Journal of Data Science,17(1). P. 37-54 ,2019
DOI:10.6339/JDS.201901_17(1).0002

A SPACIAL-TEMPORAL GAUSSIAN MIXTURE MODEL FOR ANNUAL AVERAGE $PM_{2.5}$ CONCENTRATION ANALYSIS

Chenyang Shi*, Puntipa Wanitjirattikal[2]

**Celgene Corporation, USA*

[2]*King Mongkut's Institute of Technology Ladkrabang, Thailand*

Abstract

$PM_{2.5}$ is a major air pollutant which has a high probability to cause many serious cardiopulmonary diseases, such as asthma, lung cancer, trachea cancer, bronchus cancer, etc. Up to 2014, a World Health Organization (WHO) air quality model confirmed that 92% of the population in the world lived in areas where air quality levels exceeded WHO limits (i.e., 10 $\mu g/m^3$). This indicates that $PM_{2.5}$ is still one of the most serious world-wide problems, and monitoring $PM_{2.5}$ concentrations is extremely necessary. In this paper, we proposed a easy and flexible spatial-temporal Gaussian mixture model to analyze annual average $PM_{2.5}$ concentrations. Because of the bimodal distribution of $PM_{2.5}$ concentrations, we decided for a two- component Gaussian mixture model with county-year-level spatial-temporal random effects. A Markov Chain Monte Carlo (MCMC) algorithm is used to estimating model parameters.

Keywords: Conditional autoregressive prior, Normal mixture model, $PM_{2.5}$ concentration, Spatial-Temporal random effect.

* Corresponding author: henryshichina@gmail.com

1. Introduction

Fine particles with a diameter of 2.5 μm or less ($PM_{2.5}$) is a major air pollutant which has a high probability to cause many serious cardiopulmonary diseases, such as asthma, lung cancer, trachea cancer, bronchus cancer, etc. (Monn & Becker, 1999; Cohen et al., 2005). And around 3% mortality from cardiopulmonary diseases is strongly associated with $PM_{2.5}$ (Cohen et al., 2005). Although much effort has been put into lowering $PM_{2.5}$ concentration, up to 2014, a World Health Organization (WHO) air quality model confirmed that 92% of the population in the world still lived in areas where air quality levels exceeded WHO limits (i.e., 10 $\mu g/m^3$). This indicates that $PM_{2.5}$ is still one of the most serious world-wide problems, and monitoring $PM_{2.5}$ concentration is extremely necessary.

Statistical analysis is playing a very important role in monitoring $PM_{2.5}$ concentration. So far, most statistical techniques to analyze $PM_{2.5}$ are performed revolving around two parts: 1.) Specifying the distribution of $PM_{2.5}$ data. 2.) Analyzing the spatial or temporal effects of $PM_{2.5}$. For the first part, since the distributions of $PM_{2.5}$ concentrations may differ for different regions or times, many different methods are used. Antonovsky et al. (1991) found that their data of air pollution in Borovo has multi-modal distribution, so a normal mixture model was fitted. A similar model is also used by Chu et al. (2012). Fuentes (2003) used a Bayesian model to interpolate ground measurements of pollution levels from 513 sites throughout the eastern USA. Karaca et al. (2005) used Log-logistic functions to monitor PM_{10} and $PM_{2.5}$ concentrations at a suburban site of Istanbul, Turkey. Vidale et al. (2017) used a generalized additive model to analyze the association between air pollution exposure and cardiovascular events in Como, Italy. Tian & Chen (2010) developed a semi-empirical model for predicting hourly ground-level $PM_{2.5}$ concentration in southern Ontario. Brown et al. (1994) developed and applied a multivariate approach to the spatial interpolation for analyzing air pollutant in southern Ontario, Canada. Karppinen et al. (2004) utilized a linear interpolated value in their linear regression for $PM_{2.5}$ in the City of Helsinki, Finland. Pérez et al. (2000) compared the predictions produced by multilayer neural networks, linear regression and persistence based on the data of $PM_{2.5}$ in Santiago, Chile. They found that the neural network gives the best results.

For the second part, Delamater et al. (2012) developed a Bayesian model with a temporal random effects to analyze the impact of $PM_{2.5}$ no asthma hospitalization rates in Los Angeles County. Tai et al. (2012) applied a multiple geography - based chemical transport model to understand the relationships between $PM_{2.5}$ and climate change in the United States. Zhan et al. (2017) developed a geographically-weighted gradient boosting machine by building spatial smoothing kernels to weigh the loss function for predicting $PM_{2.5}$ concentrations in China. Liu et al. (2012) developed a linear model with smooth regressions for temporal variables to evaluate the effectiveness of $PM_{2.5}$ emissions control in Beijing, China. Similar methods can be found in Li et al. (2017) and Peng et al. (2006). Wang & Fang (2016) analyzed $PM_{2.5}$ in Bohai rime, Chine, with a spatial-temporal model. They set the parameters of covariates as functions of spatial coordinates. Ma et al. (2016) developed a spatial econometric model based on spatial autoregressive model to analyze the relationship between $PM_{2.5}$ and GDP in China. Russell et al. (2017) analyzed $PM_{2.5}$ in eastern United States using a local linear penalized quantile regression.

From all above, we can see that a lot of statistical researches have been done for monitoring and analyzing $PM_{2.5}$ in many regions. Inspired by these researches, two questions occurred in our mind: 1). Can we find a random effect which can detect both of spatial random effect and temporal correlation? 2). Population and income per capita are two important social factors in air pollution studies (Ma et al., 2016; Wang & Fang, 2016), how do they influence $PM_{2.5}$ in Michigan? Motivated by these two questions, we proposed a spatial-temporal Gaussian mixture model. We analyzed a Michigan annual average $PM_{2.5}$ concentrations (2007 ~ 2011) data set, and found that the $PM_{2.5}$ concentrations in each year has a bimodal distribution. So we decided for a two-component Gaussian mixture model for our data. For spatial-temporal random effect, either a conditional autoregressive (CAR) prior or a multivariate CAR (MCAR) prior is imposed on. We adopted a deviation information criteria and developed a specific posterior predictive check for model selection and goodness-of-fit. A Markov Chain Monte Carlo (MCMC) algorithm for model parameter estimation was implemented in Winbugs 1.4.3
(http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs- project-winbugs/).

2. Data Description

Our data set contains annual average $PM_{2.5}$ concentrations, population, income per capita, and county area from 2007 to 2011 for each county in Michigan, United States. Annual average $PM_{2.5}$ concentrations, population, and income per capita vary in county and year, that is, each of annual average $PM_{2.5}$ concentrations, population, and income per capita forms a
83×5 matrix with rows for counties in Michigan, and columns years from 2007 to 2011. County areas do not vary in year. Our $PM_{2.5}$ concentrations are derived from https://www.data.gov/. Population and income per capita are derived from http://milmi.org/.

Histograms for annual average $PM_{2.5}$ concentrations in each year are shown in Figure 1. These five years have close minimum $PM_{2.5}$ concentrations, but the maximum $PM_{2.5}$ concentrations decrease to 11.48 μg/m^3 from 14.66 μg/m^3. All $PM_{2.5}$ concentrations are bimodal shaped, which implies that it is reasonable to fit a mixture model on this data (Antonovsky et al., 1991).

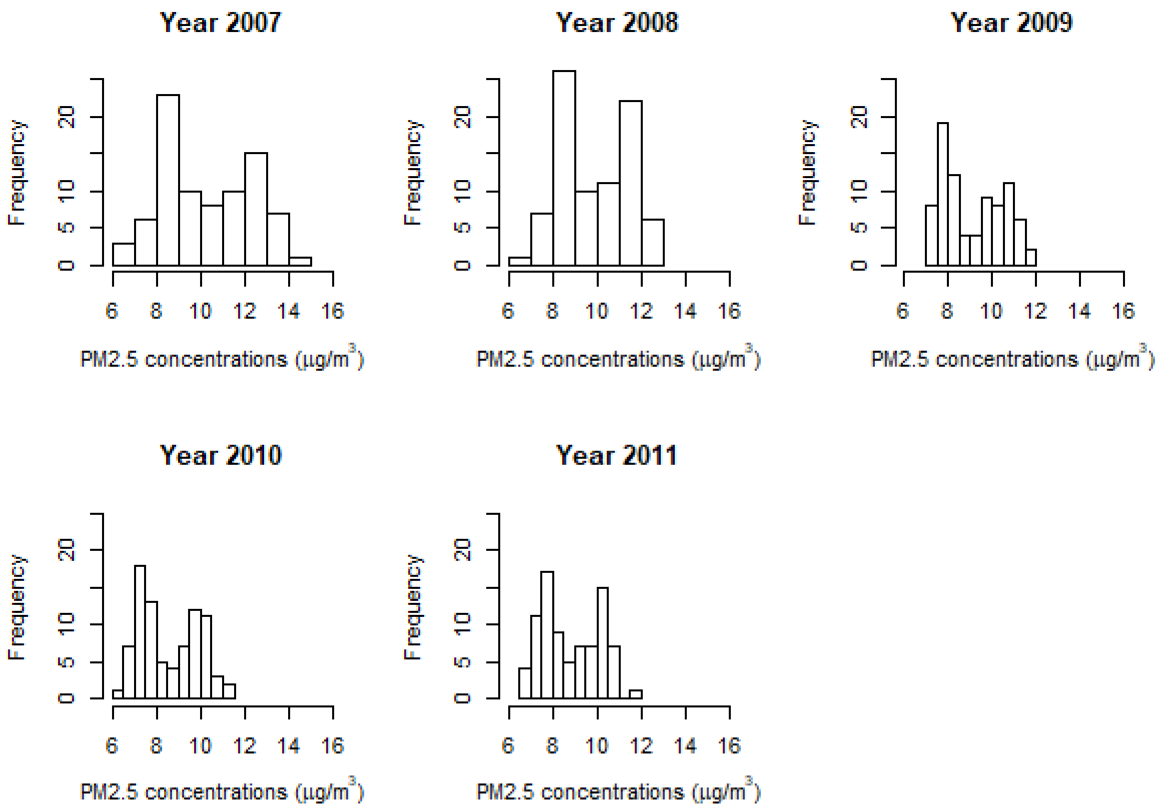

Figure 1: Histograms for Annual Average $PM_{2.5}$ Concentration (2007 2011)

For a further analysis, we mapped annual average $PM_{2.5}$ concentrations for each county in Michigan over 2007 2011 (Figure 2). We can see a geographic difference between north Michigan and south Michigan. Basically, for each year, south Michigan has higher $PM_{2.5}$ concentrations than north Michigan. Schoolcraft county has higher $PM_{2.5}$ concentrations than any other counties in upper peninsular. And we also noticed that, on the whole, $PM_{2.5}$ concentrations were decreasing from 2007 to 2011. The spatial and temporal differences indicate that a spatial-temporal analysis is deserved for our data.

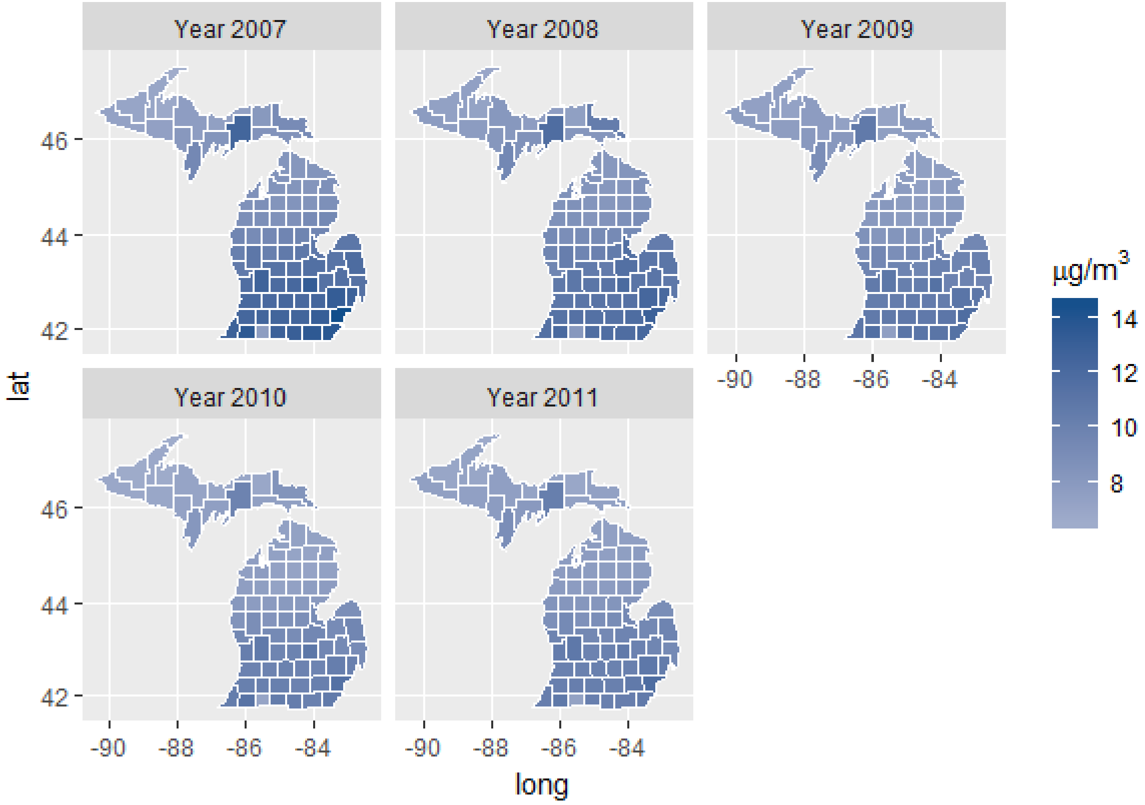

Figure 2: Annual Average $PM_{2.5}$ Concentration Maps for the counties in Michigan (2007 ~ 2011)

3. Model Method

3.1. Spatial-Temporal Gaussian Mixture Model and Its Bayesian Analysis

As mentioned in Section 2, since all histograms for $PM_{2.5}$ concentrations have bimodal shapes, and the maps of $PM_{2.5}$ concentrations for each year show spatial and temporal differences, a spatial-temporal Gaussian mixture model with two components may be a good fit for our data.

Let i denote County i, j denote Year j, and k denote Component k. The density function of our spatial-temporal Gaussian mixture model can be written as:

$$f(y_{i,j}|\theta) = \sum_{k=1}^{2} \pi_{jk} \phi(y_{ij}|\mu_{ijk}, \sigma_{jk}^2), i = 1,2,\cdots,N \ ; j = 1,2,\cdots,J \ ; k = 1,2,\cdots,K \qquad (1)$$

Where, θ is a set of all parameters, φ is a probability dense function of normal distribution, π_{jk} is the proportions for Component k, and $\sum_{k=1}^{K} \pi_{jk} = 1$, μ_{ijk} is mean, σ_{jk}^2 is variance. However, we found that when we used σ_{jk}^2 , we got some extraordinarily large variances, which may be caused by overfitting (Burnham & Anderson, 2003). Since bimodal shapes in Figure 1 were close to each other, which means the variances in different years may be very close, we took σ_k^2 instead of σ_{jk}^2.

We can model μ_{ijk} as:

$$\mu_{ijk} = \beta_{0jk} + x_{ij}^{(p)}\beta_{jk}^{(p)} + x_{ij}^{(s)}\beta_{jk}^{(s)} + f\,(area) + \xi_{ij}. \tag{2}$$

Where, $x_{ij}^{(p)}$ and $x_{ij}^{(s)}$ represent population and income per capita in County i and Year j respectively. If the values are large, we can do log transformation on both of them. *f (area)* is a smoothing function of county area. In this paper, we use a B-spline function with degree of freedoms = 5 (i.e., $f\,(area) = \sum_{h=1}^{5} a_h B_h$, B_h is B-spline basis obtained from R function *bs*(), and a_h are coefficients). ξ_{ij} represent spatial-temporal random effects.

For a Gaussian mixture model with random effects, Markov Chain Monte Carlo (MCMC) algorithm is a very effective and efficient way to estimate the parameters in the model. We give normal priors to $\beta_{0jk}, \beta_{jk}^{(p)}$, and $\beta_{jk}^{(s)}$, Dirichlet priors to π_{jk}, and inverse-gamma priors to σ_k^2, since they are conjugate. For our spatial-temporal random effect ξ_{ij}, either a CAR prior or MCAR prior can be given.

Without considering the correlations of years, a CAR prior can be imposed on ξ_{ij}:

$$\xi_{ij}|\xi_{(-i,j)}, \tau_j^2 \sim N\left(\frac{1}{m_i}\sum_{r\in\theta_i} \xi_{ij}, \frac{\tau_j^2}{m_i}\right). \tag{3}$$

Where, $\xi_{(-i,j)} = \{\xi_{(l,j)};\ l \neq i\}$, ∂_i denotes the set of neighbors for County *i, mi* denotes the number of neighbors sharing the same geographic border with County i, τ_j^2 is variance varying in years (Mariella & Tarantino, 2016; Khana et al., 2018). So our CAR prior can also be denoted by CAR(τ_j^2).

If we consider the correlations of years, then a MCAR prior can be adopted for ξ_{ij}. Let

$$\xi^T = \begin{bmatrix} \xi_{11} & \cdots & \xi_{n1} \\ \vdots & \ddots & \vdots \\ \xi_{1J} & \cdots & \xi_{nJ} \end{bmatrix} = [\xi_1, \xi_2, \cdots, \xi_n]. \tag{4}$$

Under this matrix, we have

$$\xi_i|\xi_{(-i)}, \Sigma \sim N_J\left(\frac{1}{m_i}\sum_{r\in\partial_i} \xi_r, \Sigma/m_i\right). \tag{5}$$

Where, Σ is $J \times J$ covariance matrix of the column vectors of *ξ*. So we can denote MCAR prior by MCAR(Σ). According to Brook's Lemma, (3) and (5) have:

$$\begin{aligned} \xi_i^{(1)}|\tau_j^2 &\propto \exp\left(-\frac{1}{2\tau_j^2}\xi_1^{(1)'}(M-A)\xi_j^{(1)}\right), \\ \xi^{(2)}|\Sigma &\propto \exp\left(-\frac{1}{2}\xi^{(2)'}[(M-A)\otimes\Sigma^{-1}]\xi^{(2)}\right), \end{aligned} \tag{6}$$

Where, $\xi_i^{(1)} = \left(\xi_{1j}, \xi_{2j}, \cdots, \xi_{nj}\right)'$ and $\xi^{(2)} = \left(\xi_{11}, \ldots, \xi_{n1}; \ldots; \xi_{1J}, \ldots, \xi_{nJ}\right)'$. $M = diag(m_1, m_2, \ldots, m_n)$, and A is an adjacency matrix with $a_u = 0$ and $a_{lq} = 1$ if County l and County q share the same geographic bounder, otherwise, $a_{lq} = 0$ (Neelon et al., 2014; Gelfand & Vounatsou, 2003). An introduction to Brook's lemma and a proof for (6) are provided in Appendix A.

With giving all parameters priors, a complete posterior distribution for our model is:

$$p(\theta|y) \propto \prod_{k=1}^{K} \left[\prod_{i=1}^{N}\prod_{j=1}^{J} \pi_{jk}\phi\left(y_{ij}|\mu_{ijk}, \sigma_{jk}^2\right)\right]^{I(L_{ij}=k)} p(\pi)p(\beta)p(a)p(\sigma^2)p(\xi). \tag{7}$$

Where, *p(π), p(β), p(a)*, and $p(\sigma^2)$ represents prior distributions, and we have:

$$\begin{aligned} p(\pi) &= \prod_{k=1}^{K}\prod_{j=1}^{J} Dir(1,1), \\ p(\beta) &= p(\beta_0)p\left(\beta^{(p)}\right)p\left(\beta^{(s)}\right) = \prod_{r=1}^{3}\prod_{k=1}^{K}\prod_{j=1}^{J} N\,(0, 10000), \\ \mathrm{p(a)} &= \prod_{h=1}^{5} \mathrm{N}(0, 10000), \\ \mathrm{p}(\sigma^2) &= \prod_{k=1}^{K} \mathrm{IG}(0.1, 0.01). \end{aligned} \tag{8}$$

p(ξ) denotes the exponential functions in (6) for either CAR(τ_j^2) or MCAR(Σ):

$$p(\xi) = \prod_{j=1}^{J} \exp\left(-\frac{1}{2\tau_j^2}\xi_j^{(1)'}(M-A)\xi_j^{(1)}\right),$$

or (9)

$$p(\xi) = \exp\left(-\frac{1}{2}\xi^{(2)'}[(M-A)\otimes\Sigma]\xi^{(2)}\right).$$

Lij is a latent variable sampled from a categorical distribution:

$$L_{ij} \sim Cat\left(\frac{\pi_{jk}\phi(y_{ij}|\mu_{ijk}, \sigma_k^2)}{\sum_{k=1}^{K}\pi_{jk}\phi(y_{ij}|\mu_{ijk}, \sigma_k^2))}\right). \tag{10}$$

By giving different priors to ξij, we can have three models:

Model 1: $\mu_{L_i,j} = \beta_{0jk} + x_{ij}^{(p)}\beta_{jk}^{(p)} + x_{ij}^{(s)}\beta_{jk}^{(s)} + f\,(area)$, no ξ_{ij}

Model 2: $\mu_{L_i,j} = \beta_{0jk} + x_{ij}^{(p)}\beta_{jk}^{(p)} + x_{ij}^{(s)}\beta_{jk}^{(s)} + f\,(area) + \xi_{ij}, \xi_{ij} \sim \mathrm{CAR}\left(\tau_j^2\right)$

Model 3: $\mu_{L_i,j} = \beta_{0jk} + x_{ij}^{(p)}\beta_{jk}^{(p)} + x_{ij}^{(s)}\beta_{jk}^{(s)} + f\,(area) + \xi_{ij}, \xi_{ij} \sim$ MCAR(Σ).

The above models will be implemented in Winbugs 1.4.3 which is a free statistical software. Goodness-of-fit check and model selection are done based on posterior predictive check and deviation information criteria.

3.2. Posterior Predictive Check

Posterior predictive check (PPC) is a very reliable way to check goodness- of-fit for Bayesian models. A general procedure of PPC is presented as follows:

Step 1: Estimate parameters, θ, given observed y, $\theta \leftarrow P\,(\theta|y)$

Step 2: Simulate replicated $\tilde{y}$ given θ, $\tilde{y} \leftarrow P\,(\tilde{y}|\theta)$

Step 3: Compare y and $\tilde{y}$

For our model, the PPC can be conducted as follows:

Step 1: Estimate $\theta^{(c)}$ (i.e., $\pi_{jk}^{(c)}, \beta_{jk}^{(c)}, a_h^{(c)}\ \sigma_k^{(c)}, \xi_{ij}^{(c)}$) from $p(\theta|y_{ij})$

Step 2: Simulate replicated $\tilde{y}_{ij}$= $(\tilde{y}_{ij}^{(1)}, \tilde{y}_{ij}^{(2)}, \ldots, \tilde{y}_{ij}^{(C)})'$ from $\phi(\tilde{y}|\theta^{(c)})$

Step 3: Construct a credible interval (CrI) for each observed y_{ij} with 2.5%th quantile and 97.5%th quantile of $\tilde{y}_{ij}$ (i.e., CrI= ($\tilde{y}_{2.5\%th}$, $\tilde{y}_{97.5\%th}$)). A capture rate (CR) can be calculated as: CR = (number of y_{ij} captured by CrI_{ij})/($N \times J$).

Where, c denotes the cth MCMC iteration, $c = 1, 2, \ldots, C$. Capture rates for Model 1, Model 2, and Model 3 are presented in Table 1. We can see that all capture rates are greater than 95%, which indicates that all three models are fitted well. Especially, Model 2 and Model 3 reach 100%.

Table 1: Capture Rate for Model 1, Model 2, and Model 3

Model	Capture Rate
Model 1	99.52%
Model 2	100.00%
Model 3	100.00%

3.3. Deviation Information Criteria

Deviation information criteria (DIC) is another way to check goodness- of-fit and do model selection. Let us define a deviation statistic D(θ) as:

$$D(\theta) = -2 \log L(y|\theta)$$

Where, θ denotes a set of all parameters in model, $L(y|\theta)$ denotes the likelihood of the model. Further, let $\overline{D(\theta)} = E_{\theta|y}(D(\theta)), D(\bar{\theta}) = D(E_{\theta|y}(\theta))$, and $pD = \overline{D(\theta)} - D(\theta)$. Then, a general formula of DIC is:

$$DIC = \overline{D(\theta)} + pD$$

D(θ) measures fitness, and *pD*, called effective number of parameters, measures complexity, so in a sense, DIC = ’goodness-of-fit’ + ’complexity’. Models with smaller DIC are preferable.

Usually, DIC can be computed by Winbugs, but if model contains discrete parameters (e.g., L_{ij}), then Winbugs is not able to compute the DIC for this model. We adopt DIC_3 in Celeux et al. (2006) as the DIC for our model. The first term in DIC_3 can be calculated as:

$$\overline{D(\theta)} \approx -\frac{2}{C}\sum_{c=1}^{C} \log f\left(y\middle|\theta^{(c)}\right)$$

$$= -\frac{2}{\mathrm{C}}\sum_{c=1}^{C}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{J} I\left(L_{ij}=k\right)\log\phi(y_{ij}|\mu_{ijk}^{(c)},\sigma_{jk}^{2(c)})$$

The second term in DIC3 can be calculated as:

$$pD \;=\; 2\log\hat{f}(y)$$

$$\hat{f}(y) = \frac{1}{\mathrm{C}}\sum_{c=1}^{C}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{J} I\left(L_{ij}=k\right)\log\phi(y_{ij}|\mu_{ijk}^{(c)},\sigma_{jk}^{2(c)})$$

Where, DIC_3 for Model 1, Model 2, and Model 3 in Section 2 are presented in Table 2. Compared with Model 1, DIC_3 for Model 2 and Model 3 are much improved.

Table 2: DIC3 for Model 1, Model 2, and Model 3

Model	$\overline{D(\theta)}$	pD	DIC_3
Model 1	880.9	10.3	891.2
Model 2	-1100.0	15.1	-1084.9
Model 3	-1369.5	15.7	-1353.8

4. Application to Michigan $PM_{2.5}$ Concentration Data

We applied our model methods in Section 2 to our Michigan $PM_{2.5}$ concentration data. Since population and income per capita contain very large values, we did log transformation on them. In Section 3, both Model 2 and 3 have very small DIC’s and high capture rates, but considering that it is not reasonable to ignore the correlations of years, so we decided to use Model 3 as our final model for the analysis. Additionally, Q-Q plots of the residuals of Model 3 by year are shown in Figure 3. Basically, residuals in each year have very good normality. Estimators from Model 3 are shown in Table 3. All posterior means and CrI’s are derived from Winbugs 1.4.3 with 20,000 iterations and 15,000 burn-ins. Winbugs code for Model 3 is shown in Appendix B.

Table 3 shows the estimators from Model 3. We can see that Component 2 takes up a big proportion of our model. In both components, basically, population shows a positive association with $PM_{2.5}$. Salary per capita has a positive association with $PM_{2.5}$ in Component 1, but tends to be a negative association in Component 2. Σ_{ij} evidences that correlations between any two years exist. Coefficients a_h in smoothing function are usually uninterpretable. Predicted $PM_{2.5}$ concentrations are mapped in Figure 4. Compared with true $PM_{2.5}$ concentrations in Figure 2, we can see that our predicted $PM_{2.5}$ concentrations are very accurate. North Michigan has lower $PM_{2.5}$ concentrations than south Michigan, and from 2007 to 2011, $PM_{2.5}$ concentrations were getting lower yearly.

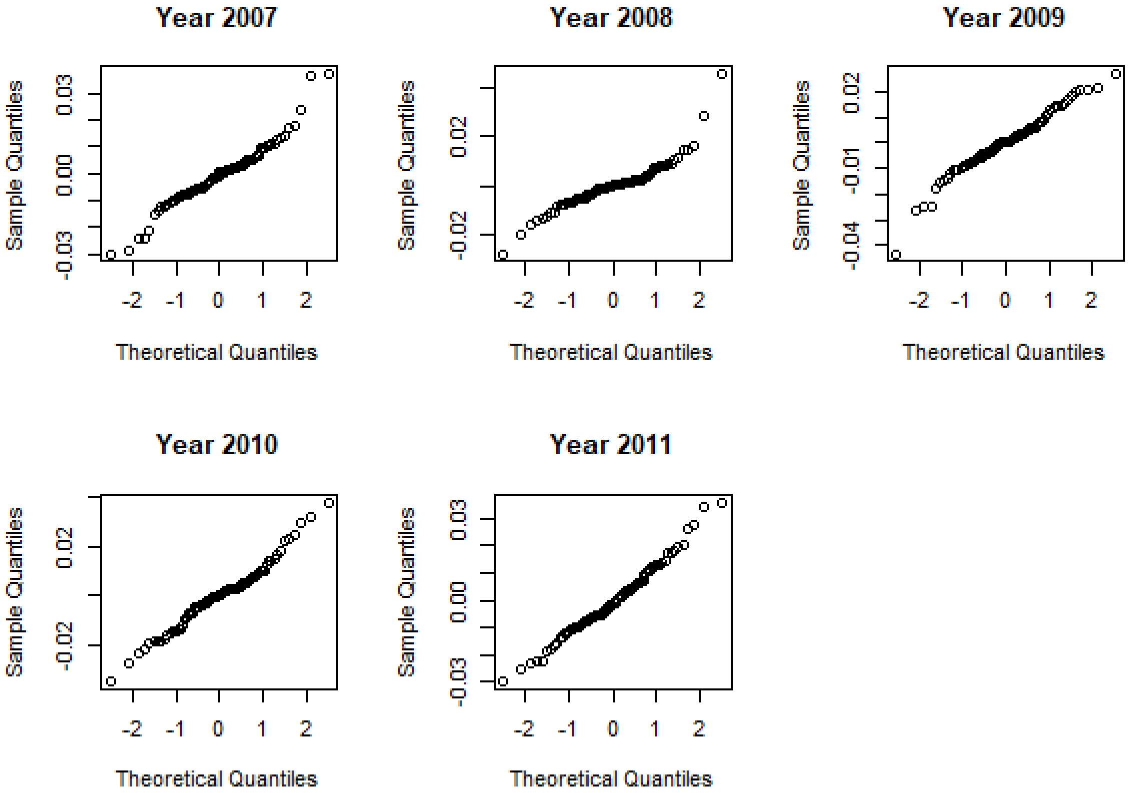

Figure 3: Q-Q Plots of Residuals in Model 2 Table 3: Posterior Estimates in Model 3

Table 3: Posterior Estimates in Model 3

	Parameter	Posterior Mean	95%CrI
Component 1	π_{11}	0.024	(0.003 , 0.067)
	π_{21}	0.054	(0.005, 0.161)
	π_{31}	0.973	(0.926 , 0.997)
	π_{41}	0.024	(0.003 , 0.066)
	π_{51}	0.027	(0.003 , 0.08)
	β_{011}	-1.421	(-1.974 , 1.969)
	β_{021}	-7.7	(-17.18 , 16.674)
	β_{031}	6.455	(1.916 , 11.1)
	β_{041}	-0.941	(-1.961 , 1.898)

β_{051}	-2.847	(-18.871 , 19.193)
$\beta_{11}^{(p)}$	0.482	(-1.438 , 1.422)
$\beta_{21}^{(p)}$	0.169	(-1.118 , 1.081)
$\beta_{31}^{(p)}$	0.298	(0.223 , 0.375)

Continued on next page

Table 3 - continued from previous page

	Parameter	Posterior Mean	95%CrI
	$\beta_{41}^{(p)}$	-0.018	(-1.459 , 1.43)
	$\beta_{51}^{(p)}$	1.032	(0.223 , 1.418)
	$\beta_{11}^{(s)}$	0.677	(-1.276 , 1.332)
	$\beta_{21}^{(s)}$	1.551	(-9.594 , 10.12)
	$\beta_{31}^{(s)}$	-0.015	(-0.514 , 0.462)
	$\beta_{41}^{(s)}$	0.946	(-1.301 , 1.344)
	$\beta_{51}^{(s)}$	0.199	(-1.29 , 1.235)
	τ_1	0.022	(0.001 , 0.004)
Component 2	π_{12}	0.976	(0.933 , 0.997)
	π_{22}	0.946	(0.84 , 0.995)
	π_{32}	0.027	(0.003 , 0.074)
	π_{42}	0.976	(0.934 , 0.997)
	π_{52}	0.973	(0.92 , 0.997)
	β_{012}	6.336	(0.433 , 12.21)
	β_{022}	6.865	(0.483 , 13.1)
	β_{032}	-1.226	(-1.997 , 1.968)
	β_{042}	7.563	(3 , 11.54)
	β_{052}	7.434	(3.033 , 11.49)
	$\beta_{12}^{(p)}$	0.364	(0.239 , 0.464)
	$\beta_{22}^{(p)}$	0.366	(0.269 , 0.453)
	$\beta_{32}^{(p)}$	-1.309	(-12.861 , 12.412)
	$\beta_{42}^{(p)}$	0.362	(0.278 , 0.438)
	$\beta_{52}^{(p)}$	0.37	(0.29 , 0.443)
	$\beta_{12}^{(s)}$	0.039	(-0.572 , 0.68)
	$\beta_{22}^{(s)}$	-0.058	(-0.681 , 0545)
	$\beta_{32}^{(s)}$	2.39	(-12.64 , 13.61)
	$\beta_{42}^{(s)}$	-0.239	(-0.673 , 0.259)
	$\beta_{52}^{(s)}$	-0.213	(-0.656 , 0.286)
	τ_2	0.002	(0.001 , 0004)
$\mathrm{a_h}$ $in\ f(area)$	a_1	-0.163	(-1.188, 0.732)
	a_2	-0.458	(-1.015, -0.037)
	a_3	-0.591	(-1.271, 0.164)
	a_4	0.361	(-0.712, 1.643)
	a_5	-0.961	(-1.708, -0.407)

Continued on next page

Table 3 - continued from previous page

	Parameter	Posterior Mean	95%CrI
$\Sigma_{ij}\ in\ MCAR(\Sigma)$	Σ_{11}	0.778	(0.559, 1.086)
	Σ_{12}	0.532	(0.357, 0.771)
	Σ_{13}	0.481	(0.338, 0.687)
	Σ_{14}	0.475	(0.33, 0.68)
	Σ_{15}	0.427	(0.294, 0.616)
	Σ_{22}	0.613	(0.442, 0.845)
	Σ_{23}	0.375	(0.252, 0.541)
	Σ_{24}	0.376	(0.249, 0.553)
	Σ_{25}	0.303	(0.191, 0.458)
	Σ_{33}	0.368	(0.261, 0.513)
	Σ_{34}	0.338	(0.236, 0.484)
	Σ_{35}	0.302	(0.21, 0.43)
	Σ_{44}	0.385	(0.276, 0.545)
	Σ_{45}	0.305	(0.209, 0.442)
	Σ_{55}	0.328	(0.234, 0.459)

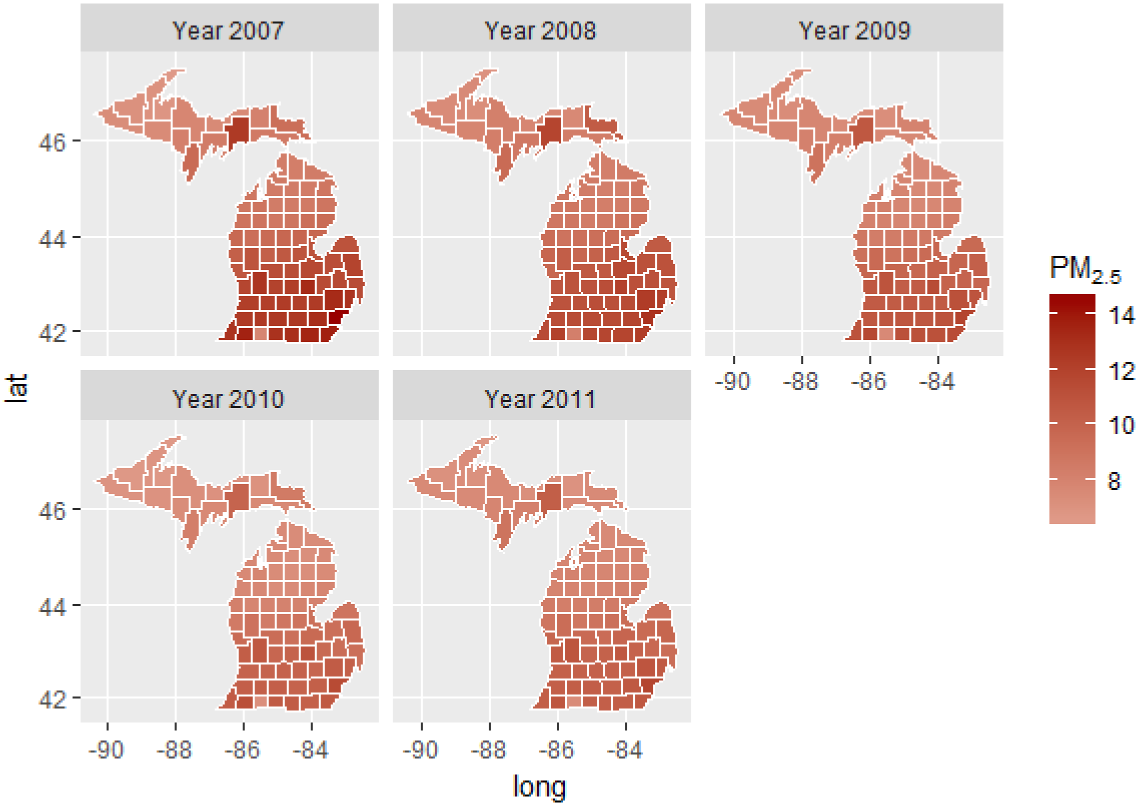

Figure 4: Spatial-Temporal Random Effects in Model 2

5. Conclusion

In this paper, we extended the work of Antonovsky et al. (1991) into a spatial-temporal Gaussian mixture model. Our method is directly inspired by Neelon et al. (2014) and Mariella & Tarantino (2016), By turning off and on the spatial-temporal random effect and giving it CAR(τ_j) prior and MCAR(Σ) prior, we end up with having three models for our annual average $PM_{2.5}$ concentrations data. We used a posterior predictive check and DIC_3 in Celeux et al. (2006) for checking goodness-of-fit and model selection. Both of Model 2 and 3 have better performance than Model 1, but Model 3 considered the correlations between any two years, so eventually, we chose Model 3 as our final model for $PM_{2.5}$ data analysis.

We applied our model to Michigan annual average $PM_{2.5}$ concentrations (2007 ~ 2011) data. To our knowledge, this is the first time to use a spatial-temporal Gaussian mixture model to analyze $PM_{2.5}$ concentration in entire Michigan. We found that population has a clearly positive association with $PM_{2.5}$ concentrations, but income per capita shows opposite signals in Component 1 and 2, which needs some further analysis.

In this paper, we assume that our Gaussian mixture model has only two components, since the histograms of our data show bimodal shapes. However, bimodal shapes can also be generated from a mixture distribution with three or more components, which will be explored in the future.

Appendix A. Brook's Lemma

If probability measure *P(x) > 0* for all x, then, for any $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$:

$$\frac{P(x)}{P(y)} = \prod_{i=1}^{n} \frac{P(x_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_n)}{P(y_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_n)}$$

Proof:

$$\begin{aligned}
\frac{P(x)}{P(y)} &= \frac{P(x_1, y_2, \dots, y_n)}{P(y_1, y_2, \dots, y_n)} \times \frac{P(x_1, x_2, \dots, x_n)}{P(x_1, y_2, \dots, y_n)} \\
&= \frac{P(x_1, y_2, \dots, y_n)/P(y_2, \dots, y_n)}{P(y_1, y_2, \dots, y_n)/P(y_2, \dots, y_n)} \times \frac{P(x_1, x_2, \dots, x_n)/P(x_1)}{P(x_1, y_2, \dots, y_n)/P(x_1)} \\
&= \frac{P(x_1 | \, y_2, \dots, y_n)}{P(y_1 | \, y_2, \dots, y_n)} \times \frac{P(x_1, x_2, \dots, x_n | x_1)}{P(x_1, y_2, \dots, y_n | x_1)} \\
&= \frac{P(x_1 | \, y_2, \dots, y_n)}{P(y_1 | \, y_2, \dots, y_n)} \times \frac{P(x_2 | x_1, y_3, \dots, y_n)}{P(y_2 | x_1, y_3, \dots, y_n)} \times \frac{P(x_3, \dots, x_n | x_1, x_2)}{P(y_3, \dots, y_n | x_1, x_2)} \\
&= \cdots
\end{aligned}$$

Keep decomposing the last term, eventually, we will get:

$$\frac{P(x)}{P(y)}=\prod_{i=1}^{n}\frac{P(x_i|x_1,\cdots,x_{i-1},y_{i+1},\cdots,y_n)}{P(y_i|x_1,\cdots,x_{i-1},y_{i+1},\cdots,y_n)}$$

Use Brook's Lemma, we can prove (6).

For CAR(τ_j^2) prior:

$$\frac{P(\xi_j)}{P(0)}=\prod_{i=1}^{n}\frac{\exp\left[-\frac{m_i}{2\tau_j^2}\left(\xi_{ij}-\frac{1}{m_i}\sum_{r<i}\xi_{\tau j}-\frac{1}{m_i}\sum_{r<i}0\right)^2\right]}{\exp\left[\frac{m_i}{2\tau_j^2}\left(0-\frac{1}{m_i}\sum_{r<i}\xi_{\tau j}-\frac{1}{m_i}\sum_{r<i}0\right)^2\right]}$$

$$=\prod_{i=1}^{n}\exp\left[-\frac{1}{2\tau^2}m_i\left(\xi_{ij}^2-\frac{2}{m_i}\sum_{r<i}\xi_{rj}\xi_{ij}\right)\right]$$

$$=\exp\left[-\frac{1}{2\tau^2}\left(\sum_{i=1}^{n}m_i\xi_{ij}^2-\sum_{i=1}^{n}\sum_{r\in\partial_i}\xi_{rj}\xi_{ij}\right)\right]$$

$$=\exp[-\frac{1}{2\tau_j^2}\xi_j'(M-A)\xi_j]$$

We have to notice that $\sum_{i=1}^{n}\sum_{r<i}\xi_{rj}\xi_{ij}=\sum_{i=1}^{n}\sum_{r>i}\xi_{rj}\xi_{ij}$, so $2\sum_{i=1}^{n}\sum_{r<i}\xi_{rj}\xi_{ij}=\sum_{i=1}^{n}\sum_{r\in\partial_i}\xi_{rj}\xi_{ij}$.

In the same way, we can prove for MCAR(Σ), since MCAR(Σ) is just an extension of CAR(τ_j^2).

Appendix B. Winbugs Code for Model 3

```
model{
for(i    in 1:N){
for(j in 1:T){
pm[i, j] ~ dnorm(mu[i, j], sigma[L[i,j]])
mu[i, j] <- step(1.5-L[i, j])*(beta0[1, j] + beta1[1, j]*population[i, j]
+ beta2[1, j]*income[i, j]) + step(L[i, j]-1.5)*(beta0[2, j]
+ beta1[2, j]*population[i, j]+ beta2[2, j]*income[i, j])
+ a[1]*z[i,1]+ a[2]*z[i,2] + a[3]*z[i,3] + a[4]*z[i,4]
+ a[5]*z[i,5] + phi[j, i]
}
}
## priors ##
for(i   in 1:N){
for(j in 1:T){

L[i, j] ~ dcat(P[j,])
}
}
for(j in 1:T){
P[j, 1:2] ~ ddirch(lambda[])
}
sigma[1] ~ dgamma(0.1, 0.01) sigma[2] ~ dgamma(0.1, 0.01) tau[1] <- 1/sigma[1]
tau[2]   <- 1/sigma[2]
for(j in    1:T){
beta0[1,  j] ~ dnorm(0,  0.0001)
beta0[2,  j] ~ dnorm(0,  0.0001)
beta1[1,  j] ~ dnorm(0,  0.0001)
beta1[2,  j] ~ dnorm(0,  0.0001)
beta2[1,  j] ~ dnorm(0,  0.0001)
beta2[2,  j] ~ dnorm(0,  0.0001)
}
for(i in 1:5){
a[i] ~ dnorm(0, 0.0001)
}
# MCAR prior
phi[1:T,1:N] ~ mv.car(adj[],weightst[],num[],Rs[,]) for(i in 1:M){weightst[i] <- 1}
Rs[1:T, 1:T] ~ dwish(I[ , ], T)
Sigma.p[1:T, 1:T] <-inverse(Rs[, ])
}
```

References

[1] Antonovsky, M. Y., Buchstaber, V., & Zelenuk, E. (1991). A statistical model of background air pollution frequency distributions. Environmental Monitoring and Assessment, 16 (3), 203–252.

[2] Brown, P. J., Le, N. D., & Zidek, J. V. (1994). Multivariate spatial interpolation and exposure to air pollutants. Canadian Journal of Statistics, 22 (4), 489–509.

[3] Burnham, K. P., & Anderson, D. R. (2003). Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.

[4] Celeux, G., Forbes, F., Robert, C. P., Titterington, D. M., et al. (2006). Deviance information criteria for missing data models. Bayesian Analysis, 1 (4), 651–673.

[5] Chu, H.-J., Yu, H.-L., & Kuo, Y.-M. (2012). Identifying spatial mixture distributions of pm2. 5 and pm10 in taiwan during and after a dust storm. Atmospheric Environment , 54 , 728–737.

[6] Cohen, A. J., Ross Anderson, H., Ostro, B., Pandey, K. D., Krzyzanowski, M., Ku¨nzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J. M., et al. (2005). The global burden of disease due to outdoor air pollution. Journal of Toxicology and Environmental Health, Part A, 68 (13-14), 1301–1307.

[7] Delamater, P. L., Finley, A. O., & Banerjee, S. (2012). An analysis of asthma hospitalizations, air pollution, and weather conditions in los an- geles county, california. Science of the Total Environment , 425 , 110–118.

[8] Fuentes, M. (2003). Statistical assessment of geographic areas of compli- ance with air quality standards. Journal of Geophysical Research: Atmo- spheres, 108 (D24).

[9] Gelfand, A. E., & Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics, 4 (1), 11–15.

[10] Karaca, F., Alagha, O., & Ertu¨rk, F. (2005). Statistical characterization of atmospheric pm 10 and pm 2.5 concentrations at a non-impacted subur- ban site of istanbul, turkey. Chemosphere, 59 (8), 1183–1190.

[11] Karppinen, A., Ha¨rk¨onen, J., Kukkonen, J., Aarnio, P., & Koskentalo, T. (2004). Statistical model for assessing the portion of fine particulate matter transported regionally and long range to urban air. Scandinavian Journal of Work, Environment & Health, 30 (2), 47–53.

[12] Khana, D., Rossen, L. M., Hedegaard, H., & Warner, M. (2018). A bayesian spatial and temporal modeling approach to mapping geographic variation in mortality rates for subnational areas with r-inla. Journal of Data Science: JDS , 16 (1), 147–182.

[13] Li, L., Wu, A. H., Cheng, I., Chen, J.-C., & Wu, J. (2017). Spatiotemporal estimation of historical pm2. 5 concentrations using pm10, meteorological variables, and spatial effect. Atmospheric Environment , 166 , 182–191.

[14] Liu, Y., He, K., Li, S., Wang, Z., Christiani, D. C., & Koutrakis, P. (2012). A statistical model to evaluate the effectiveness of pm 2.5 emissions control during the beijing 2008 olympic games. Environment International , 44 , 100–105.

[15] Ma, Y.-R., Ji, Q., & Fan, Y. (2016). Spatial linkage analysis of the impact of regional economic activities on pm 2.5 pollution in china. Journal of Cleaner Production, 139 , 1157–1167.

[16] Mariella, L., & Tarantino, M. (2016). Spatial temporal conditional auto- regressive model: A new autoregressive matrix. Austrian Journal of Statistics, 39 (3), 223–244.

[17] Monn, C., & Becker, S. (1999). Cytotoxicity and induction of proinflam- matory cytokines from human monocytes exposed to fine (pm2. 5) and coarse particles (pm10–2.5) in outdoor and indoor air. Toxicology and Applied Pharmacology, 155 (3), 245–252.

[18] Neelon, B., Gelfand, A. E., & Miranda, M. L. (2014). A multivariate spatial mixture model for areal data: examining regional differences in standard- ized test scores. Journal of the Royal Statistical Society: Series C (Applied Statistics), 63 (5), 737–761.

[19] Peng, R. D., Dominici, F., & Louis, T. A. (2006). Model choice in time series studies of air pollution and mortality. Journal of the Royal Statistical Society: Series A (Statistics in Society), 169 (2), 179–203.

[20] Pérez, P., Trier, A., & Reyes, J. (2000). Prediction of pm 2.5 concentrations several hours in advance using neural networks in santiago, chile. Atmospheric Environment , 34 (8), 1189–1196.

[21] Russell, B. T., Wang, D., & McMahan, C. S. (2017). Spatially modeling the effects of meteorological drivers of pm2. 5 in the eastern united states via a local linear penalized quantile regression estimator. Environmetrics, 28 (5), e2448.

[22] Tai, A. P., Mickley, L. J., Jacob, D. J., Leibensperger, E., Zhang, L., Fisher, J. A., & Pye, H. (2012). Meteorological modes of variability for fine particulate matter (pm 2.5) air quality in the united states: implications for pm 2.5 sensitivity to climate change. Atmospheric Chemistry and Physics, 12 (6), 3131–3145.

[23] Tian, J., & Chen, D. (2010). A semi-empirical model for predicting hourly ground-level fine particulate matter (pm 2.5) concentration in southern ontario from satellite remote sensing and ground-based meteorological measurements. Remote Sensing of Environment , 114 (2), 221–229.

[24] Vidale, S., Arnaboldi, M., Bosio, V., Corrado, G., Guidotti, M., Sterzi, R., & Campana, C. (2017). Short-term air pollution exposure and car- diovascular events: A 10-year study in the urban area of como, italy. International journal of cardiology, 248 , 389–393.

[25] Wang, Z.-b., & Fang, C.-l. (2016). Spatial-temporal characteristics and de- terminants of pm 2.5 in the bohai rim urban agglomeration. Chemosphere, 148 , 148–162.

[26] Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M. L., Shen, X., Zhu, L., & Zhang, M. (2017). Spatiotemporal prediction of continuous daily $pm_{2.5}$ concentrations across china using a spatially explicit machine learning algorithm. Atmospheric Environment , 155 , 129–139.