Estimation model of PM2.5 Concentrations Based on Spatiotemporal Adaptability and Satellite Remote Sensing

Weidong li (wdli@haut.edu.cn)

Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education, Zhengzhou, 450001 China

Liye Dong

College of Information Science and Engineering, Henan University of Technology

Linyan Bai

Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences

Research Article

Keywords: Spatiotemporal Adaptability, Satellite Remote Sensing, heterogeneity

Posted Date: December 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1131351/v1

Estimation model of PM2.5 concentrations based on spatiotemporal adaptability and satellite remote sensing

Li Weidong[1,2,*], Dong Liye[1,2], Bai Linyan[3]

ABSTRACT

Based on satellite remote sensing AOD, we can estimate and monitor the continuous changes of PM2.5, which solved the disadvantages of traditional ground station discrete monitoring. Four-dimensional spatiotemporal heterogeneity is not considered in the construction of traditional empirical regression models, such as geographically weighted regression model (GWR) and spatiotemporal geographically weighted regression model (gtwr). To solve this four-dimensional spatiotemporal nonstationarity, this article proposes and constructs a spatiotemporal adaptive fine particulate matter (PM2.5) concentration estimation model: 4D-GTWR by introducing a DEM (Digital elevation model) and time effects into a GWR model. This method solves the heterogeneity between the three-dimensional space and one-dimensional time by constructing a four-dimensional space kernel function and obtaining its weight. Based on PM2.5 ground observation data and meteorological data collected from December 2017 to February 2018 in Zhengzhou City, Henan Province, PM2.5 estimations are obtained from MODIS MYD-3K AOD data using the GWR, TWR, GTWR and 4D-GTWR models. The results showed that the MAE (mean absolute error) of the 4D-GTWR model decreased by 54.13%, 54.06% and 37.90%, compared to those of the GWR, TWR and GTWR models, respectively, and that the PM2.5 concentrations predicted by the 4D-GTWR model were closest to the measured values. The R^2 (the correlation coefficient) of the 4D-GTWR model was 0.9496, which was better than those of the GWR (R^2 =0.7761), TWR (R^2 =0.7763) and GTWR (R^2=0.8811) models. The 4D-GTWR model can not only improve the precision of PM2.5 estimations but can also reveal the four-dimensional spatial heterogeneity of PM2.5 concentrations and the differentiation of the DEM's influence on the spatial dimensions.

With the rapid development of China's economy and accelerating industrialization and urbanization in recent years, fine particulate matter (PM2.5) pollution has become the most prominent problem in China's economic development and the most concerning issue to the government and public. PM2.5 not only poses a serious threat to public health but also seriously affects urban traffic and the daily lives of citizens (Qiu et al., 2012; Wang zhenbo et al., 2015). Therefore, PM2.5, as an atmospheric pollutant with a direct impact on human lives and health, has attracted extensive attention from many scholars studying PM2.5-related inversions. Liu et al. (Liu et al., 2004) first Proposed the use of a chemical transport model to estimate PM2.5 concentrations, and the verification results showed that the correlation between the PM2.5 concentrations retrieved from the chemical transport model and data measured by the Environmental Protection Agency (EPA) of the United States was 0.81. Subsequently, Boys et al.

[1] *Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education, Zhengzhou, 450001 China;*

[2] *College of Information Science and Engineering, Henan University of Technology, Henan 450001 China;*

*3 Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences，beijing 100094 China. * Correspondence: wdli@haut.edu.cn*

(2014) used MISR and Sea WiFS AOD data to invert global PM2.5 concentrations

from 1998 to 2012. Kumar et al. (2007) established a multivariate linear regression relationship with PM2.5 by using MODIS AOD data and meteorological data obtained in Delhi, India, and the correlations between the estimated results and the actual observed values varied between 0.60 and 0.81. Bai et al. (2016)proposed using AOD, relative humidity, wind speed and temperature data as well as other indicators to invert PM2.5 concentrations. Mao et al. (2012) added AOD data to the traditional land use regression (LUR) model to estimate the PM2.5 concentrations in Florida, USA. The correlation coefficient of the model was 0.63. The multiple linear regression model added a variety of variables, making the decision coefficient of the model higher than that of the single linear regression model and improving the performance of the model.

To solve the changes in PM2.5 concentrations over time and space, researchers have considered spatially and temporally heterogenous PM2.5 concentrations. For example, Song et al. (2014)studied the spatial nonstationarity of PM2.5 concentrations in the Pearl River Delta from May 2012 to September 2013 using a geographically weighted regression (GWR) method as the research model and found that the predictive abilities of the multiple linear regression method and semiempirical model were significantly improved. Huang et al. (2010) proposed a geographically and temporally weighted regression (GTWR) method for predicting housing prices in their research area, and the results showed that the goodness of fit of the GTWR model was better than that of the GWR model, verifying the nonstationarity problem in space and time. Chu et al. (2015)used PM data collected in Taiwan from 2005 to 2009 to study the relationship between PM10 and PM2.5 using multiple linear regression models: the GWR model and the GTWR model. The results showed that the outputs of the GTWR and GWR models were relatively consistent, but the results of the GTWR model had a better fitting effect and stronger spatial and temporal interpretation abilities. Zhao Yangyang (2016)and others combined collaborative training with a spatiotemporal geographically weighted regression technique and proposed a collaborative spatiotemporal geographically weighted regression method. Taking the PM2.5 concentrations in Beijing, Tianjin and Hebei from March to July 2015 as an example, the GTWRs of different kernel functions were used to carry out comparative analysis experiments. The results showed that the method can improve the accuracy of PM2.5 concentration estimations when the number of spatiotemporal samples is insufficient. Rao Lanlan (2017)used the geographically weighted regression model and adaptive bandwidth to estimate near-surface nitrogen dioxide concentrations and compared the results with those obtained using the traditional linear regression model, time-weighted regression model and geographically weighted regression model. The results showed that the correlation between the near-surface nitrogen dioxide concentrations estimated by the GTWR model and the ground observation values was best and that this model also resulted in the smallest error. In short, the GWR and GTWR models are widely used in the fields of social economics, agricultural production, urban geography, meteorology, etc. (Danlin, 2007). In recent years, the geographically and temporally weighted regression model has been widely considered by researchers, facilitating abundant achievements in theory and

application in a wide range of application fields.

However, the GWR and GTWR models do not take into account the four-dimensional nonstationarity resulting from the considered three-dimensional space and one-dimensional time. Specific research, such as the inversion of PM2.5 concentrations, considers model parameters in the study area with four-dimensional spatiotemporal changes, and four-dimensional spatiotemporal nonstationarity exists. Therefore, in this paper, on the basis of the GWR and GTWR models, combined with PM2.5 concentration data, a 4D-GTWR model is proposed to estimate PM2.5 concentrations. First, the traditional GWR model is extended to a four-dimensional geographically and temporally weighted regression (4D-GTWR) model, and this new model is applied to estimate PM2.5 concentrations in four-dimensional space-time. Second, according to the evaluation index of the model, the goodness of fit is estimated by a comparison between the models. Finally, taking the research area in Zhengzhou, Henan Province, as an example, the accuracies of the PM2.5 inversion results output by the models were compared and analyzed.

Materials and methods

Model definition and design .Usually controlled and influenced by topography, the development process of the natural geographical environment changes, and global- or regional-scale high-resolution digital elevation models (DEMs) play an extremely important role in studies of climatic and environmental changes such as seismic geological natural disasters and atmospheric environmental pollution (li zhenhong et al., 2018). The GTWR model does not take into account the change of the nonstationarity relationship of the four-dimensional space-time with a supplementary DEM; namely, the four-dimensional spatial heterogeneity of the three-dimensional space and the one-dimensional time is not taken into account. Therefore, in this paper, a DEM is considered for integration into the calculations of spatial and temporal distances, and an analysis and calculation method of the four-dimensional spatial and temporal distances is proposed to establish a four-dimensional geographically and temporally weighted regression (4D-GTWR) model. The expression of the 4D-GTWR model is as follows:

$$y_i = \beta_{i0}(u_i, v_i, z_i, t_i) + \sum_{k=1}^{p} \beta_{ik}(u_i, v_i, z_i, t_i) x_{ik} + \xi_i, i = 1,2,\Lambda\ ,n \qquad (1)$$

where (u_i, v_i, z_i, t_i) is the four-dimensional spatial coordinate of the i-th sample point, (u_i, v_i, z_i) is the three-dimensional spatial coordinate, (t_i) is the temporal coordinate, β_{ik} is the regression coefficient parameter of each independent variable of sample point i, and ξ is the random error of sample point i, which obeys the independent and identical distribution of $\xi_i - N(0, \sigma^2)$.

Using the same calculation as that applied in the GTWR model, the regression

coefficients at the i-th sample point of the 4D-GTWR model are estimated to be $\hat{\beta}(u_i, v_i, z_i, t_i)$, as follows:

$$\hat{\beta}(u_i, v_i, z_i, t_i) = (X^T W(u_i, v_i, z_i, t_i) X)^{-1} X^T W(u_i, v_i, z_i, t_i) y \tag{2}$$

where $W(u_i, v_i, z_i, t_i)$ represents the four-dimensional spatiotemporal weight matrix. The estimated value, $\hat{y}_i$, of the dependent variable of sample point i is calculated as follows:

$$\hat{y}_i = X_i \hat{\beta}(u_i, v_i, z_i, t_i) = X_i (X^T W(u_i, v_i, z_i, t_i) X)^{-1} X^T W(u_i, v_i, z_i, t_i) y \tag{3}$$

where X_i represents the vector of line i in matrix X. Therefore, the dependent variable regression vector, $\hat{y}$, of each sample point can be calculated as follows:

$$\hat{y} = \begin{bmatrix} s_1 \\ s_2 \\ \text{M} \\ s_n \end{bmatrix} = \begin{bmatrix} X_1 (X^T W(u_1, v_1, z_1, t_1) X)^{-1} X^T W(u_1, v_1, z_1, t_1) y \\ X_2 (X^T W(u_2, v_2, z_2, t_2) X)^{-1} X^T W(u_2, v_2, z_2, t_2) y \\ \text{M} \\ X_n (X^T W(u_n, v_n, z_n, t_n) X)^{-1} X^T W(u_n, v_n, z_n, t_n) y \end{bmatrix} y = Sy \tag{4}$$

where the matrix S is the hat matrix of the 4D-GTWR model. According to the observed value y and the fitted value $\hat{y}$, the residual sum of squares (RSS) of the 4D-GTWR model can be calculated as follows:

$$RSS_{4D\text{-}GTWR} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\varepsilon}^T \hat{\varepsilon} = y^T (I - S)^T (I - S) y \tag{5}$$

where $\hat{\varepsilon} = \begin{bmatrix} y_1 \\ y_2 \\ \text{M} \\ y_n \end{bmatrix} - \begin{bmatrix} s_1 \\ s_2 \\ \text{M} \\ s_n \end{bmatrix} y = (I - S) y$ and I is the unit matrix.

Similarly, it is assumed that the fitting value $\hat{y}$ is the unbiased estimation of $E(y_i)$, namely, $E(\hat{y}_i) = E(y_i)$; thus, $E(\hat{\varepsilon}) = E(y) - E(\hat{y}) = 0$ and $\mathrm{E}(\varepsilon^{\mathrm{T}} \varepsilon) = \sigma^2 \mathrm{I}$. Then, $RSS_{4D\text{-}GTWR}$ can be expressed as follows.

$$\mathrm{E}(\mathrm{RSS}_{\text{4D-GTWR}}) = \mathrm{E}(y^T (I - S)^T (I - S) y) \tag{6}$$

$$= \mathrm{E}(\mathrm{tr}(y^T (I - S)^T (I - S) y))$$

$$= \mathrm{tr}((I - S)^T (I - S)) \mathrm{E}(y y^T)$$

$$= \sigma^2(n - 2tr(S) + tr(S^T S))$$

Therefore, the unbiased estimation, σ^2, of the random error variance in the 4D-GTWR model is as follows.

$$\sigma^2 = RSS_{4D\text{-}GTWR} / (n - 2tr(S) + tr(S^T S)) \tag{7}$$

Four-dimensional spatiotemporal kernel function

Similar to the construction of the GTWR model, the core of the 4D-GTWR model involves constructing a four-dimensional space-time kernel function to calculate the four-dimensional space-time weight. When considering the aspect of spatiotemporal proximity, it is necessary to consider the fusion of the three-dimensional spatial distance and the one-dimensional temporal distance, that is, the change in the regression coefficient with the change to four-dimensional space-time. Therefore, it is necessary to construct four-dimensional space-time kernel functions on the basis of spatial kernel functions. The following describes the construction method of the temporal distance and spatial distance of the 4D-GTWR model.

Considering that the DEM mainly has a certain impact on the spatial dimension, based on the study of spatial and temporal variations in PM2.5 in this paper, the pollution processes of PM2.5 and other pollutants undergo four-dimensional spatial and temporal changes; that is, four-dimensional nonstationary spatial and temporal changes occur. Therefore, for any number of monitoring points, the three-dimensional spatial coordinates among them differ. Considering that two- and three-dimensional spaces may have different scale effects, the 4D-GTWR model is introduced into $\oplus$ to express the scale differences between the distances. Therefore, according to the Euclidean spatial distance, the three-dimensional spatial distance d^s between monitoring stations A and B can be expressed as follows:

$$d^S = d^{S2d} \oplus d^{SDEM} \tag{8}$$

where $\oplus$ can represent various operators. On this basis, by combining the spatiotemporal distance construction method, the four-dimensional spatiotemporal distance can be expressed as follows:

$$d^{ST} = d^S \otimes d^T = (d^{S2d} \oplus d^{SDEM}) \otimes d^T \tag{9}$$

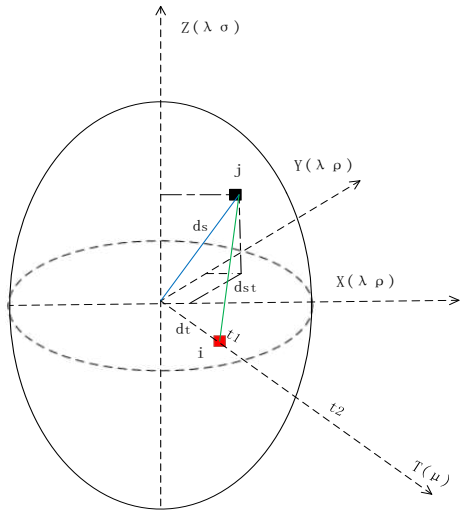
The scale-effect symbols $\otimes$ and $\oplus$ usually adopt additions; that is, a linear combination of the four-dimensional spatiotemporal distance can be used to obtain the four-dimensional spatiotemporal distance of the 4D-GTWR model as follows:

$$d^{ST} = \lambda(\varphi d^{S2d} + \delta d^{SDEM}) + \mu d^T \tag{10}$$

Among them, the λ、μ、φ、δ parameter is a four-dimensional spatiotemporal distance adjustment factor that is used to balance the scale differences among the various four-dimensional spatiotemporal distances. By using the Euclidean spatial

distance to expand the above formula, the four-dimensional spatiotemporal distances of sample points i and j can be obtained as follows:

$$d_{ij}^{ST} = \lambda\left(\varphi\sqrt{(u_i-u_j)^2+(v_i-v_j)^2}+\delta\sqrt{(z_i-z_j)^2}\right)+\mu\left(\sqrt{(ti-tj)^2}\right) \tag{11}$$



$$d_{ij}^{ST} = \lambda\left(\varphi\sqrt{(u_i-u_j)^2+(v_i-v_j)^2}+\delta\sqrt{(z_i-z_j)^2}\right)+\mu\left(\sqrt{(ti-tj)^2}\right)$$

Figure 1 Four-dimensional spatiotemporal distance diagram

A four-dimensional spatiotemporal distance diagram is shown in Figure 1. Similarly, by taking the Gaussian kernel function as an example, the four-dimensional spatiotemporal weight function of the 4D-GTWR model can be obtained as follows:

$$W_{ij}^{ST} = exp\left\{-\left(\frac{\lambda\varphi\left[(u_i-u_j)^2+(v_i-v_j)^2\right]+\lambda\delta(z_i-z_j)^2+\mu(ti-tj)^2}{(h^{4D\text{-}ST})^2}\right)\right\} \tag{12}$$

$$= \exp\left\{-\left(\frac{(u_i-u_j)^2+(v_i-v_j)^2}{\frac{(h^{4D-ST})^2}{\lambda\varphi}}+\frac{(z_i-z_j)^2}{\frac{(h^{4D-ST})^2}{\lambda\delta}}+\frac{(t_i-t_j)^2}{\frac{(h^{4D-ST})^2}{\mu}}\right)\right\}$$

$$= exp\left\{-\left(\frac{(u_i-u_j)^2+(v_i-v_j)^2}{(h^{S2d})^2}+\frac{(z_i-z_j)^2}{(h^{SDEM})^2}+\frac{(t_i-t_j)^2}{(h^{T})^2}\right)\right\}$$

$$= exp\left\{-\left(\frac{(d_{ij}^{S2d})^2}{(h^{S2d})^2}+\frac{(d_{ij}^{SDEM})^2}{(h^{SDEM})^2}+\frac{(d_{ij}^{T})^2}{(h^{T})^2}\right)\right\}$$

$$= \exp\left\{-\frac{(d_{ij}^{S2d})^2}{(h^{S2d})^2}\right\}\times\exp\left\{-\frac{(d_{ij}^{SDEM})^2}{(h^{SDEM})^2}\right\}\times\exp\left\{-\frac{(d_{ij}^{T})^2}{(h^{T})^2}\right\}$$

$$= W_{ij}^{S2d}\times W_{ij}^{SDEM}\times W_{ij}^{T}$$

where h^{S2d}、h^{SDEM}、h^{T}、$h^{4D\text{-}ST}$ represent the two-dimensional spatial bandwidth,

the spatial bandwidth of the DEM, the temporal bandwidth and the four-dimensional spatiotemporal bandwidth parameters, respectively. Similarly, W_{ij}^{S2d}、W_{ij}^{SDEM}、W_{ij}^{T} represent the two-dimensional spatial weight, the spatial weight of the DEM and the temporal weight, respectively. Therefore, the spatiotemporal weight matrix of the 4D-GTWR model is expressed as follows:

$$W(u_i,v_i,z_i,t_i)=\begin{Bmatrix} w_{i0}^{S2d}\times w_{i0}^{SDEM}\times w_{i0}^{T} & 0 & 0 & 0 \\ 0 & w_{i1}^{S2d}\times w_{i1}^{SDEM}\times w_{i1}^{T} & 0 & 0 \\ 0 & 0 & \Lambda & 0 \\ 0 & 0 & 0 & w_{in}^{S2d}\times w_{in}^{SDEM}\times w_{in}^{T} \end{Bmatrix} \quad (13)$$

Solve the optimal model

According to formula (13), if the parameter λ is assigned a value of 0, that is, if the temporal nonstationarity relationship is not considered, the TWR model is obtained. If spatial nonstationarity is not considered, the parameters μ、δ are set to 0, and a similar GWR model is obtained. A similar GTWR model can be obtained if the spatiotemporal nonstationarity relationship is considered; in this case, the parameter δ is set to 0. In the process of physical relationship modeling, the parameters λ、φ、δ、μ usually have non-zero values, and the factors that truly play a role are $\tau=\mu/\lambda$ and $\theta=\delta/\varphi$. Therefore, using a parameter-simplification strategy similar to that in the GTWR model, the 4D-GTWR model sets λ and φ to 1 to reduce the calculation of parameters; then, the model only needs to optimize its estimations of parameters μ and δ. Therefore, the modeling process of the 4D-GTWR model mainly includes the optimization of the spatiotemporal scale parameter μ, the two-dimensional and three-dimensional spatial-scale parameter δ and the four-dimensional spatiotemporal bandwidth parameter h^{4D-ST}.

Research data

Zhengzhou, the capital city of Henan Province, is an essential central city in central China, a leading comprehensive transportation hub and a core city of the Central Plains Economic Zone. It lies between 112° 42′ E-114° 14′ E and 34° 16′ N-34° 58′ N. Compared with other urban areas, the atmosphere in Zhengzhou is relatively stable. Because the air is not very fluidity, the tiny particles in the air will gather and float in the air, which is not conducive to the diffusion of pollutants, thus causing more and more severe haze pollution.

More than 30 automatic air quality monitoring stations have been established in Zhengzhou city to monitor the contents of NO_2, SO_2, NO, PM10, PM2.5 and other pollutants in the air. These stations include monitoring stations in the Gongyi, Dengfeng, Xinmi, Xinzheng, Xingyang, Zhongmou and Zhengzhou districts. These stations are mainly used to monitor the air quality in vehicle traffic environments and the impact of air pollution on pedestrians. Because the data of some stations are not available, only PM2.5 hourly mean concentration monitoring data of 31 stations (9 state-controlled monitoring stations and 22 city-controlled monitoring stations) with uniform distributions were selected in this paper. The distribution of air pollution monitoring stations is shown in Figure 2, and the basic information of the stations is shown in Table 2. The PM2.5 concentration data recorded at each site were downloaded from the website of the National Meteorological Information Center of China (http://data.cma.cn/), and some air quality source data information is shown in Table 1. Meteorological factors, including the air pressure (hPa), temperature (°C), relative humidity (%), rainfall (mm), wind speed (km/h) and wind direction (°), all have different effects on changes in PM2.5 concentrations. Because the geographic locations of the meteorological stations differ from those of the PM2.5 monitoring stations, the meteorological factor values at the PM2.5 stations are derived from the data recorded at surrounding meteorological stations via kriging interpolations. The meteorological data source is the hourly mean data provided by the Henan Meteorological Bureau. There are seven meteorological monitoring stations. The distribution of the meteorological data is shown in Figure 1, and the information of the stations is shown in Table 3. The DEM data were obtained from the Geospatial Data Cloud Platform (http://www.gscloud.cn) of the Computer Network Information Center of the Chinese Academy of Sciences, with a spatial resolution of 30 m.

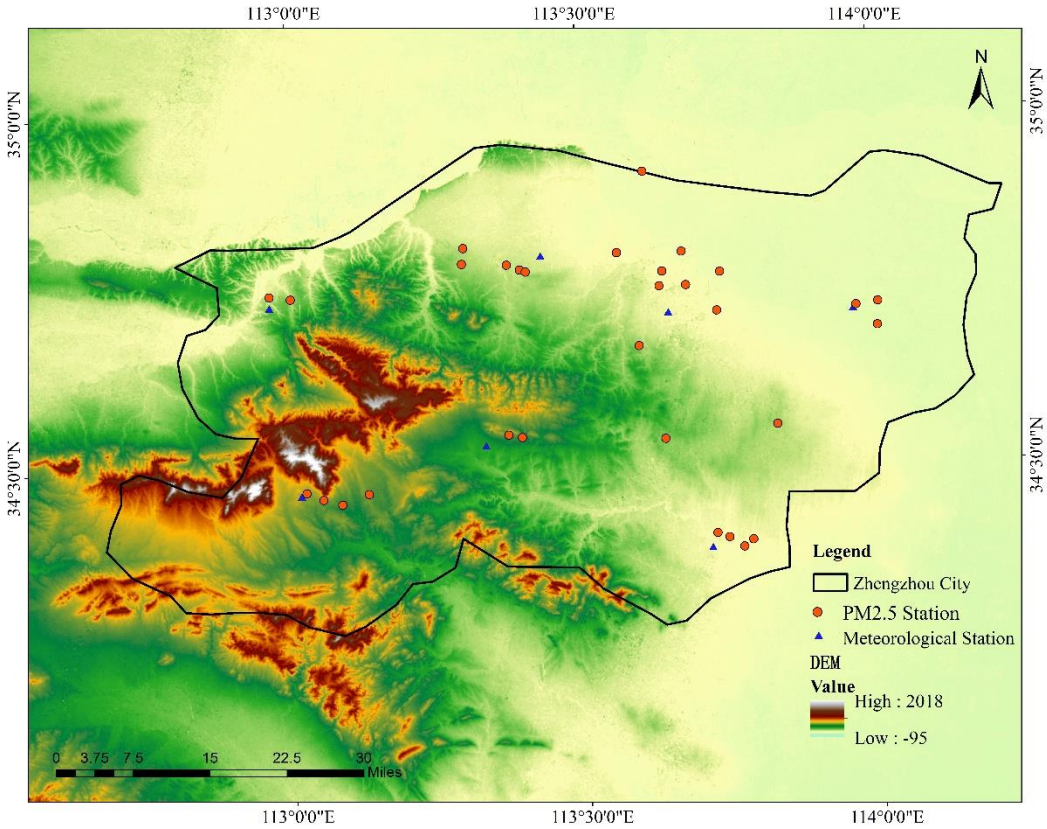


Figure 2. Air quality monitoring stations and meteorological monitoring stations distributions

Table 1. Partial Source Data of Air Quality in Zhengzhou City

City name	Station name	Station number	Date	SO_2 (ug/m^3)	SO_2/AQI (ug/m^3)	...	PM10/AQI(ug/m^3)	PM10(ug/m^3)	PM25/AQI	PM2.5 (ug/m^3)	AQI(ug/m^3)
Gongyi	EPA automatic station	41018102	2018/1/1	253	65		123	195	0	0	123
Port area	North District Construction Command	41011101	2018/1/1	36	12		128	205	140	107	140
…	…	…		…	…			…		…	

Table 2 Information of 31 air quality stations in Zhengzhou

Region name	Region coding	Station name	Station number	Longitude	Latitude	Station category
Port area	410111	North District Construction Command	41011101	113.8325	34.555	City control
Zhengzh	410100	Zhengzhou Forty-seventh	41010109	113.73984	34.772247	State

ou		Middle School				control
…	…	…	…	…	…	…

Table 3. Information of 7 meteorological stations in Zhengzhou

Station	Station number	Longitude	Latitude
Gongyi	57080	112.9666667	34.73333333
Xingyang	57081	113.4333333	34.8
Dengfeng	57082	113.0166667	34.46666667
Zhengzhou	57083	113.65	34.71666667
Xinmi	57085	113.3333333	34.53333333
Xinzheng	57086	113.7166667	34.38333333
Zhongmu	57090	113.9666667	34.71666667

MODIS 3-km (MYD 04_3K) products were selected from December 2017 to February 2018 to invert the PM2.5 concentrations. The geometric correction of the MYD 04_3K data was performed by IDL programming, in which the data are transformed into the WGS-84 geographic coordinate system and AOD data matched with concurrently measured PM2.5 station data are extracted by Arcpy. The Moderate Resolution Imaging Spectrometer (MODIS) is a common sensor used to retrieve aerosol optical thicknesses. Its scanning width is 2330 km, and it can obtain global observation data at least once a day(Barnes et al., 1998; Kaufman and Gao, 1992). MODIS has 36 spectral channels with a spectral range of 0.4-14 μm and spatial resolutions of 250 m, 500 m and 1000 m. MODIS can be used to obtain aerosol, water vapor, surface temperature and ocean data(Engel-Cox et al., 2004; Wu Haiyan et al., 2015).

MODIS data is downloaded from NASA LAADS (https://ladsweb.nascom.nasa.gov/). Some of the pre-processed observations and research data are shown in tables 4 and 5.

Table 4. Observation data

PM2.5	AOD	Humidity（%）	Pressure（hPa）	Temperature（℃）	Wind speed（m/s）	Wind direction（℃）	DEM（m）
66.5	0.16	51.93	996.16	5.86	1.39	258.43	306
24.5	0.16	51.02	995.02	5.86	1.39	258.43	334
114	0.31	70.65	1010.44	5.86	1.4	258.43	110
105	0.31	70.49	1010.08	5.86	1.39	258.43	115
27.5	0.1	22.35	985.8	-0.64	3.08	295.83	316
30.5	0.33	23	1018.15	2.72	1.84	236.47	87
20	0.18	23	1020.6	1.84	1.73	231.62	162
13.5	0.16	22.58	1003.64	2.4	1.89	154.59	306
100.5	0.44	37.48	1015.27	-0.01	1.37	169.14	110
77	0.47	37.62	1014.9	-0.01	1.37	169.14	115

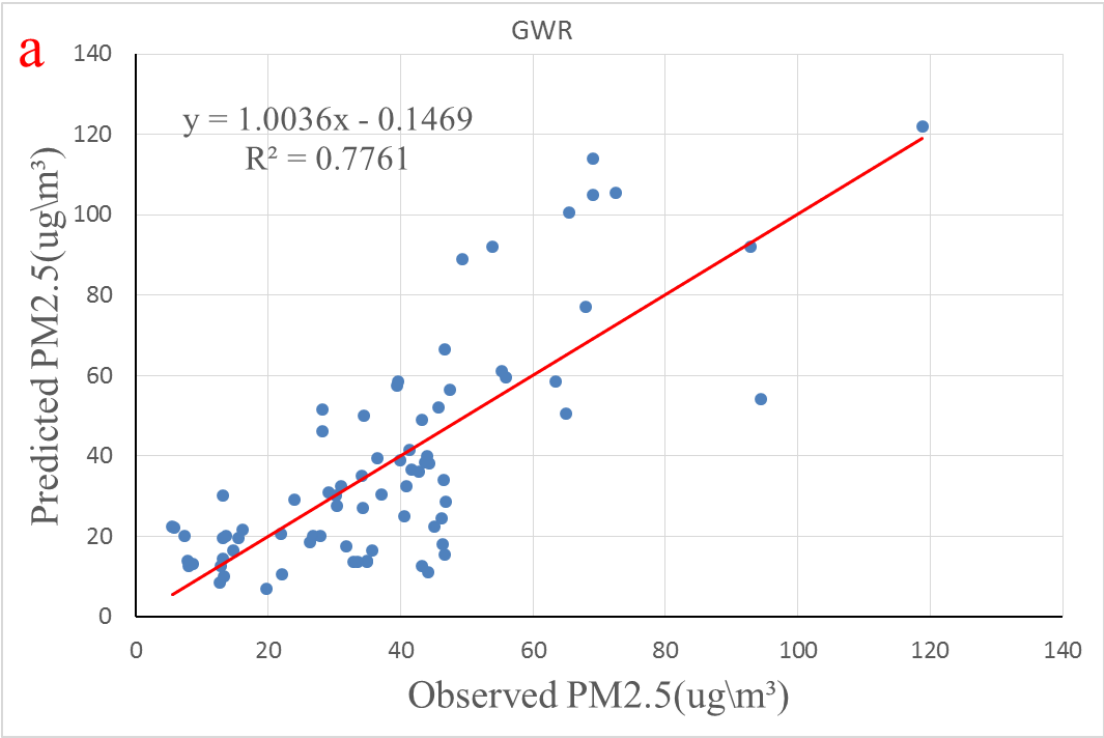
Table 5. Research data

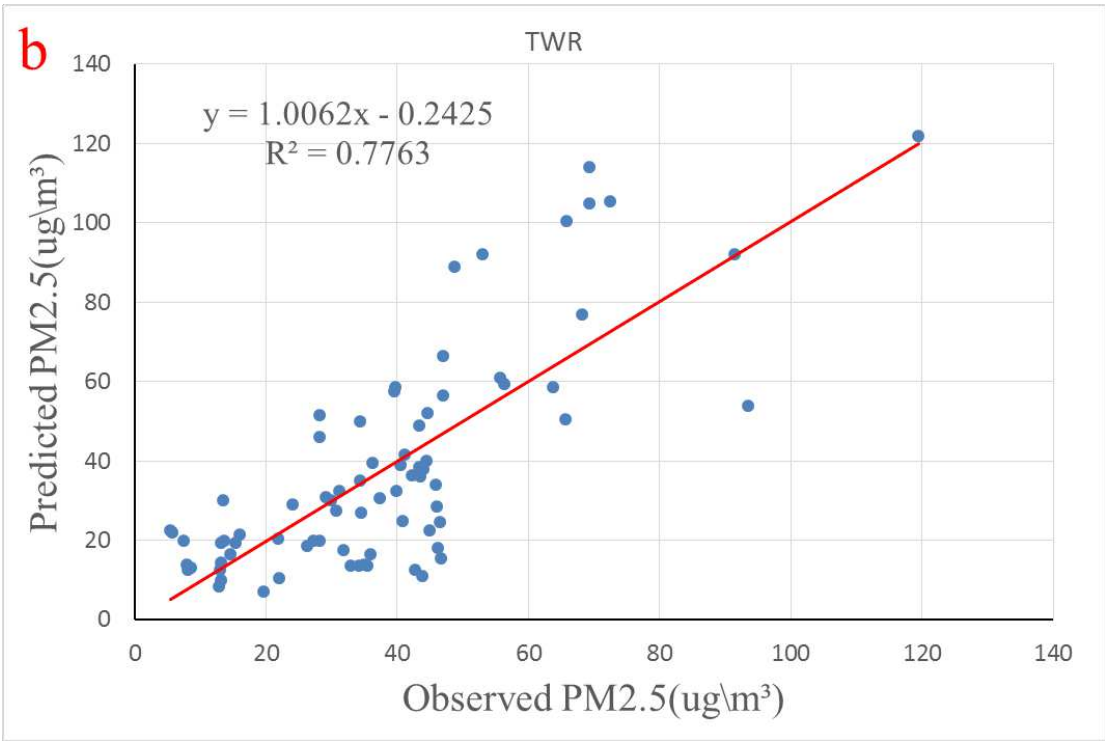
Data information / Data type	Data	Resolution	Period
remote sensing data	MYD04_3K AOD	Spatial resolution:3km	2017.12-2018.2
Air quality data	PM2.5	Temporal resolution:Hourly	2017.12-2018.2
Meteorological data	Relative humidity, Pressure,Temperature, Wind speed, Wind direction	Temporal resolution:Hourly	2017.12-2018.2
ASTER GDEM	DEM	Spatial resolution:30m	2009

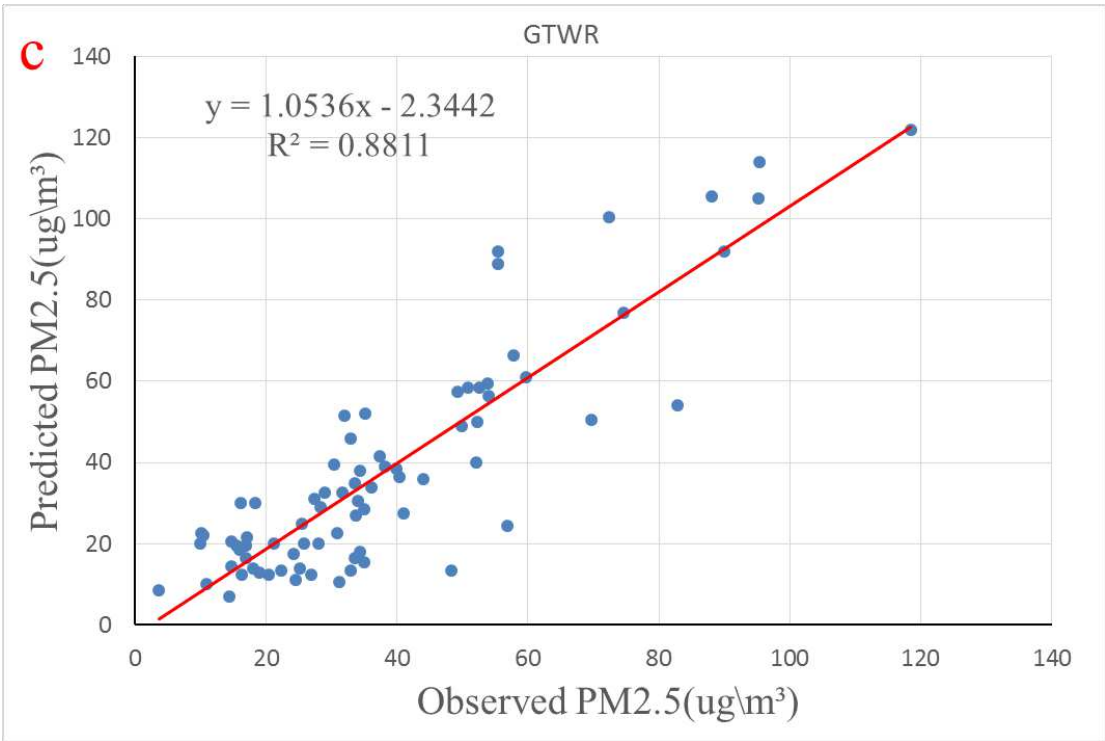
Result comparison and analysis

Comparative analysis of model results

In this paper, a total of 77 observed dataset were selected, and the GWR, TWR, GTWR and 4D-GTWR models were used to estimate and predict PM2.5 concentrations. Moreover, the correlation coefficient (R^2), root mean square error (RMSE) and mean absolute error (MAE) were selected to evaluate the accuracy of the models. The Pearson correlation coefficients (R^2) between the predicted PM2.5 concentrations by the GWR, TWR, GTWR and 4D-GTWR models and the observed values are 0.7761, 0.7763, 0.8811, and 0.9496; the RMSE values are 17.2230, 17.2159, 12.9579, and 8.5931 g/m^3; and the MAE values are 12.9705, 12.9517, 9.5815, and 5.9498, respectively. The RMSE and MAE values showed gradually decreasing trends, indicating that the model accuracy gradually improved, as shown in Tables 6-10 and Figure 3.


a
GWR
y = 1.0036x - 0.1469
R² = 0.7761
Predicted PM2.5(ug\m³)
Observed PM2.5(ug\m³)
0
20
40
60
80
100
120
140


b
TWR
y = 1.0062x - 0.2425
R² = 0.7763
Predicted PM2.5(ug\m³)
Observed PM2.5(ug\m³)
0
20
40
60
80
100
120
140



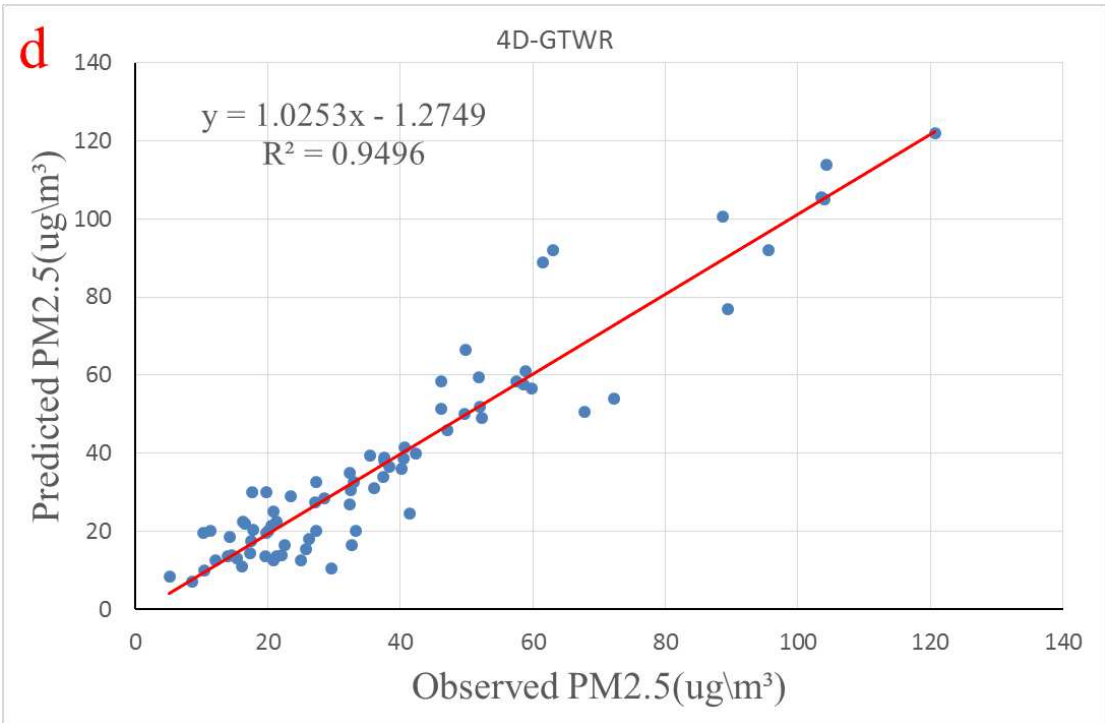


Figure 3. Scatter plots of PM2.5 predicted values and observed values:（a）GWR, (b)TWR,（c）GTWR, and（d）4D-GTWR

Table 6. Comparison of evaluation indicators for each model

Parameter / Model	MAE	RMSE	R^2

GWR	12.9705	17.2230	0.7761
TWR	12.9517	17.2159	0.7763
GTWR	9.5815	12.9579	0.8811
4D-GTWR	5.9498	8.5931	0.9496

Table 7. GWR model（bandwidth =2.2895）

Parameter	Min	LQ	Med	UQ	Max
Intercept（β_0）	35.96	36.2	36.73	37.79	38.8
AOD（β_1）	62.31	62.86	64.42	64.86	65.77
Relativehumidit（β_2）	44.62	44.87	45.25	45.55	45.76
Pressure（β_3）	-14.67	-14.59	-14.54	-14.45	-14.44
Temperatur（β_4）	-19.91	-19.57	-19.12	-18.88	-18.8
Wind speed（β_5）	-0.3	0.01	0.25	0.68	0.77
Wind direction（β_6）	-18.89	-17.98	-17.34	-17.25	-16.43
RMSE	17.2230				
MAE	12.9705				
R^2	0.7761				

Table 8 TWR model（bandwidth =0.0331）

Parameter	Min	LQ	Med	UQ	Max
Intercept（β_0）	-3.65	0.54	8.01	97.69	115.05
AOD（β_1）	45.81	53.16	86.36	87.68	93.24
Relativehumidit（β_2）	19.5	29.5	42.95	57.75	70.88
Pressure（β_3）	-50.87	-41.77	-17.65	-15.16	-5.69
Temperatur（β_4）	-34.7	-33.18	-6.93	25.82	30.8
Wind speed（β_5）	-4.63	-4.23	0.42	12.6	17.46
Wind direction（β_6）	-81.88	-67	-7.66	-5.06	-3.46
RMSE	17.2159				
MAE	12.9517				
R^2	0.7763				

Table 9. GTWR model（bandwidth =2.2895）

Parameter	Min	LQ	Med	UQ	Max
Intercept（β_0）	-3.74	0.66	7.88	98.21	115.43
AOD（β_1）	44.14	51.77	86.09	87.5	94.27
Relativehumidit（β_2）	19.02	29.23	42.41	57.36	70.82
Pressure（β_3）	-50.8	-41.93	-17.86	-14.84	-5.38
Temperatur（β_4）	-34.99	-33.32	-6.52	26.84	31.09
Wind speed（β_5）	-5.23	-4.53	0.2	14.17	19.39
Wind direction（β_6）	-81.54	-66.68	-8.17	-5.11	-3.29
RMSE	12.9579				
MAE	9.5815				
R^2	0.8811				

Table 10. 4D-GTWR model（bandwidth =0.0063）

Parameter	Min	LQ	Med	UQ	Max
Intercept（β_0）	-95.32	-4.94	20.73	189.66	236.77
AOD（β_1）	-63.6	16.8	76.75	104.22	300.05
Relativehumidit（β_2）	-12.04	14.5	45.18	71.59	127.67
Pressure（β_3）	-135.56	-83.47	-23.54	8.99	68.78
Temperatur（β_4）	-213.84	-68.69	-4.96	13.39	64.7
Wind speed（β_5）	-34.46	-1.43	12.02	22.2	101.46
Wind direction（β_6）	-127.86	-104.26	-29.8	4.31	36.52
RMSE	8.5931				
MAE	5.9498				
R^2	0.9496				

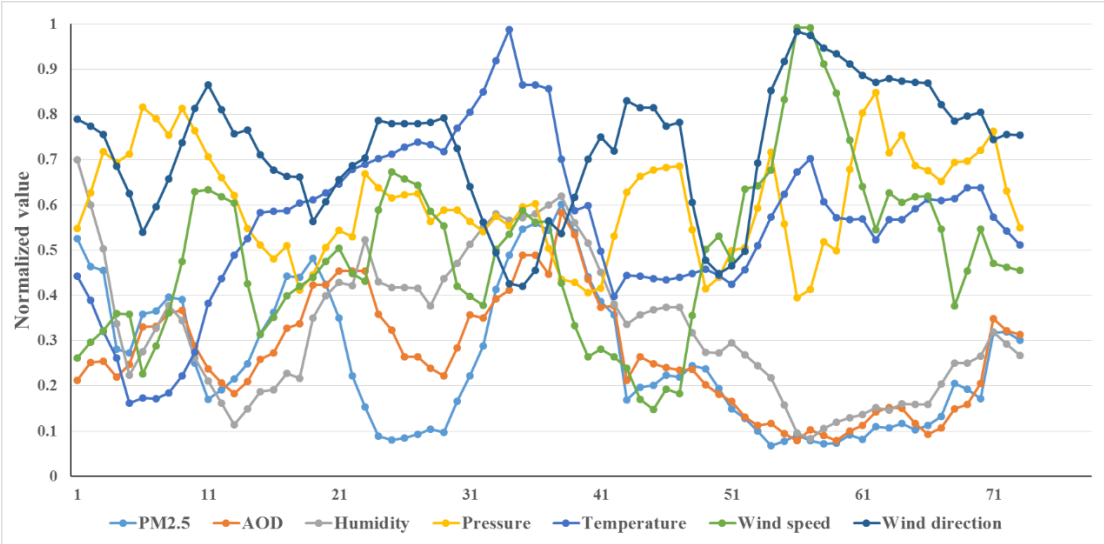


Figure 4. Variations in the 5-point sliding average trends of air quality and meteorological data after normalization

Tables 7-10 show the minimum (Min), lower-quartile (LQ), median (Med), upper-quartile (UQ) and maximum (Max) values of the coefficients of variables in the GWR, TWR, GTWR and 4D-GTWR models, respectively. The variations in the coefficients of variables at are described through time and space. In terms of the median values, the coefficients of the AOD, relative humidity and wind speed are positive, indicating that they are positively correlated with PM2.5, while the coefficients of the air pressure, air temperature and wind direction are negative, indicating that they are negatively correlated with PM2.5. These results are consistent with the variation trends of the variables shown in Figure 4. The estimation accuracies of the GWR and TWR models are comparable, probably because the experimental data have 31 coordinate points in space and the time span is only 3 months, meaning that the effects of temporal and spatial changes are relatively balanced. The estimation accuracies of the GTWR and 4D-GTWR models are significantly higher than those of the GWR and TWR models. Therefore, the fitting ability of the 4D-GTWR model is the best relative to the GWR, TWR and GTWR models for the four-dimensional

spatiotemporal nonstationary relationship. At the same time, the four-dimensional spatiotemporal nonstationarity with the DEM is verified to be more significant than that with separate temporal and spatial nonstationarity, indicating that the three-dimensional spatiotemporal distance plays an important role in the calculation of the four-dimensional spatiotemporal distance in the 4D-GTWR model and further shows that the DEM factor has an important influence on changes in the PM2.5 concentrations. At the same time, it is verified that the effect of four-dimensional space-time non-stationarity with DEM is better than that of space-time non-stationarity.

PM2.5 estimation results

The parameter coefficients of each model inversion used to obtain the PM2.5 concentrations are obtained for Zhengzhou City, Henan Province, taking February 26, 2018 as an example, as shown in Table 11. Then, the spatial and temporal prediction distributions (at a 3-km resolution) of the PM2.5 concentration inversion output by each model are shown in Figure 5, Figure 6, Figure 7 and Figure 8. Overall, the spatial distribution of PM2.5 is patchy and has obvious regional distribution characteristics. Moderate and high concentrations of PM2.5 are mainly distributed in the eastern part of Zhengzhou City, including in Zhongmou County and Xinzheng City. The PM2.5 concentrations in these regions are generally higher than 75 μg/m^3. In other areas, the pollution is relatively low, with PM2.5 values generally less than 75 μg/m^3, and the air quality is good. By comparing the area proportion of PM2.5, five pollution levels inverted by each model were compared with the source data obtained from the PM2.5 stations, as shown in Table 12 and Figure 9. It can be concluded that the 4D-GTWR model has the best inversion effect at fine, moderate and severe pollution levels, but the differences among models at the good and light pollution levels are not obvious. Compared with other models, the 4D-GTWR model inversion predicts a larger range of PM2.5. By comparing the source PM2.5 data, the 4D-GTWR model inversion is shown to predict PM2.5 concentrations that are closer to the true values than those output by other models

Table 11. Parameter coefficients of PM2.5 estimation from each model

Parameter / Model	Intercept	AOD	Relative humidity	Temperature	Wind speed	Wind direction
GWR	36.9816	64.1896	45.2309	-19.2096	0.2958	-17.5103
TWR	49.2732	70.2264	41.6096	-4.2468	4.6244	-36.5351
GTWR	49.0748	70.4542	41.3849	-3.9460	4.3895	-36.3840
4D-GTWR	69.9718	79.6533	43.7280	-21.6426	14.2754	-46.7785

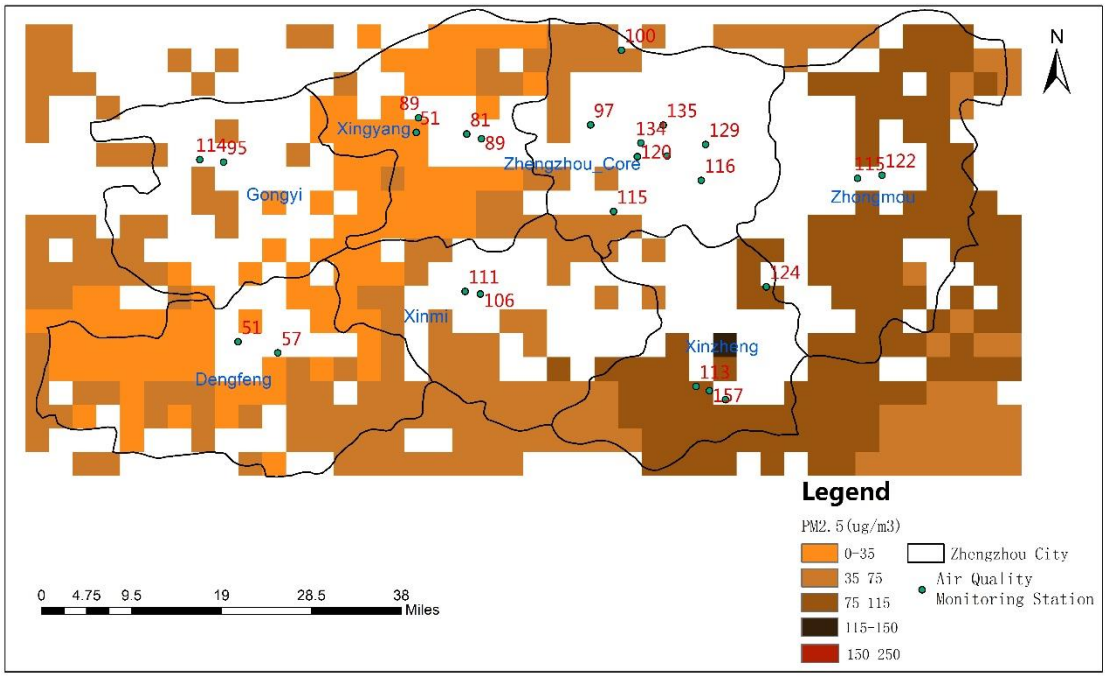


Figure 5. PM2.5 concentration distribution predicted by the GWR model

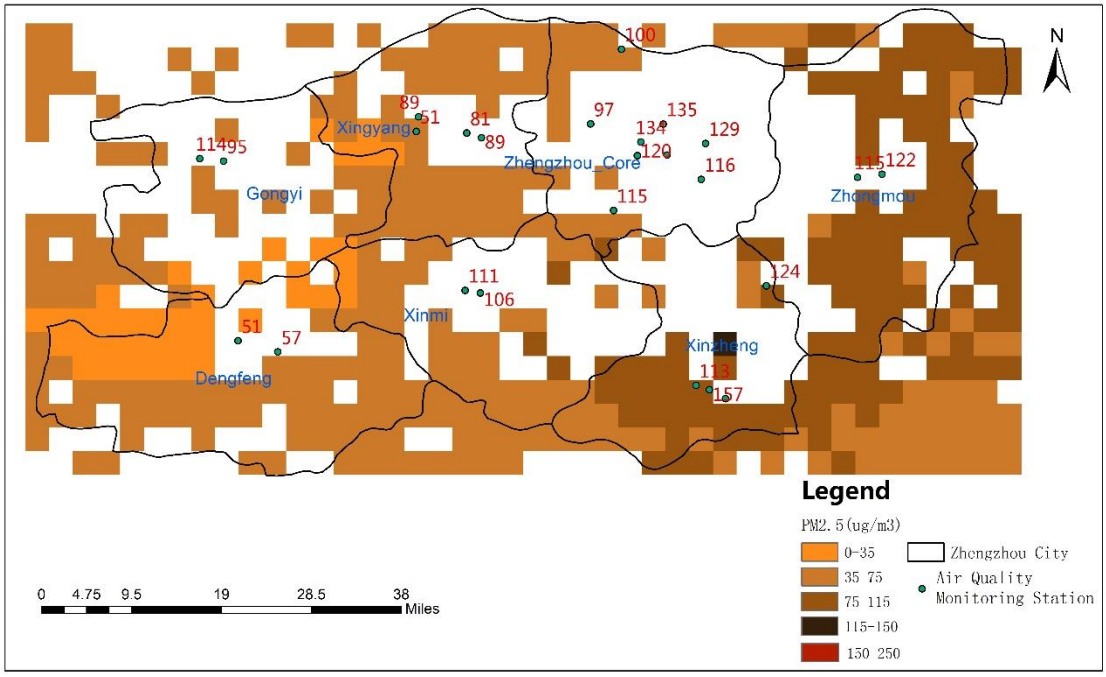


Figure 6. PM2.5 concentration distribution predicted by the TWR model

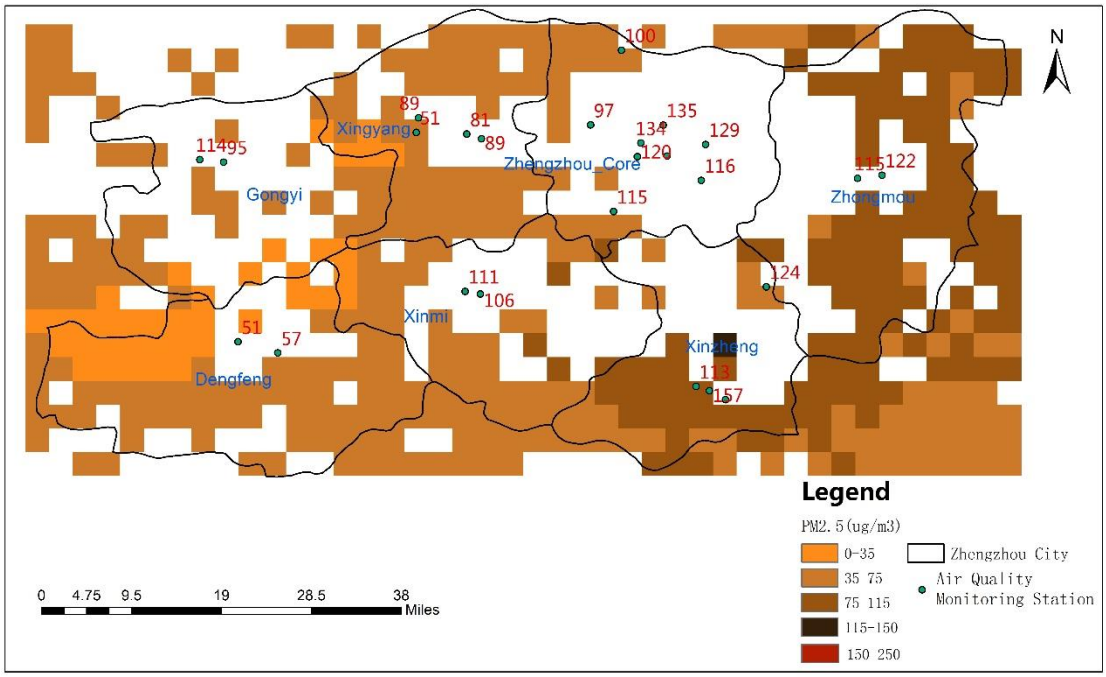


Figure 7. PM2.5 concentration distribution predicted by the GTWR model

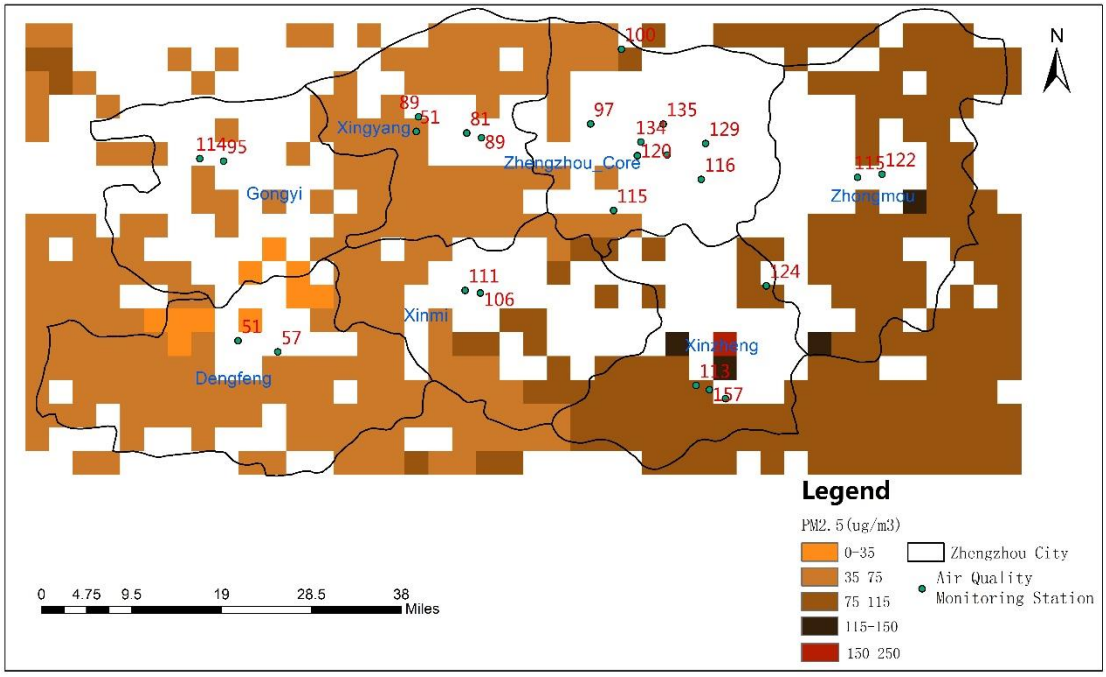


Figure 8. PM2.5 concentration distribution predicted by the 4D-GTWR model

Table 12. Statistical Table of PM2.5 Pollution Level Area Proportion

pollution level / Model	Fine (0-35ug/m^3)	Good (35-75ug/m^3)	Slight pollution (75-115ug/m^3)	Moderate pollution (115-150ug/m^3)	Heavy pollution (150-250ug/m^3)
TWR	23.73%	47%	29.03%	0.23%	0
GWR	8.99%	62.67%	28.11%	0.23%	0

GTWR	8.76%	63.36%	27.65%	0.23%	0
4D-GTWR	2.3%	52.76%	43.78%	0.92%	0.23%

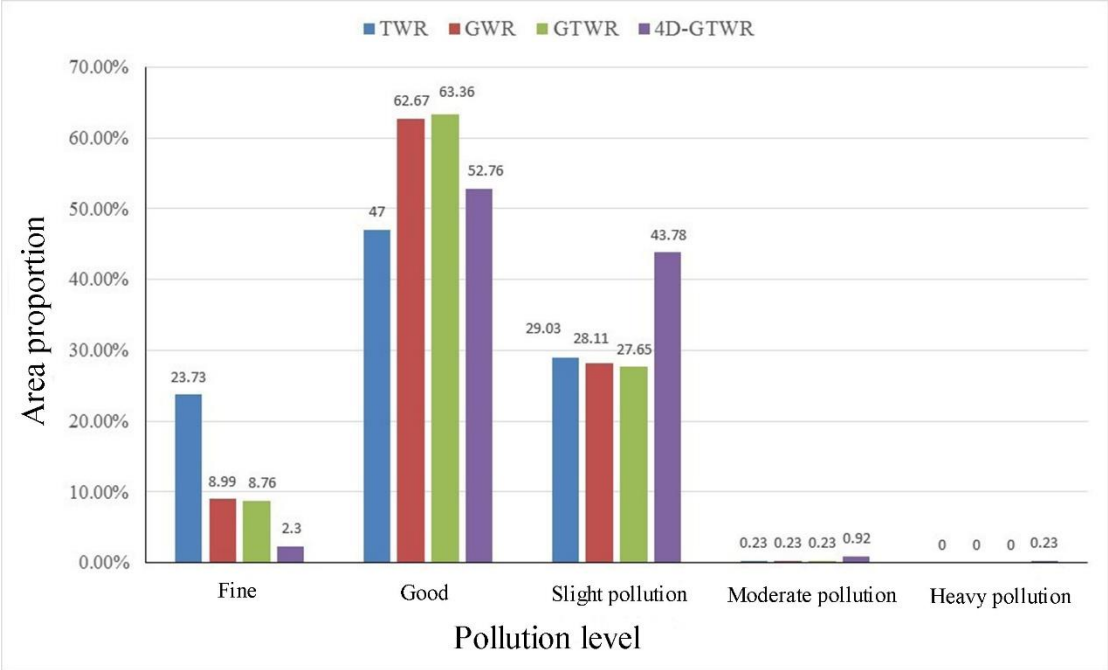


Figure 9 Statistics of the area proportions of PM2.5 pollution levels

Conclusion

In this paper, six variables were selected as the influencing factors of PM2.5 concentrations, including the AOD, relative humidity, temperature, wind speed, wind direction and air pressure. Based on inversion methods used for PM2.5 concentrations in China and globally, GWR, TWR and GTWR models were constructed, and a 4D-GTWR model was proposed. The PM2.5 concentration and its influencing factors in Zhengzhou City, Henan Province, were analyzed as an example. The main conclusions are as follows.

(1) Combined with the basic theory and estimation methods of the GWR, TWR, GTWR and 4D-GTWR models, we can further grasp the spatiotemporal differentiation among different influencing factors. In the analysis and research of PM2.5 concentration inversions, the proposed model can better reflect the impacts of different factors on PM2.5 concentration inversions in the spatial and temporal dimensions with a high goodness of fit.

(2) A four-dimensional geographically and temporally weighted regression method is proposed in this study. The GWR and GTWR models usually use Eurospatial distance measurements. This paper introduces a DEM to construct a four-dimensional spatial distance measurement that can better reflect the four-dimensional spatial distribution characteristics of PM2.5 and provide a more real spatial distance to analyze the actual situation more effectively. Empirical research was carried out on 77 groups of measured data from December 2017 to February 2018 in Zhengzhou City, Henan Province. The goodness of fit R^2, MAE and RMSE values of the models were used as evaluation criteria for analysis and comparison. The

results showed that the 4D-GTWR model was superior to the traditional GTWR and GWR models in which the Euclidean distance is used. In the analysis and study of PM2.5 concentration inversions, the four-dimensional spatiotemporal variation in PM2.5 can be explained more objectively using the proposed model. Therefore, it is meaningful to introduce four-dimensional spatiotemporal nonstationarity into the GWR model, and the 4D-GTWR model provides an effective method for estimating the mass concentration of PM2.5.

However, this study still has some limitations that warrant further study. For example, only the DEM data of 31 sites and PM2.5 concentration data covering 3 months are used in this study, so the four-dimensional spatiotemporal heterogeneity deficiency may lead to the degradation of the performance of the 4D-GTWR model.

Data and codes availability statement

The data and codes that support the findings of this study are available in [“figshare”] with the identifier(s) [data “doi: 10.6084/m9.figshare.9974168”].

Reference

1.Akaike H. A New Look At the Statistical Model Identification. IEEE Trans on Automatic Control, 1975; Volume 19, pp.716-723.

2.Bai Y, Wu L, Qin K, et al. A Geographically and Temporally Weighted Regression Model for Ground-Level PM2.5 Estimation from Satellite-Derived 500 m Resolution AOD. Remote Sensing, 2016; Volume 8,pp.262.

3.Barnes W, Pagano T S, Salomonson V V. Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1. IEEE Trans on Geoscience & Remote Sensing, 1998; Volume 36, pp.1088-1100.

4.Boys B L, Martin R V, Van Donkelaar A , et al. Fifteen-Year Global Time Series of Satellite-Derived Fine Particulate Matter. Environmental Science & Technology, 2014; Volume 48, pp.11109-11118.

5.Chu H J, Huang B, Lin C Y. Modeling the spatio-temporal heterogeneity in the PM10-PM2.5 relationship. Atmospheric Environment, 2015; Volume 102, pp.176-182.

6.Engel-Cox J A, Holloman C H, Coutant B W, et al. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. Atmospheric Environment, 2004; Volume 38, pp.2495-2509.

7.Fotheringham A S, Brunsdon C, Charlton M. Geographically Weighted

Regression: the Analysis of Spatially Varying Relationships. American Journal of Agricultural Economics, 2004; Volume 86, pp.554-556.

8.Huang B, Wu B, Barry M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. International Journal of Geographical Information Science, 2010; Volume 24, pp.383-401.

9.Kaufman Y J, Gao B C. Remote-Sensing of Water-Vapor in the near Ir from Eos/Modis. IEEE Trans Geoscience & Remote Sensing, 1992; Volume 30, pp.871-884.

10.Kumar N, Chu A, Foster A. An empirical relationship between PM2.5 and aerosol optical depth in Delhi Metropolitan. Atmospheric Environment, 2007; Volume 41, pp.4492-4503.

11.Li Z H, Li P, Dong D. Research Progress of Global High Resolution Digital Elevation Models. Geomatics and Information Science of Wuhan University, 2018; Volume 43, pp.1927-1942.

12.Liu Y, Park R J, Jacob D J, et al. Mapping annual mean ground - level PM2.5 concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. Journal of Geophysical Research Atmospheres, 2004; Volume 109, pp.1-10.

13.Mao L, Qiu Y, Kusano C , et al. Predicting regional space–time variation of PM2.5with land-use regression model and MODIS data. Environmental Science and Pollution Research, 2012; Volume 19, pp.128-138.

14.Qiu H, Yu I T, Tian L, et al. Effects of Coarse Particulate Matter on Emergency Hospital Admissions for Respiratory Diseases: A Time-Series Analysis in Hong Kong. Environmental Health Perspectives, 2012; Volume 120, pp.572-576.

15.Rao L L. Estimating Ground-level NO2 Concentrations Based on Geographically and Temporally Weighted Regression Model. Thesis [Master]. China University of Mining and Technology, 2017.

16.Song W, Jia H, Huang J, et al. A satellite-based Geographically Weighted Regression model for regional PM2.5 estimation over the Pearl River Delta region in China. Remote Sensing of Environment, 2014; pp.154:1-7.

17.Wang Z B, Fang C L, Xu G, et al. Spatial-temporal characteristics of the PM2.5 in China in 2014. Acta Geographica Sinica, 2015; Volume 70, pp.1720-1734.

18.Wu H Y, Lu Y H. Aerosol optical thickness in Nanning City determined by inversion calculation based on Dark Target method. Environmental Protection and Technology, 2015; Volume 21, pp.34-53.

19.Yu D L. Modeling Owner-Occupied Single-Family House Values in the City of Milwaukee: A Geographically Weighted Regression Approach. Mapping Sciences and Remote Sensing, 2007; Volume 44, pp.16.

20.Yu, J.Z., Tung, J.W.T., Wu, A.W.M., et al. Abundance and seasonal characteristics of elemental and organic carbon in Hong Kong PM10. Atmospheric Environment, 2004; Volume 38, pp.1511-1521.

21.Zhao Y Y, Liu J P, Yang Y, et al. An approach of co-training geographically and temporally weighted regression to estimate PM2.5 concentration. Science of Surveying and Mapping, 2016; Volume 41, pp.172-178.