Study on Spatial and Temporal Distribution Characteristics of PM$_{2.5}$ in the Main Urban Area of Nanjing and Influencing Factors

Zhao Jindong$^1$, Yang Li$^2$* and Ren Haiyang$^3$

$^1$School of Economics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
$^2$School of Geography and Bioinformatics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

$^*$Email: 1219125020@njupt.edu.cn

Abstract. The study used multiple linear regression models and the inverse distance weight interpolation method to construct a regression analysis model of PM$_{2.5}$ concentration influencing factors based on the measured data of 9 PM$_{2.5}$ concentration monitoring points in Nanjing from 2012 to 2019, in order to clarify the spatial and temporal distribution characteristics of PM$_{2.5}$ pollution in the main urban area of Nanjing. The results show that industrial land area, arable land area, and the density of food and beverage outlets all contribute to PM$_{2.5}$ concentration, while forest land area has a certain inhibitory effect on the increase of PM$_{2.5}$ concentration; simulation results show that PM$_{2.5}$ concentration in Nanjing's main urban area is decreasing year by year, and the spatial distribution of PM$_{2.5}$ is primarily influenced by the density of food and beverage outlets. The modeling findings reveal that PM$_{2.5}$ concentrations in Nanjing's main urban region are decreasing year by year, with a "low north and high south" pattern dominating the spatial distribution.

1. Introduction

The natural environment is the foundation for human survival and growth. Human beings have put great pressure on the natural environment in recent years as a result of accelerating industrialization and urbanization, which has had a greater impact on ambient air quality [1]. Since the end of 2011, hazy weather has continued to develop in most areas of China, substantially affecting regular production and life, and PM$_{2.5}$, the principal pollutant in the haze, has increasingly reached the public eye.

Scholars in the China and abroad have conducted extensive research and analysis on PM$_{2.5}$ in terms of spatial and temporal distribution characteristics [2-4], transport paths [5-6], chemical composition [7-9], health risks [10-11], influencing factors [12-13], and concentration prediction [14-16], and other topics, resulting in a wealth of research findings. Among the studies on the factors influencing PM$_{2.5}$ concentrations in Nanjing, meteorological factors [17-18] were studied more than natural and socioeconomic factors; among the studies on the spatial and temporal distribution characteristics of PM$_{2.5}$, the majority of them only analysed data from air quality monitoring stations directly to obtain the distribution characteristics of PM$_{2.5}$ [19]. However, because of the uneven distribution of air
quality monitoring points in Nanjing, many regions lack accurate PM$_{2.5}$ monitoring data, so it is difficult to directly obtain the spatial distribution characteristics of PM$_{2.5}$ concentrations in Nanjing.

Nanjing is an important central city and comprehensive transportation hub in eastern China, but high PM$_{2.5}$ concentration has become a prominent problem of air pollution in the city. This study takes into account both natural and socioeconomic factors, first establishing a multiple linear regression model to investigate the influencing factors of PM$_{2.5}$ concentrations, and then using the inverse evolution method and inverse distance weight interpolation, analyzing the spatial and temporal distribution characteristics of PM$_{2.5}$ concentrations in Nanjing's main urban area. It also suggests a new direction for PM$_{2.5}$ research in the future.

2. Data and methods

2.1 Study area
Nanjing is the capital of Jiangsu Province, which is located in China's southeast and is part of the Yangtze River Delta. Nanjing's climate is humid north subtropical, with four distinct seasons and plenty of rain. Nanjing has a total size of 6,582.31 square kilometers, with the urban region accounting for more than 70% of it. Nanjing's topography is dominated by low mountains and hills, with ring-shaped mountains, striped mountains, and skip-shaped basins as the city's key features, with an overall topography of "three sides encircled by mountains and one side by water." [20] According to the 2019 Nanjing Environmental Status Bulletin, the number of days in 2019 when the ambient air quality in Nanjing's built-up region exceeded the secondary level was 255, with a 69.9 percent attainment rate it did not meet the secondary standard was 110. (among which, 97 days were lightly polluted, 12 days were moderately polluted, and one day was heavily polluted).

Due to the uneven distribution of air quality monitoring stations in Nanjing, the study area of this paper is the eight districts in the main city of Nanjing, including Gulou District, Xuanwu district, Qinhuai district, Jianye district, Qixia district, Yuhuatai district, Pukou district, and Jiangning district, to ensure the accuracy of the analysis results.

2.2 Data collection
The study took into account a wide range of environmental and socioeconomic elements that could affect PM$_{2.5}$ [21-23], such as the area of various land types (industrial, forest, construction, grassland, water bodies, and arable land), urban traffic, and restaurants.

Nanjing Environmental Monitoring Center Station provided PM$_{2.5}$ air quality data. There are nine air quality monitoring sites in Nanjing, including Xianlin University City, Xuanwu District, Zhonghua Gate, Ruijin Road, Maigaoqiao, Shanxi Road, Pukou, and Olympic Sports Center. This study collected daily hour-by-hour data for the nine monitoring sites from 2012 to 2019.

PM$_{2.5}$ concentrations in cities are inversely proportional to the distribution of land-use types. To see if the layout of land use influences PM$_{2.5}$ concentrations in Nanjing's main urban area, we used Landsat8 remote sensing images from the geospatial data cloud from 2014 to 2019 and a combination of supervised classification and visual interpretation through ArcGIS10.2 and ENVI5.3 software to calculate the areas of various land use types within the buff.

Urban traffic is also a major source of PM$_{2.5}$. In this study, we selected to represent the influence of roads on PM$_{2.5}$ concentrations by using the entire length of roads within a given range of the monitoring station. Restaurant emissions are also a significant source of urban air pollution. The density of restaurant sites within the monitoring site's buffer zone indicates the effect of restaurants on PM$_{2.5}$ concentrations.

2.3 Research method

2.3.1 Multiple linear regression model. Multiple linear regression models are typically used to investigate the relationship between changes in an explanatory variable dependent on multiple
explanatory variables; if the relationship between the two can be represented in a linear form, multiple linear models can be developed for analysis. The environmental and social elements that may influence PM\(_{2.5}\) concentration were thoroughly investigated in this study, and the explanatory variables were first standardized using SPSS23.0 software. Then a multiple linear regression model with PM\(_{2.5}\) concentration was created. The multiple regression linear model takes the following general form.

\[
Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{m-1} X_{m-1} + \varepsilon
\]  

In the formula, \(Y\) is the dependent variable, \(X_1, X_2, \ldots, X_{m-1}\) is the explanatory variable, \(\beta_0, \beta_1, \beta_2, \ldots, \beta_{m-1}\) represents the coefficient corresponding to the explanatory variable, \(\varepsilon \sim N(0, \sigma^2)\) represents random error [24].

2.3.2 Inverse distance weighted interpolation. The inverse distance weight interpolation (IDW) approach is based on Tobler's theorem. It works based on finding the weighted average of the measured values of each point near the unmeasured end to interpolate. The spatial autocorrelation principle states that the closer entities or phenomena are in space, the bigger their weights obtained at the nearest point are. As a result, IDW has substantial dependence on spatial location in the proximity range for interpolation error. The general phrase is as follows.

\[
Z_0 = \frac{\sum_{i=1}^{n} z_i / d_i^r}{\sum_{i=1}^{n} 1 / d_i^r}
\]  

\(Z_0\) represents the estimated value of point \(O; Z_i\) represents the value of control point \(i\); \(d_i\) represents the distance between \(Z\) of control point \(i\) and point \(O\); \(n\) represents the number of control points employed in the estimation; and \(r\) represents the given Power [25].

To acquire the geographical distribution of PM\(_{2.5}\) concentrations in Nanjing's main urban area, the PM\(_{2.5}\) concentrations at unmonitored places were first predicted using a regression model, and then the PM\(_{2.5}\) concentration data were analyzed using the inverse distance weight interpolation method.

3. Results and discussion

3.1 PM\(_{2.5}\) concentration modeling and model verification

3.1.1 Bivariate correlation analysis. To analyze the relationship between each influencing factor and PM\(_{2.5}\) concentration in this study, the correlation analysis of each influencing factor with PM\(_{2.5}\) concentration was performed independently using SPSS23.0 software. The results were calculated as follows.

\(\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
\text{Impact factor} & \text{Water area} & \text{Forest area} & \text{Industrial land area} & \text{Cultivated area} & \text{Grassland area} & \text{Construction land area} & \text{City traffic} & \text{Density of participation} \\
\hline
r & 0.012 & -0.590 & 0.703 & 0.385 & -0.009 & 0.334 & 0.449 & 0.524 \\
P & 0.944 & 0.000 & 0.000 & 0.021 & 0.960 & 0.046 & 0.006 & 0.000 \\
\hline
\end{array}\)

According to Table 1, among the influencing factors, the area of forest land, the area of industrial land, the area of cultivated land, and the area of construction land are significantly correlated with PM\(_{2.5}\) concentration, as are urban traffic and the density of food and beverage outlets. In contrast, the area of water bodies and grassland do not reach significant levels. Among the influencing factors that were significantly correlated with PM\(_{2.5}\) concentration, industrial land area, arable land area, construction land area, urban traffic, and density of food and beverage outlets were all positively correlated with PM\(_{2.5}\) concentration, with the industrial land area having the highest correlation; forest land area was negatively correlated with PM\(_{2.5}\) concentration. The influencing elements that did not
achieve a substantial connection with PM$_{2.5}$ concentration were excluded based on the above screening criteria of significant correlation.

### 3.1.2. Model construction and analysis

Stepwise linear regression was performed using SPSS23.0 software between the influencing factors and PM$_{2.5}$ concentration, and stepwise linear regression can effectively avoid multicollinearity between independent variables [26], with the following model.

$$Y = -68.677 - 0.317X_1 + 8.364X_2 + 2.075X_3 + 0.117X_4$$  \hspace{1cm} (3)

In the formula, $X_1$, $X_2$, $X_3$, $X_4$ denote the area of forest land, area of industrial land, area of cultivated land and density of food and beverage sites within the 5000 m buffer zone of the monitoring site, respectively, and the parameters and fits of the model are as follows.

#### Table 2. Results of regression model

| model | $R^2$ | Adj-$R^2$ | $F$  | $P$  |
|-------|-------|-----------|------|------|
|       | 0.727 | 0.692     | 20.622 | 0.000 |

#### Table 3. Regression model coefficient table

|                          | $\beta$  | SE     | $t$    | $P$    | VIF |
|--------------------------|----------|--------|--------|--------|-----|
| Constant                 | -68.677  | 22.753 | -3.018 | 0.005  |     |
| Forest area              | -0.317   | 0.068  | -4.690 | 0.000  | 2.747|
| Industrial land area     | 8.364    | 1.626  | 5.143  | 0.000  | 3.079|
| Cultivated area          | 2.075    | 0.490  | 4.239  | 0.000  | 2.870|
| Density of dining spots  | 0.117    | 0.117  | 4.651  | 0.000  | 3.427|

According to Table 2, the regression model’s adjusted R-squared is 0.692, indicating a good fit. The regression equation and coefficients passed the test at the significance level of 0.05, and the regression model did not exhibit multicollinearity.

The coefficients of each explanatory variable in the regression model show that woodland area has a suppressive effect on PM$_{2.5}$ concentration, i.e., an increase in the woodland area will lead to a decrease in PM$_{2.5}$ concentration; the coefficients of industrial land area, arable land area, and restaurant point density are positive, indicating that the density of catering points increases the concentration of PM$_{2.5}$; according to the regression coefficients, the influence of industrial land area on PM$_{2.5}$ concentration is more significant than that of cultivated land area.

According to Table 3, the coefficients of each explanatory variable in the regression model show that woodland area has a suppressive effect on PM$_{2.5}$ concentration, i.e., an increase in the woodland area will lead to a decrease in PM$_{2.5}$ concentration; the coefficients of industrial land area, arable land area, and restaurant point density are positive, indicating that the density of catering points increases the concentration of PM$_{2.5}$; according to the regression coefficients, the influence of industrial land area on PM$_{2.5}$ concentration is more significant than that of cultivated land area.

According to the collected data, the area of industrial land and the area of cultivated land in Nanjing both decreased to some extent from 2012 to 2019, with the area of industrial land decreasing more and the density of restaurant sites fluctuating only slightly; the area of forest land in Nanjing increased to a greater extent. As a result, in recent years, the reduction of industrial and arable land areas, as well as the development of forest land area in Nanjing's main urban area, has played a significant role in lowering PM$_{2.5}$ concentrations.

### 3.2 Simulation of the spatial distribution of PM2.5 concentration

The model was cross-validated to further verify its predictive capacity, and the cross-validation scatter plot obtained is given in Figure 1.
According to Figure 1, the overall multiple regression linear model has a high prediction accuracy, and the mean square error between predicted and measured PM$_{2.5}$ concentrations at monitoring points is 6.27$\mu$g/m$^3$, with a prediction accuracy of 85.67 percent, indicating that PM$_{2.5}$ concentrations at unmonitored points can be simulated using this model.

The goal of this research is to look at the spatial distribution of PM$_{2.5}$ in Nanjing's main urban area, and the data used are PM$_{2.5}$ concentration values from nine monitoring points in Nanjing between 2012 and 2019. Because of the uneven distribution of monitoring stations, additional data processing is required to reduce errors. In order to analyze the spatial distribution of PM$_{2.5}$ concentrations in the main urban area of Nanjing, 29 unmonitored points covering basically the entire main urban area of Nanjing were selected and the PM$_{2.5}$ concentration data of these 29 points were simulated according to the regression model already fitted above. Subsequently, the PM$_{2.5}$ concentration data from these selected points and the PM$_{2.5}$ concentration monitoring points in Nanjing were interpolated using inverse distance weights to estimate the spatial distribution of PM$_{2.5}$ concentrations in the main urban area of Nanjing.

The spatial distribution of PM$_{2.5}$ concentration was calculated using ArcGIS10.2 software's "inverse distance weight interpolation" approach.
According to Figure 2, in terms of time, the PM$_{2.5}$ concentration in Nanjing's main urban region has decreased from 2014 to 2019. In 2014, the average PM$_{2.5}$ concentration in Nanjing's major urban area reached 78.22 μg/m$^3$, much exceeding the national level concentration limit; in 2016, the PM$_{2.5}$ concentration in Nanjing reduced dramatically compared to 2014. In general, the PM$_{2.5}$ concentration in Nanjing's main urban area has declined significantly in recent years. The difference between PM$_{2.5}$ concentrations in various locations has gradually narrowed, indicating that the air environmental quality in Nanjing's main urban area has improved.

The distribution of PM$_{2.5}$ concentrations in Nanjing's main urban area follows a pattern of "low in the north and high in the south." The locations with the highest concentrations are primarily centered in Jiangning District's south, Pukou District's south, and some areas in the city center. The high concentration of PM$_{2.5}$ in Jiangning and Pukou districts is primarily due to the concentration of factories and arable land in these locations. Although the distribution of industrial areas and arable land in the downtown region is not concentrated, the higher density of food and beverage sites, as well as the more concentrated distribution of population, lead to higher PM$_{2.5}$ concentrations in the area. PM$_{2.5}$ concentrations are lower in locations where woodlands are more densely distributed, such as Zhongshan Scenic Area, Tangshan Scenic Area, and Laoshan National Forest Park. This is because forest vegetation purifies the air. The air pollutants created in these places are significantly lower than in other areas due to fewer human activities and a greater distance from huge chemical facilities. In summary, the distribution of industrial regions, arable land, green spaces, and the density of restaurant sites is directly associated with PM$_{2.5}$.

4. Conclusions and recommendations

The regression model better fit the variation in PM$_{2.5}$ concentration in Nanjing's main urban region. According to the regression model's coefficients, it was discovered that the amount of industrial land, the area of cultivated land, and the density of restaurant sites were the key factors leading to an increase in PM$_{2.5}$ concentration. In contrast, forest land had a slight inhibitory effect.

The spatial distribution of PM$_{2.5}$ concentration in Nanjing's major urban area follows the pattern of "low in the north and high in the south." It is closely related to the distribution of industrial land, arable land, green land, and restaurant density.

From a chronological standpoint, the decreasing trend of PM$_{2.5}$ concentration in the main urban area of Nanjing from 2012 to 2019 indicates that air pollution in the main urban area of Nanjing has been treated more effectively. The reduction of industrial land area and arable land area and the increase of forest land area have all played an essential role in the decrease of PM$_{2.5}$ concentration in the main urban area of Nanjing.
References

[1] DENMAN A R, CROCKETT R, GROVES KIRKBY C, et al. The value of seasonal correction factors in assessing the health risk from domestic radon—a case study in Northamptonshire, UK [J]. Environment International, 2007,33(1):34-44.

[2] Huang Yalin, Liu Chao, Zeng Kefeng, et al. The temporal and spatial distribution characteristics of PM$_{2.5}$ in Wuhan from 2013 to 2014 and its relationship with meteorological conditions [J]. Journal of Eco-Environment, 2015,24(08):1330-1335.

[3] Chen Youliang, Tao Tianhui, Liu Xinyi. Temporal and Spatial Distribution of Air Quality in Urban Agglomeration around Poyang Lake [J]. Environmental pollution and prevention, 2018,40(12):1418-1424.

[4] Hu Wen, Yang Guangbin, An Hongfeng, et al. Study on the spatial interpolation and distribution of atmospheric PM$_{2.5}$ and PM10 concentrations in Guiyang in winter [J]. Environmental pollution and prevention, 2019,41(05):579-583.

[5] Yang Junchao, Chen Zhiming, Zhi Guorui, et al. Analysis on Transmission Path and Source of PM$_{2.5}$ in Winter in Baise City [J]. Environmental monitoring management and technology, 2020, (01):1-7.

[6] Duan Shiguang, Jiang Nan, Yang Liuming, et al. Analysis of Transmission Path and Potential Source of Atmospheric PM$_{2.5}$ in Zhengzhou City in Winter [J]. Environmental science, 2019,40(01):86-93.

[7] Zhang Yiqin, Wang Jie, Gao Jian, et al. Characteristics and source analysis of PM$_{2.5}$ components in the eastern part of the North China Plain during the heating period from 2016 to 2017 [J]. Environmental science, 2019,40(12):5202-5212.

[8] Wu Ming, Wu Dan, Xia Junrong, et al. Pollution characteristics and source analysis of PM$_{2.5}$ chemical components in Chengdu in winter [J]. Environmental science, 2019,40(01):76-85.

[9] Chen Quan, Li Can, Wei Xiaoqing, et al. Study on Characteristics of Indoor PM$_{2.5}$ Pollution Components and Health Risks in Zhuzhou City in Winter [J]. Environmental pollution and prevention, 2019,41(09):1088-1093.

[10] Chen Zhangqi. Pollution characteristics, sources and health risk assessment of PAHs in atmospheric PM$_{2.5}$ in Quanzhou [J]. Earth and environment, 2019,47(03):275-282.

[11] Fu Conghui, Wang Wenjun, Tang Jian, et al. Research on the Spatial Population Distribution of PM$_{2.5}$ Health Risks——Taking Shenzhen as an Example [J]. China Soft Science, 2014(09):78-91.

[12] Sun Jianru, Zhong Yun. Economic Analysis of the Factors Influencing PM$_{2.5}$ in Big Cities in my country——An Empirical Study Based on City-level Panel Data [J]. Ecological Economy, 2015,31(03):62-65+77.

[13] Zhang Shuping, Han Lijian, Zhou Weiqi, et al. Analysis of meteorological factors affecting PM$_{2.5}$ in winter [J]. Acta Ecologica Sinica, 2016,36(24):7897-7907.

[14] Liu D,Sun K.Short-term PM$_{2.5}$ forecasting based on CEEMD-RF in five cities of China.[J].Environmental science and pollution research international,2019,26(32).

[15] Park Y,Kwon B,Heo J,et al.Estimating PM$_{2.5}$ concentration of the conterminous United States via interpretable convolutional neural networks.[J].Environmental pollution (Barking,Essex :1987),2020,256.

[16] Marsha A,Larkin N K.A statistical model for predicting PM$_{2.5}$ for the western United States[J].Journal of the Air & Waste Management Association,2019,69(10).

[17] Huang Jun, Guo Shengli, Wang Xi. Analysis of pollution characteristics and influencing factors of PM$_{2.5}$ and PM$_{10}$ in winter and spring in Nanjing [J]. Environmental Engineering, 2015,33(12):69-74.

[18] Wan Xin, Wang Lei, Zhang Yanan, et al. The influence of meteorological factors on the mass concentration of atmospheric PM$_{2.5}$ in the southern suburbs of Nanjing [J]. Jiangsu Forestry Science and Technology, 2015,42(06):1-4+18.
Zhang Wei, Guo Shengli, Shen Fuzhen, Luo Qi. The distribution characteristics of PM$_{2.5}$ and PM$_{10}$ concentration in Nanjing area and the relationship with relevant meteorological conditions [J]. Science Technology and Engineering, 2016, 16(04): 124-129.

Han Xiuxiu. Distribution characteristics and source analysis of atmospheric PM$_{2.5}$ and its components in Nanjing area [D]. Nanjing Normal University, 2018.

Yu Jing, Shang Erping. Spatial Correspondence of PM$_{2.5}$ Concentrations of Main Land Types under the Rapid Urban Development——Taking Shenyang as an Example [J]. Urban Development Research, 2013, 20(09): 128-130+144.

Li Xiaoyu, Wu Lin, Zou Chao, et al. Analysis of vehicle emission characteristics on main and secondary arterial roads in Nanjing based on real-time traffic data [J]. Environmental Science, 2017, 38(04): 1340-1347.

Liu Mengxiong, Wang Yin. Research on the relationship between the distribution of restaurants and the spatial distribution of PM$_{2.5}$ in Xi’an [J]. Environmental Protection and Circular Economy, 2017, 37(10): 52-55.

Len Jianfei, Gao Xu, Zhu Jiaping. Application of multiple linear regression statistical forecasting model [J]. Statistics and Decision, 2016(07): 82-85.

Liu Guangmeng, Wang Yunjia, Wang Yun. Analysis of the influence of inverse distance weight interpolation factor on interpolation error [J]. China Science and Technology Paper Online, 2010, 5(11): 879-884.

Xu Gang, Jiao Limin, Xiao Fengtao, Zhang Xin. Land use regression model to simulate the spatial distribution of PM2.5 concentration in Beijing-Tianjin-Hebei [J]. Arid Land Resources and Environment, 2016, 30(10): 116-120.