Spatio-temporal modeling of fine particulate matter

Sujit K. SAHU, Alan E. GELFAND and David M. HOLLAND *

Studies indicate that even short-term exposure to high concentrations of fine atmospheric particulate matter ($PM_{2.5}$) can lead to long-term health effects. In this paper, we propose a random effects model for $PM_{2.5}$ concentrations. In particular, we anticipate urban/rural differences with regard to both mean levels and variability. Hence we introduce two random effects components, one for rural or background levels and the other as a supplement for urban areas. These are specified in the form of spatio-temporal processes. Weighting these processes through a population density surface results in nonstationarity in space. We analyze daily $PM_{2.5}$ concentrations in three Midwestern U.S. states for the year 2001. A fully Bayesian model is implemented, using MCMC techniques, which enables full inference with regard to process unknowns as well as predictions in time and space.

Key Words: Hierarchical model; Markov chain Monte Carlo; nonstationary spatio-temporal process; separable process.

1   Introduction

Particulate matter (PM) has been linked to widespread public health effects, including a range of serious respiratory and cardiovascular problems, and to reduced visibility in many parts of the United States, see the Environmental Protection Agency (EPA) report (2004) and relevant pages on http://cfpub.epa.gov/ncea/cfm/partmatt.cfm. Additionally, these particles and their components interact in ways that contribute to elevated concentrations of other air pollutants and stress to vegetation and ecosystems. In 1997, the U.S. EPA promulgated new regulations that established National Ambient Air Quality Standards (NAAQS) for particulate matter (PM) with aerodynamic diameters less than 2.5μm ($PM_{2.5}$). As part of the program to implement these standards, a network of ambient mass monitoring sites was established. In 2003, over 950 sites were in operation in the U.S. These sites are located primarily in populated regions, measuring pollution where people live and work. Most of these sites are used to evaluate compliance with particulate air quality standards; other sites are located away from urban areas to characterize transport, background concentrations, and visibility levels.

PM can be emitted directly or formed in the atmosphere. Generally, $PM_{2.5}$ contains particles formed in the atmosphere from gaseous emissions. Examples include sulfates formed from sulfur dioxide (SO_2) emissions, nitrates formed from NO_x emissions, and carbon formed from organic gas emissions. Regional and local PM concentrations are affected by emissions, topography, land cover, and a variety of processes affecting the rates of conversion of gases to particles. The general pattern

*Sujit K. Sahu is senior lecturer, School of Mathematics, S^3RI, University of Southampton, Southampton, UK. (Email: S.K.Sahu@maths.soton.ac.uk). Alan E. Gelfand is Professor, Institute of Statistics and Decisions, Duke University, Durham, NC, USA (Email: alan@stat.duke.edu). David M. Holland is senior statistician, U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC, USA (Email: holland.david@epa.gov).

of air movement across the U.S. also influences particle levels. During summer, generally, a south to north or northeast transport direction is found over the eastern U.S.

The demand for spatial models to assess regional progress in air quality has grown rapidly over the past decade. For improved environmental decision-making, it is imperative that such models enable spatial prediction to reveal important gradients in air pollution, offer guidance for determining areas in non-attainment with air standards, and provide air quality input to models for determining individual exposure to air pollution. Spatial prediction has the potential to suggest new perspectives in the development of emission control strategies and to provide a credible basis for resource allocation decisions, particularly with regard to network design.

Space-time modeling of pollutants has some history including, e.g., Guttorp *et al.* (1994), Haas (1995) and Carroll *et al.* (1997). In recent years, hierarchical Bayesian approaches for spatial prediction of air pollution have been developed (Brown *et al.*, 1994; Le *et al.*, 1997; Sun *et al.*, 2000). Zidek *et al.* (2002) developed predictive distributions on non-monitored PM_{10} concentrations (particles with diameters less than $10\mu m$) in Vancouver, CA. They noted the under-prediction of extreme values in the pollution field, but their methodology provided useful estimates of uncertainties for large values. Cressie *et al.* (1999) compared kriging and Markov-random field models in the prediction of PM_{10} concentrations around Pittsburgh, PA. Kibria *et al.* (2000) developed a multivariate spatial prediction methodology in a Bayesian context for the prediction of $PM_{2.5}$ in Philadelphia, PA. This approach used both $PM_{2.5}$ and PM_{10} data at monitoring sites with different start-up times. Shaddick and Wakefield (2002) propose short term space-time modeling for PM_{10}.

Smith *et al.* (2003) proposed a spatio-temporal model for predicting weekly averages of $PM_{2.5}$ and other derived quantities such as annual averages within three southeastern states. The $PM_{2.5}$ field is represented as the sum of semi-parametric spatial and temporal trends, with a random component that is spatially correlated, but not temporally. These authors apply a variant of the expectation-maximization (EM) algorithm to account for high percentages of missing data. Sahu and Mardia (2005) present a short-term forecasting analysis of $PM_{2.5}$ data in New York City during 2002. Within a Bayesian hierarchical framework, they model the spatial structure with principal kriging functions and the time component is modeled by a vector random-walk process.

In the application considered here, we address a frequent criticism of spatial prediction using air pollution data from large-scale monitoring networks. Many of these networks were designed to capture peak pollution levels within urban, highly populated areas. For this situation, there is a potential for over-prediction within sparsely monitored rural areas with misleading prediction errors. This paper presents a hierarchical space-time model that introduces two spatio-temporal processes, one capturing rural or background effects, the second adding extra variability for urban/suburban locations. We also consider the relationship of population density to fine particulate matter and incorporate non-stationary spatial and temporal covariance structure. We model weekly average concentrations to gain insight to the spatial variability and uncertainties of $PM_{2.5}$ concentrations over short time periods. These weekly averages can be aggregated to seasonal and annual summaries, while accounting for the variance of the assumed space-time process.

The structure of the paper is as follows. In Section 2, we provide a description of the data with summary statistics and exploratory graphics. Section 3 presents the spatio-temporal model that accounts for trend in space and time and random urban and rural effects. The full Bayesian model is presented with fitting and computational issues deferred to an appendix. Bayesian prediction methods are developed in Section 4. Section 5 presents our approach for model selection and estimation results. A brief Section 6 offers conclusions and future work.

2

2 Exploratory analysis

2.1 Data used in the study

Average daily fine particulate data at 124 irregularly-spaced urban and rural monitoring sites located in the Midwestern region of the U.S. (Illinois, Indiana, and Ohio) are used in this study. These sites are part of EPA's Federal Reference Method fine particulate monitoring network that began operation in 1999. This analysis considers data collected in 2001 which was the only year for which we had reliable data at the time we started this project. The sampling frequencies vary from site to site; most sites have measurements recorded every three days, others have continuous daily records, and others have data every six days. Given the disparate sampling frequencies across sites, we aggregated the data into weekly averages based on all available measurements within a week as is often employed in handling monitoring data, see e.g. Smith *et al.* (2003). This misalignment results in differing numbers of measurements for different sites and for different weeks, hence potential heterogeneity in variance. However, more than 90% of the weekly averages were formed by the same number of daily observations and informal exploratory analysis provided comfort with the homoscedasticity assumption. If heteroscedasticity was a concern, the weekly variances could be inversely scaled by the respective number of measurements averaged over. For the above data, with interest in learning about space-time behavior, high temporal resolution is desired and weekly aggregation is the finest resolution we can comfortably achieve.

We set aside data from ten sites for validation purposes. These sites are chosen at random; see the next sub-section (paragraph starting with Table 1) for more details. We model data from the remaining 114 sites, see Figure 1 for their locations. Out of these 5928 (=114×52) atomic data points 274 observations (about 4.6%) were missing at random. In our modeling approach, we employ information about these monitoring sites such as location coordinates and an urbanization classifier with classifications as rural, urban, or suburban. Also utilized was census tract population data.

2.2 Summary tables and graphs

Figure 1 shows the 114 monitoring sites in the three states (again, Illinois, Indiana and Ohio). Initial analysis of the data showed that variances increase as the mean level increases. In particular, Figure 2 shows means and variances on the original scale of measurement, on the square root scale, and on the log scale. The square root scale seems most attractive and is consistent with an earlier analysis of fine particulate matter in the southeast U.S. by Smith *et al.* (2003). So, in the sequel we model on the square-root scale. However, we report all predictions on the original scale for ease of interpretation and comparison. Hence, below, we let $Z(\mathbf{s}_i, t)$ denote the square-root of the observed $PM_{2.5}$ concentration (measured in $\mu g/m^3$ units) at site $\mathbf{s}_i$ and at time t where $i = 1, \ldots, n = 114$ and $t = 1, \ldots, T = 52$.

In the supplied dataset, each of the 114 sites received a (subjective) classification to location-type which had three categories, namely, rural, suburban and urban. As mentioned above, it is hypothesized that the pollution levels in rural areas are lower than those in urban and suburban areas. We present the mean and variance of the response (after omitting the missing values) categorized by the location type in Table 1. The table lends some support to this hypothesis since it shows that the suburban and urban sites have higher mean levels than the rural ones. Also, while the variances on the square root scale are comfortably homogeneous across location types, the implication is that, on the original scale, variance will increase with the square of the mean, suggesting smaller variance in rural locations. Since there is not much difference between the suburban and urban areas, in Section 3 we collapse the suburban and urban categories. We let $\alpha(\mathbf{s}_i)$ denote the indicator of the urban sites, i.e., $\alpha(\mathbf{s}_i) = 1$ if the site $\mathbf{s}_i$ is a suburban or urban site, =0 otherwise. In choosing ten sites for validation we have selected one rural site and nine urban sites at random from the available sites.

3

Table 1: The overall means ($\mu g/m^3$) and the variances of square root of $PM_{2.5}$ categorized by location type.

	Number	Mean	Var
Rural	12	3.53	0.72
Suburban	57	3.88	0.69
Urban	45	3.93	0.75
All	114	3.87	0.73

As part of our analysis we wish to learn whether there is a relationship between population density and concentration levels. Hence, a population density surface was obtained at high resolution (population per square mile in the year 2001) using census tract data. For the sampled sites, population density varied between 0 and 24059.15. The minimum population density was observed at a site situated in the middle of a forest while the maximum density was for a site in Chicago, Illinois. In attempting to use population density to explain $PM_{2.5}$ measurements (Section 3) the wide range (with associated skewness) in the density values led us to work with normalized square-root transformed population densities defined as $p(\mathbf{s}) = \sqrt{p'(\mathbf{s})/\max p'(\mathbf{s})}$, where $p'(\mathbf{s})$ denotes the population density at site $\mathbf{s}$ and the maximization is performed over all the 124 sampled and a set of 3210 prediction sites. Thus we have $0 \leq p(\mathbf{s}) \leq 1$ for all $\mathbf{s}$. The 3210 prediction sites form a grid of locations in the three states where we will eventually perform spatial predictions.

Next, we partitioned the $p(\mathbf{s}_i)$ for the observed sites into two categories: low-population areas and high-population areas (using the cut point 0.25 on the transformed 0–1 scale so that the low-population sites had populations less than 25% of that of the most populated site). The mean responses for the two categories on the square root scale were 3.82 and 3.89 with variances 0.78 and 0.71, respectively. Comparison with Table 1 shows that population density does not separate the pollution levels in the same way that location types do. The explanation lies in the fact that there are commercially-zoned suburban and urban areas with very low population densities but with high pollution levels. Thus we keep both population density (as a continuous variable) and urban indicator as explanatory variables in our model. In fact, we introduce a coefficient for population density/urban indicator interaction. However, this raises an issue with regard to prediction. We have $p(\mathbf{s})$ at high resolution (i.e. for all 3210 predictive sites) but we have $\alpha(\mathbf{s})$ at only the 124 sampling sites. We return to this point in Section 4.

Figure 3 reveals the strong seasonal patterns of the $PM_{2.5}$ measurements across weeks and across months separately for the set of 12 rural sites and the set of 102 urban sites. Higher summer fine particulate levels likely result from increased levels of sulfate particles which are formed from the reaction of SO_2 and other gases in the atmosphere. However, we also observe high concentrations of fine particulate in January, and attribute this result to high concentrations of nitrate particles that result from NO_x emissions. Further, winter particulate concentrations in the Midwest are likely influenced by the re-entrainment of road salt into the atmosphere.

2.3 Empirical variograms

We informally investigate the spatial variation using an empirical variogram derived from the data. First, we find the residuals after fitting a linear model with explanatory variables month, urban indicator, population density, and the interaction effect between urban indicator and population density. This is the linear model for the mean surface we employ in this paper; more justification for adopting this model is given in Section 3.

In creating the empirical variogram for the residual variation, we use the geodetic distance between two locations with given latitudes η_1 and η_2 and longitudes ν_1 and ν_2 (converted to radians).

The geodetic distance is the distance at the surface of the Earth considered as a sphere of radius $R = 6371$ kilometers. We use the formula d $= 6371 \arccos(\Xi)$ (km) where $\Xi = \sin(\eta_1)\sin(\eta_2) + \cos(\eta_1)\cos(\eta_2)\cos(\nu_1 - \nu_2)$.

We first obtain the variogram cloud of points using the the average over 52 weeks of the squared differences between the residuals for all possible pairs of locations. Then we fitted a smooth curve using the S-Plus function `lowess`. The variogram clouds along with the fitted curves are plotted in Figure 4. Panel (a) is for the 12 rural sites, panel (b) is for the 102 suburban and urban sites, and panel (c) is for all the 114 sites. In panel (d) only the fitted curves from the previous three panels have been plotted to enhance readability. These empirical variograms are merely suggestive and show evidence of spatial variation. All three plots suggest finite values for the corresponding sill and range parameters. There is some indication of difference between the variogram of the rural sites (top panel) and that of the remaining sites (middle panel); we take this into account in our modeling effort in the next section where we employ two different spatio-temporal processes, one for highly populated suburban and urban areas and the other for rural areas.

3 Modeling details

3.1 Mean structure

The model developed here is applicable for spatio-temporal data recorded at n sites $\mathbf{s}_i$ ($\in \mathbb{R}^2$), $i = 1, \ldots, n$, over a period of T equally spaced time points, $1, 2, \ldots, T$. In our case $n = 114$ and $T = 52$. To capture seasonal effects we define monthly seasonal indicators, $u(t, m)$ as follows:

$$u(t, m) = \begin{cases} 1 & \text{if the time } t \text{ is in the } m\text{th month} \\ 0 & \text{otherwise,} \end{cases}$$

for $t = 1, \ldots, 52$, $m = 1, \ldots, 12$. These monthly dummies are motivated from discussion around Figure 3 and these are intended to capture the seasonal variation, see the residual analysis in Section 5.2. The mean function $\mu(\mathbf{s}_i, t)$ is given by

$$\mu(\mathbf{s}_i, t) = \beta_0 + \beta_1\, p(\mathbf{s}_i) + \beta_2\, \alpha(\mathbf{s}_i) + \beta_3\, \alpha(\mathbf{s}_i) \times p(\mathbf{s}_i) + \sum_{m=2}^{12} \gamma_m u(t, m). \tag{1}$$

The function in (1) provides a spatial model and makes full use of the available covariate information. We did not examine a trend surface component (there are no anticipated directional gradients), electing to introduce spatio-temporal random effects instead. The random effects capture any inadequacy in the mean model (1). Since β_0 is included in (1), for identifiability purposes we do not include $u(t, 1)$ in the model. Let $\mathbf{x}(\mathbf{s}_i, t) = \big(1, p(\mathbf{s}_i), \alpha(\mathbf{s}_i), \alpha(\mathbf{s}_i) \times p(\mathbf{s}_i), u(t, 2), \ldots, u(t, 12)\big)'$ denote the covariate values and seasonal dummies and let the 15×1 vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_2, \ldots, \gamma_{12})'$. Then, $\mu(\mathbf{s}_i, t) = \mathbf{x}(\mathbf{s}_i, t)'\boldsymbol{\beta}$.

We assume the hierarchical structure:

$$Z(\mathbf{s}_i, t) = Y(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t), \ i = 1, \ldots, n, \ \ t = 1, \ldots, T, \tag{2}$$

where $Y(\mathbf{s}, t)$ is a space-time process and the error term $\epsilon(\mathbf{s}_i, t)$ is a white noise process and specifically assumed to follow $N(0, \sigma_\epsilon^2)$ independently. In principle, σ_ϵ^2 could evolve in time but with our short time scale - weekly intervals - we take it to be constant. Furthermore, as noted in Section 2.1, if heteroscedasticity was a concern, σ_ϵ^2 can be inversely weighted by the number of observations averaged to form $Z(\mathbf{s}_i, t)$.

The space-time process $Y(\mathbf{s}_i, t)$ is expressed as

$$Y(\mathbf{s}_i, t) = \mu(\mathbf{s}_i, t) + w(\mathbf{s}_i, t) + p(\mathbf{s}_i)\, v(\mathbf{s}_i, t), \tag{3}$$

5

where $w(\mathbf{s}, t)$ and $v(\mathbf{s}, t)$ are independent zero mean spatio-temporal processes, termed below as 'rural' and 'urban' processes respectively. The quantity $w(\mathbf{s}_i, t) + p(\mathbf{s}_i)\, v(\mathbf{s}_i, t)$ can be treated as a nonparametric spatially varying temporal trend. Block averages of this over specific regions yield regional trend parameters, providing high resolution adjustment to the γ_m. Averaging over time, e.g., months or the year, provides local spatial adjustment at that temporal scale.

For convenience, for each of the processes, we adopt a separable covariance structure, (see e.g. Mardia and Goodall, 1993). That is,

$$\mathrm{Cov}\big\{w(\mathbf{s}_i, t'),\ w(\mathbf{s}_j, t)\big\} \quad = \quad \sigma_w^2\, \rho_{sw}(\mathbf{s}_i - \mathbf{s}_j; \phi_{sw})\, \rho_{tw}(t' - t; \phi_{tw}), \tag{4}$$

$$\mathrm{Cov}\big\{v(\mathbf{s}_i, t'),\ v(\mathbf{s}_j, t)\big\} \quad = \quad \sigma_v^2\, \rho_{sv}(\mathbf{s}_i - \mathbf{s}_j; \phi_{sv})\, \rho_{tv}(t' - t; \phi_{tv}). \tag{5}$$

In addition, the four ρ's are taken to be exponential covariance functions, i.e., $\rho(d; \phi) = \exp(-\phi d)$. We acknowledge the simplification associated with a separable specification but the additive form in (3) implies that the space-time component does not have a separable covariance function. (See De Iaco, Myers, and Posa, 2001 in this regard.) Also, we have explored Matèrn choices for the ρ's with different values for the smoothness parameter, ν, e.g., $\nu = 1$ and 1.5. We observed no gain in predictive performance by imposing additional smoothness on the surfaces, thus we use $\nu = 0.5$ (corresponding to the exponential model) for the balance of this paper.

From (1), β_0 is the intercept associated with the rural sites with $\beta_0 + \beta_2$ being the intercept for the urban/suburban sites. Similarly, β_1 is the slope associated with rural sites while $\beta_1 + \beta_3$ is associated with urban/suburban sites. Analogously, our motivation in introducing two random processes in (3) is to view $w(\mathbf{s}, t)$ as a *rural* or background zero mean process while $v(\mathbf{s}, t)$ adds *urban/suburban* spatio-temporal uncertainty, weighted by population density. Hence, $\mathrm{Var}(Y(\mathbf{s}, t)|p(\mathbf{s})) = \sigma_w^2 + p^2(\mathbf{s})\sigma_v^2$ so variance increases with population density. Note that, if we did not seek to force such order in the variance structure we could employ the form $\{1 - p(\mathbf{s})\}w(\mathbf{s}, t) + p(\mathbf{s})v(\mathbf{s}, t)$. In fact, we examined this specification and found it to fit much worse than our choice in (3).

Denote the unknown parameters by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_w^2, \sigma_v^2, \sigma_\epsilon^2)'$. We assume that, a priori, the β's are independent with distribution $N(0, A^2)$ with A^2 taken to be large. For the three variance parameters σ_ϵ^2, σ_w^2 and σ_v^2 we assume independent proper inverse gamma prior distributions, $IG(a, b)$ (with $a > 1$, hence, mean $b/(a - 1)$) to avoid having an improper posterior distribution. In fact, in our numerical example we set $a = 2$ and $b = 1$ so that the resulting prior distribution has mean 1 and infinite variance. We experimented with many other values of a and b; the parameter estimates and model choice criteria values were insensitive to moderate changes in the values of a and b (even when tending toward impropriety) because of the large amount of data providing information for the variance components.

Let $\boldsymbol{\phi} = (\phi_{sw}, \phi_{sv}, \phi_{tw}, \phi_{tv})'$ denote the decay parameters in the covariance function. Ideally, $\boldsymbol{\phi}$ would be estimated within the Bayesian model as well. However, under weak prior distributions, a Gibbs sampler for $\boldsymbol{\phi}$ and the σ^2's is often poorly behaved due to weak identifiability and extremely slow-mixing of the associated Markov chains. Moreover, the posterior distribution resulting from a limiting non-informative prior distribution can be improper, see e.g. Berger *et al.* (2001). Also, in a standard classical inference setting, Zhang (2004) shows that it is not possible to individually estimate all the parameters, (e.g. $\sigma_\epsilon^2, \sigma_w^2, \sigma_v^2$, and $\boldsymbol{\phi}$ in our model) consistently in a typical model for spatial data with a covariance function belonging to the Matern family. Lastly, difficulties in estimating $\boldsymbol{\phi}$ are also exacerbated by the large number of locations - time point combinations we work with here as well as the desire to do spatial prediction over a large spatial grid (3210 sites). The spatial interpolation using Bayesian methods becomes a much more computationally manageable task if $\boldsymbol{\phi}$ is kept fixed throughout the MCMC run. (More details are provided in Section 4.) In view of the above, we explore a four dimensional grid of $\boldsymbol{\phi}$ values and use a validation criterion to choose $\boldsymbol{\phi}$, see Section 5.

3.2 Joint posterior details

The log-likelihood is written as:

$$l(\boldsymbol{\theta}, \mathbf{W}, \mathbf{V}; \mathbf{Z}) \propto -\frac{N}{2}\log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{Z(\mathbf{s}_i, t) - \mu(\mathbf{s}_i, t) - w(\mathbf{s}_i, t) - p(\mathbf{s}_i)v(\mathbf{s}_i, t)\right\}^2,$$

where $\mathbf{Z}$ denotes all the data; $\mathbf{W}$ and $\mathbf{V}$ are defined below, and $N = nT$.

Let Σ_{sw} and Σ_{tw} denote the spatial and temporal correlation matrices of the $w(\mathbf{s}, t)$ process, Σ_{sv} and Σ_{tv} denote the correlation matrices of the $v(\mathbf{s}, t)$ process. That is, for $i, j = 1, \ldots, n$ and $k, l = 1, \ldots, T$, we have:

$$(\Sigma_{sw})_{ij} = \rho_{sw}(\mathbf{s}_i - \mathbf{s}_j; \phi_{sw}), \quad (\Sigma_{tw})_{kl} = \rho_{tw}(k - l; \phi_{tw}), \quad \text{and}$$

$$(\Sigma_{sv})_{ij} = \rho_{sv}(\mathbf{s}_i - \mathbf{s}_j; \phi_{sv}), \quad (\Sigma_{tv})_{kl} = \rho_{tv}(k - l; \phi_{tv}).$$

Let us define

$$W_{T\times n} = \begin{pmatrix} w_{\mathbf{s}_1,1} & w_{\mathbf{s}_2,1} & \cdots & w_{\mathbf{s}_n,1} \\ w_{\mathbf{s}_1,2} & w_{\mathbf{s}_2,2} & \cdots & w_{\mathbf{s}_n,2} \\ \vdots & \vdots & \vdots & \vdots \\ w_{\mathbf{s}_1,T} & w_{\mathbf{s}_2,T} & \cdots & w_{\mathbf{s}_n,T} \end{pmatrix} \text{ and } V_{T\times n} = \begin{pmatrix} v_{\mathbf{s}_1,1} & v_{\mathbf{s}_2,1} & \cdots & v_{\mathbf{s}_n,1} \\ v_{\mathbf{s}_1,2} & v_{\mathbf{s}_2,2} & \cdots & v_{\mathbf{s}_n,2} \\ \vdots & \vdots & \vdots & \vdots \\ v_{\mathbf{s}_1,T} & v_{\mathbf{s}_2,T} & \cdots & v_{\mathbf{s}_n,T} \end{pmatrix}.$$

We concatenate the matrices W and V to vectors $\mathbf{W}$ and $\mathbf{V}$ whence the prior specification for the $w(\mathbf{s}, t)$ process is now seen to be

$$\mathbf{W} \sim N\left(\mathbf{0}, \sigma_w^2\, \Sigma_{sw} \otimes \Sigma_{tw}\right)$$

where $\otimes$ denotes the Kronecker product. Similarly the prior specification for the $v(\mathbf{s}, t)$ process is

$$\mathbf{V} \sim N\left(\mathbf{0}, \sigma_v^2\, \Sigma_{sv} \otimes \Sigma_{tv}\right).$$

Assembling the pieces, the joint posterior distribution is given by:

$$\begin{aligned}
\log\left\{\pi(\boldsymbol{\theta}, \mathbf{W}, \mathbf{V}|\mathbf{z})\right\} \quad \propto \quad & \log(\pi(\boldsymbol{\beta})) + \log\left\{\pi(\sigma_\epsilon^2, \sigma_w^2, \sigma_v^2)\right\} - \frac{N}{2}\log(\sigma_\epsilon^2) \\
& -\; \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{z(\mathbf{s}_i, t) - \mu(\mathbf{s}_i, t) - w(\mathbf{s}_i, t) - p(\mathbf{s}_i)v(\mathbf{s}_i, t)\right\}^2 \\
& -\; \frac{N}{2}\log(\sigma_w^2) - \frac{T}{2}\log|\Sigma_{sw}| - \frac{n}{2}\log|\Sigma_{tw}| - \frac{1}{2\sigma_w^2}\mathbf{W}'(\Sigma_{sw}^{-1} \otimes \Sigma_{tw}^{-1})\mathbf{W} \\
& -\; \frac{N}{2}\log(\sigma_v^2) - \frac{T}{2}\log|\Sigma_{sv}| - \frac{n}{2}\log|\Sigma_{tv}| - \frac{1}{2\sigma_v^2}\mathbf{V}'(\Sigma_{sv}^{-1} \otimes \Sigma_{tv}^{-1})\mathbf{V},
\end{aligned}$$

where $\pi(\boldsymbol{\beta})$ and $\pi(\sigma_\epsilon^2, \sigma_w^2, \sigma_v^2)$ denote the prior distributions. This model is fitted using a Gibbs sampler. Development of the full conditional distributions with related discussion is provided in the Appendix.

4 Prediction details

4.1 A spatial model for urban indicator

Spatial prediction at location $\mathbf{s}'$ and time t' is based upon the predictive distribution of $Z(\mathbf{s}', t')$ described in (6) and (7) below. In (6) the mean function depends on the population density $p(\mathbf{s}')$, the urban-rural indicator $\alpha(\mathbf{s}')$ of the site $\mathbf{s}'$ and the seasonality of time t'. The seasonal indicators are obtained by examining the time t' for which we want the predictions. The population densities $p(\mathbf{s}')$ are obtained from census-tracts as we have discussed earlier in Section 2. However, as noted there, the $\alpha(\mathbf{s}')$ are only supplied for the 124 monitoring sites. They are not known for the grid of 3210

prediction locations adopted here. As a result we propose a model for $\alpha(\mathbf{s})$ through $p(\mathbf{s})$. If, a high resolution rural/urban surface were available, we would avoid this additional model specification.

The urban indicator is viewed as a latent variable and we require a model for $\alpha(\mathbf{s})$ for spatial prediction at a new site $\mathbf{s}$. In fact, the collection of urban indicators can be viewed as a binary process over the region. A natural way to view the binary process is through a latent or clipped Gaussian field (see, e.g., De Oliveira, 2000). In the present context, we would set $\alpha(\mathbf{s}) = 1$ if $u(\mathbf{s}) \geq 0$ where $u(\mathbf{s})$ is a Gaussian process with mean $a_0 + a_1 p(\mathbf{s})$, variance equal to 1, and correlation function $\rho(\mathbf{s} - \mathbf{s}'; \phi_u)$. This process encourages $u(\mathbf{s})$ to be close to $u(\mathbf{s}')$ if $\mathbf{s}$ is close to $\mathbf{s}'$, and thus a spatial pattern emerges in the $\alpha(\mathbf{s})$ map. Note that this model would also be fitted using a Gibbs sampler but that this is apart from fitting the main model in Section 3. In fact, the output from the prediction for $\alpha(\mathbf{s})$ can be stored separately and introduced when it is time to do the prediction. This model has been fitted and optimized for our data set in the usual way; we omit the details for brevity.

4.2 Weekly predictions

The modeling in Sections 3 and 4.1, due to the presence of the space-time process specification, allows us to interpolate the spatial surface at any time point t' which can be in the past or the future. More precisely, using (2) and (3), for a new location $\mathbf{s}'$ at time t', $Z(\mathbf{s}', t')$ is conditionally independent of $\mathbf{z}$ given $w(\mathbf{s}', t')$, $v(\mathbf{s}', t')$ and $\alpha(\mathbf{s}')$ with its distribution given by

$$Z(\mathbf{s}', t') \sim N\left(\mu(\mathbf{s}', t') + w(\mathbf{s}', t') + p(\mathbf{s}')v(\mathbf{s}', t'), \ \sigma_\epsilon^2\right). \tag{6}$$

The posterior predictive distribution of $Z(\mathbf{s}', t')$ is obtained by integrating over the unknown parameters in (6) with respect to the joint posterior distribution, that is,

$$
\begin{aligned}
\pi\left(Z(\mathbf{s}', t')|\mathbf{z}\right) \ = \ & \int \pi\left(Z(\mathbf{s}', t')|w(\mathbf{s}', t'), v(\mathbf{s}', t'), \alpha(\mathbf{s}'), \boldsymbol{\beta}, \sigma_\epsilon^2\right) \\
& \pi\left(w(\mathbf{s}', t')|\mathbf{W}, \phi_{sw}, \phi_{tw}, \sigma_w^2\right) \ \pi\left(v(\mathbf{s}', t')|\mathbf{V}, \phi_{sv}, \phi_{tv}, \sigma_v^2\right) \\
& \pi(\boldsymbol{\theta}, \mathbf{W}, \mathbf{V}|\mathbf{z}) \, dw(\mathbf{s}', t') \, dv(\mathbf{s}', t') \, d\mathbf{W} \, d\mathbf{V} \, d\boldsymbol{\theta}.
\end{aligned}
\tag{7}
$$

When using MCMC methods to draw samples from the posterior, the predictive distribution (7) is sampled by composition; draws from the posterior, $\pi(\boldsymbol{\theta}, \mathbf{W}, \mathbf{V}|\mathbf{z})$ enable draws for $w(\mathbf{s}', t')$ and $v(\mathbf{s}', t')$ and thus draws for $Z(\mathbf{s}', t')$. To report the predictions on the original scale, we simply work with the square of the predictive realizations drawn from (7).

In (7) we need to generate the random variables $w(\mathbf{s}', t')$ and $v(\mathbf{s}', t')$ conditional on the posterior samples at the observed locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and at the time points $1, \ldots, T$. For this we have:

$$\begin{pmatrix} w(\mathbf{s}', t') \\ \mathbf{W} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \ \sigma_w^2 \begin{pmatrix} 1 & \Sigma'_{sw}(\mathbf{s} - \mathbf{s}') \otimes \Sigma'_{tw}(\mathbf{t} - t') \\ \Sigma_{sw}(\mathbf{s} - \mathbf{s}') \otimes \Sigma_{tw}(\mathbf{t} - t') & \Sigma_{sw} \otimes \Sigma_{tw} \end{pmatrix}\right]$$

where $\Sigma_{sw}(\mathbf{s} - \mathbf{s}')$ is an $n \times 1$ column vector with the ith entry given by $\sigma(\mathbf{s}_i - \mathbf{s}') = \rho_{sw}(\mathbf{s}_i - \mathbf{s}'; \phi_{sw})$ and $\Sigma_{tw}(\mathbf{t} - t')$ is a $T \times 1$ column vector with the kth entry given by $\sigma(k - t') = \rho_{tw}(k - t'; \phi_{tw})$. Hence,

$$w(\mathbf{s}', t')|\mathbf{W} \sim N\left(\sum_{j=1}^{n}\sum_{k=1}^{T} b_{jk}(\mathbf{s}', t')w(\mathbf{s}_j, k), \ \sigma_w^2 C(\mathbf{s}', t')\right) \tag{8}$$

where

$$b_{jk}(\mathbf{s}', t') = \sum_{i=1}^{n}\sum_{l=1}^{T} \sigma(\mathbf{s}_i - \mathbf{s}')\sigma(l - t')(\Sigma_{sw}^{-1})_{ij}(\Sigma_{tw}^{-1})_{lk} \tag{9}$$

and

$$C(\mathbf{s}', t') = 1 - \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{T}\sum_{k=1}^{T} \sigma(\mathbf{s}_i - \mathbf{s}')\sigma(l - t')(\Sigma_{sw}^{-1})_{ij}(\Sigma_{tw}^{-1})_{lk}\sigma(\mathbf{s}_j - \mathbf{s}')\sigma(k - t'). \tag{10}$$

8

The conditional mean and variance are very computationally expensive to calculate due to the dimensionality of $\mathbf{W}$ and the very large number of sites $\mathbf{s}'$ for which we require predictions. However, by fixing the decay parameters $\boldsymbol{\phi}$, the quantities $b_{jk}(\mathbf{s}', t')$ and $C(\mathbf{s}', t')$ given in (9) and (10) need only be calculated once; no updating is required in the MCMC. In fact, to perform predictions for each week t' we calculate and store the 5928×3210 array $b_{jk}(\mathbf{s}', t')$ and the 3210 $C(\mathbf{s}', t')$ before starting the MCMC iterations. Again, we see the advantage of fixing $\boldsymbol{\phi}$ and then using model choice afterwards. Were we to sample $\boldsymbol{\phi}$ we would have to re-calculate $b_{jk}(\mathbf{s}', t')$ and $C(\mathbf{s}', t')$ at each MCMC iteration. The conditional distribution $v(\mathbf{s}', t')|\mathbf{V}$ is obtained similarly. In our example, for a given week, we will have to sample 3210 $w(\mathbf{s}', t')$'s and 3210 $v(\mathbf{s}', t')$'s, each conditioning on a 5928 ($=114 \times 52$) dimensional joint distribution.

4.3 Annual predictions

It is of interest to provide an annual prediction surface for the entire region. At first thought, this may seem to be a relatively straightforward problem of averaging the weekly predictions for the 52 weeks. However, the computational burdens are huge because of the need to compute the conditional distribution (8) (as well as the corresponding one for $v(\mathbf{s}', t')|\mathbf{V}$) 3210×52 times. The storage burden would be considerable as well; for L posterior draws, we would need to store a $3210 \times 52 \times L \times 2$ array. For, say, $L = 6000$ this is more than two billion numbers. We adopt the following approximation.

First, consider the annual average on the square-root scale $\bar{Z}(\mathbf{s}') = \frac{1}{52} \sum_{k=1}^{52} Z(\mathbf{s}', t)$. Analogous to (7), we can sample its predictive distribution since

$$\bar{Z}(\mathbf{s}') \sim N\left(\bar{\mu}(\mathbf{s}') + \bar{w}(\mathbf{s}') + p(\mathbf{s}')\,\bar{v}(\mathbf{s}'),\ \frac{\sigma_\epsilon^2}{52}\right),\tag{11}$$

where

$$\bar{\mu}(\mathbf{s}') = \frac{1}{52} \sum_{k=1}^{52} \mu(\mathbf{s}', k),\ \ \bar{w}(\mathbf{s}') = \frac{1}{52} \sum_{k=1}^{52} w(\mathbf{s}', k)\ \ \text{and}\ \ \bar{v}(\mathbf{s}') = \frac{1}{52} \sum_{k=1}^{52} v(\mathbf{s}', k).$$

As in Section 4.2 the predictive distribution for $\bar{Z}(\mathbf{s}')$ at a new site $\mathbf{s}'$ will require the conditional distributions $\bar{w}(\mathbf{s}')|\mathbf{W}$ and $\bar{v}(\mathbf{s}')|\mathbf{V}$. These conditional distributions are obtained respectively from the joint distributions $(\bar{w}(\mathbf{s}'), \mathbf{W})$ and $(\bar{v}(\mathbf{s}'), \mathbf{V})$. Simple arguments based on sufficiency show that the conditional distributions depend only upon $\bar{\mathbf{w}} = (\bar{w}(\mathbf{s}_1), \ldots, \bar{w}(\mathbf{s}_n))'$ and $\bar{\mathbf{v}} = (\bar{v}(\mathbf{s}_1), \ldots, \bar{v}(\mathbf{s}_n))'$. Straightforward calculation yields that

$$\text{Cov}\{\bar{w}(\mathbf{s}_i),\ \bar{w}(\mathbf{s}_j)\} = \sigma_w^2 \frac{\mathbf{1}'\Sigma_{tw}\mathbf{1}}{52^2} \rho_{sw}(\mathbf{s}_i - \mathbf{s}_j; \phi_{sw}).$$

Now we obtain that

$$\bar{w}(\mathbf{s}')|\bar{\mathbf{w}} \sim N\left[\Sigma_{sw}'(\mathbf{s} - \mathbf{s}')\Sigma_{sw}^{-1}\bar{\mathbf{w}},\ \ \sigma_w^2 \frac{\mathbf{1}'\Sigma_{tw}\mathbf{1}}{52^2}\left\{1 - \Sigma_{sw}'(\mathbf{s} - \mathbf{s}')\Sigma_{sw}^{-1}\Sigma_{sw}'(\mathbf{s} - \mathbf{s}')\right\}\right].\tag{12}$$

The conditional distribution $\bar{v}(\mathbf{s}')|\bar{\mathbf{v}}$ is obtained similarly.

The annual predictions on the square-root scale, $\bar{Z}(\mathbf{s}')$, are now obtained using equations (12) and (11), and the predictive distribution $\pi\left(\bar{Z}(\mathbf{s}')|\mathbf{z}\right)$ which is obtained from the equation analogous to (7).

The annual predictions on the original scale are given by:

$$\bar{Z}_{\text{orig}}(\mathbf{s}') = \frac{1}{52} \sum_{k=1}^{52} Z^2(\mathbf{s}', k) = \frac{1}{52} \sum_{k=1}^{52} \left\{Z(\mathbf{s}', k) - \bar{Z}(\mathbf{s}')\right\}^2 + \bar{Z}^2(\mathbf{s}').\tag{13}$$

We can use (13) to approximate the annual predictions as follows. We obtain the annual predictions on the square-root scale, $\bar{Z}(\mathbf{s}')$ using the details given above. The square of these predictions provide

the second term in the right hand side of (13). We then obtain an approximation to the first term of this identity by noting that

$$\frac{1}{52}\sum_{k=1}^{52}\left\{Z(\mathbf{s}',k)-\bar{Z}(\mathbf{s}')\right\}^2 \approx \frac{1}{R}\sum_{j=1}^{R}\left\{Z(\mathbf{s}',t_{k_j})-\bar{Z}_R(\mathbf{s}')\right\}^2$$

where R is a number much less than 52 and the sequence t_{k_j} is an equally spaced sub-sequence of $\{1,\ldots,52\}$ and $\bar{Z}_R(\mathbf{s}) = \sum_{j=1}^{R} Z(\mathbf{s},t_j)/R$. In our implementation we take R=4, and take the four weeks equally spaced across the year with one week in each calendar quarter. Results were similar for R=6.

To provide more detail, we obtain for each l, $l = 1,2,\ldots,L$, a realization of each term on the right hand side of (13), hence approximate realizations $\bar{Z}_{l,\mathrm{orig}}(\mathbf{s}'), l = 1,2,\ldots,L$, from the predictive distribution of $\bar{Z}_{\mathrm{orig}}(\mathbf{s}')$. Such realizations enable us to obtain an annual mean surface and an estimate of the variability in the surface at each location.

5 Summary of analysis

5.1 Model choice

We first select the decay parameters ϕ using a validation criterion described below. Recall that we have set aside data from 10 sites of which one is rural and nine are urban sites. In Section 5.3 below, we examine spatial prediction for all 52 weeks for each of these sites. However, in choosing ϕ, we focus on week 20 (May 14–21) and week 46 (November 11–17) later in this section. We choose these two weeks because the $PM_{2.5}$ concentrations in week 20 are near the overall average and the concentrations in week 46 are larger with greater variability, see the boxplots for these two weeks in Figure 3. We consider the validation mean-square error

$$\mathrm{VMSE} = \frac{1}{n_v}\sum_{i=1}^{10}\sum_{k=1}^{2}\left(Z(\mathbf{s}'_i,t_k)-\hat{Z}(\mathbf{s}'_i,t_k)\right)^2 I(Z(\mathbf{s}'_i,t_k))$$

where $I(Z(\mathbf{s}'_i,t_k)) = 1$ if $Z(\mathbf{s}'_i,t_k)$ has been observed and 0 otherwise; $t_1 = 20$ and $t_2 = 46$ corresponding to weeks 20 and 46; $n_v = \sum_{i=1}^{10}\sum_{k=1}^{2} I(Z(\mathbf{s}'_i,t_k))$; $\mathbf{s}'_i, i = 1,\ldots,10$ are the locations of the 10 validation sites; and $\hat{Z}(\mathbf{s}'_i,t_k)$ is the predicted value at location $\mathbf{s}'_i$ and at time t_k, see Section 4.2 for methods to obtain these. We calculate VMSE for different values of ϕ to choose an optimal value. An alternative is to use a criterion based on Mahalanobis distance developed in Sahu and Mardia (2005).

We searched for the optimal ϕ in a four dimensional grid. For the spatial decay parameters ϕ_{sw} and ϕ_{sv} we considered the the values 0.01, 0.02, 0.03, 0.05, and 0.10. These values correspond to approximate ranges of 300, 150, 100, 60, and 30 kilometers, respectively. Note that we use the relationship $\exp(-\phi d) \approx 0.05$ to determine the range for a given value of the decay parameter ϕ. To determine the optimal ϕ_t we searched among the values 1.5, 1, 0.75; these correspond to time dependence of 2, 3 and 4 weeks, respectively for the temporal processes.

We calculated the VMSE, for all 225 ($=5^2 \times 3^2$) combinations of ϕ values. The combination $\phi_{sw} = 0.02$, $\phi_{sv} = 0.05$, $\phi_{tw} = 1.0$ and $\phi_{tv} = 1.0$ provided the smallest estimated VMSE. We have found that VMSE is not sensitive to the choice of the decay parameters near these best values. As a result, although it is possible to further refine the grid in a neighborhood of the best value we do not explore beyond our grid here.

Given the decay parameters, to compare different choices for $Y(\mathbf{s},t)$ in (3), we use the formal predictive model choice criterion (PMCC) (see e.g. Laud and Ibrahim, 1995; Gelfand and Ghosh,

1998) given by:

$$PMCC = \sum_{i=1}^{n} \sum_{k=1}^{T} \left[\left\{ Z_{\mathrm{obs}}(\mathbf{s}_i, t_k) - E\left(Z_{\mathrm{pred}}(\mathbf{s}_i, t_k)\right) \right\}^2 + \mathrm{Var}\left(Z_{\mathrm{pred}}(\mathbf{s}_i, t_k)\right) \right], \qquad (14)$$

where $Z_{\mathrm{pred}}(\mathbf{s}_i, t_k)$ is a replicate observation corresponding to the observed value of $Z(\mathbf{s}_i, t)$, denoted by $Z_{\mathrm{obs}}(\mathbf{s}_i, t_k)$, under the assumed model, see equation (7). In (14), the first term on the right hand side is obviously a goodness-of-fit term (G), the second can be viewed as a penalty term (P) in the sense that large predictive variances make a model less desirable. In the summation we have excluded the missing observations.

For instance, PMCC allows us to evaluate the improvement made by the proposed model over a simpler linear regression model. By removing the spatio-temporal processes from (3) we can re-write the model (2) as a linear regression model

$$Z(\mathbf{s}_i, t) = \mu(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t), \ i = 1, \ldots, n, \ \ t = 1, \ldots, T.$$

This is our base model and our proposed model with the spatio-temporal processes $w(\mathbf{s}, t)$ and $v(\mathbf{s}, t)$ is the full non-stationary spatio-temporal model. Comparing these two, we find that PMCC for the base model is 3293.74 (G) + 3295.54 (P) = 6589.28 while for the full spatio-temporal model it is 141.06 (G) + 618.67 (P) =759.73. Though the PMCC criterion is not readily calibrated, it is evident that the spatio-temporal model provides a huge improvement over the linear regression model. There are a variety of intermediate models between the base model and the proposed model. In particular, as the discussion in Section 5.2 reveals, with few rural sites in our dataset, the model where $p(\mathbf{s})$ is set to 0 is indistinguishable from the model in (3).

5.2   Model diagnostics

The objective of this section is to provide various residual plots to check model adequacy. For instance, in linear models a residual against the fitted values plot is used to check many assumptions including homoscedasticity. This plot, provided in the top panel of Figure 5, shows no curvature or pattern and confirms that the homoscedasticity assumption is acceptable; there are only a few extreme values.

The middle panel of Figure 5 plots the residuals against the month and it shows roughly equal variability of the residuals for all the months. We have calculated the variance of the plotted residuals for each month. The ratio of the largest to the smallest variance is 2.2 which further supports the conclusion that there is no unusually high residual variability in any particular month. The bottom panel of Figure 5 plots the residuals against the scaled population density values. Like the preceding ones this does not show any pattern as well.

Finally, in Figure 6 we provide time series plot of the residuals for 15 sites chosen at random. All the plots have the same Y-axis scale as is often done in residual analysis. For improved clarity we have provided individual plots, another way to see those is in a single graph with multiple time series plots with the same Y-axis scale. This plot confirms that there is no temporal variation left in the residuals. We have checked this for all the remaining sites as well and the same conclusion holds for all of them.

All these residual plots correspond to data used in fitting the model. These show that the model fits the data well. However, these do not guarantee accuracy in out of sample predictions which we consider next.

5.3   Validation

We return to the validation for the data from the ten randomly chosen hold-out sites, see Section 2.1, with locations numbered 1 to 10 in Figure 1. We do this for all the 52 weeks. The 95% prediction

intervals are plotted in Figure 7. The observed data are also super-imposed in the plot. Site 10 in Illinois is the only rural site, this respects the proportion of rural sites used in modeling. Out of the 520 observations to be validated 15 were missing. Of the 505 95% predictive intervals which we can compare with actual observations, 27 fail to include the observed value. That is, 94.7% of the 95% prediction intervals are correct, an indication that the model is performing as it should for out of sample predictions.

5.4   Parameter estimates

The parameter estimates associated with the spatio-temporal model in Section 5.1 are given in Table 2. From (1), since β_1 is significantly positive, higher population density explains higher concentrations of $PM_{2.5}$. The significance of β_2 asserts a higher urban intercept than rural intercept. Offsetting this, the negative β_3 shows a smaller slope for urban areas compared with rural areas. While the urban line operates over a different domain for $p(\mathbf{s})$ than the rural line, this diminished slope is expected. Viewed as an interaction term, some antagonism between urban/rural and population density is anticipated due to the commercial zones in urban/suburban areas with low population density. Apart from June (γ_6), the estimates of the monthly seasonal effects are all significant with January as the base month. This is in accord with the descriptive Figure 3 of seasonal effects. The posterior standard deviations of these parameters are roughly similar with differences attributable to missing observations and the varying number of weeks in a month. The fact that these effects are significant in the presence of spatio-temporal random effects is sufficient evidence that their inclusion will improve prediction.

The estimates of the variance components σ_ϵ^2, σ_w^2 and σ_v^2 show that more variation is explained by the spatio-temporal effects than the pure error process $\epsilon(\mathbf{s}, t)$. It is also seen that the latent $w(\mathbf{s}, t)$ process explains most of the spatial variability. The additional process $v(\mathbf{s}, t)$ captures only a small portion of the overall spatial variability. This is expected since from the previous paragraph, we find a significant mean difference between rural and non-rural sites but as seen in Table 1 there is only a small difference in variability between the rural and urban sites. Moreover, with monitoring data that contains a higher proportion of rural sites, the inclusion of urban and rural processes would be expected to improve prediction.

5.5   Prediction

In the interest of space, we illustrate weekly predictions for weeks, week 20 (May 14–21) and week 46 (November 11–17) as selected above in Section 5.1. For prediction, we recall that there will be variation in the mean level across the region as well as a variance associated with the prediction at a given location. The predicted mean surfaces of $PM_{2.5}$ are given in Figure 8. Despite some similarities, there are large differences with regard to the locations of highs and lows between the two maps. Also, there is more spatial variation in the predicted means for week 46 than for week 20.

The standard deviations of the above weekly predictions are given in Figure 9. These maps show spatial pattern in standard deviations. As expected, they are higher for locations that are far away from the monitoring stations. In addition, they tend to be larger where mean levels are larger. Since the mean levels were larger for week 46, the standard deviations are larger for week 46. In fact, the actual values of these predictive standard deviations are also comparable to the standard deviations of the data, see the mean-variance plot on the original scale (top-panel) of Figure 2. For both weeks, the urban areas are seen to have higher $PM_{2.5}$ concentrations than the rural areas. Ohio has generally higher $PM_{2.5}$ concentrations in comparison to the other two states, perhaps due to the high concentration of industry in this state.

We show the annual prediction surface (posterior mean and standard deviation) in Figure 10. As expected, the annual averaging smooths the surface in comparison to the weekly surfaces given

in Figure 8. Still, urban areas have the highest predicted concentrations and many areas exceed the level of the national air quality $PM_{2.5}$ standard (CFR, 2002) of 15 $\mu g/m^3$. Also, we observe typical patterns of smaller standard errors near the monitoring sites.

6 Conclusions and further work

We have proposed and fitted a flexible nonstationary spatio-temporal model for weekly $PM_{2.5}$ concentrations for Ohio, Indiana and Illinois. In particular, we have allowed for a shift in mean level and a variability increase as we go from rural to urban sites. Apart from estimation of model parameters, we have shown how to implement both weekly and annual prediction.

Future work will seek to compare this three-state region with other regions of the U.S. Eventually, we would like to develop annual national maps under this modeling but this will require suitable approximations to enable feasible computation. A natural extension of our modeling would consider a non-separable form for the covariance functions for the $w(\mathbf{s}, t)$ and $v(\mathbf{s}, t)$ processes. Also our modeling could be specified to allow for temporally varying variance in the pure error component. It could also be extended to simultaneous modeling of $PM_{2.5}$ along with, say PM_{10} or perhaps ozone. It is anticipated that there will be association in levels of these pollutants and thus we could improve prediction for individual pollutants using data for all of them. A multivariate spatio-temporal process model will be required.

7 Appendix

Straightforward calculation yields the following complete conditional distributions:

$$\frac{1}{\sigma_\epsilon^2} \sim G\left(\frac{N}{2} + a, b + \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{z(\mathbf{s}_i, t) - \mu(\mathbf{s}_i, t) - w(\mathbf{s}_i, t) - p(\mathbf{s}_i)v(\mathbf{s}_i, t)\right\}^2\right),$$
$$\frac{1}{\sigma_w^2} \sim G\left(\frac{N}{2} + a, b + \mathbf{W}'(\Sigma_{sw}^{-1} \otimes \Sigma_{tw}^{-1})\mathbf{W}\right),$$
$$\frac{1}{\sigma_v^2} \sim G\left(\frac{N}{2} + a, b + \mathbf{V}'(\Sigma_{sv}^{-1} \otimes \Sigma_{tv}^{-1})\mathbf{V}\right).$$

The complete conditional distributions of $\boldsymbol{\beta}$, $w(\mathbf{s}_i, t)$, $v(\mathbf{s}_i, t)$ are normal with appropriate parameters. The complete conditional distribution of β_j is normal with mean

$$\sigma_{\beta_j}^2\left[\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}\sum_{t=1}^{T}x_j(\mathbf{s}_i, t)\left\{z(\mathbf{s}_i, t) - \mu(\mathbf{s}_i, t) + \beta_j x_j(\mathbf{s}_i, t) - w(\mathbf{s}_i, t) - p(\mathbf{s}_i)v(\mathbf{s}_i, t)\right\}\right],$$

and variance

$$\sigma_{\beta_j}^2 = \left(\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}\sum_{t=1}^{T}x_j(\mathbf{s}_i, t)^2 + \frac{1}{L^2}\right)^{-1}, j = 0, \ldots, 2 + M,$$

where $x_j(\mathbf{s}_i, t)$ is the jth element of the vector

$$\mathbf{x}(\mathbf{s}_i, t) = \left(1, p(\mathbf{s}_i), \alpha(\mathbf{s}_i), \alpha(\mathbf{s}_i) \times p(\mathbf{s}_i), u(t, 2), \ldots, u(t, M)\right)'.$$

This sampling scheme updates the components of $\boldsymbol{\beta}$ one after another. We, however, sample $\boldsymbol{\beta}$ in a single block for faster convergence. We omit the details for brevity.

We sample the random effects $w(\mathbf{s}_i, t)$ and $v(\mathbf{s}_i, t)$ en-bloc as follows. Let $\mathbf{w}_{*j}$ denote the jth column of W. The prior complete conditional distribution of $\mathbf{w}_{*j}$ for $j = 1, \ldots, n$ given all other columns $i \neq j, i = 1, \ldots, n$ is normal with mean $\boldsymbol{\zeta}_j$ and covariance Λ_j where

$$\boldsymbol{\zeta}_j = -\sum_{i \neq j, i=1}^{n}\frac{(\Sigma_{sw})_{ij}^{-1}}{(\Sigma_{sw})_{jj}^{-1}}\mathbf{w}_{*i}, \text{ and } \Lambda_j = \sigma_w^2\frac{1}{(\Sigma_{sw})_{jj}^{-1}}\Sigma_{tw}.$$

13

The likelihood contribution for $\mathbf{w}_{*j}$ is also normal with

$$\text{mean} = \boldsymbol{\xi}_j = \mathbf{z}_{*j} - \boldsymbol{\mu}_{*j} - p_j \mathbf{v}_{*j}, \text{and covariance} = \sigma_\epsilon^2 I,$$

where I is the identity matrix of appropriate order. The posterior complete conditional distribution is now seen to be normal with mean

$$\boldsymbol{\chi}_j = \Omega_j \left(\frac{1}{\sigma_\epsilon^2} \boldsymbol{\xi}_j + \Lambda_j^{-1} \boldsymbol{\zeta}_j \right) \quad \text{and covariance } \Omega_j = \left(\frac{I}{\sigma_\epsilon^2} + \Lambda_j^{-1} \right)^{-1}.$$

The complete conditional distribution of $\mathbf{v}_{*j}$ is obtained as follows. (We re-use the same intermediate symbols for notational simplicity.) The prior complete conditional distribution of $\mathbf{v}_{*j}$ given all other columns is normal with mean $\boldsymbol{\zeta}_j$ and covariance Λ_j where

$$\boldsymbol{\zeta}_j = - \sum_{i \neq j, i=1}^{n} \frac{(\Sigma_{sv})_{ji}^{-1}}{(\Sigma_{sv})_{jj}^{-1}} \mathbf{v}_{*i}, \text{ and } \Lambda_j = \sigma_v^2 \frac{1}{(\Sigma_{sv})_{jj}^{-1}} \Sigma_{tv}.$$

Two cases arise depending on the value of $p_j = p(\mathbf{s}_j)$. If $p_j = 0$ then there is no likelihood contribution for $\mathbf{v}_{*j}$ as it does not appear in the likelihood and we simply simulate it from the above prior conditional distribution. We now consider the case of $p_j > 0$. The likelihood contribution for $\mathbf{v}_{*j}$ is normal with

$$\text{mean} = \boldsymbol{\xi}_j = \frac{(\mathbf{z}_{*j} - \boldsymbol{\mu}_{*j} - \mathbf{w}_{*j})}{p_j}, \quad \text{and covariance} = \frac{\sigma_\epsilon^2}{p_j^2} I,$$

for the sites j for which $p_j > 0$. The posterior complete conditional distribution is now seen to be normal with mean

$$\boldsymbol{\chi}_j = \Omega_j \left(\frac{p_j^2 \boldsymbol{\xi}_j}{\sigma_\epsilon^2} + \Lambda_j^{-1} \boldsymbol{\zeta}_j \right) \quad \text{and covariance } \Omega_j = \left(I \frac{p_j^2}{\sigma_\epsilon^2} + \Lambda_j^{-1} \right)^{-1}.$$

Note that this is well defined even if $p_j = 0$.

ACKNOWLEDGMENT

REFERENCES

Berger, J. O., de Oliveira, V. and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, **96**, 1361–1374.

Brown, P. J., Le, N. D., Zidek, J. V. (1994). Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics*, **22**, 489–510.

Carroll, R. J., Chen, R., George, E. I., Li, T.H., Newton, H.J., Schmiediche, H. and Wang, N. (1997). Ozone exposure and population density in Harris County, Texas. *Journal of the American Statistical Association*, **92**, 392–404.

Cressie, N., Kaiser, M. S., Daniels, M. J., Aldworth, J., Lee, J., Lahiri, S. N., Cox, L. (1999). Spatial Analysis of Particulate Matter in an Urban Environment. In *GeoEnv II: Geostatistics for Environmental Applications*, (Eds. J. Gmez-Hernndez, A. Soares, R. Froidevaux)) Kluwer:Dordrecht, pp 41–52.

CFR (2002). Code of Federal Regulations, Chapter 40, Part 50, Appendix N, pp 121–127.

De Iaco, S., Myers, D. E., and Posa, D. (2001). Space-time analysis using a generalized product-sum model, *Statistics and Probability Letters*, **52**, 21–28.

De Oliveira, V. (2000). Bayesian Predicition of Clipped Gaussian Fields. *Computational Statistics and Data Analysis*, **34**, 299–314.

Gelfand, A. E. and Ghosh, S. K. (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika*, **85**, 1–11.

Guttorp, P., Meiring, W. and Sampson, P. D. (1994). A Space-time Analysis of Ground-level Ozone Data. *Environmetrics*, **5**, 241–254.

Haas, T.C. (1995). Local Prediction of a Spatio-Temporal Process With an Application to Wet Sulfate Deposition. *Journal of the American Statistical Association*, **90**, 1189–1199.

Kibria, B., Golam, M. Sun, L., Zidek, J. V., Le, N.D. (2002). Bayesian spatial prediction of random space-time fields with application to mapping $PM_{2.5}$ exposure. *Journal of the American Statistical Association*, **97**, 112–124.

Laud, P. W. and Ibrahim, J. G. (1995). Predictive Model Selection. *Journal of the Royal Statistical Society*, B, **57**, 247–262.

Le, N. D., Sun, W., Zidek, J. V. (1997). Bayesian multivariate spatial interpolation with data missing by design. *Journal of the Royal Statistical Society*, Series B, **59**, 501–510.

Mardia, K. V. and Goodall, C., (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In *Multivariate Environmental Statistics*. (Eds. G. P. Patil and C. R. Rao). Amsterdam: Elsevier, pp 347–386.

Sahu, S. K. and Mardia, K. V. (2005). A Bayesian Kriged-Kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society*, Series C, **54**, 223–244.

Shaddick, G. and Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society*, Series C. **51**, 351–372.

Smith, R. L., Kolenikov, S. and Cox, L. H. (2003). Spatio-Temporal modelling of $PM_{2.5}$ data with missing values. *Journal of Geophysical Research-Atmospheres*, **108** D24 9004, doi:10.1029/2002JD002914.

Sun L., Zidek, J. V., Le, N. D. and Ozkaynak, H. (2000). Interpolating Vancouver's daily ambient PM10 field. *Environmetrics*, **11**, 651–663.

U. S. Environmental Protection Agency (2004). Particulate Matter Research Program, Five Years of progress. U. S. Enviromental Protection Agency, Office of Research and Development, Washington DC 20460, EPA 600/R-04/058

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99**, 250–261.

Zidek, J. V., Sun, L., Le, N., Ozkaynak, H.(2002). Contending with space-time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM_{10} field. *Environmetrics*, **13**, 595–613.

Table 2: The estimates of the parameters.

	mean	sd	95% interval
β_0	4.218	0.080	(4.045, 4.369)
β_1	0.374	0.049	(0.282, 0.471)
β_2	0.758	0.350	(0.0934, 1.442)
β_3	-0.910	0.351	$(-1.599, -0.237)$
γ_2	-0.645	0.080	$(-0.797, -0.490)$
γ_3	-0.718	0.081	$(-0.872, -0.554)$
γ_4	-1.163	0.095	$(-1.343, -0.954)$
γ_5	-0.786	0.094	$(-0.991, -0.600)$
γ_6	0.016	0.091	$(-0.153, 0.209)$
γ_7	-0.772	0.075	$(-0.924, -0.612)$
γ_8	-0.684	0.082	$(-0.839, -0.520)$
γ_9	-1.134	0.107	$(-1.354, -0.915)$
γ_{10}	-1.057	0.087	$(-1.220, -0.888)$
γ_{11}	-0.962	0.091	$(-1.146, -0.774)$
γ_{12}	-1.023	0.082	$(-1.177, -0.862)$
σ_ϵ^2	0.066	0.003	(0.061, 0.072)
σ_w^2	0.395	0.012	(0.372, 0.419)
σ_v^2	0.062	0.009	(0.046, 0.081)
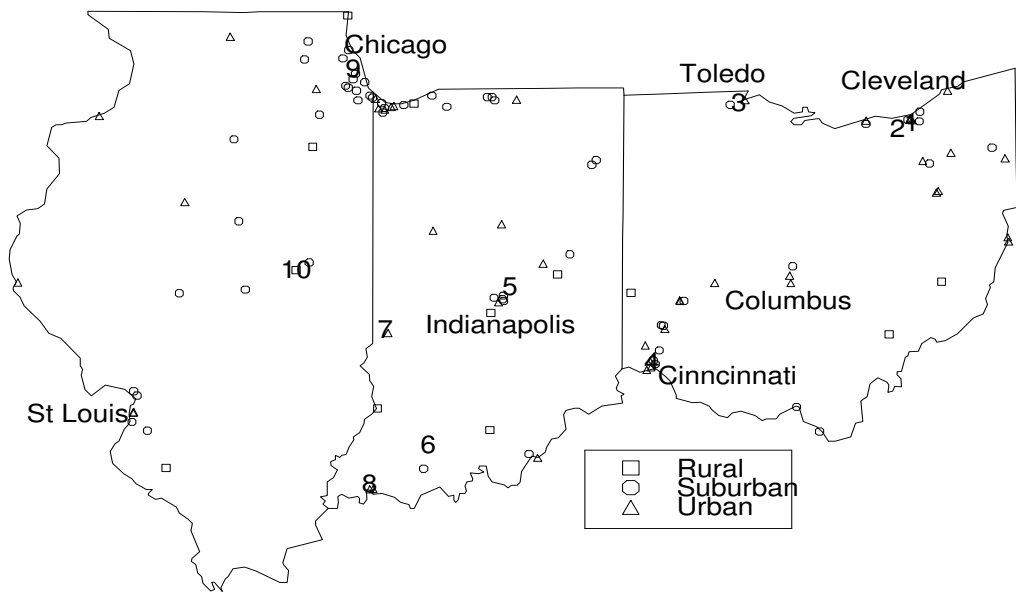
Figure 1: The 114 monitoring sites in Illinois, Indiana and Ohio. The additional 10 validation sites are numbered 1 to 10.
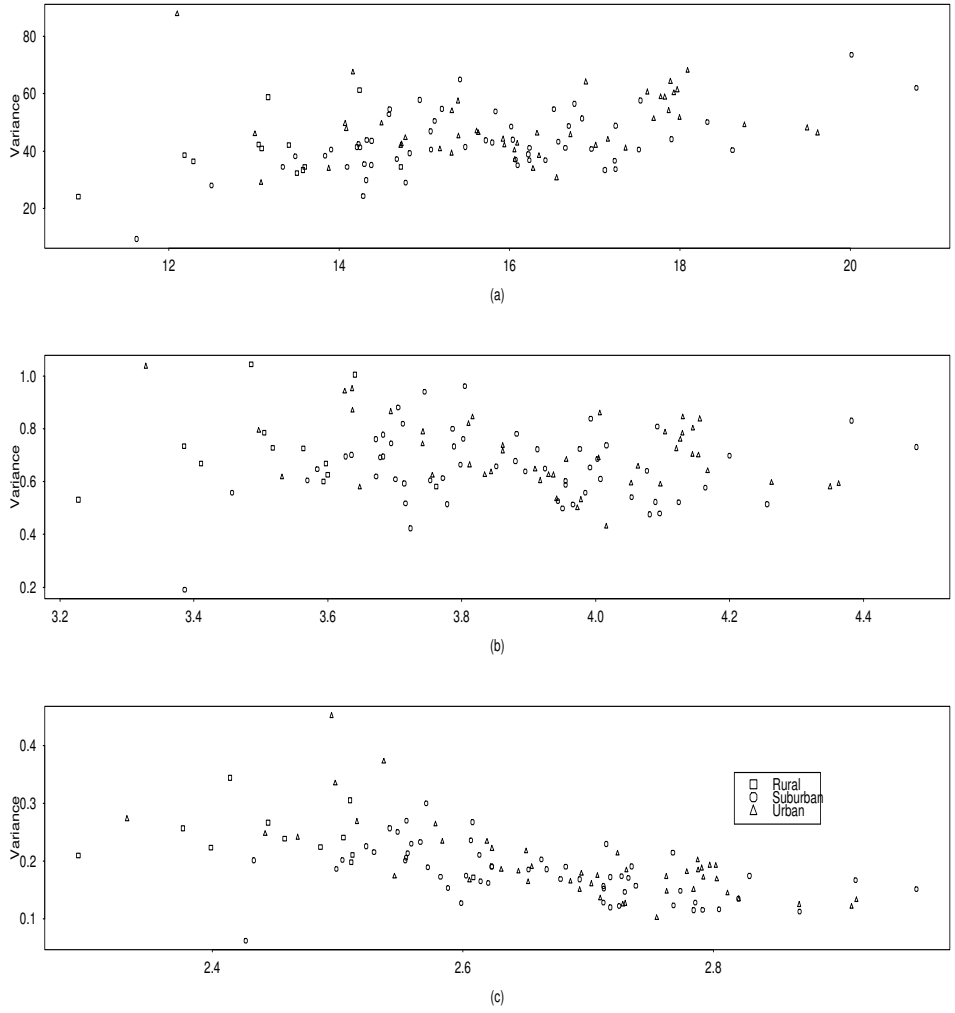
Figure 2: The variance against the mean of PM$_{2.5}$ concentration levels in each of the 114 sites. Panel (a) is on the original scale, panel (b) is on the square root scale and panel (c) is on the log scale.
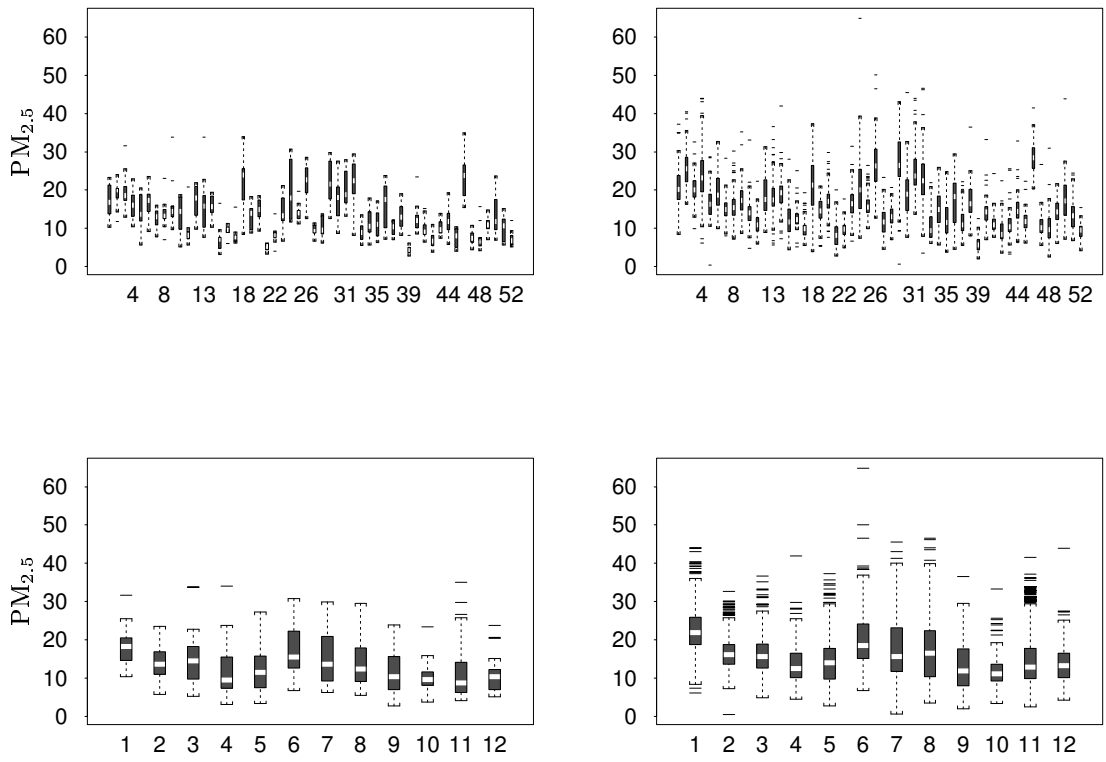
Figure 3: Boxplot of the PM$_{2.5}$ concentrations by weeks and months. The first column is for the 12 rural sites while the second column is for the 102 suburban and urban sites. The weeks, labelled in the two plots in the first row, denote the last week of the months.

20
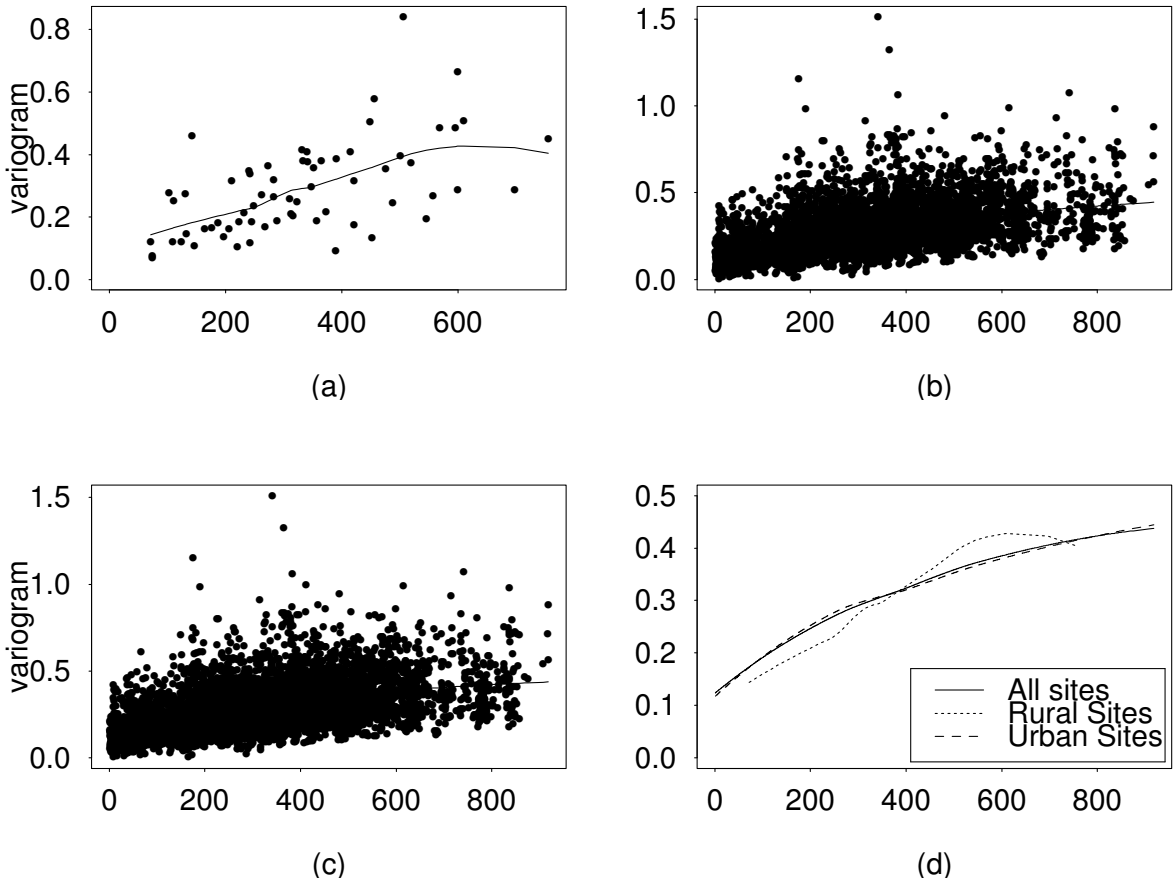
Figure 4: Variogram clouds and their lowess fits: (a) rural sites, (b) urban sites and (c) all sites. The three lowess curves from panels (a)-(c) have been plotted again in panel (d).
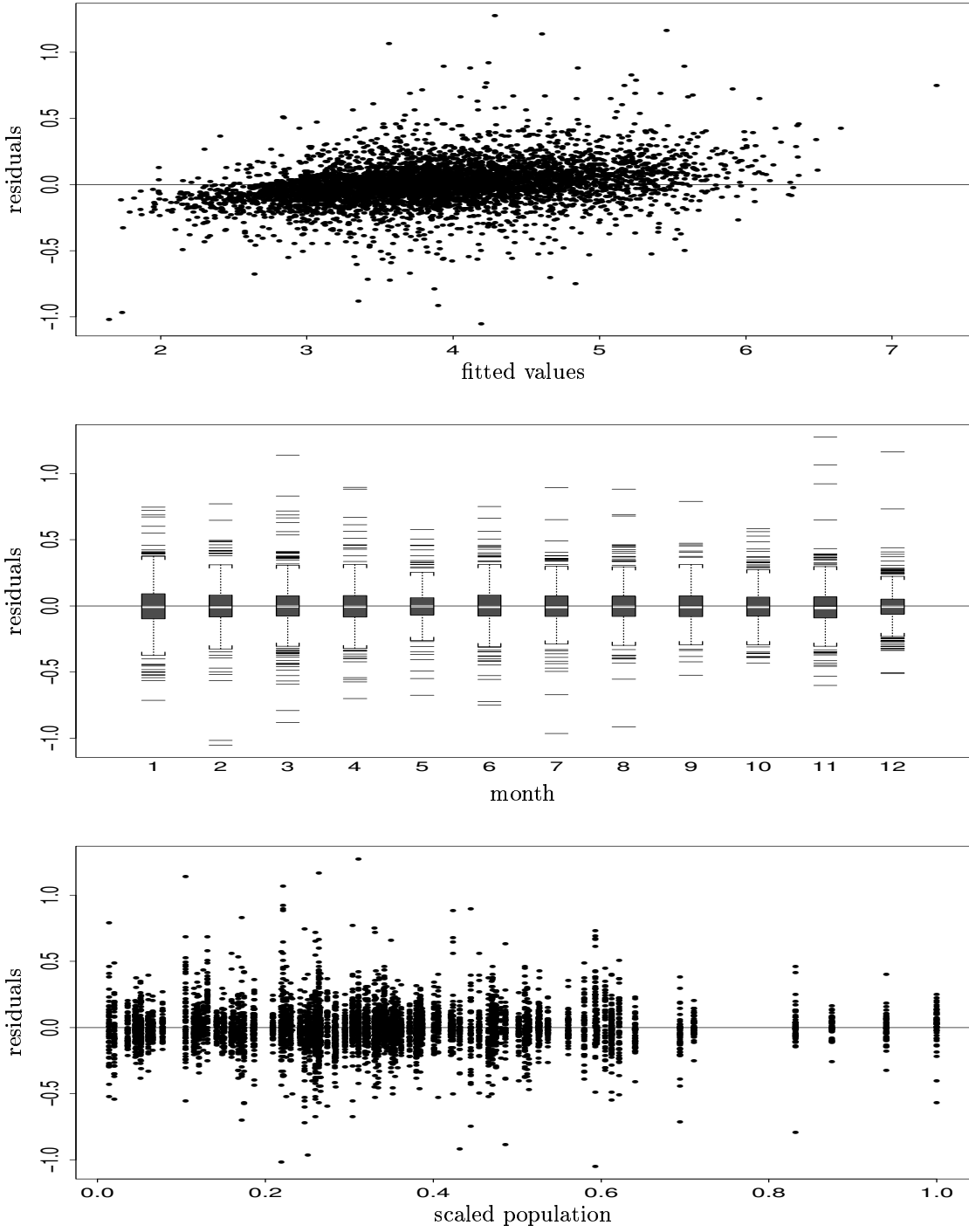
Figure 5: Plot of the residuals against: (a) the fitted values (top panel), (b) the months (middle panel), and (c) the population density (bottom panel).
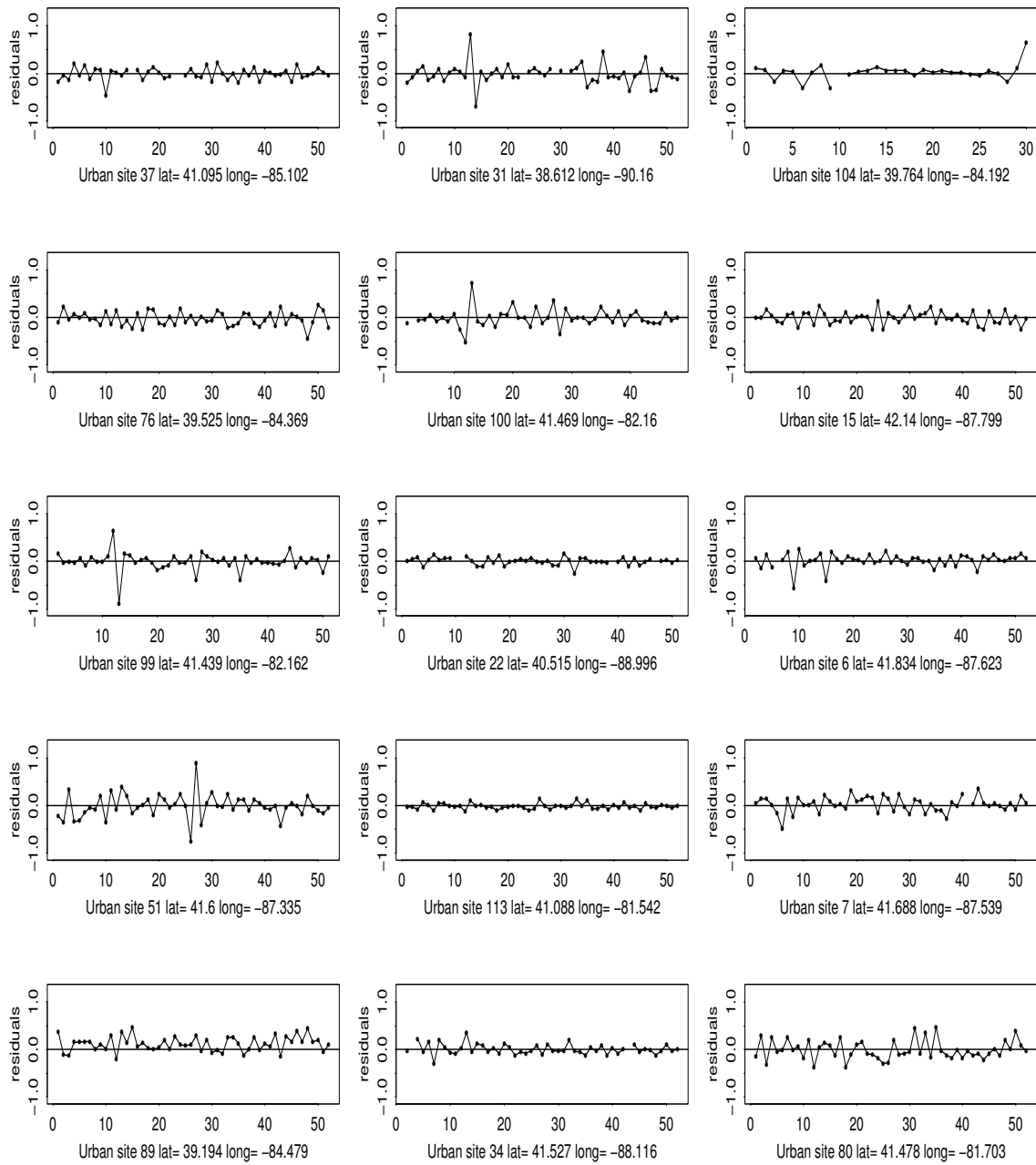
Figure 6: Time series plot of the residuals at 15 randomly chosen sites.
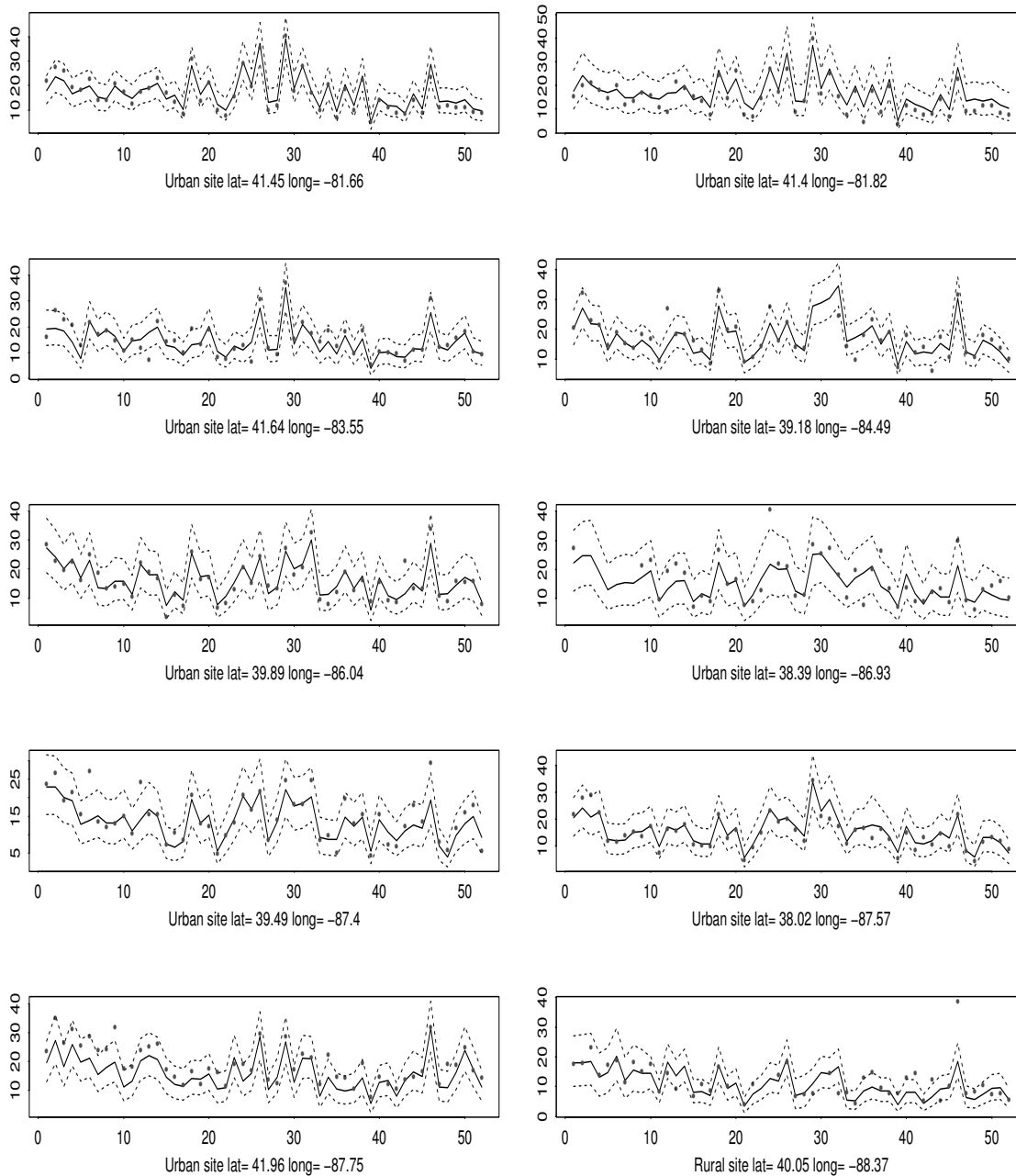
Figure 7: Validation plots for the 10 reserved sites. The observations are the points in the graph, the predictions are plotted as solid lines and the 95% prediction intervals are plotted as dotted lines.

(a)

(b)
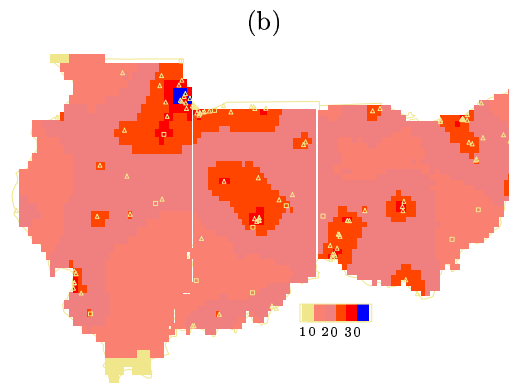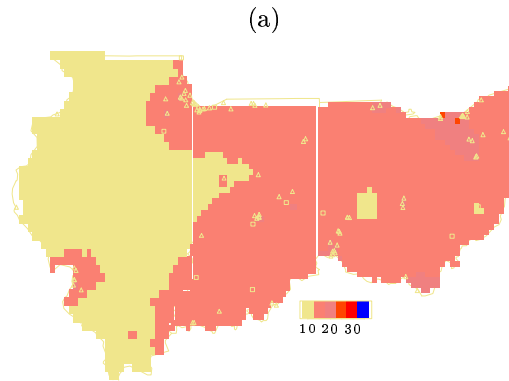
Figure 8: The predicted PM$_{2.5}$ levels (a) for week 20 (May 14–21) and (b) for week 46 (November 11–17).

(a)

(b)
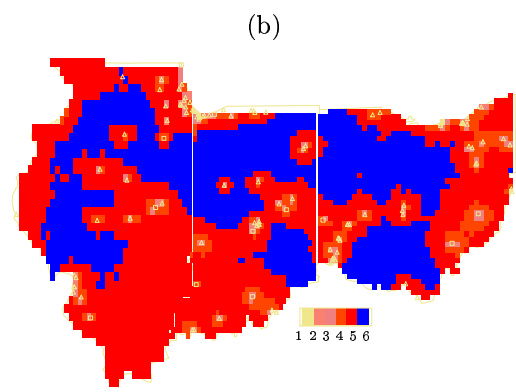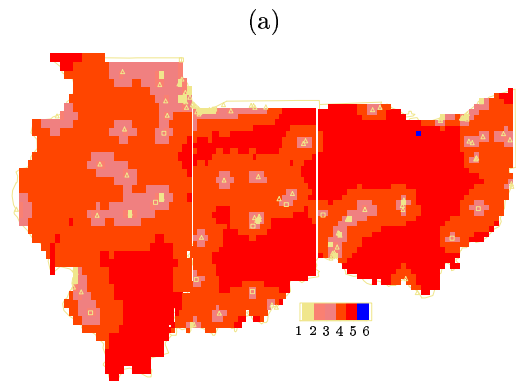
Figure 9: The standard deviations of predictions (a) for week 20 and (b) for week 46.
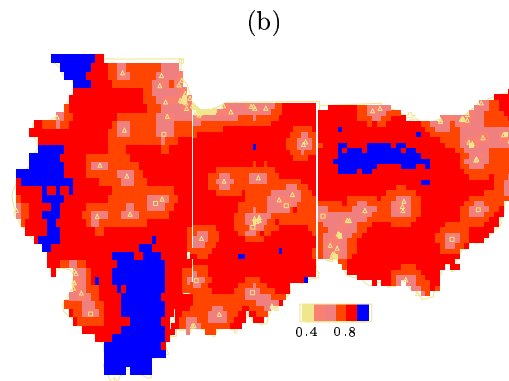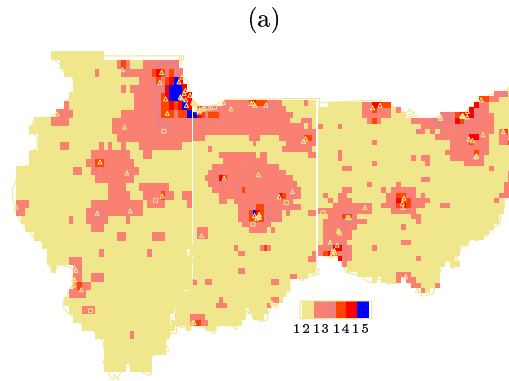
26

(a)

(b)

Figure 10: (a) Predicted annual average PM$_{2.5}$ levels and (b) their standard deviations.