SPATIOTEMPORAL ANALYSIS METHOD OF URBAN ENVIRONMENTAL FACTORS ALONG STREETS CONSTRAINED BY ROAD NETWORK

Lujin Hu [1]*
[1]Beijing University of Civil Engineering and Architecture, School of Geomatics and Urban Spatial Informatics
Beijing ,China - hulujin@bucea.edu.cn

KEY WORDS: Spatiotemporal Prediction, Road Network Constraints, Urban Environmental Factors, Kriging Interpolation, Geographically Weighted Regression.

ABSTRACT:

Since people and vehicles in the city are mostly concentrated in the area along the road, there are few researches on the spatiotemporal analysis of environmental factors in the street area. This paper mainly focuses on the spatial and temporal analysis theory of environmental factors based on geographically weighted regression model taking PM2.5 as an example, breaking through the temporal and spatial analysis method of environmental factors along the street constrained by the road network, a spatiotemporal analysis and prediction based on the weighted impact of the road network buffer area and neighboring stations is proposed. Taking the distribution of PM2.5 in Beijing as an example, an experiment was conducted to analyze the spatial and temporal characteristics of PM2.5 along the street to verify the accuracy and reliability of the method proposed in this paper. Further improve the geospatial scale of the spatiotemporal analysis of environmental factors to achieve more refined spatiotemporal prediction of environmental factors.

1. INTRODUCTION

Environmental factors are the core elements of urban environmental problems which generally refer to the basic units that constitute the next-level environmental configuration such as temperature, precipitation, humidity, and wind speed are all climatic factors, as well as abiotic environmental factors. Most of the environmental factors studied by urban middle school scholars mainly include air quality and weather conditions that are closely related to people's lives. The urban streets are often more polluted than other areas due to the pollutants produced by transportation and people's production and life. Therefore, the spatiotemporal analysis of environmental factors of urban streets is particularly important.

The research on air quality issues mainly focuses on the two aspects of time dimension and space dimension for research and prediction:

Study the law of air quality changes from different temporal and spatial scales: for example, from the national scale (Xiao Yue et al., 2017; Wang Zhenbo et al., 2015), using a few areas as the research scope (Li Wenjie et al., 2012), Focusing on the study of developed regions (Wu Ying et al., 2011), and using typical cities as examples (Li Xiangyang et al., 2011; Huang Yalin et al., 2015; Zhang Hong et al., 2020) and other regional scales to study the characteristics of regional air quality changes ; and study the temporal and spatial distribution characteristics of air quality in different regions on the time scale of multi-year variation (Huang Yalin et al., 2015; Yizaitiguli and Waili et al., 2020; Wu Yiling et al., 2020).

Use different data and different methods to explore the temporal and spatial characteristics of air quality in China: GIS technology is used to explore the temporal and spatial distribution of air quality across the country and the Pearl River Delta region (Liu Yongwei, 2013; Zhang Baochun, 2011), Therefore, it is found that the air pollution in my country has the characteristics of seasonal variation and spatial convergence; There is also a time series-based processing model that finds that the air pollution in my country's cities also has the characteristics of regional distribution; and based on the gray correlation model, it is found that some provinces and cities in my country also have the characteristics of spatial distribution.

To sum up, domestic and foreign researchers have made some achievements in air quality and conducted in-depth research on its prediction. Therefore, this paper will take the air quality of Beijing as an example, use 34 collection sites in Beijing, select daily data in the selected range, and add road network constraints to conduct spatiotemporal analysis of environmental factors.

2. STUDY AREA

Taking the air quality of Beijing as an example, this paper studies the refined spatiotemporal analysis of urban street environmental factors under the constraints of the road network. The research area of this paper takes the entire area of Beijing as the research area, and mainly collects the surrounding areas of 34 monitoring stations provided by the Beijing Air Quality Historical Data Website and the collected data as the research source to conduct analysis experiments mainly focusing on factors such as PM2.5. The study area and 34 monitoring sites are shown in Figure 1.

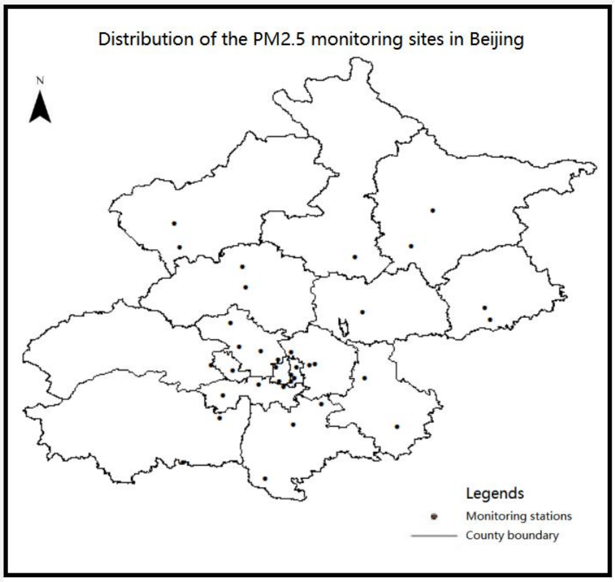

Figure 1. Location of 34 monitoring sites in Beijing

3. RESEARCH METHODS AND PROCESSES

3.1 Basic Method

3.1.1 Spatial interpolation method of urban environment factor in Kriging: In this paper, since there are fewer open sites of environmental factors in Beijing, it is necessary to use the spatial interpolation method to obtain the distribution of environmental quality in the entire Beijing area. Therefore, spatial interpolation is the core and foundation of this paper for analysis in spatial dimensions.

Spatial interpolation is often used to convert discrete point measurement data into continuous data surfaces by interpolation methods for comparison with the distribution patterns of other spatial phenomena, with two algorithms: spatial interpolation and spatial extrapolation. Mainly from the sampling point data to the entire research involved area and the application of predicting unknown samples with known samples. The spatial interpolation algorithm is to infer the unknown point data of the same area through the data of known points; The spatial extrapolation algorithm is to infer the data of other regions from the data of known regions.

There are many methods of spatial interpolation, but each of them has more or less unsolvable problems, such as boundary interpolation method needs to have a certain law and is at the boundary, trend surface method requires data to be linear, regression model requires multiple parameter variables and other methods that require constraints such as a large number of data points to support, etc., and the problems of these methods cannot be solved well, in order to solve these problems, there are French geographical mathematicians and South African mining engineers who have studied an optimized interpolation method. This method fully absorbs the idea of geostatistics, takes spatial statistics as the theoretical basis, mainly focuses on the determination of weight coefficients, overcomes the problem that the error in interpolation is difficult to analyze, so that the interpolation function is in the best state, because it provides the best linear unbiased prediction of the variable values at a specific point, and uses random surfaces for better description, so it is widely used in various predictive mapping and other fields, becoming an important part of GIS software geostatistical interpolation. This is Kriging interpolation.

There are two commonly used kriging methods: ordinary kriging and pan-kriging. Ordinary kriging is the most common and widely used method of kriging interpolation, and it is also the default method.

Pan-kriging assumes that there is a trend of coverage in the data, for example, a prevailing wind that can be modeled by a deterministic function (polynomial). The polynomial is subtracted from the original measurement point, and the autocorrelation is modeled by random error. After fitting the model with random errors, the polynomial is added back to the prediction to produce meaningful results before making the prediction. It should only be used if you understand that there is a trend in your data and can provide scientific judgment to describe pan-kriging.

For ordinary kriging, let the observed value of the regional variation over a series of sample points be $Z(x_1), Z(x_2), \ldots\ldots, Z(x_n)$, the estimate of a grid point in the area, as shown in (1):

$$Z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{1}$$

In this formula, the values $Z(x_0)$ for the points to be fixed, n for the number of measurements, and n for the observations of the sample points are the weights, as shown in the system of equations (2):

$$\begin{cases} \sum_{i=1}^{n} \lambda_i \gamma(x_i, x_j) - \mu = \gamma(x_i, x_0) \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{2}$$

Among them, the covariance $\gamma(x_i, x_j)$ between the sample points of the station, the covariance $\gamma(x_i, x_0)$ is between the sample points of the station and the interpolation point, and the μ is Lagrange multiplier. The spatial structure of the difference data is described by the semivariogram, and its expression is shown in (3):

$$\nu(\hbar) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left(Z(x_i) - Z(x_i + h)\right)^2 \tag{3}$$

Among them, for the number of experimental data pairs divided by distance segments, according to the characteristics of the experimental variogram, the appropriate theoretical variogram model is selected, and the experimental variogram graph obtained according to the experimental semivariogram is determined, so as to determine a reasonable theoretical model of the variogram.

The Kriging interpolation method takes into account the position relationship between the observed point and the estimated point, and also takes into account the relative position relationship between an observation point, so the interpolation effect is much better than other methods such as inverse distance weights when the points are scarce. There are different kriging interpolation methods in GIS, and the corresponding types of kriging interpolation methods that meet the scope of application should be selected according to the different data.

3.1.2 The principle of spatiotemporal prediction method based on geographically weighted regression model: Spatial statistics have two major characteristics that are different from classical statistics: one is spatial correlation, the other is spatial heterogeneity, of which Moran exponents and the like can be used to quantify spatial correlation, and the other is geographically weighted regression that can be used to quantify spatial heterogeneity. Geographically weighted regression (GWR) is a spatial analysis technique that is widely used in geography and related disciplines related to spatial analysis. The model generally uses local regression equations based on each point in the spatial extent to explore the spatial changes of the object of study at a certain scale and to predict future outcomes based on the drivers associated with it. Because the model takes into account the local effects of objects in space, it has higher accuracy than other linear regression methods. In spatial analysis, observational data is generally sampled in geographic locations, but as geographic locations change, the relationships and structures of variables change, which is called "spatial nonstationarity" in GIS.

This kind of spatial non-stationarity is common in spatial data, such as the PM2.5 concentration used in this study, so if the traditional linear regression model is used to analyze this spatial data, it is difficult to obtain satisfactory results, because the global model assumes that the relationship between variables has "isotropy" before analysis, and the results obtained are only some kind of "average" in the study area. Such a global "average" does not involve the spatial level, so it is necessary to adopt a new local regression method to deal with this property of the spatial data itself. Therefore, the Geographically Weighted Regression (GWR) model has been proposed and widely used by researchers.

In traditional regression analysis, such as the basic least squares model (OLS), the basic assumption is that the relationship of the independent variable y to the dependent variable x_n remains stable over the region. The general common formula is shown in (4):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \quad (4)$$

A geographically weighted regression model is an extension of a normal linear regression model that embeds the spatial position of the data into the regression equation as shown in (5):

$$y = \beta_0(u_i, \ v_i) + \sum_{k=1}^{p} \beta_k (u_i, v_i) x_{ik} + \varepsilon \quad (5)$$

Among them, $(u_i, \ v_i)$ is the coordinate of the sampling point i, is the k-th regression parameter on the sampling point i, which is a function of the geographical location, and is generally obtained in the estimation process by using the weight function method.

Geographic weighting is actually the same as any other regression analysis, and the field of study needs to be defined in the first place. In general, this field can also be an entire field that includes research data. Then, the most important thing is to calculate the attenuation function with the different spatial positions of each element, this function is a continuous function, and using this attenuation function, the general spatial position of each element can be obtained. Value elements with coordinate information This function takes the weight value and merges the value into the regression equation. It is precisely because this attenuation function derives different weights that geographically weighted regression analysis is so different. Using these formulas, it is possible to calculate all the sample points point by point, and when calculating each sample point, the other samples involved in the calculation will also be given different weights according to the different spatial relationships with the sample point, so that the correlation regression coefficient of each different sample can be finally obtained. Finally, by interpreting these coefficients, the entire analysis process of geographically weighted regression analysis is completed.

3.2 Method implementation

Data processing of Environmental Factors: During the data collection process, due to the occurrence of a series of accidental factors such as networks and devices, it often leads to missing values in the collected data. In order to improve the accuracy of information processing efficiency, the collected raw data is preprocessed before statistical and modeling analysis. Missing values are common in most datasets, so the handling of missing values directly affects the final outcome of subsequent models. The choice of processing method is mainly based on the importance of the missing value attribute and the distribution of missing values.If it is a numerical data with a small loss rate and low importance, it can be simply filled in according to the distribution of the data, such as the data distribution is uniform, and the average value is entered; If the data distribution is skewed, the median value is filled; If the loss rate is as high as 95% or more and the importance is low, you can consider deleting it.

In the data used in this paper, each of the 34 monitoring stations in Beijing will have a small number of missing values at a certain time, and a large number of missing data sites ranging from 12 hours to half a month are mainly concentrated in Tongzhou, Shunyi, Changping, Fengtai Garden, Liulihe, Yongle Dian and Qianmen 7 stations.

Air quality data and meteorological data involve unexpected situations such as site missing measurements or site maintenance, and missing values are generated at a certain point in time or time period (that is, there is no observation data), and this article will use the average of their adjacent time nodes as the observation data for that time point. For the entire station missing data for this day, this paper will take the average value of the corresponding time of the station this month as the data of each moment of the station to fill in the missing value.

3.2.1 Data processing of Roads and Monitoring Sites: Road network processing: To refine the spatiotemporal analysis of environmental factors along urban streets, it is necessary to conduct research around the road and obtain data around the road. The types of roads in cities generally include: highways, national roads, provincial roads, county roads, township roads, main roads, secondary roads, ramps, branch lines, internal roads, machine farming roads, rural roads, small roads, etc. The road network processing is mainly buffering After the area is created, multiple polygon features will be formed, and the created buffer zone will be fused at the same time, so as to obtain the road network polygon features without overlap.

Creation of Thiessen polygons based on detection sites: Since the experimental data in this paper are the concentration data of each site, the spatial and temporal prediction based on geographically weighted regression needs to operate on the surface elements. The point features of each monitoring site are processed by Thiessen polygon into polygon features. The processing flow can be shown as Figure 2.

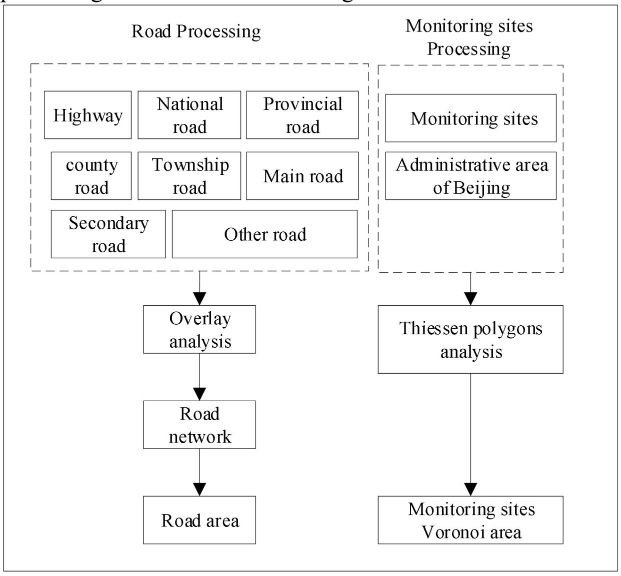

Figure 2. Data processing flow

3.2.2 Calculation of Weights for Neighboring Sites Considering the Influence of Neighboring Sites: To take into account the influence of adjacent stations on the road network, this paper firstly analyzes the neighbors of 34 monitoring stations, obtains the monitoring stations near each point, and assigns each attribute value of its adjacent stations.

Proximity analysis, also known as the nearest neighbor analysis, works by determining the distance between each of the input features and the nearest of the adjacent features of the search radius. After the neighbor analysis, the latitude and longitude of the adjacent stations, the adjacent distance, and the value of each environmental factor corresponding to the adjacent stations will be obtained.

The weight calculation needs to use the centroid coordinates of the road buffer area, and use the field calculator to calculate the centroid coordinates of the road network, respectively, and calculate the distance from the coordinates of the station where the polygonal area where the road belongs and the coordinates of the adjacent stations obtained in the previous step. The formula is (6) shown:

$$b = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{6}$$

The weight is calculated according to the distance between the two stations and the road network, and the formula is shown in (7):

$$a = \frac{\frac{1}{b_1}}{\frac{1}{b_1}+\frac{1}{b_2}} \tag{7}$$

Among them, b1 and b2 are the distances between the adjacent stations and the centroid of the road network buffer.

The weight is multiplied by the existing numerical values of each environmental factor, and the weight value of each environmental factor is calculated.

3.2.3 Spatiotemporal Prediction of Environmental Factors based on Station Weights and Road Network Buffer Area Constraints: In this paper, after the neighbor analysis is carried out on the station, the data of the adjacent stations is obtained, and each data of the adjacent stations is assigned to the road network buffer surface, so that the road network surface attributes are complete. Use the nearest neighbor analysis to obtain weights for calculating the distance between neighboring stations and the road network. The road network buffer surface and Thiessen polygons are then superimposed and analyzed to obtain the road network analysis area for geographically weighted regression prediction. Finally, the calculated weight values are used to perform spatial and temporal prediction by geographically weighted regression analysis, and the characteristics of the prediction results are analyzed, the data analysis flow can be shown as Figure 3.

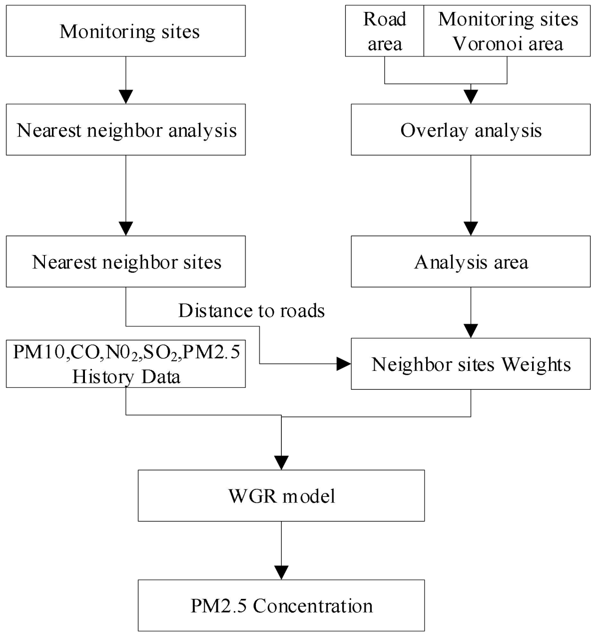

Figure 3. PM2.5 Spatiotemporal Prediction flow chart

3.3 Method process

In this paper, the application of spatiotemporal analysis and prediction of the environment factors studied based on the geographically weighted regression model is operated in ArcGIS software.

First of all, the data required for the experiment are sorted out, and other elements related to the elements of this study need to be prepared for modeling analysis as explanatory variables in the analysis of the geographically weighted regression model, and the more accurate the analysis of the dependent variable elements of the study is the greater the number of explanatory variables, this paper mainly takes the PM2.5 concentration in January 2020 as an example for the analysis experiment, collects the PM10 、 CO data that is also an environmental factor, fills in the missing values respectively, and takes the average value as the experimental data.

Second, because the model uses polygon features, Thiessen polygon processing is performed on site data that is point features.

Finally, the predictions are made using a geographically weighted regression model, and the predicted raster map is output for analysis. After obtaining a geographically weighted regression model, you can export the predictions as raster maps for more intuitive observational analysis. The accuracy of the output of geographically weighted regression is the size of the residual values of each region in the output attribute table, and the smaller the absolute value of the residual value, the higher the accuracy of the prediction.

4. EXPERIMENT ANALYSIS

4.1 Spatial Analysis

First of all, the PM2.5 concentration in Beijing is to be analyzed spatially, and the data have been preprocessed by missing values in this paper. Considering that the results may be inaccurate due to other factors in the single day of data, this paper will interpolate kriging on two scales of one day and one month, and use the toolbox function of ArcMap to analyze the average data of PM2.5 concentrations in 34 sites in Beijing from January 1 to January 31, 2020. The two interpolation processes and the results are as follows:

Single-day spatial interpolation: PM2.5 for each monitoring site within 24 hours recorded on January 1, 2020.

The concentration data were filled with missing values, and the PM2.5 concentration monitoring data of each monitoring site on that day was averaged to obtain the average PM2.5 concentration of 34 stations for one day.

Then adding the map of Beijing as the basemap, add coordinates to form 34 monitoring sites after adding the Figure 1 table, import the PM2.5 concentration table of each site in a single day, take the Beijing county boundary vector map with the added data as the basemap, and use Beijing as the processing range for kriging interpolation analysis.

After the above settings are com pleted, the interpolation can obtain the single-day PM2.5 distribution of 34 monitoring stations in Beijing. As shown in Figure 7:

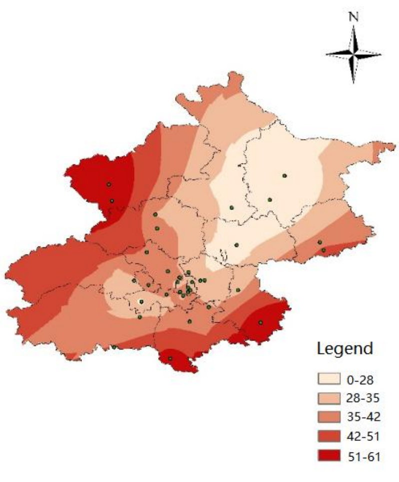

Figure 4. Predicted distribution of average PM2.5 concentration in 2020 January 1.

4.2 Temporal Analysis

The analysis of the ARMIA model obtained the prediction graph of PM2.5 concentration at each station in Beijing, due to the small time scale, the predicted value obtained is basically the trend chart of the approximate rise or fall, and the predicted value is compared with the real value, and it is found that the prediction of the ARIMA model is basically accurate, and the upward or downward trend is more qualified for the approximate trend of the prediction. The comparison chart takes the East Fourth Station as an example, as shown in Figure 12; The kriging interpolation of the predicted values is generally consistent for seven days, and the distribution of the predicted values is generally the same for seven days, and the distribution chart of February 1, 2020 is shown in Figure 5 below:

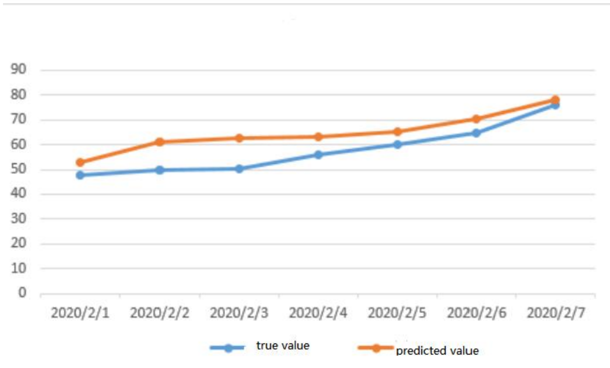

Figure 5. Beijing Dongsi Railway Station January 7 forecast value compared with the real value.

Take the average daily PM2.5 concentration for the whole year of 2020 as the study variable, The predicted value of PM2.5 concentration from February 1 to 18, 2020 is compared with the real value, and it is found that the periodical data after differential stabilization and seasonal decomposition is more accurate than the prediction of single-month data, and the trend and cyclical change law can be predicted. The comparison of the predicted value and the real value within 18 days is shown in Figure 6.

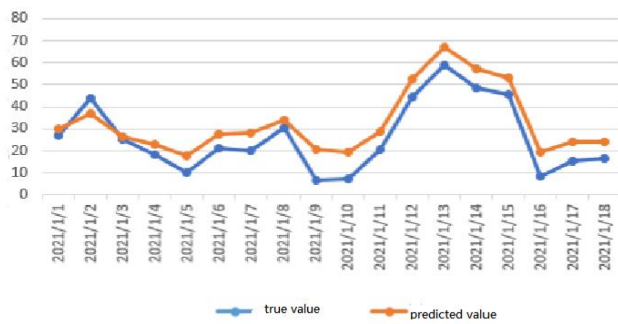

Figure 6. Beijing Dongsi Railway Station forecast value compared with the real value.

Judging from the experimental results, the timing analysis and prediction of PM2.5 concentration as the main research goal in Beijing in the time dimension are good, and the prediction is accurate, and the seasonal change law and trend can be seen. After the study of PM2.5 as the main air pollutant for the whole year of 2020, the analysis and prediction experiment of the time dimension of its concentration can be seen that the average concentration value of PM2.5 in Beijing in 2020 is between 30-50, and there is no obvious upward or downward trend in the measured data for the whole year.

4.3 Spatiotemporal analysis of PM2.5 concentration under the constraints of the road network

Spatiotemporal analysis and prediction of PM2.5 pollutant concentrations in January 2020.Geographically weighted regression analysis and prediction requires other relevant variables as explanatory variables for analysis. Therefore, in addition to the existing January 2020 PM2.5 pollution concentration, the PM10, CO,NO_2、SO_2 and concentration data of the same period will also be collected, and missing values will be processed. After the Thiessen polygon processing, the pollutant concentration data of the nearest monitoring station to the road network is obtained through the buffer proximity analysis of the road network. The selected qualified influencing factors are subjected to a spatial and temporal analysis of the environmental factors under the constraints of the road network based on the geographically weighted regression method. Since the road network is too dense, the overall distribution map is not obvious. 21 shows:

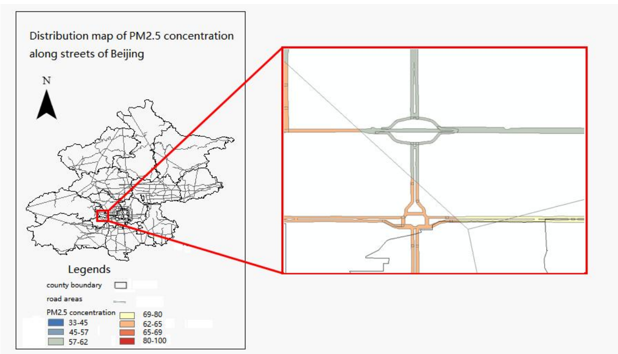

Figure 7. Predicted distribution of average PM2.5 concentration under the constraints of Beijing's road network

It can be seen from the prediction results that under the constraints of the road network, the trend is high in the south and low in the north. The average concentration of pollutants in the southern part of Beijing is the highest, and the pollutant concentration in the northern part is the lowest, decreasing in a step-like manner from south to north. After the road network is weighted, it can be refined to the distribution of streets. The concentration of each street is different. In areas with high concentrations, there will be streets with lower concentrations. The area shown in the figure is a high-level area in the middle, and the street shown in the figure is Xizhimen North Street. Compared with the area in which it is located, the concentration of this street is relatively low, while the concentration of the adjacent Xicheng Fuxing Road is higher than that of the area in which it is located.

5. CONCLUSION AND DISCUSSION

In this paper, through the weighted analysis of adjacent stations and the prediction based on the constraints of the road network, the refined analysis and prediction of the environmental factors along the urban street is completed. And the experimental analysis is carried out with the monitoring data of 34 stations in Beijing. The experiment shows that the method in this paper is more accurate in the analysis and prediction under the condition of the larger and more accurate data, and it can predict and analyze the spatial and temporal distribution of environmental factors in the block. In the future, we will consider adding more influencing factors, such as traffic flow, meteorological conditions, vegetation coverage and other factors as explanatory variables to predict and analyze environmental factors in time and space, and study the size of their respective influences, so as to further achieve higher-precision time and space of environmental factors along the street.

REFERENCES

HUANG Yalin, LIU Chao, Zeng Kefeng, DING lei,CHENG Shenggao. Spatio-temporal Distribution of PM2.5 in Wuhan and Its Relationship with Meteorological Conditionsin[J]. Ecology and Environmental Sciences, 2015, 024(008):1330-1335.

Li Wenjie， Zhang Shihuang， Gao Qingxian， et al. Relationship between Temporal-Spatial Distribution Pattern of Air Pollution Index and Meteorological Elements in Beijing, Tianjin and Shijiazhuang [J].Resources Science， 2012, 34(8):1392-1400.

LI Xiangyang，DING Xiaomei，GAO Hong，ZHU Zhenhua. Characteristics of air pollution index in typical cities of North China [J].JOURNAL OF ARID LAND RESOURCES AND ENVIRONMENT，2011,25(3):96-101.

LIU Yongwei，RUN Qingwu，HUANG Jie，et al. Spatio-temporal distribution of API based on GIS in China [J]. Ecology and Environmental Sciences，2013(8):1386-1394.

WANG Zhenbo， FANG Chuanglin， XU Guang， PAN Yuepeng. Spatial-temporal characteristics of the PM2.5 in China in 2014 [J]. Acta Geographica Sinica， 2015, 70(11):1720-1734.

WU Yiling, LI Chengming, DAI Zhaoxin, WU Zheng. Spatial and temporal distribution characteristics and influencing mechanisms of air quality index in urban areas [J]. Bulletin of Surveying and Mapping, 2020, No.517(04):84-89.

Wu Ying，Ji Dongsheng，Song Tao，et al. Characteristics of Atmospheric Pollutants in Beijing,Zhuozhou,Baoding and Shijiazhuang During the Period of Summer and Autumn [J].Chinese Journal of Environmental Science ， 2011,32(9):2741-2749.

Xiao Yue. Spatiotemporal pattern of air quality and its influencing factors in China [D]. Southwest University,2018.

YIZAITIGULI Waili,WANG Mingli,YANG Jianjun,HUANG Wenjun,LI Yanan. Spatial and Temporal Characteristics of PM2. 5 and PM10 in Urumqi City from 2015 to 2018.

ZANG Baochun， CHEN Yanjun， LI Weijian， et al.The research of spatial and temporal characteristics on air quality level in Pearl Delta base on GIS [J]. Ecology and Environmental Sciences，2011,20(4):600-605.

Zhang Hong. spatial and temporal distribution characteristics of air quality and perdiction of PM2.5 concentration in Beijing-Tianjin-Hebei region [D]. Harbin Normal University,2020.