and the conversion rate was 100%.

Gaussian Smoke Plume Model Modified with Multiple Factors

In the actual prediction process of $PM_{2.5}$, interference conditions (Hosseini et al., 2016), such as air humidity, gravity sedimentation, and earth surface reflection, exist. In order to solve the inaccurate diffusion prediction of traditional Gaussian smoke plume models under the actual conditions of several influencing factors, the modified models were established in view of the three largest influencing factors (air humidity, gravity sedimentation and earth surface reflection) such that the Gaussian smoke plume model satisfied the various constraints, and the accuracy of the predicted $PM_{2.5}$ concentration in air could be increased.

(1) Air humidity factor:
When air humidity is high, during fog, rain, and snow events, some particles in the air are cleaned, and soluble gases will be incorporated into the water; when air humidity is higher, some $PM_{2.5}$ particles will suffer erosion or sedimentation, leading to changes in the diffusion region (Elperina et al., 2016). It was assumed that the coefficient of $PM_{2.5}$ was washed or subsided as φ. The relationship between the coefficient and moisture impact strength is as follows:

$$\varphi = aI^b \quad (5)$$

I signifies the strength of the effect of moisture; a and b are parameters, and their experience values were usually taken.

Considering erosion or sedimentation, the source point strength was modified as follows:

$$Q'(x) = \exp(-\frac{\varphi x}{k}) \quad (6)$$

k denotes the average wind speed. The adjusted source point strength can be expressed as follows:

$$Q_A(x) = Q(x) \times Q'(x) \quad (7)$$

$Q(x)$ signifies the original source point strength.

According to the change of the source point, the Gaussian smoke plume model in Eq. (4) was modified and can be represented as follows:

$$C(x, y, z, H(m)) = \frac{Q_A(x)}{2\pi k \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right)$$
$$\left[\exp\left(-\frac{(z - H(m))^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z + H(m))^2}{2\sigma_z^2}\right)\right] \quad (8)$$

When $z = 0$, the concentrations distribution of ground $PM_{2.5}$ could be obtained:

$$C(x, y, 0, H(m)) = \frac{Q_A(x)}{2\pi k \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right)$$
$$\exp\left(-\frac{H(m)^2}{2\sigma_z^2}\right)(1 + \alpha) \quad (9)$$

When $z = y = 0$, the concentrations distribution of $PM_{2.5}$ at the ground axis direction could be obtained:

$$C(x, 0, 0, H(m)) = \frac{Q_A(x)}{2\pi k \sigma_y \sigma_z} \exp\left(-\frac{H(m)^2}{2\sigma_z^2}\right)(1 + \alpha) \quad (10)$$

The Gaussian smoke plume model was modified by regarding the influence of erosion or sedimentation as a weight to make the model obtain accurate simulation results of $PM_{2.5}$ diffusion in a high-humidity environment.

(2) Gravity sedimentation factor
In actual observation scenarios, $PM_{2.5}$ particle diffusion is influenced by rain wash as well as sedimentation due to gravity (Ristic et al., 2015). For sedimentation caused by gravity, sedimentation velocity generally depends on the combined action of the particles' own gravity and the acting force of air on particles. In the gravitational theory, it could be expressed as follows by the Stoke formula:

$$V_s = \frac{\rho g D^2}{\mu} \quad (11)$$

ρ denotes the density of $PM_{2.5}$ particles; g is acceleration due to gravity; D is the diameter of $PM_{2.5}$ particles; μ is the air reacting force coefficient; V_s signifies the sedimentation velocity of particles under the action of gravity.

Under the effect of downward speed produced by gravity

sedimentation, after time t, the descending distance of $PM_{2.5}$ particles on X axis is as follows:

$$V_s t = \frac{V_s x}{\mu} \tag{12}$$

Within t time in the X-axis direction, the result of changes in height is $H - (V_s x)/\mu$; according to the diffusion theory and momentum transfer theory, the amount of ground sedimentation produced by the height change is as follows:

$$W_d = V_s \times C(x, y, 0) \tag{13}$$

Under the action of ground sedimentation volume, the ground $PM_{2.5}$ concentration distribution was correspondingly modified:

$$C(x, y, z, H(m)) = \frac{Q_A(x) W_d}{2\pi k \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right)$$
$$\left[\exp\left(-\frac{(z - H(m))^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z + H(m))^2}{2\sigma_z^2}\right)\right] \tag{14}$$

(3) Earth surface reflection factor:

Earth surface reflection is another major factor influencing $PM_{2.5}$ particles. When $PM_{2.5}$ particles undergo washing or gravity sedimentation on a building surface or the ground, some $PM_{2.5}$ particles are rebounded into the air; these particles could still cause secondary diffusion (Juodis *et al.*, 2016). The rebounded $PM_{2.5}$ particles also require further weighting, and therefore, the original Gaussian smoke plume model was modified. It was assumed that the ground rebound coefficient in the rebound model was α; for a point P in space, pollutants can be considered as the sum of two parts; one part is $PM_{2.5}$ concentrations produced by sedimentary volume, and the other part is $PM_{2.5}$ concentrations produced by the ground reaction. The sum of the two parts is the $PM_{2.5}$ particle concentrations at current point P:

$$C(x, y, z, H(m)) = \frac{Q_A(x) W_d}{2\pi k \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right)$$
$$\left[\exp\left(-\frac{(z - H(m))^2}{2\sigma_z^2}\right) + \alpha \exp\left(-\frac{(z + H(m))^2}{2\sigma_z^2}\right)\right] \tag{15}$$

In addition, the bounce produced by building surfaces need to be taken into account. Buildings generally produce tilting bounce. Assuming that the rebounded coefficient is β, the revised $PM_{2.5}$ particle concentration at point in the space is as follows:

$$C(x, y, z, H(m)) = \frac{Q_A(x) W_d}{2\pi k \sigma_y \sigma_z}$$
$$\left[\exp\left(-\frac{x^2}{2\sigma_z^2}\right) + \beta \exp\left(-\frac{(x + 2H(m))^2}{2\sigma_z^2}\right)\right] \tag{16}$$

By combining with Eqs. (15) and (16), a condition of $\alpha = 0$, $\beta = 0$ implies that ground/building surface pollutants were completely absorbed; in contrast, $\alpha = 1$, $\beta = 1$ implies that the ground/building surface pollutants were completely rebounded.

The modified Gaussian smoke plume model could successfully simulate the diffusion of $PM_{2.5}$ considering air humidity, gravity sedimentation, and earth surface reflection in actual situations. In addition, the diffusion area in the three-dimensional space could be simulated through the model. Nevertheless, in addition to the model building of diffusion scope, the $PM_{2.5}$ value at each sampling point of the diffusion also needed to be estimated, which is influenced by many factors.

EVOLUTION AND PREDICTION OF $PM_{2.5}$ CONCENTRATION BASED ON SVM

According to the correlation analysis of $PM_{2.5}$, the concentration level was influenced by many factors, generally CO_2, SO_2, O_3, H_2O, temperature, and humidity. Thus, it was difficult to obtain good generalization through a mathematical model (Zhang *et al.*, 2016). In the study, from the viewpoint of machine learning, nine important influencing factors of $PM_{2.5}$ were considered as feature vectors; then, prediction and regression were performed with machine learning such that the $PM_{2.5}$ concentration in the area would automatically evolve over time. Based on statistical rules, there was no obvious linear relationship between the air pollution factors and $PM_{2.5}$, and various nonlinear relationships further complicated model depiction. On this basis, SVM was used for training the influencing factors. $PM_{2.5}$ evolution was predicted using a data training model. The model could use the kernel method to describe all kinds of nonlinear relations, had good adaptability to different samples, and was suitable for solving various nonlinear problems.

SVM is a type of machine learning classifier based on the statistical learning method (Hou *et al.*, 2016). Mapping low dimensional input vectors to high dimensional feature space through nonlinear mapping allows the feature vector to always be split by the hyperplane in high dimensional space. For this purpose, the classification surface of the data requires optimization, namely, the maximum interval between the data on the classification surface. The optimal classification plane $w \times x + b = 0$ was analyzed using SVM, where w denotes the weight, x denotes the support vector, and b is additive bias. Solving the optimal classification plane means solving the quadratic programming problem of w and b:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 \tag{17}$$

$$s.t. y_i(\omega \cdot x_i + b) \geq, i = 1, 2, \ldots, n \tag{18}$$

By solving the optimization problem of quadratic programming, the optimization of weights w and b could be solved through the Lagrange multiplier method and

KKT conditions. Then, the optimal results of the sample classification could be obtained.

In fact, almost all of the practical problems cannot meet the linear separable data set x, and linear inseparable data sets cannot be classified only through linear classification. On this basis, SVM combines slack variables and allows some samples to be wrongly classified. In this case, SVM can be used to solve the nonlinear separable data sets. In addition, as shown in Fig. 2, such data sets could not be solved through slack variables because classification according to the linear classification plane would lead to more incorrectly classified samples, resulting in poorer final classifier performance. Therefore, in SVM, the concept of kernel function was adopted, and the linear inseparable data sets were changed through the kernel function to the high dimensional space, through which mapped original data such that the linear classification plane can be effectively used for classification.

The quadratic optimization problem of SVM could be converted as follows by changing the kernel function:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{m} \xi_i \tag{19}$$

$$s.t. y_i(\omega \cdot \varphi(x_i) + b) \geq 1 - \xi_i \tag{20}$$

$$\xi_i \geq 0, i = 1, 2, 3, \ldots, n \tag{21}$$

ξ_i denotes the slack factor, φ_x denotes the kernel function, and C is the penalty coefficient. Similarly, the optimal solution of quadratic programming could be obtained using the Lagrange multiplier method and KKT conditions, and then the nonlinear SVM classification plane could be expressed as follows:

$$f(x) = \text{sgn}\left[\sum_{i,j=1}^{m} a_i y_i K(x_i, x_j) + b \right] \tag{22}$$

$K(x_i, x_j)$ denotes the kernel function. Based on the idea of kernel function, we could easily predict PM$_{2.5}$ diffusion using SVM with multiple kernel functions.

In general, multiple characteristics are required for the prediction and classification with SVM, and the prediction of PM$_{2.5}$ evolution can be completed by building related characteristics of PM$_{2.5}$. According to the related studies on the evolution process of PM$_{2.5}$, influencing factors that have significant relationships with PM$_{2.5}$ generally include this study, according to the literature, nine influencing factors air, temperature, humidity, pressure, and wind speed. In were selected for the prediction of PM$_{2.5}$ (SO$_2$, NO$_2$, O$_3$, PM$_{10}$, CO, temperature, humidity, air pressure and wind speed) from many factors. We input the 9D features to the SVM in the form of vectors, and predicted the PM$_{2.5}$ evolution by building SVM training and prediction processes. For the kernel function, the suitable high-dimensional space RBF kernel function was selected:

$$f(x) = \text{sgn}\left\{ \sum_{i=1}^{n} a_i \exp\left\{ -\frac{x_i - x}{\sigma^2} \right\} + b \right\} \tag{23}$$

In the process of prediction, the collected training set was assumed to be $\{x_i, y_i, i = 1, 2, 3, \ldots, 9\}$, and nine samples were considered. In the training set, x denotes the feature vector composed of the nine influencing factors, and y signifies the specific PM$_{2.5}$ concentration value obtained from every sample. They were regarded as the tag for regression prediction in actual use; regression could be completed through the input feature vector in the trained model. The training process of SVM was aimed at finding an optimal separating hyperplane $w^T x + b = 0$, on which the 9D feature vectors provided by x could be distinguished to the highest degree. In fact, when the data had low degree of differentiation, SVM provided the kernel function to solve the problem. The original feature could be converted to high dimensional space through different kernel functions. Under the conversion function, characteristics had stronger linear separability, and the generalization of different categories of data was increased.

Because the 9 feature vectors extracted in the study were independently and identically distributed in space and time, and various variables almost had no relationship, they could be effectively distinguished through the radial basis function. Then, the prediction and regression of PM$_{2.5}$ could be completed through the SVM separating hyperplane. Thus, the SVM classification kernel function in the paper was selected as the radial basis function.
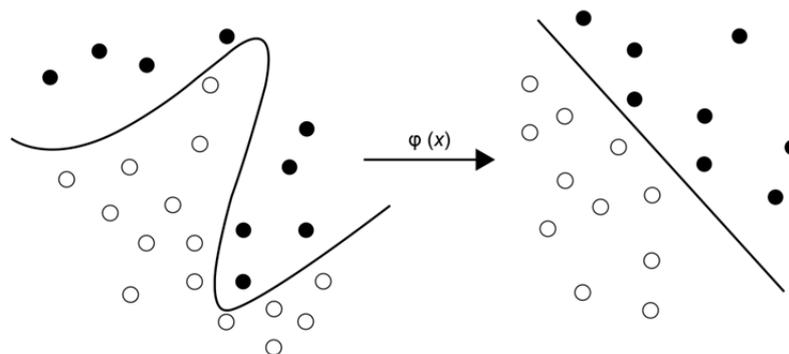

Fig. 2. Changing nonlinear data set and higher dimensional linear separable data sets through the kernel function.

RESULTS AND DISCUSSION

Experimental Scenario and Data

PM$_{2.5}$ data from Wuhan were taken as the analysis object. Wuhan, Hubei Province, is located in the middle and lower reaches of the Yangtze River at 30.5931°N, 114.3054°E, and covers an area of 8000 km^2, of which urban area accounts for around 800 km^2 and forest cover accounts for more than 25%. Because its year-round climate is temperate continental climate and subtropical monsoon climate, it receives adequate annual rainfall, sufficient sunlight, and four distinct seasons. It is hot in summer and cold in winter and has high annual rainfall. There are nine PM$_{2.5}$ monitoring sites in Wuhan, which publish real-time monitoring data every hour. The monitoring data include SO$_2$, NO$_2$ and O$_3$, PM$_{10}$, CO concentrations required for the SVM prediction in this study. Each index required by PM$_{2.5}$ regression and prediction could be obtained by combining with the real-time monitoring indexes of temperature, humidity, air pressure, and wind speed. Fig. 3 shows the distribution of PM$_{2.5}$ monitoring stations in Wuhan.

Experimental Process and Parameter Solving

(1) Solving diffusion parameters of the improved Gaussian smoke plume model: When solving the parameters of the improved Gaussian plume model, atmospheric stability was first required. In general, atmospheric stability includes six levels, namely, extremely unstable, unstable,

slightly unstable, moderately stable, stable, and highly stable. In the process of analysis and calculation of the experimental data, parameters were built in the highly stable atmospheric environment because Wuhan has moderate latitude and better atmospheric stability (Liu *et al.*, 2015). In addition, it was assumed that the diffusion parameters on the direction of *x* and *y* were the same. Only the two parameters σ_y, σ_z along the direction of *y* and *z* were actually used. The diffusion coefficient was determined by the following equation:

$$\sigma_y = \gamma_1 X^{\varepsilon 1}, \ \sigma_z = \gamma_2 X^{\varepsilon 2} \tag{24}$$

γ, ε is the effect of wind speed on PM$_{2.5}$ particle diffusion, and *X* is the observation data at the current time point.

According to the data of the PM$_{2.5}$ monitoring station in January and February 2015, the improved Gaussian smoke plume model was solved. When solving the Gaussian plume model, we needed to consider the proposed three major factors, namely, rain wash, gravity settling, and surface reflection. Rain wash is more complex than the other factors in the specific calculation process, so it was excluded in the experiment. Therefore, gravity settling was first considered in the experiment. According to Eqs. (11–13), the key parameter W_d in gravity settling factor was calculated. On this basis, considering the surface reflection factor, the key parameters α and β of the surface reflection factor were calculated according to Eq. (15). Finally, Q' and $H(m)$ of the Gaussian model under ideal conditions were calculated
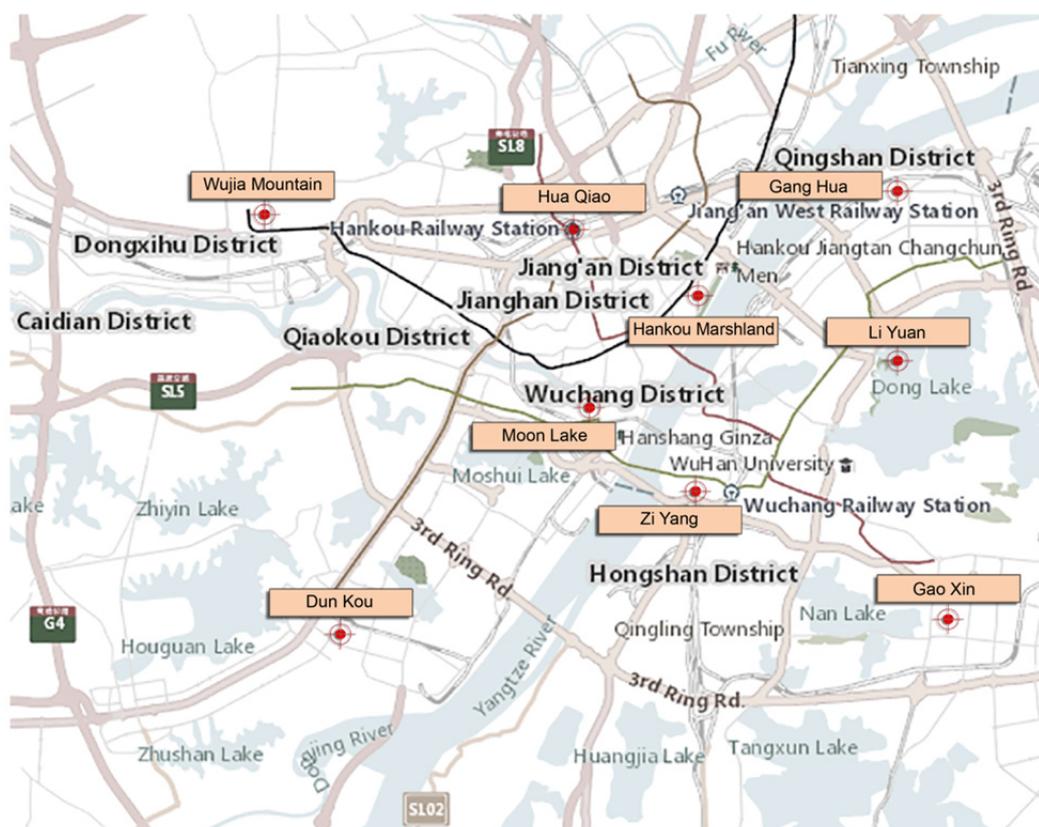

Fig. 3. Distribution of PM$_{2.5}$ monitoring stations in the urban areas of Wuhan.

according to Eqs. (1)–(4); according to the parameter values of the two key factors, the adjusted Q' value and $H(m)$ were calculated through Eqs. (14) and (16). The parameter solving results are given in Table 1. Parameters γ_1 and γ_2 were obtained through observation, and parameters ε_1 and ε_2 were obtained through calculation based on Eq. (24).

(2) Solving PM$_{2.5}$ source strength: The collected mass fraction was first converted to concentration because the actual diffusion prediction involved the prediction of PM$_{2.5}$ concentrations. The equation below gives the conversion process (Zhu *et al.*, 2017):

$$C_P = BP_{LO} + \frac{(IAQI_P - IAQI_{LO}) \times (BP_{Hi} - BP_{LO})}{IAQI_{Hi} - IAQI_{LO}} \quad (25)$$

$IAQI_P$ signifies air quality partial of PM$_{2.5}$; C_P is the mass density of PM$_{2.5}$; BP_{Hi} is the air quality subindex and the high value of nearby pollution concentration limit of the corresponding C_P; BP_{LO} denotes low value of nearby pollution concentration limit of the corresponding C_P; $IAQI_{Hi}$ signifies the air quality partial corresponding to BP_{Hi}; and $IAQI_{LO}$ is the air quality partial corresponding to BP_{LO}. After the abovementioned transformation, the recorded values of PM$_{2.5}$ from the stations could be converted into the mass density. The location of the collection stations is shown in Fig. 3; formatting data are provided to simplify the calculation of C_P, and C_P was calculated using data from all the collection stations. Through the total mass concentrations of PM$_{2.5}$ calculated using the abovementioned process, source intensity Q could be expressed with unit time per unit volume of mass concentration of PM$_{2.5}$ (Ma *et al.*, 2016). We assumed a C unit mass concentration within t time and V volume. Then, the source intensity Q could be solved with C_P:

$$Q = \frac{C_P}{t} \times V \quad (26)$$

It was assumed that $t = 1$ h, $V = 10$ km^3.

(3) Calculation of the distance between collection stations: The distance between collection stations is closely related to the diffusion of PM$_{2.5}$. According to the latitude and longitude relationship between each collection station, the shortest distance between each collection station could be calculated from the great-circle distance. Assuming that $(\varphi_1, \lambda_1), (\varphi_2, \lambda_2)$ denotes the latitude and longitude of two collection stations, the great circle distance between two collection stations on the sphere could be built through the definition of the haversine function and cosine of two angles.

$$\Delta\hat{\sigma} = 2\arcsin(\sqrt{\sin^2(\Delta\varphi/2) + \cos\varphi_s \cos\varphi_f \sin^2(\Delta\varphi/2)}) \quad (27)$$

After the above parameter solving, source transformation, and distance calculation, the diffusion results of PM$_{2.5}$ pollution in the 9 monitoring points could be computed. Table 2 shows the diffusion results and statistical value.

As shown in Table 2, Yuehu, Huaqiao, Qingshan, Ziyang, and Donghu in heavy industry area of Wuhan had higher PM$_{2.5}$ concentration, and more obvious diffusion of PM$_{2.5}$. Moreover, Wujiashan, Jiangtan, as well as other scenic and living areas had lower concentrations of PM$_{2.5}$, and smaller diffusion range, being consistent with the basic law of the environment. In addition, remote areas, such as Zhuankou, which has lower local population density and belonged to an agricultural area, had the lowest PM$_{2.5}$ concentration without any diffusion of PM$_{2.5}$. In short, the Gaussian smoke plume model established in the paper could effectively analyze different environmental conditions and different concentrations of PM$_{2.5}$ and diffusion. The model built in this study considers gravity settling, building surface rebound, and the rain wash, so it could ensure true and reliable results corresponding to actual conditions, and adaptability to various cases. The statistical results of the nine monitoring

Table 1. Results solving parameters of the improved Gaussian plume model.

Index	γ_1	ε_1	γ_2	ε_2	α	β	Q'	W_d	$H(m)$
PM$_{2.5}$ Value1	0.92	0.05	0.78	0.06	0.11	0.08	32.5	84.7	[0, 1 k]
PM$_{2.5}$ Value2	0.88	0.07	0.53	0.37	0.07	0.03	33.7	89.2	[1 k, 10 k]
PM$_{2.5}$ Value3	0.74	0.13	0.32	2.4	0.03	0.01	34.6	87.6	> 10 k

Table 2. Diffusion results and statistical values of PM$_{2.5}$ pollution at nine monitoring points.

	Wujia Mountain	Moon Lake	Hua Qiao	Hankou Marshland	Gang Hua	Li Yuan	Zi Yang	Dun Kou	Gao Xin
Wujia Mountain	0	10.12	8.11	6.03	7.12	4.58	7.12	8.43	11.9
Moon Lake	9.42	0	4.33	5.51	6.01	3.12	5.87	6.41	5.18
Hua Qiao	7.43	4.12	0	8.51	6.71	5.98	4.45	6.98	6.55
Hankou Marshland	5.87	5.63	8.34	0	5.11	4.59	3.87	4.03	6.43
Gang Hua	7.34	6.33	6.39	5.38	0	4.72	4.39	5.87	6.73
Li Yuan	4.51	2.45	5.56	4.32	5.31	0	5.03	6.01	7.93
Zi Yang	7.23	6.74	4.39	3.64	4.56	5.21	0	1.85	2.34
Dun Kou	8.33	7.42	7.43	3.87	6.34	6.26	2.19	0	6.48
Gao Xin	12.1	4.35	6.31	7.34	6.93	8.43	2.73	8.73	0

points in Wuhan show that there was a certain difference in the values of mutual testing areas. For example, the test value was 10.12 from Wujiashan to Yuehu, but it was 9.42 from Yuehu to Wujiashan. In general, such small differences are caused by factors such as temperature, humidity and wind direction. These minor factors should be taken into consideration in the process of follow-up studies so as to increase the reliability of the results.

(4) Predicting $PM_{2.5}$ evolution with SVM: The indexes of SO_2, NO_2 and O_3, PM_{10} and CO could be obtained from $PM_{2.5}$ monitoring stations. By combining with the real-time parameters such as temperature, humidity, air pressure, and wind speed provided by the Wuhan Meteorological Agency, multidimensional characteristics could be generated with the nine real-time parameters. Then, through SVM training with a month as the unit, $PM_{2.5}$ diffusion and evolution could be predicted for the follow-up months. Finally, the evolution prediction results using the SVM model were compared with the actual values. Fig. 3 shows the comparison of the prediction values and real values of concentration evolution for $PM_{2.5}$ monitoring stations in Wuhan within eight follow-up months. In addition, a contrasting experiment was conducted under the same feature vectors using the linear regression model (Di *et al.*, 2016). Finally, the evolution prediction results through SVM model and the linear regression model were compared with real results. Fig. 4 shows the comparison of the evolution prediction results of the SVM model and the linear regression model with the real results. Furthermore, in the experiment, the concentration evolution of $PM_{2.5}$ at the nine monitoring stations in Wuhan was predicted within the eight follow-up months.

As the comparison results show, there were no significant differences between the predicted value and actual value of $PM_{2.5}$ evolution from March to May, while there were more differences from June to October. The climate, temperature, rainfall, and humidity in Wuhan changed slightly from March to May and had little impact on $PM_{2.5}$. With the change of temperature and humidity, coupled with the impact of monsoon winds, the hypothesis of the prediction research was nullified. Thus, the gap between evolution prediction value and the actual value gradually increased. As shown by the comparison of the evolution prediction results of the linear regression model and SVM model, the SVM model showed clearly higher precision than the linear regression model towards the last few months. With longer prediction time, the prediction model requires higherdimension features, which would result in poorer fitting of the machine learning model, and eventually decrease the prediction precision. However, in the SVM model, we adopted the RBF kernel, which could cope with higher dimension features and provided better characteristics and robustness of prediction. Nevertheless, because there were insufficient samples in SVM modeling, it was assumed that samples under various climates throughout the year could be obtained. Accordingly, the SVM modeling was performed with $PM_{2.5}$ data for the whole year to obtain better results. The SVM model built in this study had certain generalizations and can better predict $PM_{2.5}$ evolution.

CONCLUSION

With rapid industrialization, haze has become an increasingly serious threat in China. To effectively control haze, the $PM_{2.5}$ concentration must first be decreased. A large number of studies have shown that the formation of
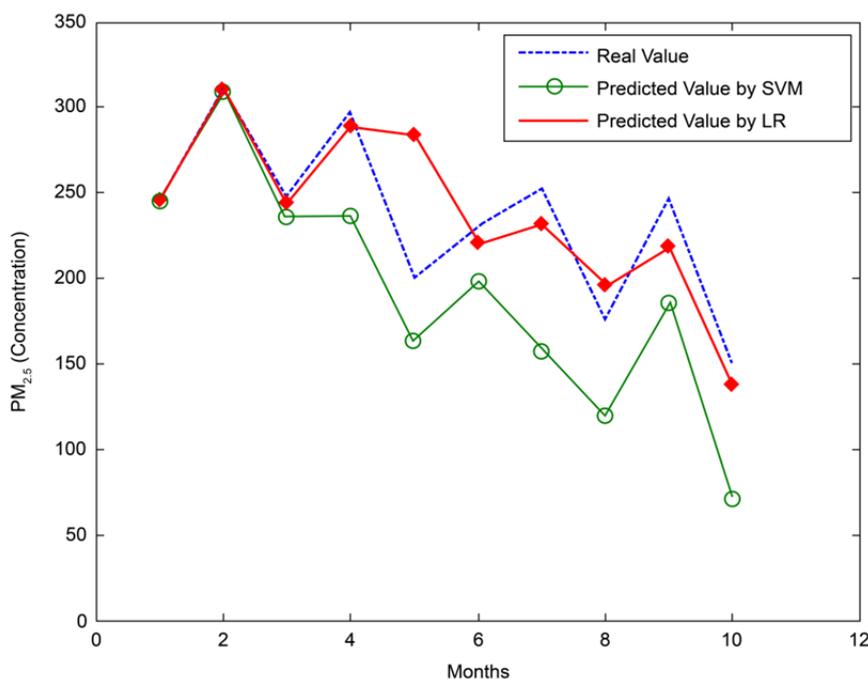

Fig. 4. Comparison of the prediction values and real values of $PM_{2.5}$ evolution in the monitoring stations in Wuhan within 8 months.

$PM_{2.5}$ is complicated and mutually influenced by several factors. In order to more effectively simulate and predict $PM_{2.5}$ evolution, the Gaussian smoke plume model was rebuilt based on the practical situation and a consideration of multiple influential factors. Feature vectors were constructed by selecting the 9 factors with the largest influence on $PM_{2.5}$, and an SVM model was established. The model achieved good diffusion simulation and good evolution prediction. However, because various factors influence $PM_{2.5}$, the currently built Gaussian smoke plume model and the SVM model still cannot effectively and accurately simulate and predict $PM_{2.5}$ values. Therefore, more factors should be considered in future work. In addition, the Gaussian smoke plume model and SVM model should be modified to ensure that the simulation and prediction of $PM_{2.5}$ are closer to the actual measured values.

ACKNOWLEDGEMENTS

REFERENCES

Atkinson, R.W., Kang, S., Anderson, H.R., Mills, I.C. and Walton, H.A. (2014). Epidemiological time series studies of $PM_{2.5}$ and daily mortality and hospital admissions: A systematic review and meta-analysis. *Thorax* 69: 660–665.

Bell, M.L., Ebisu, K., Leaderer, B.P., Gent, J.F., Lee, H.J., Koutrakis, P., Wang, Y., Dominici, F. and Peng, R.D. (2014). Associations of $PM_{2.5}$ constituents and sources with hospital admissions: analysis of four counties in Connecticut and Massachusetts (USA) for persons ≥ 65 years of age. *Environ. Health Perspect.* 122: 138–144.

Cao, Z., Yu, S. and Xu, G. (2016). Multiple adaptive kernel size KLMS for Beijing $PM_{2.5}$ prediction, In 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, Vancouver, pp. 1403–1407.

Chafe, Z.A., Brauer, M., Klimont, Z., Van Dingenen, R., Mehta, S., Rao, S., Riahi, K., Dentener, F. and Smith, K.R. (2015). *Household cooking with solid fuels contributes to ambient $PM_{2.5}$ air pollution and the burden of disease.* Dissertation, University of British Columbia.

Di, Q., Koutrakis, P. and Schwartz, J.A. (2016). Hybrid prediction model for $PM_{2.5}$ mass and components using a chemical transport model and land use regression. *Atmos. Environ.* 131: 390–399.

Elperin, T., Fominykh, A. and Krasovitov, B. (2016). Effect of raindrop size distribution on scavenging of aerosol particles from Gaussian air pollution plumes and puffs in turbulent atmosphere. *Process Saf. Environ. Prot.* 102: 303–315.

Gan, G.W. (2012). Atmospheric Environment impact assessment of acid gas release flare in oil refinery—Base on single-sourced gaussian plume model. *Technol. Dev. Chem. Ind.* 41: 44–48.

Gao, M., Saide, P.E. and Xin, J. (2017). Estimates of health impacts and radiative forcing in winter haze in eastern China through constraints of surface $PM_{2.5}$ predictions. *Environ. Sci. Technol.* 51: 2178–2185.

Hosseini, B. and Stockie, J.M. (2016). Bayesian estimation of airborne fugitive emissions using a Gaussian plume model. *Atmos. Environ.* 141: 122–138.

Hou, W., Li, Z., Zhang, Y., Xu, H., Zhang, Y., Li, K., Li, D., Wei, P. and Ma, Y. (2014). Using support vector regression to predict PM_{10} and $PM_{2.5}$. *IOP Conf. Ser.: Earth Environ. Sci.* 17: 012268.

Hu, X., Waller, L.A. and Belle, J.H. (2015). Estimation of $PM_{2.5}$ concentrations in the conterminous US using MODIS data and a three-stage model. AGU Fall Meeting Abstracts 2015, AGU, San Francisco.

Huang, G.J., Guo, H.M., Chen, J.P. and Fang, S.M. (2014). Research of $PM_{2.5}$ in the air. *Math. Pract. Theory* 44: 59–65.

Huang, J., McQueen, J. and Wilczak, J. (2017). Improving NOAA NAQFC $PM_{2.5}$ predictions with a bias correction approach. *Weather Forecasting* 32: 407–421.

Juodis, L., Filistovič, V., Maceika, E. and Remeikis, V. (2016). Analytical dispersion model for the chain of primary and secondary air pollutants released from point source. *Atmos. Environ.* 128: 216–226.

Kim, S.Y., Olives, C. and Sheppard, L. (2017.) Historical prediction modeling approach for estimating long-term concentrations of $PM_{2.5}$ in cohort studies before the 1999 implementation of widespread monitoring. *Environ. Health Perspect.* 125: 38–46.

Kioumourtzoglou, M.A., Schwartz, J.D., Weisskopf, M.G., Melly, S.J., Wang, Y., Dominici, F. and Zanobetti, A. (2016). Long-term $PM_{2.5}$ exposure and neurological hospital admissions in the northeastern United States. *Environ. Health Perspect.* 124: 23–29.

Kloog, I., Zanobetti, A. and Nordio, F. (2015). Effects of airborne fine particles ($PM_{2.5}$) on deep vein thrombosis admissions in the northeastern United States. *J. Thromb. Haemost.* 13: 768–774.

Li, W.L., Wu, H.Y. and Chen, Y. (2014). $PM_{2.5}$ Diffusion problem research based on Gaussian model in Wuhan city. *Chin. High Technol. Lett.* 24: 1153–1159. (in Chinese)

Liu, S.Y. and Ding, X.M. (2015). On forecast evaluation of $PM_{2.5}$ diffusion in Xi`an. *J. Southwest China Normal Univ.* 40: 78–83.

Liu, X., Godbole, A., Lu, C., Michal, G. and Venton, P. (2015). Optimisation of dispersion parameters of Gaussian plume model for CO_2 dispersion. *Environ. Sci. Pollut. Res. Int.* 22: 18288–18299.

Ma, Z., Hu, X. and Sayer, A.M. (2016). Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004–2013. *Environ. Health Perspect.* 124: 184–192.

Ristic, B., Gunatilaka, A., Gailis, R. and Skvortsov, A. (2015). Bayesian likelihood-free localisation of a biochemical source using multiple dispersion models. *Signal Process.* 108: 13–24.

You, W., Zang, Z. and Zhang, L. (2016). National-scale estimates of ground-level $PM_{2.5}$ concentration in China using geographically weighted regression based on 3 km

resolution MODIS AOD. *Remote Sens.* 8: 184.

Zhan, Y., Luo, Y. and Deng, X. (2017). Spatiotemporal prediction of continuous daily PM$_{2.5}$ concentrations across China using a spatially explicit machine learning algorithm. *Atmos. Environ.* 155: 129–139.

Zhang, C.J., Dai, L.J. and Ma, L.M. (2016). Rolling forecasting model of PM$_{2.5}$ concentration based on support vector machine and particle swarm optimization. International Symposium on Optoelectronic Technology and Application. International Society for Optics and Photonics 2016.

Zhang, Y.Y., Miao, G.H., Yan, J.S., Wang, J.L., Wang, Z.Q. and Wang, D.H. (2013). The factors affecting PM$_{2.5}$ multivariate statistical analysis and prediction. *Resour. Econ. Environ. Prot.* 11: 135–136. (in Chinese)

Zhou, L., Xu, X.D., Ding, G.A., Miao, Q.J. and Wei, F.Y. (2003). The correlation factors and pollution forecast model for PM$_{2.5}$ concentration in Beijing area. *Acta Meteorol. Sin.* 61: 761–768. (in Chinese)

Zhu, H., Liu, X., Wang, Q. and Sun, J. (2017). *The simulation of release consequence with a modified Gaussian plume model.* In *Fire science and technology 2015.* Harada, K., Matsuyama, K., Himoto, K., Nakamura, Y. and Wakatsuki, K. (Eds.), Springer, Singapore, pp. 445–454.