# Estimation of PM2.5 Spatial Distribution Based on Kriging Interpolation

## Lirong Deng

[1]School of Electrical and Electronic Engineering Department, North China Electric Power University, Baoding 071003, China.

1064843655@qq.com

**Keywords:** PM2.5 spatial distribution, kriging interpolation, Matlab.

**Abstract.** The air quality in Beijing City is getting more and more attention. Aimed at analyzing distribution of PM2.5 concentration, this study adopts thirty-two observation sites PM2.5 concentration data in six days from Beijing City Environmental Protection Bureau. By kriging interpolation method, 100×100 unknown points' concentrations have been predicted and PM2.5 spatial distribution drawings have been plotted. The results show that PM2.5 concentration takes the gradient distribution. South area is higher than north while west area is higher than east. Further, we discussed reasons for this distribution.

## Introduction

Fine particles are called PM2.5, referring to the ambient air particulate matter diameter is less than or equal to 2.5 micron. The smaller the particle diameter is, the deeper the respiratory tract it goes in. It has a direct effect on the lung ventilation function. In 2012 the environmental NGO friends of nature announced air quality rankings in China's major cities, in which Beijing ranked the bottom third[1]. In the air pollution indexes, PM2.5 pollution problem is significant in Beijing. Therefore, our research is to study Beijing PM2.5 spatial distribution in March. We screen concentration information of each PM2.5 monitoring point, plot PM2.5 distribution drawings and conduct some analysis.

## Data and Methods

**Selection of Monitoring sites and Data.** According to Beijing environmental protection monitoring center website [2], we get thirty-two monitoring spots information, including every spot geographical position, which is shown in Fig.1. We collect PM2.5 24 hours daily average concentration with every monitoring spot above during March 7th and March 13th in 2014. These mean values can be simple and effective to reflect a day in the situation of pollutant concentration.

**Methods.** Since our data is from each monitoring station, PM2.5 concentration is discrete. Spot data need to be interpolated to surface data to characterize the entire region. Commonly used spatial interpolation, such as inverse distance weighted methods, trend surface method and the Tyson polygon are difficult to analyze errors. However, geostatistical interpolation with its solid theoretical foundation in spatial statistics, not only analyzes interpolation error, but also conducts the error of point-wise estimation.

As one of main contents of geostatistics, kriging interpolation is a method of unbiased optimal estimation for regionalized variables, which is based on variation function theory and structural analysis [3]. So kriging interpolation is selected to plot PM2.5 distribution drawings

**Kriging Estimators.** In our study area, there are 32 measured points. We assume that these data obey normal distribution. For any unknown point to be estimated, its estimator is expressed as a linear combination of those effective sample values.

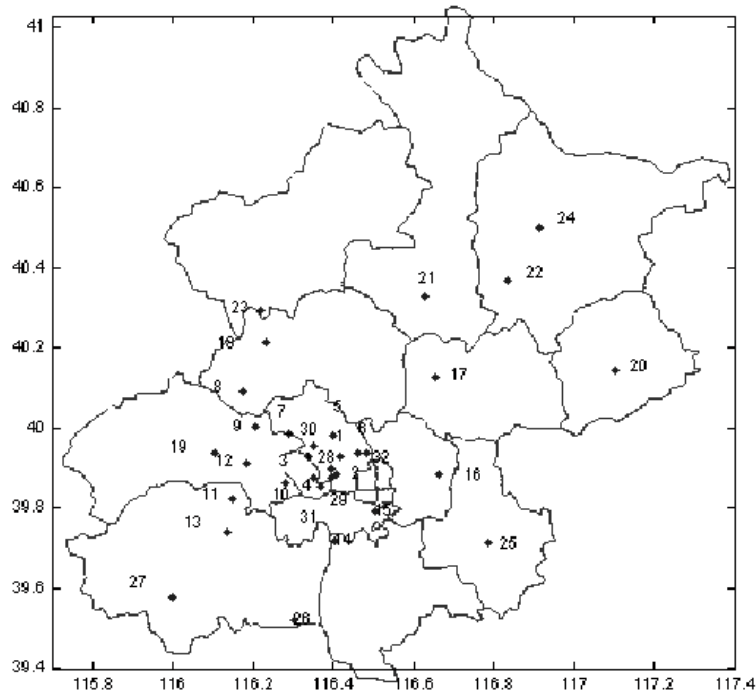$$\hat{Z}(s_i) = \sum_{j=1}^{32} \lambda_j Z(s_j) \tag{1}$$

Fig.1 Observation sites of PM2.5 concentration in Beijing

where $s_i$ is predicted position; $Z(s_i)$ is measured value of $i^{th}$ point; $\lambda_j$ is unknown weight of $Z(s_j)$. $\lambda_j$ depends on the spatial relationship fitting model between $s_i - s_j$ distances and measured values of $s_i$. In order to ensure the model is unbiased estimation,

$$\sum_{j=1}^{32} \lambda_j = 1. \tag{2}$$

$\hat{Z}(s_i)$ could be calculated under the condition of ensuring the kriging variance minimum.

**Semivariogram Function Model.** In the space modeling of the measured points, variance analysis begins with empirical semivariogram figure[4]. The formula is

$$S(d_h) = 0.5 \times \frac{1}{N} \sum_{j=1}^{N} \{Z(s_i) - Z(s_j)\}^2 \tag{3}$$

where $h$ is step column unit. $d_h$ is average distance between predicted point and measured points in column $h$; $N$ is the number of measured points in column h. In our algorithm, we set $h$ to $0.05^\circ$. Thus, there are 26 column units from $116^\circ$ to $117.2^\circ$. Figure 2 shows the empirical semivariogram function.
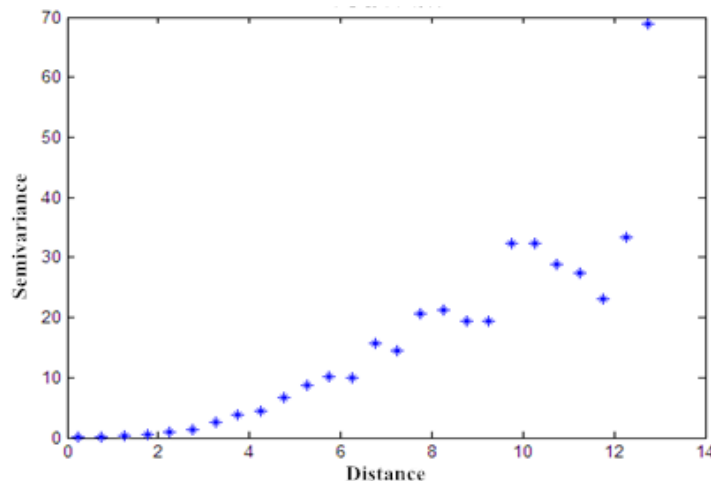


Fig.2 Semivariogram function.

A fitting model needs to be decided in accordance with the semivariogram function. After compare several function models, such as circular, spherical, exponential, Gaussian and linear, we choose Gaussian model as fitting curve, shown in Fig. 3. From the image, our semivariogram function has the very good data fitting relations with Gaussian function.
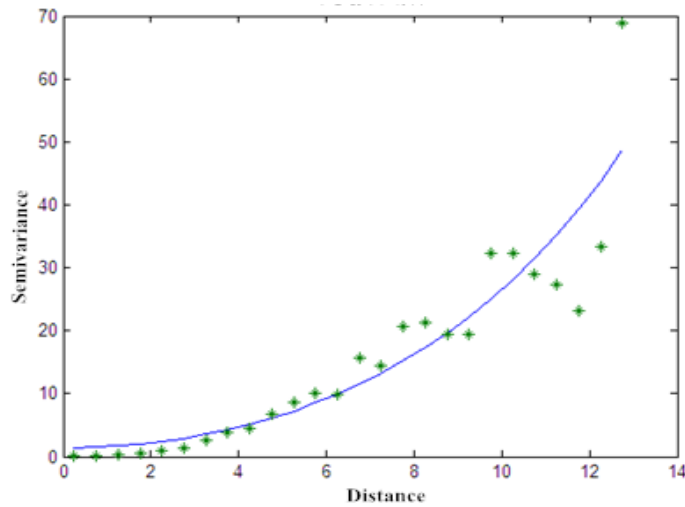


Fig.3 Semivariogram function with Gaussian fitting

## Algorithm Design

The algorithm was designed by MATLAB, and detailed process has shown below.

Step 1. Calculate each column unit semivariogram values according to formula (3);

Step 2. Fitting semivariogram curve with Gaussian function.

Step 3. For each unknown point $(x_i, y_i)$, solve $dis_{ij}(j = 1,2,\cdots,32)$. $dis_{ij}$ means distance from unknown point $i$ to measured point $j$. Substitute $dis_{ij}$ in the fitting semivariogram curve and get $sem_{ij}$.
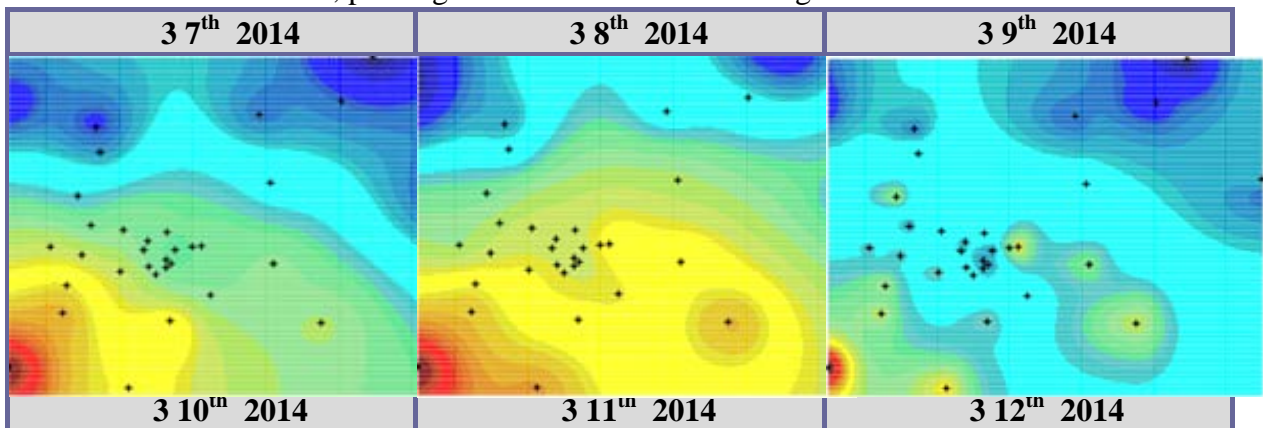
Step 4. Solve equations:

$$sum_i = \sum_{j=1}^{32} \frac{1}{sem_{ij}} \tag{4}$$

$$\lambda_{ij} = \frac{1}{sum_i \times semi_{ij}} \tag{5}$$

Step 5. Get $Z(s_i)$ through Eq (1).

After substituting data from 32 observation points to our algorithm, we predict $100 \times 100$ unknown points PM2.5 concentration, plotting PM2.5 distribution drawings from March $7^{th}$ to March $12^{th}$.
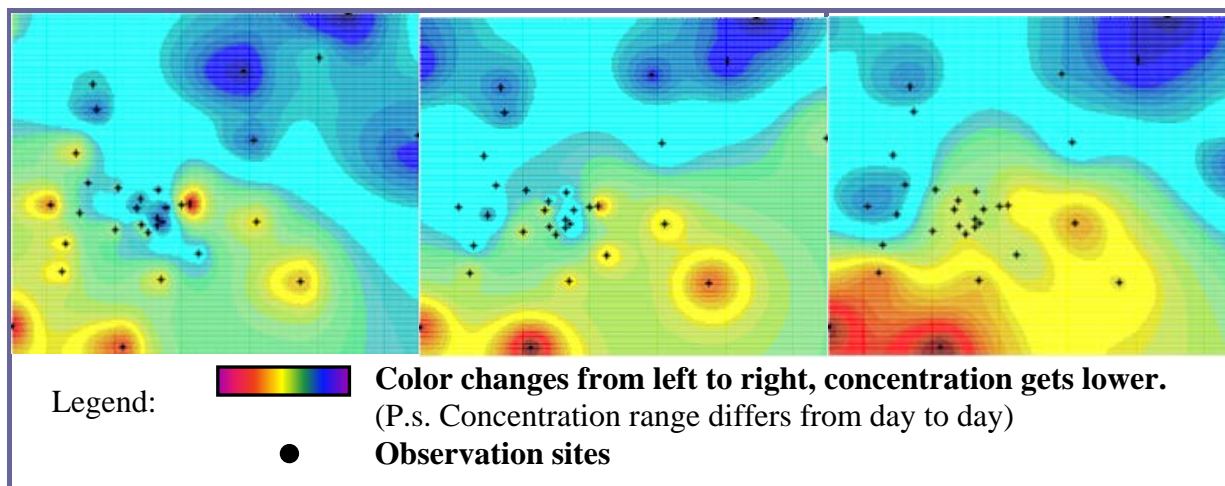
Fig. 4 PM2.5 spatial distribution in Beijing

## Summary

In this study, kriging interpolation has been used to make PM2.5 spatial concentration in Beijing. From the view of Fig.3, we get a good Gaussian fitting of semivariogram function, which proves that our model is reliable. And we can draw some conclusions as follows:

From the overall point of view, PM2.5 concentration takes the gradient distribution. South area is higher than north while west area is higher than east. Specifically speaking, the southeast and southwest borders have highest value, shown in the picture with color red and orange; the south and urban have high value, shown with yellow and green; north area is overall good with color blue.

Compared to the corresponding weather from March 7th to March 12[th][5], we find that during these days, weather changes from day to day, such as sunny, haze, drizzle, cloudy, etc. However, no matter what weather it is, the PM2.5 concentration distribution looks like the same.

The reason why Beijing has such king of spatial distribution is various factors influence jointly, mainly topography, wind direction, discharge and atmospheric stability [6]. Beijing is surrounded by mountains on three sides and monsoon climate is significant, so the mountain valley wind forms. The hilly areas have undulating terrain, which is conducive to form wind gap and makes the particles quickly dispersed. Therefore, this kind of topography and climate are reasons that cause differences between the north and the south.

## References

[1] Junhong Yu. PM2.5 pollution in Beijing city governance countermeasure research. MS., Capital University of Economics and Business, China 2013. p.1.

[2] Information on: http://www.bjmemc.com.cn

[3] Qing Ji, Ming Yu. Parameters selection of the annual mean temperature spatial interpolation method based on collaborative kriging interpolation. Journal of Capital Normal University. (2010) No. 3, p.81-87.

[4] Information on: http://blog.sina.com.cn/s/blog_4aa4593d01014gwc.html

[5] Information on: http://weather.mipang.com/beijing/3yuefen.html

[6] Wenhong Bao, Li Wang, Li Sun, etc. Analysis of influencing factors of atmospheric pollutants diffusion and impact trend. Baosteel Technology. (2002): p.72-73.