Analysis of Air Quality Characteristics Based on Information Diffusion Technology in Beijing, China

He ji*, Chen Haitao*†, Duan Chunqing**, Chen Xiaonan*** and Wang Wenchuan*

*School of Water Resources, North China University of Water Resources and Electric Power, Zhengzhou City Henan Province, 450045, PR China

**Beijing Water Affairs Center for Suburbs, Haidian District Beijing, 100073, PR China

***Construction and Administration Bureau of South-to-North Water Diversion Middle Route Project, Haidian District Beijing, 100038, PR China

†Corresponding Author: Chen Haitao; zzchenhaitao@126.com

ABSTRACT

To study the characteristics of air quality and the relationship between air quality and weather factors, based on daily meteorological data from 2016 to 2019 in Beijing using information diffusion technology, the probability distribution of air quality index in different seasons and the development trend of air quality have been studied, and the relationship between weather factors and air quality discussed. The results show that: 1) According to the air quality, the order of the four seasons is summer, spring, autumn and winter. In summer, the frequency of moderate air pollution and above is about 2.54%, and the frequency of serious air pollution is about 0%. In winter, the frequency of moderate air pollution and above is 17.83%, and the frequency of serious air pollution is 2.93%. 2) The air quality of Beijing has been improving in recent years, which shows that with the strengthening of air pollution control efforts, certain results have been achieved. 3) Quantitative analysis of the relationship between winter air quality index and temperature and wind in Beijing shows that the degree of air pollution in winter increases with the increase of temperature and decreases with the increase of wind force. The frequency of mild air pollution and above is about 8.91% when the daily maximum temperature is below 0°C and 48.78% when the daily maximum temperature is above 9°C. The frequency of mild air pollution and above is about 45.17% when the daily maximum wind force is level 0, and 20.89% when the daily maximum wind force is level 3 and above. Examples show that the information diffusion technology can make full use of the location information of the sample points by transforming the traditional sample data points into fuzzy sets, and achieves good results in frequency statistics and trend fitting. The model established in this paper has the value of popularization and application.

INTRODUCTION

People all over the world have suffered or are suffering from the haze, including China, the United States, Germany, Japan and so on. Haze pollution has become an important threat to the global environment. Along with the rapid development of economy in China, air pollution problems arise and lead to frequent haze in different degrees, which not only affects people's lives and health but also makes the environmental problems more and more serious in the future.

Air pollution and haze weather do great harm to human beings. Air pollution directly leads to a significant increase in the incidence of respiratory diseases, affecting physical and mental health. Haze pollution leads to reduced visibility, increased traffic accidents, affecting daily activities such as travel. Haze causes serious losses to agricultural production, aquaculture production and tourism development, and affects ecosystems.

In the autumn and winter of 2016, China's air pollution became more serious. The haze pollution problem extended from some areas to all parts of the country, covering 25 provinces. More than 100 large and medium-sized cities had haze weather of varying degrees, which not only affected the sea, land and air traffic conditions but also caused direct harm to people's physical and mental health. As the capital of China, Beijing is the political, cultural and economic centre of the whole country. However, in recent years, haze weather has occurred frequently. It is necessary to analyse the characteristics of air quality in-depth, which can provide a decision-making basis for the prediction and prevention of haze pollution weather.

The London smog event in Britain and the Los Angeles photochemical smog event in the United States appeared earlier than the haze weather in China and were the first to suffer from haze pollution. The results show that sulphur oxides and dust produced by coal combustion in the process

of industrial development were the direct causes of London smog events, and the accumulation of air pollutants with time due to the existence of inversion layer was the indirect cause (Davis et al. 2002). Primary pollutants from motor vehicles and chemical plants and secondary pollutants from photochemistry caused photochemical smog in Los Angeles, USA (Hinton et al. 2006). Minguillon (2012) used a positive definite matrix to analyse the main components and formation factors of PM2.5 in Switzerland. Through a comparative study of PM10 in residential and industrial areas in Calcutta, India, Chalouolakou et al. (2003) found that soot and motor vehicle emissions had the greatest impact on haze pollution in the region. With the continuous occurrence of haze weather in China, many Chinese scholars have conducted in-depth studies on the weather characteristics during the occurrence of haze. Wang (2011) found that the sharp increase of PM2.5 concentration in Beijing in 2008 was due to the slow movement of atmospheric clouds, which made the air unable to circulate rapidly, and the accumulation of atmospheric pollutants caused haze pollution. Wang (2002) studied the distribution and evolution of visibility in Beijing under haze pollution. Jiang et al. (2017) studied the meteorological factors of haze weather with the worst visibility in Zhengzhou City. The results showed that the meteorological conditions during haze pollution were low-level wind, relatively stable atmospheric structure and low-level inversion layer. Qian et al. (2006) analysed the time distribution characteristics of fog and haze weather in Guangdong Province, and found that the visibility of fog and haze weather is different from seasonal variation.

There are also many studies on haze prediction. Jian et al. (2012) used the autoregressive integral moving average model to predict the PM10 concentration in heavy traffic areas of Hangzhou, China. Dong et al. (2009) proposed a hidden Markov model to predict high PM2.5 concentration in haze-polluted weather. Li (2017) predicted PM2.5 concentration by multivariate statistical methods. Su et al. (2008) established a haze prediction model based on the Chi-square test and BP neural network. Mishra et al. (2015) combined with neural network and fuzzy logic regression to predict PM2.5 concentration in haze weather in India.

For a long time, many scholars have done many studies on the causes, hazards and control of haze, and have achieved rich results. At present, there are many characteristic analysis models of haze, such as regression, statistics and neural networks and so on. When there are contradictory samples or small samples in data processing, the effect of traditional methods is unsatisfactory. Information diffusion technology is a data processing method rising in recent years. By transforming data points into fuzzy sets, the location information of samples can be fully utilized, especially for data analysis

under incomplete conditions of samples (Huang 2018, Chen 2012). Regional air quality data are usually short in length and not abundant in quantity. The non-linear relationship regression and stable probability distribution among factors can be well realized by using information diffusion technology.

Taking Beijing as a typical research area and air quality index as research index, this paper uses information diffusion technology to analyse the evolution characteristics of air quality in different seasons in Beijing, calculates the probability distribution of haze occurrence in different seasons, and studies the evolution trend of air quality and air quality in recent years. To provide a scientific basis for prevention, early warning and control of local air pollution, the relationship between air pollution and temperature, wind force factors is discussed.

DATA AND METHODS

Survey of Research Area

Beijing is the capital of China, located in the northwest end of the North China Plain, adjacent to Tianjin in the east, surrounded by Hebei Province in the west, south and north. The total area of the city is about 16410 km^2, of which the mountainous area accounts for about 60% and the plain area accounts for about 40%. Beijing has 16 districts under its jurisdiction, of which 6 are in the urban area, namely Dongcheng District, Xicheng District, Haidian District, Fengtai District, Chaoyang District and Shijingshan District. There are 10 suburban districts, including Yanqing District, Huairou District, Miyun District, Changping District, Shunyi District, Pinggu District, Mentougou District, Tongzhou District, Fangshan District and Daxing District (Fig. 1). The climate in Beijing is a typical warm temperate humid continental monsoon climate. The seasonal variation of precipitation is great. The precipitation is mostly concentrated in summer and the annual average precipitation is about 585 mm.

Beijing is the political, economic and cultural centre in China, which attracts worldwide attention. However, in recent years, haze weather has frequently appeared. In January 2013, Beijing experienced the most haze intrusion in 59 years. It experienced four haze attacks for 25 days. The maximum PM2.5 index reached 755. According to the air pollution index published by relevant departments, the pollution situation has reached the "serious pollution" level. Since the beginning of winter in 2015, the haze has appeared frequently, lasting for up to two weeks. In December, the "red" warning of fog was issued. Haze has become a common weather condition in Beijing, causing huge adverse effects on the citizens. Although all parties have made great efforts to control the haze, no significant effect has been achieved.

Fig. 1: Sketch of administrative divisions in Beijing.

Data Sources

The assessment of air quality grade is based on the technical regulations of the Environmental Air Quality Index (AQI) (HJ633-2012) in China. Air pollution index is divided into six grades: 0-50, 51-100, 101-150, 151-200, 201-300 and more than 300. The bigger the index and the higher the level, the more serious the pollution is, and the more obvious the impact on human health is.

When the air pollution index belongs to the interval (0, 50], the air quality level is 1, and the air quality is excellent. There is no air pollution. All kinds of people can move normally.

When the air pollution index belongs to the interval (50, 100], the air quality level is 2, and air quality is good. Air quality is acceptable, but some pollutants may have a weak impact on the health of a very small number of abnormally sensitive people. It is suggested that abnormally sensitive people should reduce outdoor activities.

When the air pollution index belongs to the interval (100, 150], the air quality level is 3, and the air quality condition belongs to mild pollution. The symptoms of susceptible people are slightly aggravated, and the irritation symptoms are found in healthy people. It is suggested that children, the elderly people and patients with heart and respiratory diseases should reduce intensive outdoor exercise.

When the air pollution index belongs to the interval (150, 200], the air quality level is 4, and the air quality condition belongs to moderate pollution. The haze will further aggravate the symptoms of susceptible people and may affect the heart and respiratory system of healthy people. It is suggested that patients with diseases should avoid high-intensity outdoor exercise and that the general population should reduce outdoor exercise.

When the air pollution index belongs to the interval (200, 300], the air quality level is 5, and the air quality condition belongs to heavy pollution. The symptoms of patients with heart disease and pulmonary disease are significantly aggravated, exercise tolerance is reduced, and symptoms are common in healthy people. It is suggested that children, the elderly and patients with heart disease and pulmonary disease should stay indoors, stop outdoor exercise, and the general population should reduce outdoor exercise.

When the air pollution index is more than 300, the air quality level is 6, and the air quality is seriously polluted. It is suggested that children, the elderly and patients should stay indoors to avoid physical exhaustion and the general population should avoid outdoor activities.

ANALYSIS MODEL OF AIR QUALITY CHARACTERISTICS BASED ON INFORMATION DIFFUSION TECHNOLOGY

Air Quality Index Frequency Distribution Based on Information Diffusion

Information diffusion is a new way to study function approximation by using the method of fuzzy sets. The advantage of this method is that its estimation accuracy is higher than that of the traditional histogram. In particular, when the overall hypothesis of a given sample is not guaranteed to be correct

and the sample size is small, this new method is very useful to avoid the artificial destruction of the data structure and make full use of the information carried by each knowledge sample point.

Suppose that the discourse domain is U. A single observed sample point can diffuse the information it carries to all the points in the universe according to the following formula:

$$f_i(u_j) = \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{(x_i - u_j)^2}{2h^2}\right] \qquad \ldots(1)$$

Where, h is the information diffusion coefficient, which can be calculated by the maximum value a and minimum value b in the sample and the number of sample points n.

$$h = \begin{cases} 0.8146(b-a), & n=5; \\ 0.5690(b-a), & n=6; \\ 0.4560(b-a), & n=7; \\ 0.3860(b-a), & n=8; \\ 0.3362(b-a), & n=9; \\ 0.2986(b-a), & n=10; \\ 2.6851(b-a)/(n-1), & n \geq 11 \end{cases} \qquad \ldots(2)$$

Set,

$$C_i = \sum_{j=1}^{m} f_i(u_j) \qquad \ldots(3)$$

The membership function of the corresponding fuzzy subset is:

$$\mu_{x_i}(u_j) = \frac{f_i(u_j)}{C_i} \qquad \ldots(4)$$

Where $\mu_{x_i}(u_j)$ is called normalized information diffusion of sample point x_i.

Let,

$$q(u_j) = \sum_{i=1}^{n} \mu_{x_i}(u_j) \qquad \ldots(5)$$

$$Q = \sum_{j=1}^{m} q(u_j) \qquad \ldots(6)$$

the frequency value of the sample point at u_j is

$$p(u_j) = \frac{q(u_j)}{Q} \qquad \ldots(7)$$

Where, $p(u_j)$ is the exceeding probability of u_j.

According to the air quality index data in spring (March to May), summer (June to August), autumn (September to November) and winter (December to January of the following year), the probability distribution of air quality index in each season can be analysed by using the above information diffusion method.

Trend Analysis of Air Quality Based on Information Diffusion Approximate Reasoning

Regression analysis based on information diffusion technology is not only easy to calculate, but also can better reflect the overall trend of the data changes. Without grouping data artificially, it can automatically fit local trend changes and finally get a smooth, fluctuating regression curve.

Approximate reasoning based on information diffusion is a non-linear regression model between two variables. The main steps are as follows:

(1) Suppose that there are L samples $\{(x_1,y_1),(x_2,y_2),\ldots,(x_l,y_l)\}$. According to the distribution of sample values, the discrete domains of input variable x and output variable y can be determined respectively $U=\{u_1,u_2\ldots,u_s\}$, $V=\{v_1,v_2,\ldots,v_r\}$. (x_i,y_i) can be transformed into a fuzzy set according to the following formula:

$$A_i : \mu_{x_i}(u) = \exp\left[-\frac{(u-x_i)^2}{2h_x^2}\right], \quad u \in U \qquad \ldots(8)$$

$$B_i : \mu_{y_i}(v) = \exp\left[-\frac{(v-y_i)^2}{2h_y^2}\right], \quad v \in V \qquad \ldots(9)$$

Where, h_x and h_y are information diffusion coefficients of input and output samples respectively, calculated according to equation (2).

Table 1: Probability distribution of air quality in each season.

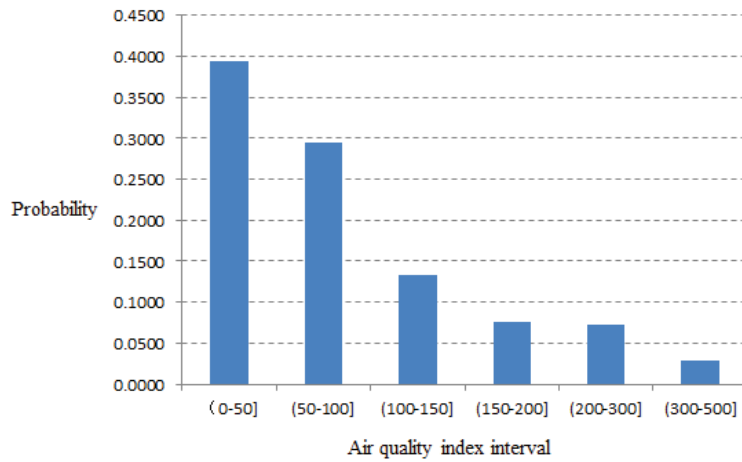| Index interval	Winter	Spring	Summer	Autumn
(0, 50]	0.3933	0.1882	0.2752	0.3669
(50, 100]	0.2952	0.4135	0.5184	0.3171
(100, 150]	0.1332	0.2333	0.1811	0.1901
(150, 200]	0.0757	0.0999	0.0254	0.0637
(200, 300]	0.0733	0.0470	0.0000	0.0586
(300, 500]	0.0293	0.0181	0.0000	0.0037

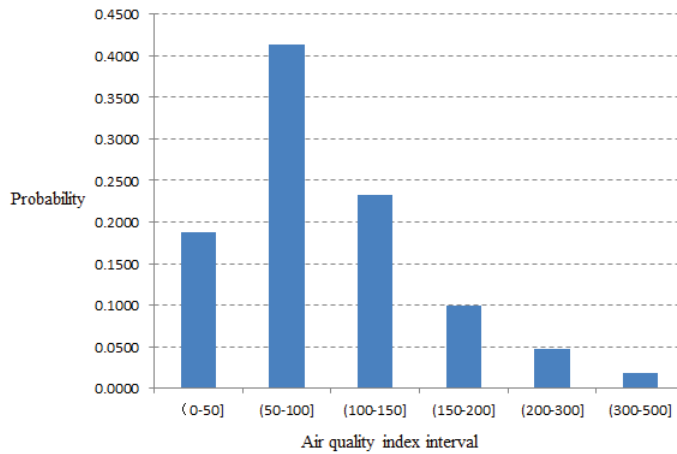Fig. 2: Probability distribution of air quality in winter.

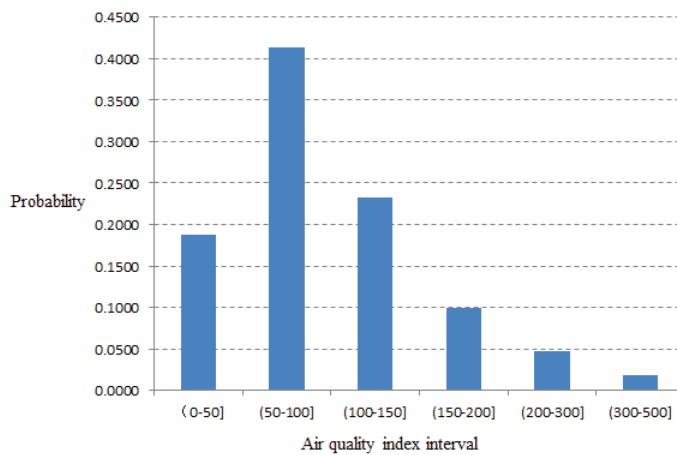Fig. 3: Probability distribution of air quality in spring

Fig. 4: Probability distribution of air quality in summer.

(2) According to the fuzzy set A_i and B_i transformed from the sample point (x_i, y_i), the fuzzy relations R_i can be obtained, which is calculated by the following formula:

$$\mu_{R_i}(u,v) = \mu_{A_i}(u)\mu_{B_i}(v) \qquad \ldots(10)$$

(3) For input data x_0, the estimated value of output y_0 is deduced. Converting x_0 to a fuzzy set according to the following equation:

$$\mu_{x_0}(u_j) = \begin{cases} 1-|x_0-u_j|/d, & |x_0-u_j| \leqslant d \\ 0, & |x_0-u_j| > d \end{cases} \qquad \ldots(11)$$

Where, $d=|u_1-u_2|$.

(4) According to the fuzzy relation R_i and the transformed fuzzy set from x_0, fuzzy reasoning is carried out as follows:

$$\mu_{y_0} = \sum_u \mu_{x_0}(u)\mu_{R_i}(u,v) \qquad \ldots(12)$$

Let v' satisfy the following function relation:

$$\mu_{y_0}(v') = \max_{v \in V}\{\mu_{y_0}(v)\} \qquad \ldots(13)$$

v' is called an estimated value of y_0, and the membership of v' is taken as the weight w_i.

(5) According to the above method, l estimated values of y_0 and corresponding weights are obtained for each group of samples. The final y_0 estimate is calculated by weighted average:

$$\bar{y}_0 = \left(\sum_{i=1}^{l} w_i\hat{y}_i\right) \Big/ \left(\sum_{i=1}^{l} w_i\right) \qquad \ldots(14)$$

The nonlinear regression relationship between air quality index and time is established by the information diffusion approximate reasoning method to describe the changing trend of air condition.

ANALYSIS OF AIR QUALITY CHARACTERISTICS IN BEIJING

Analysis of Air Quality Index Distribution in Different Seasons

According to the daily air quality data in different seasons, the air quality distribution in each season is analysed by information diffusion technology. The probability distribution of air quality grades in each season is shown in Table 1 and Figs. 2-6.

According to the calculation results, the air quality is the best in summer and the worst in winter. The air quality index in summer is less than 200, and there is no heavy pollution or serious pollution. The frequency of moderate pollution is about 2.54% in summer. In winter, the frequency of serious pollution is 2.93%, and the frequency of moderate air pollution and above is 17.83%.

Analysis of Change Trend of Air Quality in Winter

Air quality is the worst in winter. The trend of winter air quality in Beijing was studied by information diffusion technology. After fitting the daily air quality data in winter, the trend of air quality was obtained.

After processing the original data by information diffusion technology, it can better reflect the changing trend of air quality. From the overall trend, the air quality index shows a decreasing trend, that is to say, air quality has improved to a certain extent.

Correlation Analysis Between Air Quality Index and Meteorological Factors

Based on the winter air quality data in Beijing, the relationship between air quality index and daily maximum tempera-
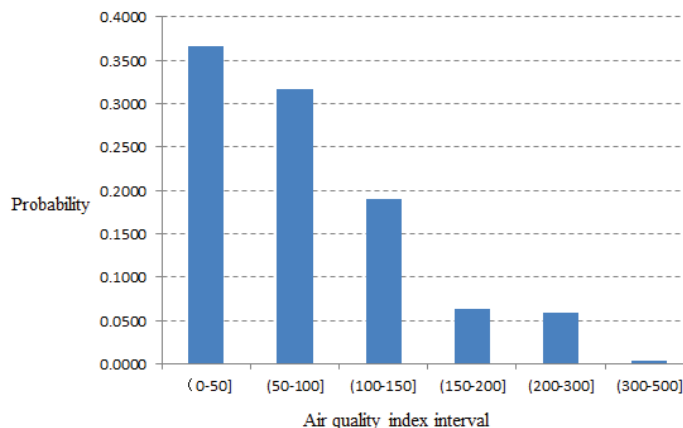

Fig. 5: Probability distribution of air quality in autumn.

Table 2: Probability distribution of air quality in different daily maximum temperature levels.

Air quality level	Temperature level 1	Temperature level 2	Temperature level 3	Temperature level 4
(0, 50]	0.6656	0.4188	0.2879	0.2350
(50, 100]	0.2453	0.3476	0.2853	0.2770
(100, 50]	0.0324	0.1086	0.1578	0.2752
(150,200]	0.0562	0.0410	0.1037	0.1035
(200, 00]	0.0005	0.0596	0.1135	0.0930
(300, 00]	0.0000	0.0244	0.0518	0.0161

ture and daily maximum wind power was analysed by using information diffusion technology. Because of the uncertainty of the correlation, this paper studies the correlation through probability analysis. Firstly, meteorological factors are graded. The daily maximum temperature is divided into 1~4 grades according to the distribution range of the measured data, and the daily maximum wind force is divided into 0~5 grades according to the relevant meteorological classification standards. The higher the grade value of meteorological factors is, the greater the corresponding value is. Then, the probability distribution of air quality corresponding to different meteorological factor levels is calculated, and the uncertainty relationship between air quality and meteorological factors is analysed from the perspective of probability. According to the calculation, the distribution of winter air quality index in Beijing at different temperature levels is shown in Table 2:

As can be seen from the data in the table above, the higher the winter temperature is, the worse the air quality is. When the air temperature belongs to grade 1 (0°C or below), the distribution of the air quality index is mainly excellent and good. No heavy pollution or serious pollution has occurred. The frequency of mild air pollution or above is about 8.91%.

When the air temperature belongs to grade 4 (9°C or above), the frequency of mild air pollution or above is 48.78%, and heavy pollution and serious pollution occur more frequently, reaching 11%.

The distribution of winter air quality index in Beijing at different wind force levels is shown in Fig. 7:

As can be seen from the data in Fig. 7 that the higher the winter wind force is, the better the air quality is. The frequency of mild air pollution and above is about 45.17% when the daily maximum wind force is level 0, and 20.89% when the daily maximum wind force is level 3 and above.

CONCLUSIONS

Information diffusion technology was applied to the study of air quality characteristics. The example shows that the information diffusion technology can make full use of the position information of the data by fuzzing the sample data and obtaining a good result of probability calculation and regression analysis. Based on the analysis of air quality in Beijing by information diffusion model, the following characteristics are found: The probability of air pollution is
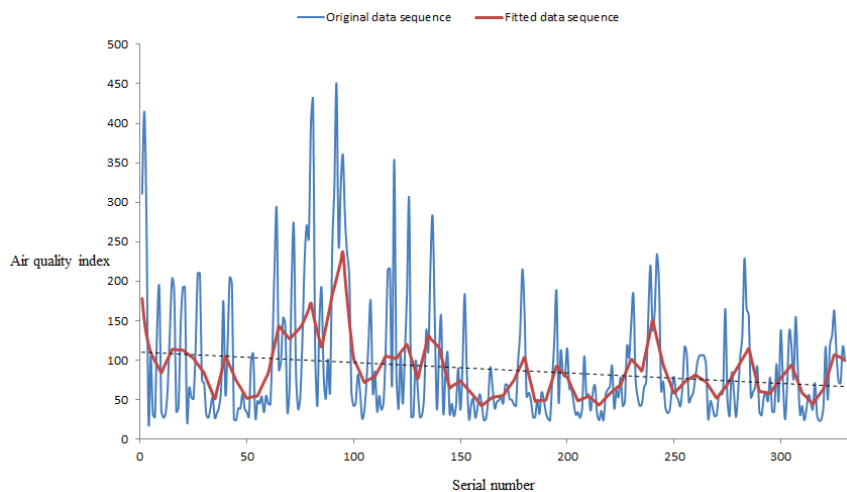

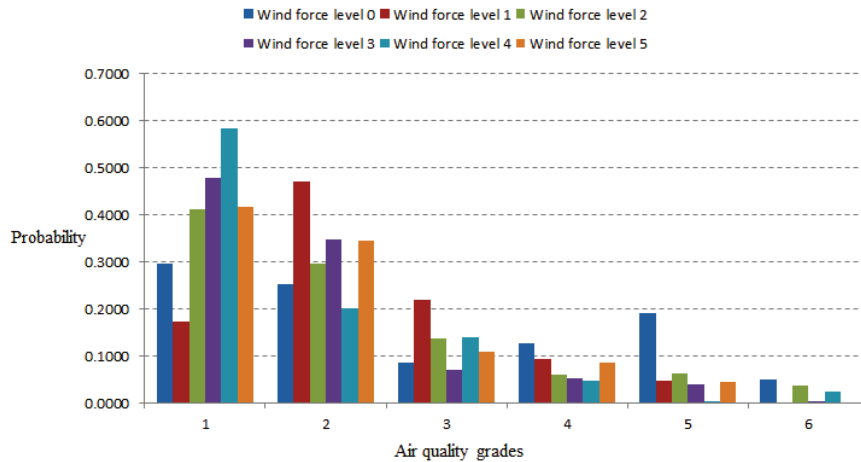Fig. 6: Trend of air quality in winter.

Fig. 7: Probability distribution of air quality under different wind force levels.

Fig. 7: Probability distribution of air quality under different wind force levels.

the highest in winter. In recent years, air quality has been improving to some extent. The air quality of Beijing in winter has a certain relationship with meteorological factors such as air temperature and wind force. The higher the air temperature is, the smaller the wind force is, the greater the probability of air pollution.

ACKNOWLEDGEMENT

REFERENCES

Chaloulakou, A., Kassomenos, P., Spyrellis, N., Demokritou, P. and Koutrakis, P. 2003. Measurements of PM 10, and PM 2.5, particle concentrations in Athens, Greece. Atmospheric Environment, 37(5): 649-660.

Chen, X.N., Jia, Q., Gao, Y., Liu, S. and Duan, C.Q. 2012. Analysis on wheat drought risk based on information diffusion technology in Xi'an city. Journal of North China Institute of Water Conservancy and Hydroelectric Power, 33(3): 20-22.

Davis, D.L., Bell, M.L. and Fletcher, T. 2002. A look back at the London smog of 1952 and the half century since. Environmental Health Perspectives, 110(12): 374-375.

Dong, M., Kuang, Y. and Kuang, Y. 2009. PM2.5 concentration prediction using hidden semi-Markov model based times series data mining. Expert Systems with Applications, 36(5): 9046-9055.

Hinton, G.E. and Salakhutdinov, R.R. 2006. Reducing the dimensionality of data with neural networks. Science, 313(5786): 504-507.

Huang, C.F. 2018. Improving estimation of typhoon risk by using information diffusion model. Systems Engineering - Theory & Practice, 38(9): 2315-2325.

Jian, L., Zhao, Y., Zhu, Y.P., Zhang, M.B. and Bertolatti, D. 2012. An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China. Science of the Total Environment, 426(2): 336-245.

Jiang, Q.J., Li, H., Wang, S.P. and Fu, C.L. 2017. Study on causes and countermeasures of haze in Zhengzhou. Guangzhou Chemical Industry, 45(16): 136-137.

Li, H.M. 2017. Prediction of haze weather based on cubic exponential smoothing model. Journal of the Environmental Management College of China-EMCC, 27(3): 52-55.

Minguillón, M.C., Querol, X., Baltensperger, U. and Prévôt, A.S.H. 2012. Fine and coarse PM composition and sources in rural and urban sites in Switzerland: local or regional pollution. Science of the Total Environment, 428(12): 191-202.

Mishra, D., Goyal, P. and Upadhyay, A. 2015. Artificial intelligence based approach to forecast PM 2.5, during haze episodes: A case study of Delhi, India. Atmospheric Environment, 102(2): 239-248.

Qian, J.P., Huang, F., Du, J., Wang, G., Jin, A. and Peng, L. 2006. Spatial-temporal characteristic of meteorological visibility under fog and haze condition in Guangdong Province. Ecology and Environment, 15(6): 1324-1330.

Wang, B.B. 2011. The analysis of the pollution characteristics and influence factors of PM2.5 in Beijing during the Olympic games. Shangdong Normal University, pp. 82-90.

Wang, J.Z., Xu, X.D. and Yang, Y.Q. 2002. A study of characteristics of urban visibility and fog in Beijing and the surrounding area. Journal of Applied Meteorological Science, 13(Suppl.): 160-169.