Journal of Information and Computing Science
J. Info. Comput. Sci., Vol. 18, No. 2, pp. 117-128

DOI: 10.4208/JICS-2023-008
December 2023

Prediction of $PM_{2.5}$ Concentration in Beijing Based on Bayesian Hierarchical Autoregressive Spatio-Temporal Model

Jing Wang[1] and Chunzheng Cao[1,*]

[1] *School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China*

Abstract. Here, a hierarchical autoregressive spatio-temporal model under the Bayesian framework is proposed to address the simultaneous multi-site $PM_{2.5}$ prediction. The true daily average concentration of $PM_{2.5}$ is regarded as a potential spatio-temporal process, then the temporal correlation is described by the first-order autoregressive process and the spatial correlation is captured based on the Matérn process, which greatly improves the efficiency in dimension reduction and synchronous prediction. In addition, meteorological factors such as daily maximum temperature, relative humidity and wind speed are used as explanatory variables to improve the prediction accuracy. The combination of Bayesian method and MCMC can realize parameter estimation and prediction process due to the model's hierarchical structure. The empirical analysis of daily $PM_{2.5}$ concentration in Beijing shows that the proposed model has good interpolation or prediction performance in both spatial and temporal dimensions.

AMS subject classifications: 62C10, 60J10

Key words: Bayesian method, Hierarchical model, Autoregressive, Spatio-temporal model, $PM_{2.5}$ prediction, Markov Chain Monte Carlo (MCMC).

1 Introduction

As one of the main air pollutants, $PM_{2.5}$, due to its small particle size, can be directly inhaled by the human body, and has a long residence time in the atmosphere and a long transportation distance, so it has a great impact on human health and atmospheric environmental quality. Medical studies have shown that too high concentration of $PM_{2.5}$

Translated from *Journal of Nanjing University of Information Science & Technology*, 2023, 15(1): 34-41.
*Corresponding author. *Email addresses:* **jwang_jane@163.com** (J. Wang), **caochunzheng@163.com** (C. Cao).

will not only lead to an increase in the incidence and mortality of cardiopulmonary diseases [1], but also affect the cardiovascular system, nervous system and immune system of the human body [2-3], and even have toxic effects on genetic materials at different levels such as chromosomes and DNA, causing cancer and birth defects [4-5].

Research on $PM_{2.5}$ includes data collection methods, mechanisms, causes and influencing factors [6-7]. From a statistical point of view, $PM_{2.5}$ concentration in a region over a period of time is regarded as a typical spatio-temporal data set, and relevant research focuses on spatial interpolation and short-term or long-term prediction in time. The space-time Kriging method [8-9] is a popular method for spatial interpolation of $PM_{2.5}$, which can realize linear and unbiased optimal estimation of unobserved locations based on the spatio-temporal position relationship and spatio-temporal variation characteristics of spatio-temporal data, while the prediction of $PM_{2.5}$ in time dimension can be made using mechanism analysis or statistical modeling methods. Mechanism analysis methods mainly model the physicochemical processes of the generation, conversion and diffusion of air pollutants, such as CMAQ model [10]. The statistical modeling method is to capture the characteristics of the data to obtain the change rule of pollutant concentration, including Multivariable Linear Regression (MLR) [11], Generalized Additive Model (GAM) [12-13], as well as various extension models of statistical learning models such as BP neural network [14] and Long Short-Term Memory (LSTM) [15-16]. Compared with mechanism analysis method, statistical method relies less on pollution source data, transmission mode and physical mechanism, and focuses more on the law of data itself. Quantitative analysis has more advantages in accuracy, and is a powerful tool for processing complex data.

In recent years, many studies have focused on the spatio-temporal characteristics and statistical inference of $PM_{2.5}$ concentration. For example, Cheam et al. [17] applied EM algorithm to the inference of parametric spatiotemporal mixed model to cluster air quality data. Based on the semi-parametric spatiotemporal model, Clifford et al. [18] use Gaussian Markov random field to approximate the spatial random effect and non-parametric time trend, and make Bayesian inference to predict the concentration of atmospheric particulate matter. These studies focus more on the flexibility of models and calculations and do not take into account the meteorological variables that play an important role in triggering air pollution. Some studies have also developed spatio-temporal models containing meteorological variables and applied them to spatio-temporal prediction [19-20]. For example, Wan et al. [21] conducted a comprehensive study on $PM_{2.5}$ concentration in Beijing by establishing a fine parametric statistical model, analyzed the spatio-temporal dependent structure of $PM_{2.5}$ concentration and made a prediction. However, when dealing with large-scale data, especially multi-site synchronous prediction, such spatio-temporal models will face excessive computational complexity.

In this paper, a Bayesian Hierarchical Autoregression (BHAR) spatio-temporal model was established for the average daily $PM_{2.5}$ concentration of 35 air quality monitoring points in Beijing, based on the Bayesian framework, stratified model theory

and meteorological factors. The model has three advantages: First, the hierarchical structure is used to describe the clear correlation of variables and the spatial and temporal structure; Second, using Bayesian method can achieve the purpose of parameter estimation and multi-site synchronous prediction at the same time; Third, it can forecast the locations with meteorological information in addition to the existing air quality monitoring points, and solve the problem of sparse distribution of air quality monitoring points in some areas. The BHAR space-time model fits the temporal and spatial correlation simultaneously in the underlying space-time process, and achieves dimensionality reduction, which solves the problem of high computational complexity of the traditional space-time model. Further, with the help of sp-Timer package [22] in R software, Markov Chain Monte Carlo (MCMC) algorithm is used to estimate and predict parameters of the model.

2 Preliminary data analysis

The research area of this paper is Beijing, located at $115.7°\sim117.4°E$, $39.4°\sim41.6°N$, with a high terrain in the northwest and low terrain in the southeast. The west, north and northeast are surrounded by mountains on three sides, and the southeast plain gradually slopes toward the Bohai Sea, as shown in Figure 1.

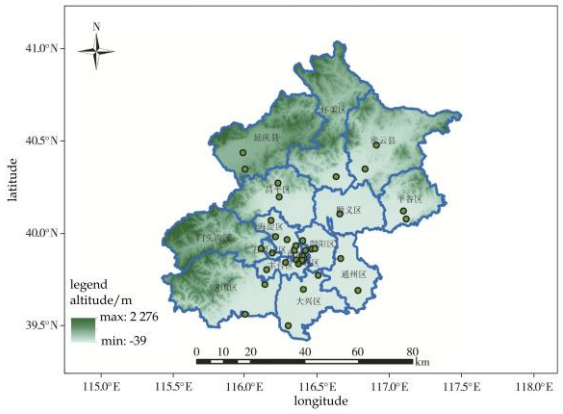

Figure 1: Map of Beijing's topography and its air quality monitoring stations (green dots)

Beijing had 27 air quality monitoring points, but eight $PM_{2.5}$ monitoring points were added in 2012. Since the air quality release platform (http://zx.bjmemc.com.cn) of Beijing Municipal Ecological and Environmental Monitoring Center updated the names of air quality monitoring points in Beijing from January 23, 2021, the monitoring points were re-classified according to the six districts of the city, the southeast, the northeast, the southwest and the northwest (Table 1). At the same time, considering that $PM_{2.5}$ pollution mostly occurs in winter [23], this paper collected $PM_{2.5}$ mass concentration ($\mu g/m^3$) data of 24 h per day from February 1, 2021 to March 31, 2021 from 35 monitoring sites in Beijing for subsequent analysis. Based on the collected data, this paper models and predicts the 24 h daily average mass concentration of $PM_{2.5}$(daily average mass

concentration).

Table 1: Air quality monitoring stations in Beijing

Region	Monitoring point
City 6 Districts	Dongcheng Dongfourth, Dongcheng Temple of Heaven, Xicheng Official Garden, Xicheng, Longevity West Palace, Chaoyang Olympic Sports Center, Chaoyang Agriculture Exhibition Hall
	Haidian Wanliu, Haidian Sijiqing, Fengtai small tun, Fengtai Yungang, Shijingshan Ancient city, Shijingshan Old Mountain
Southeast	Tongzhou Yongshun, Tongzhou Dongguan, Daxing Huang Village, Daxing Old Palace, Yizhuang Development Zone, southeast of Beijing regional point
Northeast	Huairou Town, Huairou New Town, Miyun Town, Miyun New Town, Pinggu Town, Pinggu New Town, Shunyi New Town, Shunyi North Xiaoying
Northwest	Changping Town, Changping Nanshao, Dingling (comparison point), Yanqing Xiadu, Yanqing Shihe Camp

In order to preliminarily explore the temporal and spatial distribution characteristics of $PM_{2.5}$ concentration in Beijing, the average $PM_{2.5}$ mass concentration of each hour in a day and each day in a week at all stations was calculated by hour, and then the box plot was drawn. It can be observed from Figure 2a that the mass concentration of $PM_{2.5}$ shows a trend of first decreasing and then rising throughout the day, gradually decreasing from early morning until 14:00, which may be related to the terrain and winter climate conditions, which lead to thermal inversion and weak wind in the morning and evening, preventing the diffusion of pollutants [24]. At the same time, it can also be found that the mass concentration of $PM_{2.5}$ has a small increase in the morning and evening peak hours, which may be affected by automobile exhaust. Figure 2b shows that the variation of $PM_{2.5}$ in a week is also regular. With the continuous improvement of human activities in a week, the $PM_{2.5}$ mass concentration also gradually increases, while on Mondays and rest days, the $PM_{2.5}$ mass concentration is significantly lower.

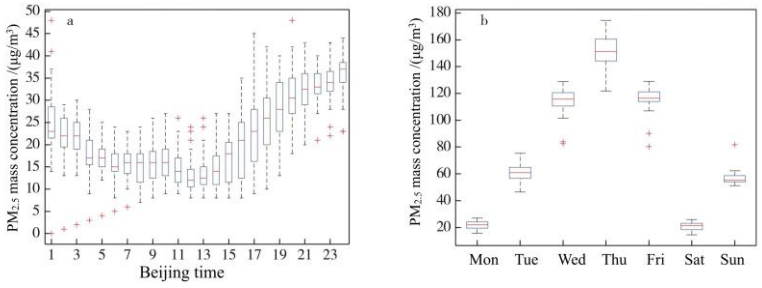

Figure 2: Diurnal (a) and weekly (b) variation of $PM_{2.5}$ concentration averaged by 35 air quality monitoring stations in Beijing from February to March 2021

Next, the spatial distribution characteristics of $PM_{2.5}$ concentration in Beijing are analyzed. First, the spatial distribution map of the average $PM_{2.5}$ mass concentration in Beijing in February and March 2021 was drawn. It can be observed from Figure 3 that the mass concentration of $PM_{2.5}$ in the northern mountainous area is relatively low, among which Yanqing County has the lowest concentration, which may be affected by the valley topography, while the high mass concentration of $PM_{2.5}$ in the central urban area is

obviously reasonable.

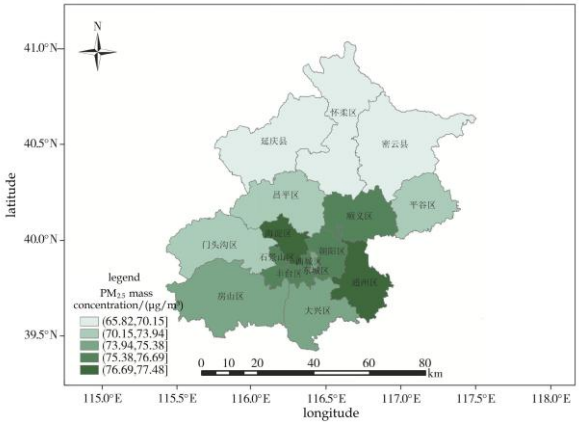

Figure 3: Distribution of average $PM_{2.5}$ concentration in Beijing from February to March 2021

Then the spatial autocorrelation of $PM_{2.5}$ mass concentration distribution was analyzed with the help of global Moreland index. The global Moran index formula is

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)\sum_{i=1}^{n}(y_i-\bar{y})^2}, \quad (1)$$

where w_{ij} is the distance weight, representing the weighted distance between sites i and j, and y_i represents the $PM_{2.5}$ mass concentration value. After calculation, the global Molan index $I = 0.2$ of $PM_{2.5}$ mass concentration is positive, and the P-value of significance test is 3.9×10^{-6}, indicating that there is a significant positive correlation and spatial aggregation of $PM_{2.5}$ mass concentration distribution in 35 air quality monitoring points in Beijing from February to March.

Further collected daily wind speed (m/s), relative humidity (%), and maximum temperature (˚C) of three meteorological stations in Yanqing, Miyun, and Beijing in the Beijing area from February 1, 2021 to March 31, 2021, based on the data collected from the China Meteorological Data Network (http://data.cma.cn). The three meteorological variables are briefly recorded as WS, RH and MT respectively, and detailed summary information is shown in Table 2. In order to analyze the correlation between meteorological variables and $PM_{2.5}$ mass concentration, the three weather stations were first matched with the nearest air quality monitoring points. The three air quality monitoring points matched with Yanqing, Miyun and Beijing meteorological stations are Yanqing Xiadu, Huairou New City and Daxing Huangcun. Spearman correlation coefficients of $PM_{2.5}$ mass concentration and three meteorological variables were calculated respectively, and the results are shown in the last column of Table 2. The results show that there is a strong positive correlation between the relative humidity and $PM_{2.5}$. The relative humidity in Beijing is relatively low in February-March, when the chemical polymerization of air pollutants causes the increase rate of $PM_{2.5}$ to be higher

than the decrease rate of $PM_{2.5}$ caused by sedimentation. Relative humidity has a positive effect on $PM_{2.5}$, that is, the mass concentration of $PM_{2.5}$ increases with the increase of relative humidity. Wind speed was negatively correlated with $PM_{2.5}$, while temperature was weakly correlated with $PM_{2.5}$.

Table 2: Summary of and correlation coefficients between daily $PM_{2.5}$ concentrations and meteorological variables

variable	mean value	minimum	maximum	Spearman correlation coefficient
$\rho(PM_{2.5})/(\mu g/m^3)$	74.20	3.25	296.42	
WS /(m/s)	1.85	0.40	4.80	-0.44***
RH /%	50.69	15.30	91.00	0.68***
MT /°C	12.05	-0.20	25.60	0.26***

Note :*** means $p < 0.001$.

3 BHAR space-time model

3.1 Model Establishment

Assume that $Z(s,t)$ represents the actual observed $PM_{2.5}$ mass concentration of the station s at time t, and the corresponding true concentration value is described by a potential random process $Y(s,t)$, both of which satisfy the following measurement error model:

$$Z(s,t) = X^T(s,t)\beta + Y(s,t) + \varepsilon(s,t), \tag{2}$$

where: $s = s_1, s_2, \cdots, s_n$ is the geographical location of n sites; $t = 1,2, \cdots, T$ is time (d); $X(s,t)$ represents the p-dimensional meteorological variable, i.e. $X(s,t) = \left(x_1(s,t), x_2(s,t), \cdots, x_p(s,t)\right)^T$; β is the regression coefficient; $\varepsilon(s,t)$ is the error term and is usually assumed to be a white noise process, i.e. $\varepsilon(s,t) \sim \mathrm{GP}(0, \sigma_\varepsilon^2)$. In spatial statistics, σ_ε^2 is often referred to as the nugget value. When the distance of sampling points is 0, the semi-variance function value should also be 0. However, due to measurement error and spatial variation, when the two sampling points are very close, the corresponding semi-variance function value is not 0, that is, the nugget value exists.

Establish a first-order autoregressive model for the potential pollutant emission level $Y(s,t)$ [22]:

$$Y(s,t) = \rho Y(s,t-1) + \eta(s,t), \tag{3}$$

where $\eta(s,t)$ is a residual random term used to describe the spatiotemporal random effects of potential pollutant emission levels. It is assumed that $\eta(s,t)$ is independent in time but satisfies the Gaussian process $\mathrm{GP}(0, \Sigma_\eta)$ in space, where $\Sigma_\eta = \sigma_\eta^2 S_\eta$, σ_η^2 is the variance that does not vary with space, S_η represents the spatially dependent covariance

matrix, which is usually described by the Matern family correlation function [25]. At this time, the covariance matrix of $\eta(s,t)$ is $n \times n$ dimension instead of $nT \times nT$ dimension, which realizes dimension reduction and simplifies the calculation. The general form of the Matérn family correlation function is

$$\kappa(u;\varphi,v) = \frac{1}{2^{v-1}\Gamma(v)}\left(2\sqrt{v}u\varphi\right)^{v} K_v\left(2\sqrt{v}u\varphi\right), \varphi > 0, v > 0, \quad (4)$$

In the command, $u = \|s_i - s_j\|$ indicates the distance between the monitoring point s_i and s_j. In this case, the Euclidean distance is selected. φ is used to control the decay rate of the spatial correlation, i.e. the greater the distance u, the faster the decay rate. v is the parameter controlling the smoothness degree; K_v is the second Bessel function of the v order. When $v = 0.5$, the Matérn family correlation function degenerates to an exponential correlation function, i.e. $\kappa(u;\varphi) = exp(-\varphi u)$; when $v = 3/2$, $\kappa(u;\varphi) = (1+\varphi u)\, exp(-\varphi u)$; when $v \to \infty$, the Matérn family correlation function degenerates into a Gaussian process function, i.e. $\kappa(u;\varphi) = exp(-\varphi^2 u^2)$.

In summary, for the measured data, the structure of the BHAR space-time model is as follows:

$$Z_t = X_t\beta + Y_t + \varepsilon_t, \quad (5)$$

$$Y_t = \rho Y_{t-1} + \eta_t, \quad (6)$$

In the command, $Z_t = \left(Z(s_1,t),\cdots,Z(s_n,t)\right)^T$, $Y_t = \left(Y(s_1,t),\cdots,Y(s_n,t)\right)^T$, $\varepsilon_t = \left(\varepsilon(s_1,t),\cdots,\varepsilon(s_n,t)\right)^T$, $\eta_t = \left(\eta(s_1,t),\cdots,\eta(s_n,t)\right)^T$, and $\varepsilon_t \sim N(0,\sigma_\varepsilon^2 I_n)$, $\eta_t \sim N\left(0,\sigma_\eta^2 S_\eta\right)$. According to the hierarchical model structure, the BHAR spatio-temporal model can be divided into three layers: the first layer represents the distribution of the original data under the conditions of given potential spatio-temporal processes and parameters; the second layer represents the distribution of potential processes given the parameters $Y_t|\Theta$; the third layer represents the prior distribution of the introduced parameters or hyperparameters. The processes in the second layer can add different levels of interpretation [26], the first level represents the real potential process, and the second level describes the spatiotemporal random effects of the potential process.

3.2 Parameter estimation and prediction

The parameter to be estimated in the BHAR space-time model is $\Theta = \left\{\beta,\rho,\sigma_\varepsilon^2,\sigma_\eta^2,\varphi,v\right\}$, which is estimated by MCMC method. All other parameters except ϕ and ν have conjugate prior distributions, $\beta,\rho,\sigma_\varepsilon^2,\sigma_\eta^2$ can be obtained after the given prior distributions, and the parameters are further estimated by Gibbs sampling method. Fixed $v = 0.5$, the Metropolis-Hastings (MH) algorithm is used to estimate φ.

The prediction of $Z(s,t)$ can be divided into three categories: one is to predict the value of unknown monitoring point s_0 at known time t; the second is to predict the value of known monitoring point s at unknown time point t_0; the third is to predict the value of unknown monitoring point s_0 at unknown time point t_0. The first kind of prediction is spatial interpolation, and the second and third kinds of prediction are in time.

Firstly, spatial interpolation is introduced. At unknown monitoring point s_0,

equation (1) can be used to obtain:

$$Z(s_0,t) = X^T(s_0,t)\beta + Y(s_0,t) + \varepsilon(s_0,t), \quad (7)$$

where $Y(s_0,t) = \rho Y(s_0,t-1) + \eta(s_0,t)$. Obviously, $Y(s_0,t)$ can only be determined by the $Y(s_0,\cdot)$ order of all time points prior to t, and includes $Y(s_0,0)$. $Y(s_0,0)$ can be calculated based on the prior distribution of the initial condition Y_0. Of course, if Y_0 is specified as a fixed constant, then $Y(s_0,0)$ can also be thought of as the same constant [19], so for simplicity, Y_0 is usually chosen as a fixed value.

The prediction of $Z(s_0,0)$ is generally based on the posterior distribution $\pi(Z(s_0,t)|Z)$, which can be obtained by integrating the joint posterior distribution:

$$\pi(Z(s_0,t)|Z) = \int \pi(Z(s_0,t)|Y(s_0,t),\sigma_\varepsilon^2) \times \pi(Y(s_0,t)|\Theta,Y) \times \pi(\Theta,Y|Z)dY(s_0,t)dYd\Theta, \quad (8)$$

In the above formula, Z and Y represent the values of the known time t and the monitoring point s respectively. The estimate of the predicted value $Z(s_0,0)$ was obtained by the MCMC component sampling method as follows:

1) Random sample $\Theta^{(j)}$, $Y^{(j)}$ from the posterior distribution $\pi(\Theta,Y|Z)$;

2) Sample $Y^{(j)}(s_0,t)$ from the posterior distribution $\pi\left(Y(s_0,t)\middle|\Theta^{(j)},Y^{(j)}\right)$;

3) Sample $Z^{(j)}(s_0,t)$ is extracted from the posterior distribution $\pi\left(Z(s_0,t)\middle|Y^{(j)}(s_0,t),\sigma_\varepsilon^{2(j)}\right)$.

In terms of time dimension, the prediction process is similar to the spatial interpolation process. For a certain site s (including existing monitoring points or any designated position as the monitoring point), the forward time prediction can also be realized based on the posterior distribution of $Z(s,T+1)$ according to the MCMC sampling method similar to spatial interpolation. The main difference with spatial interpolation is that the prediction in the time dimension needs to simulate $Y(s,T+1)$ from a marginal distribution with zero mean and variance $\sigma^2\eta S_\eta$, rather than a conditional distribution. Since all the information at the observation point has been used to obtain $Y(s,T)$ from time *0* to time *T*, at future time $T+1$, there is no new information available for the conditional distribution except for the new value of the regression term $X(s,T+1)$.

4 Case Analysis

Combined with the spatial distribution of 35 air quality monitoring points in Beijing, 9 of them were selected as the spatial verification set, and the two days of March 30 and March 31, 2021 were selected as the time verification set, and the data of the remaining 26 monitoring points were used as the training set to fit the model. The R software package spTimer is used to realize the calculation process of parameter estimation and prediction simultaneously. From the parameter estimation table (Table 3), it can be seen that the 95% confidence interval of the estimation of regression coefficients β_1 and β_3 for the two variables WS and MT contains zero points, so it is not significant. Among the meteorological variables, only the regression coefficient β_2 of relative humidity RH is significant and positive, which is consistent with the results of preliminary data analysis,

and further indicates that the relative humidity RH in Beijing has a significant positive impact on $PM_{2.5}$ from February to March.

Table 3: MCMC parameter estimation of BHAR model

parameters	mean value	median	standard deviation	95% confidence interval
β_0	3.5571	3.5468	0.6790	[2.2342,4.9009]
β_1	0.1084	0.1089	0.0759	[-0.0399,0.2542]
β_2	0.0174	0.0174	0.0078	[0.0016,0.0329]
β_3	-0.0003	-0.0007	0.0253	[-0.0503,0.049]
ρ	0.4144	04146	0.0234	[0.3678,0.4596]
σ_ε^2	0.0063	0.0063	0.0003	[0.0058,0.0069]
σ_η^2	5.9494	5.8522	0.8581	[4.5854,7.7547]
ϕ	0.0020	0.0020	0.0003	[0.0015,0.0026]

In order to evaluate the prediction performance of BHAR spatio-temporal model, two measurement indexes, root mean square error RMSE and mean absolute error MAE, were used to compare the error between the predicted data and the original data. The formula for RMSE and MAE is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Z}_i - Z_i\right)^2}, \tag{9}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{Z}_i - Z_i\right|. \tag{10}$$

Firstly, the spatial interpolation of 9 monitoring points of the spatial verification set is carried out, and the BHAR space-time model can realize the synchronous prediction of 9 monitoring points. For comparison, the gam function in the mgcv package of R software was further used to fit the GAM model to the training set data, predict the $PM_{2.5}$ mass concentration of each station in the spatial verification set, and calculate the above two measurement indicators at the same time. Sorek-Hamer et al. [12] improved the prediction effect of $PM_{2.5}$ mass concentration by taking advantage of GAM's ability to fit the nonlinear relationship between the explanatory variable and the explained variable. The comparison results are shown in Table 4. It can be seen that both RMSE and MAE of the BHAR space-time model are about 1/3 of that of GAM, which proves that the spatial interpolation effect of the BHAR space-time model proposed in this paper is consistently superior to that of GAM.

Further, synchronous time prediction was made for the monitoring points in the spatial verification set and the training set respectively, and the average daily $PM_{2.5}$ mass concentration on March 30 and 31 was predicted. LSTM was selected as the main comparison model, and the conventional ARIMA model was selected as the baseline comparison model. The obtained measurement indicators are listed in Table 5. It can be seen that ARIMA model has the worst prediction effect, followed by LSTM model, and BHAR model has the best time prediction effect. The BHAR spatio-temporal model

proposed in this paper is used to model all monitoring points as a whole, and the correlation between space and time is fully considered, so the prediction results with high accuracy are obtained.

Table 4: Comparison of spatial interpolation performance

Model	RMSE	MAE
BHAR	12.45	8.16
GAM	34.8	24.97

Table 5: Comparison of prediction performance in time dimension

Model	RMSE	MAE
BHAR	10.12	8.68
LSTM	12.22	11.48
ARIMA	26.81	24.85

5 Summary

The BHAR spatiotemporal model established in this paper takes the $PM_{2.5}$ data of a region as a spatial process of time series, fits the temporal and spatial correlation characteristics of $PM_{2.5}$ mass concentration on the whole, and realizes the short-term prediction function of $PM_{2.5}$ spatial interpolation and time for specific sites, and the prediction effect is better than that of GAM and LSTM. This model is not only suitable for the prediction of $PM_{2.5}$ mass concentration, but also can be extended to other air quality ground monitoring data, such as PM_{10} and O_3 concentrations. Modeling under the Bayesian framework is more inclusive to the uncertainty of the model, and the prior distribution given in advance can fuse the expert knowledge and improve the prediction accuracy. Further, the establishment of a hierarchical model can more clearly depict the underlying space-time process inside the data, and enhance the interpretability of the model. At the same time, the hierarchical structure of the model also makes the inference process such as parameter estimation and prediction more convenient.

In this paper, wind speed, humidity and temperature are selected as explanatory variables to improve the actual forecasting effect of the model. In order to simplify the model, fixed coefficient is used in this paper, but the impact of meteorological variables on $PM_{2.5}$ may vary with time and space, so the fitting variable coefficient model will be considered in the subsequent study. In this paper, the first order autoregression is used to describe the correlation in time dimension, which achieves good numerical results and reduces the computational complexity. In practical application, we can select the appropriate autoregressive order according to the data characteristics and model selection method. In addition, the Matern kernel function used in this model to describe

spatial correlation adopts homogeneous Euclidean distance. With more geographical details, non-homogeneous Euclidean distance or other non-Euclidean distance can be considered to capture more real spatial correlation.

Acknowledgments

This work is supported by the Natural Science Foundation of Jiangsu Province (Grant No. BK20191394) and Major Project of National Social Science Foundation (Grant No. 16ZDA-047).

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] C. A. III. Pope, R. T. Burnett, M. J. Thun, et al., Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *JAMA*, 2002, 287(9): 1132-1141.

[2] Y. M. Guo, Y. P. Jia, X. C. Pan, et al., The association between fine particulate air pollution and hospital emergency room visits for cardiovascular diseases in Beijing, China, *Sci. Total Environ.*, 2009, 407(17): 4826-4830.

[3] Y. M. Guo, L. Q. Liu, J. M. Chen, et al., Association between the concentration of particulate matters and the hospital emergency room visits for circulatory diseases: a case-crossover study, *Chin. J. Epidemiol.*, 2008, 29(11): 1064-1068.

[4] A. Valavanidis, K. Fiotakis, and T. Vlachogianni. Airborne particulate matter and human health: toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms, *J. Environ. Sci. Health, Part C*, 2008, 26(4): 339-362.

[5] J. M. Samet, D. M. DeMarini, and H. V. Malling. Do airborne particles induce heritable mutations? *Science*, 2004, 304(5673): 971-972.

[6] S. Guo, M. Hu, M. L. Zamora, et al., Elucidating severe urban haze formation in China, *Proc. Natl. Acad. Sci. USA*, 2014, 111(49): 17373-17378.

[7] Y. L. Sun, Z. F. Wang, W. Du, et al., Long-term real-time measurements of aerosol particle composition in Beijing, China: seasonal variations, meteorological effects, and source analysis, *Atmos. Chem. Phys.*, 2015, 15(17): 10149-10165.

[8] Y. M. Lu, L. Wang, A. G. Qiu, et al., PM2.5 spatial interpolation method based on local weighted linear regression model, *Sci. Surv. Mapp.*, 2018, 43(11): 79-84, 91.

[9] P. Gething, P. Atkinson, A. Noor, et al., A local space-time Kriging approach applied to a national outpatient malaria dataset, *Comput. Geosci.*, 2007, 33(10): 1337-1350.

[10] D. Byun, and K. L. Schere. Review of the governing equations, computational algorithms, and

other components of the models-3 community multiscale air quality (CMAQ) modeling system, *Appl. Mech. Rev.*, 2006, 59(2): 51-77.
[11] Y. R. Li, J. X. Wang, T. T. Han, et al., Using multiple linear regression method to evaluate the impact of meteorological conditions and control measures on air quality in Beijing during APEC 2014, *Environ. Sci.*, 2019, 40(3): 1024-1034.
[12] M. Sorek-Hamer, A. W. Strawa, R. B. Chatfield, et al., Improved retrieval of PM2.5 from satellite data products using non-linear methods, *Environ. Pollut.*, 2013, 182C: 417-423.
[13] S. Yu, G. N. Wang, L. Wang, et al., Estimation and inference for generalized geoadditive models, *J. Am. Stat. Assoc.*, 2020, 115(530): 761-774.
[14] Y. Bai, Y. Li, X. X. Wang, et al., Air pollutants concentrations forecasting using back propagation, *Atmos. Pollut. Res.*, 2016, 7(3): 557-566.
[15] Q. P. Zhou, H. Y. Jiang, J. Z. Wang, et al., A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network, *Sci. Total Environ.*, 2014, 496: 264-274.
[16] S. N. Bai, and X. L. Shen, PM2.5 prediction based on LSTM recurrent neural network, *Comput. Appl. Softw.*, 2019, 36(1): 67-70, 104.
[17] A. S. M. Cheam, M. Marbac, and P. D. McNicholas, Model-based clustering for spatiotemporal data on air quality monitoring, *Environmetrics*, 2017, 28(3): e2437.
[18] S. Clifford, S. Low-Choy, M. Mazaheri, et al., A Bayesian spatiotemporal model of panel design data: airborne particle number concentration in Brisbane, Australia, *Environmetrics*, 2019, 30(7): e2597.
[19] O. Nicolis, M. Díaz, S. K. Sahu, et al., Bayesian spatiotemporal modeling for estimating short-term exposure to air pollution in Santiago de Chile, *Environmetrics*, 2019, 30(7): e2574.
[20] L. Padilla, B. Lagos-Álvarez, J. Mateu, et al., Space-time autoregressive estimation and prediction with missing data based on Kalman filtering, *Environmetrics*, 2020, 31(7): e2627.
[21] Y. T. Wan, M. Y. Xu, H. Huang, et al., A spatio-temporal model for the analysis and prediction of fine particulate matter concentration in Beijing, *Environmetrics*, 2021, 32(1): e2648.
[22] K. S. Bakar, and S. K. Sahu. spTimer: spatio-temporal Bayesian modeling using R, *J. Stat. Softw.*, 2015, 63(15): 1-32.
[23] L. S. Liang, J. L. Jing, A. N. Wang, et al., Spatial-temporal distribution characteristics of PM2.5 concentrations in Beijing-Tianjin-Hebei region in winter from 2014 to 2019, *J. Guilin Univ. Technol.*, 2020, 40(4): 788-797.
[24] C. Wang, Y. Shi, Y. Jing, et al., Spatial and temporal distribution characteristics of PM2.5 in Beijing-Tianjin-Hebei region based on remote sensing data, *Adm. Technol. Environ. Monit.*, 2020(1): 37-41.
[25] M. S. Handcock, and M. L. Stein, A Bayesian analysis of Kriging, *Technometrics*, 1993, 35(4): 403-410.
[26] A. E. Gelfand. Hierarchical modeling for spatial data problems, *Spatial Stat.*, 2012, 1: 30-39.