Hierarchical Bayesian models for space-time air pollution data

Sujit K. Sahu *

June 14, 2011

Abstract

Recent successful developments in Bayesian modeling and computation for large and complex data sets have led to a step change in the analysis of space-time air pollution data. Accurate predictions and inferences, as a result of joint modeling of spatial and temporal dependence, are being made even for summaries and aggregates in time and/or space. Modelers are increasingly benefiting from their ability to reducing uncertainty in the inference statements for the aggregates, in addition, by combining information from several sources in a hierarchical Bayesian framework. The information sources may include actual observed data from several heterogeneous monitoring networks, outputs of community numerical models, meteorological observations, land use surfaces and power station emission volumes. The best statistical model for the particular problem at hand recognizes the relative contribution of the available sources and decides on their optimum role in it. In this chapter we develop a hierarchical auto-regressive Bayesian model for space-time air pollution data and illustrate the benefits of modeling with a real data example on monitoring ozone pollution. We report substantial gains in predictive mean square error for the proposed model over some other currently available competing modeling methods.

1 Introduction

The clean air act, amended in 1990 by the legislators in the United States of America (USA), requires two types of air quality standards to be maintained for six most important air pollutants: ozone, particulate matter, carbon monoxide, nitrogen oxides, sulfur dioxide, and lead. The first, *primary* standards, set limits to protect public health, including the health of "sensitive" populations such as asthmatics, children, and the elderly. The *secondary* standards set limits to public welfare, provide protection against decreased visibility, damage to animals, crops, vegetation, and buildings. The website *http://epa.gov/air/criteria.html* provides the current values of these standards. Out of the six pollutants, particulate matter and ozone have received the most attention in the literature and these two are the focus of this chapter, although the modeling methods are applicable to other pollutants as well.

To monitor compliance to the air quality standards and to evaluate exposure to air pollution the United States Environmental Protection Agency (USEPA) has developed several sparse networks of monitoring stations covering the whole of the United States (US). Data obtained from these sparse networks must be processed using stochastic spatial and spatio-temporal models to make valid inference for air pollution

*Sujit K. Sahu is Senior Lecturer, School of Mathematics, Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, UK (Email: S.K.Sahu@soton.ac.uk).

levels at particular sites, such as urban areas, based on rigorous statistical methods. In fact, 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 quality 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.

Particulate matter is a complex mixture of extremely small particles and liquid droplets and is harmful to human health when inhaled. There are two variations of particulate matter. Particles with diameters less than 10 micrometers (μm) are called PM_{10} and those with diameters less than 2.5μm are called $PM_{2.5}$. PM_{10} are found near roadways and dusty industries and $PM_{2.5}$ are found in smoke and haze. A lot of modeling effort has gone into analyzing spatio-temporal behavior of particulate matter, below we provide a brief review. Cressie *et al.* (1999) compare kriging and Markov-random field models in the prediction of PM_{10} concentrations around the city of Pittsburgh. Sun *et al.* (2000) develop a spatial predictive distribution for the space-time response of daily ambient PM_{10} in Vancouver, Canada. Kibria *et al.* (2000) develop a multivariate spatial prediction methodology in a Bayesian context for the prediction of $PM_{2.5}$ in the city of Philadelphia. 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}. Zidek *et al.* (2002) develop predictive distributions on non-monitored PM_{10} concentrations in Vancouver. Smith *et al.* (2003) propose a spatio-temporal model for predicting weekly averages of $PM_{2.5}$ and other derived quantities such as annual averages within three southeastern states in the US. Sahu and Mardia (2005) present a short-term forecasting analysis of $PM_{2.5}$ data in New York City during 2002. Sahu *et al.* (2006) consider modeling of $PM_{2.5}$ by mixing two processes one for the rural background areas and the other for the urban areas. Cocchi *et al.* (2007) develop hierarchical Bayesian model for daily average PM_{10} concentration levels. Pollice and Lasinio (2010) develop a Bayesian kriging based method for estimating daily PM_{10} surfaces.

Ground-level ozone is a pollutant that is a significant health risk, especially for children with asthma and vulnerable adults with respiratory problems. It also damages crops, trees and other vegetation. It is a main ingredient of urban smog. Some early references on ozone modeling include Cox and Chu (1992), Brown *et al.* (1994), Guttorp *et al.* (1994) and Carroll *et al.* (1997), and Thompson *et al.* (2001). Porter *et al.* (2001) report on the estimation of trends in ozone concentrations adjusted for meteorological variables at individual monitoring sites. Zhu *et al.* (2003) relate ambient ozone and pediatric asthma emergency room visits in Atlanta using hierarchical regression methods for spatially mis-aligned data. Huerta *et al.* (2004) model hourly readings of concentrations of ozone jointly with air temperature for data from Mexico City. Cocchi *et al.* (2005) follow the approach of Huang and Smith (1999) by using a tree based partitioning of daily maxima ozone concentrations and assumed these maxima are Weibull distributed. McMillan *et al.* (2005) propose a regime switching model for ozone forecasting using meteorological variables as covariates and they illustrate using data from April to September in 1999 over a spatial domain covering Lake Michigan. Sahu *et al.* (2007) deal with misalignment between ozone data and meteorological information. Sahu *et al.* (2009b) develop a hierarchical space-time model for daily 8-hour maximum ozone concentration data covering much of the eastern United States. Dou *et al.* (2010) compares two Bayesian methods for modeling hourly ozone concentration levels. Berrocal *et al.* (2010&2011) propose various downscaling approaches by regressing the observed point level ozone concentration data on grid cell level computer model output with spatially varying regression co-efficients specified through a Gaussian process.

Several authors have developed generic models for analyzing spatio-temporal data. Research in this area dates back to Cressie (1994) and Goodall and Mardia (1994). More recent articles in this area include: Kyriakidis and Journal (1999), Stroud *et al.* (2001), Wikle and Cressie (1999), Wikle (2003), Gelfand *et al.* (2005), Cressie *et al.* (2010). A recent book, Cressie and Wikle (2011), provides a very comprehensive

review of both classical and Bayesian methods for analysing space-time data.

The format of the remainder of this Chapter is as follows. In Section 2 we develop the hierarchical autoregressive model based on our recent work. We also provide an introduction to Gaussian processes. Spatial prediction and forecasting methods, including their estimation in an iterative Markov chain Monte Carlo (MCMC) computation setup, are provided in Section 3. An illustration of the modeling methods is given in Section 4 using daily maximum 8-hour average ozone concentration levels observed in three mid-western states, Illinois, Indiana and Ohio in 2006. A few summary remarks are provided in Section 5. The Appendix outlines the full conditional distributions needed for setting up the MCMC.

2 Hierarchical Models

2.1 Models for Data

Point level air pollution data at location $\mathbf{s}$ and at time t is denoted by $Z(\mathbf{s}, t)$, after any transformation, if necessary. Air pollution data are often modeled on the square-root scale which encourages normality and stabilizes the variance, see e.g. Sahu *et al.* (2007), although the log-transformation is also used sometimes. Model fitting statistics, i.e. goodness-of-fit, model diagnostics, parameter estimates and their uncertainty measures are reported on the modeled scale. However, the validations and predictions are reported on the original scale for ease of communication to the practitioners and the end users.

We assume that $Z(\mathbf{s}, t)$ is univariate and the spatial reference vector $\mathbf{s}$ is two dimensional describing the latitude-longitude pair (or its equivalent Northing and Easting coordinates for example) and the time index t is discrete. We also assume that $Z(\mathbf{s}, t)$ is observed at n monitoring sites denoted by $\mathbf{s}_i$, $i = 1, \ldots, n$, say and at T time points so that $t = 1, \ldots, T$. The time unit is typically an hour or a day, although coarser units such as month or year are also used depending on the specific modeling objectives.

The first stage of the hierarchy assumes the measurement error model:

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

where $Y(\mathbf{s}_i, t)$ is the true underline spatio-temporal process and the error term $\epsilon(\mathbf{s}_i, t)$ is a white noise process, assumed to follow the $N(0, \sigma^2_\epsilon)$ distribution. In the spatial statistics literature σ^2_ϵ is often called the nugget effect which quantifies variation of the data points measured at locations that are very small distances apart. In principle, σ^2_ϵ could evolve in time but in many applications it is treated as a constant for the sake of parsimony. This first stage specification is advantageous in handling missing data in the Bayesian modeling setup with MCMC computation since, any missing data $Z(\mathbf{s}_i, t)$ is simply simulated from the $N(Y(\mathbf{s}_i, t), \sigma^2_\epsilon)$ distribution as implied by (1) at each MCMC iteration. The specification for $Y(\mathbf{s}_i, t)$ is provided in the next stage.

The space-time process $Y(\mathbf{s}_i, t)$ is assumed to have a systematic mean component, $\mu(\mathbf{s}_i, t)$ that may depend on past values and relevant covariates. A first order auto-regressive model given by $\rho Y(\mathbf{s}_i, t-1)$ can be used to model dependence on past values. This model will be appropriate when there is high auto-correlation present between the successive temporal values at any particular site. Additional auto-regressive terms can also be considered if the first order model is inadequate in modeling the temporal dependence. Those terms, however, may not remain significant when other model components such as the covariate effects are introduced. In this chapter we will only consider the first order auto-regressive process for the sake of parsimony.

The mean function, $\mu(\mathbf{s}_i, t)$, can be further enriched by a set of p, say, relevant spatially and temporally varying covariates $\mathbf{x}(\mathbf{s}_i, t) = (x_1(s_i, t), \ldots, x_p(s_i, t))'$. Note that some of these can only vary spatially and some others may only vary temporally. The covariate effect can be assumed to be spatially varying by assuming a spatially varying p-dimensional co-efficient process $\boldsymbol{\beta}(\mathbf{s})$. This model allows the possibility of

particular covariates making local adjustments to the mean function. A suitable prior process must be specified for $\boldsymbol{\beta}(\mathbf{s})$. Below we discuss a Gaussian process prior often used in practical problems. In our illustration in Section 4, however, we will use a fixed $\boldsymbol{\beta}$ for all sites $\mathbf{s}$.

The third and final component in $Y(\mathbf{s}_i, t)$ is assumed to be a residual random intercept, $w(\mathbf{s}_i, t)$, varying in both space and time. Having modeled the temporal dependence by an auto-regressive process, we can assume $w(\mathbf{s}_i, t)$ to be a temporally independent zero mean Gaussian process with a specified covariance function. This independence assumption simplifies the computation since covariance matrices of only order $n \times n$ need to be worked with instead of the full $nT \times nT$ matrices. Wikle and Cressie (1999) suggest an alternative specification for $w(\mathbf{s}_i, t)$ using orthonormal basis functions in space and random mean zero variables in time.

In summary, the second stage model specification is given by:

$$Y(\mathbf{s}_i, t) = \rho Y(\mathbf{s}_i, t-1) + \mathbf{x}(\mathbf{s}_i, t)'\boldsymbol{\beta} + w(\mathbf{s}_i, t). \tag{2}$$

Note that the auto-regressive component in time and the regression term compete against each other to provide alternative explanations of data. These together also compete against the explanation provided by the assumed spatial correlation structure. Because of this, a practical model fitting exercise can be thought to weigh-up these three alternative sources of information for choosing the best possible mixture of model components for explaining the data. Of-course, formal Bayesian model choice criteria can be adopted to compare specific models of interest, i.e. the model without the regressors and so on. For models based on a top-level Gaussian distribution there are several predictive Bayesian model choice criteria such as the predictive model choice criteria (PMCC), see e.g. Sahu *et al.* (2009b) for an illustration with the PMCC. In this chapter, however, we do not consider such model choice criteria any further and instead use the significance of parameter estimates to decide whether to include them in the model.

The auto-regressive model requires a specification for the initial condition $\mathbf{Y}_0' = (Y(\mathbf{s}_1, 0), \ldots, Y(\mathbf{s}_n, 0))$. There are two possible alternative specifications for $\mathbf{Y}_0$, e.g. (i) treat it as a fixed constant where $Y(\mathbf{s}_i, 0)$ is set at the overall mean of location $\mathbf{s}_i$, (ii) assign a prior distribution with mean $\boldsymbol{\mu}_0$ and covariance matrix Σ_0. In the latter case, the elements of $\boldsymbol{\mu}_0$ can be taken as the site-wise means, but there may be several possibilities for treating Σ_0. For example, it may be assumed to be a diagonal matrix with a large value 10^4, say for each diagonal entry corresponding to the assumption of a flat prior. Alternatively, elements of Σ_0 can be specified using a Gaussian covariance function discussed in the next sub-section. In our illustration, we treat $\mathbf{Y}_0$ as fixed for convenience and simplicity.

2.2 Gaussian Processes

Often Gaussian processes are assumed as components in spatial and spatio-temporal modeling. These stochastic processes are defined over a continuum, e.g. a spatial study region and specifying the resulting infinite dimensional random variable is often a challenge in practice. Gaussian processes are very convenient to work in these settings since they are fully defined by a mean function, say $\mu(\mathbf{s})$ and a valid covariance function, say $C(s, s^*) = \text{Cov}(w(s), w(s^*))$ which is required to be positive definite. A covariance function is said to be positive definite if the covariance matrix, implied by that covariance function, for a finite number of random variables belonging to that process is positive definite. Below we provide a family of valid positive definite covariance functions.

Gaussian processes are often preferred in spatial modeling because of the attractive distribution theory associated with them. All finite dimensional distributions of Gaussian processes are multivariate normals. Hence, joint distribution of data observed at any finite set of locations (or the associated random effects) is multivariate normal. Moreover, kriging or spatial prediction at yet unobserved locations conditionally on the observed data is facilitated by means of a conditional distribution which is also normal. This convenient distribution theory is very attractive for spatial prediction in the context of modern, fully model-based spatial

analysis within a Bayesian framework. The spatial predictive distributions are easy to compute and simulate from in an iterative MCMC framework.

We now turn to the specification of a valid covariance function. There is a substantial literature on this. Chapter 2 of Banerjee *et al.* (2004) provides a thorough discussion on this topic and the related concepts of stationarity, isotropy, and separability. Briefly, a process is defined to be *weakly stationary* if the covariance between a pair of random variables depends only on the separation vector between the two locations and not on the actual locations where those are observed. An *isotropic* covariance function only depends upon the distance between any two locations and not on the direction. Thus any pair of random variables observed at any two locations will have the same covariance as any other pair of random variables observed at any other locations separated by same distance. Separability is a concept used in modeling multivariate spatial data including spatio-temporal data. A *separable covariance function* in space and time is simply the product of two covariance functions one for space and the other for time.

The Matèrn family of covariance functions provides a very general choice and is given by:

$$C(u) = \sigma^2 \frac{1}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}u\phi)^{\nu} K_{\nu}(2\sqrt{\nu}u\phi), \;\; \phi > 0, \nu \geq 1, u > 0 \tag{3}$$

where $K_{\nu}(\cdot)$ is the modified Bessel function of second kind and of order ν, see e.g. Abramowitz and Stegun (1965, Chapter 9). Popular special cases of the Matèrn family are: (i) $\nu = 1/2$ corresponding to the exponential model $C(u) = \sigma^2 \exp(-\phi u)$ and (ii) $\nu = 3/2$ which leads to $C(u) = \sigma^2(1 + \phi u)\exp(-\phi u)$ and (iii) Gaussian, $C(u) = \sigma^2 \exp(-\phi^2 u^2)$ when $\nu \to \infty$.

The minimum value of u for which $C(u) \approx 0$ is defined as the *range* in spatial statistics literature. Note that, for the exponential covariance function, $C(u)$ can be exactly 0 only when u is very large, in other words ∞. To avoid this value of infinite range when our study region is a finite domain (in the sense that the maximum distance between any two locations is finite), we often calculate the range as that value of the distance u for which $C(u)$ is very small, i.e. 0.01 or 0.05. In this chapter, we shall illustrate with the exponential covariance function for which we define the range as $-\log(0.05)/\phi \approx 3/\phi$.

2.3 Joint Posterior Distribution

Define $\mathbf{Z}_t = (Z(\mathbf{s}_1, t), \ldots, Z(\mathbf{s}_n, t))'$, and $\mathbf{Y}_t = (Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_n, t))'$. Let X_t denote the $n \times p$ matrix having the ith row as $\mathbf{x}(\mathbf{s}_i, t)'$. It is convenient to write the joint posterior distribution using $\mathbf{Z}_t$, $\mathbf{Y}_t$ and X_t. To facilitate this, we now re-write the hierarchical model specifications using these vectors and matrices as follows. The first model equation is obtained from (1):

$$\mathbf{Z}_t = \mathbf{Y}_t + \boldsymbol{\epsilon}_t, \tag{4}$$

for $t = 1, \ldots, T$, where $\boldsymbol{\epsilon}_t = (\epsilon(\mathbf{s}_1, t), \cdots, \epsilon(\mathbf{s}_n, t))'$. From (2) we have:

$$\mathbf{Y}_t = \rho\, \mathbf{Y}_{t-1} + X_t\boldsymbol{\beta} + \mathbf{w}_t, \tag{5}$$

for $t = 1, \ldots, T$, where $\mathbf{w}_t = (w(\mathbf{s}_1, t), \ldots, w(\mathbf{s}_n, t))'$.

For the measurement error model in (4) we have that $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma^2_{\epsilon} I_n)$, $t = 1, \ldots, T$, independently, where $\mathbf{0}$ is the vector with all elements zero and I_n is the identity matrix of order n. For the spatially correlated error we assume that $\mathbf{w}_t$ follows the GP independently with the covariance function $\sigma^2_w \rho_w(\mathbf{s}_i - \mathbf{s}_j; \phi_w)$. We take $\rho_w(\mathbf{s}_i - \mathbf{s}_j; \phi_w) = \exp(-\phi_w d(\mathbf{s}_i, \mathbf{s}_j))$ where $d(\mathbf{s}_i, \mathbf{s}_j)$ is the distance between sites $\mathbf{s}_i$ and $\mathbf{s}_j$, $i, j = 1, \ldots, n$. This GP assumption implies that $\mathbf{w}_t \sim N(\mathbf{0}, \Sigma_w)$, $t = 1, \ldots, T$ where Σ_w has elements $\sigma_w(i, j) = \sigma^2_w \exp(-\phi_w d(\mathbf{s}_i, \mathbf{s}_j))$. For future use, we define S_w by the relation $\Sigma_w = \sigma^2_w S_w$.

Let $\boldsymbol{\vartheta}_t = \rho\, \mathbf{Y}_{t-1} + X_t\boldsymbol{\beta}$, for $t = 1, \ldots T$. Further, let $\boldsymbol{\theta}$ denote all the parameters, $\boldsymbol{\beta}$, ρ, $\sigma^2_{\epsilon}, \phi_w$, and σ^2_w. Let $\mathbf{v}$ denote all the augmented data, $\mathbf{Y}_t$ and the missing data, denoted by $z^*(\mathbf{s}_i, t)$, for $i = 1, \ldots, n$,

$t = 1, \ldots, T$, and $\mathbf{z}$ denote all the observed non-missing data $z(\mathbf{s}_i, t)$, for $i = 1, \ldots, n$, $t = 1, \ldots, T$. The log of the posterior distribution, denoted by $\log \pi(\boldsymbol{\theta}, \mathbf{v}|\mathbf{z})$, can be written as

$$
\begin{aligned}
&-\tfrac{nT}{2}\log(\sigma^2_\epsilon) - \tfrac{1}{2\sigma^2_\epsilon}\textstyle\sum_{t=1}^{T}(\mathbf{Z}_t - \mathbf{Y}_t)'(\mathbf{Z}_t - \mathbf{Y}_t) \\
&-\tfrac{nT}{2}\log(\sigma^2_w) - \tfrac{T}{2}|S_w| - \tfrac{1}{2\sigma^2_w}\textstyle\sum_{t=1}^{T}(\mathbf{Y}_t - \boldsymbol{\vartheta}_t)'S_w^{-1}(\mathbf{Y}_t - \boldsymbol{\vartheta}_t) \\
&+\log\left(\pi(\rho, \boldsymbol{\beta}, \sigma^2_\epsilon, \sigma^2_w, \phi_w)\right)
\end{aligned}
$$

where $\pi(\rho, \boldsymbol{\beta}, \sigma^2_\epsilon, \sigma^2_w, \phi_w)$ denotes the prior distribution, and $|S_w|$ denotes the determinant of S_w. We assume that *a-priori* $\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma^2_\beta I_p)$, and to have a flat prior distribution we take $\sigma^2_\beta = 10^4$. The auto-regressive coefficient ρ is also specified as the $N(0, 10^4)$ distribution, but restricted in the interval $I(0 < \rho < 1)$, so that this distribution is essentially flat. The inverse of the variance components, $1/\sigma^2_\epsilon$, $1/\sigma^2_w$, are assumed to follow $G(a, b)$ independently, where the distribution $G(a, b)$ has mean a/b. In our implementation we take $a = 2$ and $b = 1$ to have a proper prior specification for each of these variance components since improper prior distributions may lead to improper posterior distributions.

For the correlation decay parameter, ϕ_w we assume an independent uniform prior distribution in the interval $(0.001, 1)$. This corresponds to a value of spatial range between 3 and 3000 distance units (often taken as kilometers or miles). This prior distribution is appropriate for modeling air pollution data observed in a variety of study regions, e.g. a city where the maximum distance between any two locations is only a few kilometers or a substantial part of the eastern US where the maximum distance is approximately 3000 kilometers. Clearly, the endpoints of the prior interval can be changed to accommodate the spatial range taking any meaningful value in a particular practical problem.

3 Prediction Details

We first develop the methods for spatial interpolation of the air pollution levels at a new location $\mathbf{s}_0$ and any time t, $t = 1, \ldots, T$. Details for one step-ahead forecasting at time $t = T + 1$ are given at the forecasting subsection below. Spatial interpolation at location $\mathbf{s}_0$ and time t is based upon the predictive distribution of $Z(\mathbf{s}_0, t)$ given in the model equations (1) and (2). According to (1), $Z(\mathbf{s}_0, t)$, has the distribution:

$$
Z(\mathbf{s}_0, t) \sim N\left(Y(\mathbf{s}_0, t),\ \sigma^2_\epsilon\right) \tag{6}
$$

and

$$
Y(\mathbf{s}_0, t) = \rho\, Y(\mathbf{s}_0, t-1) + x(\mathbf{s}_0, t)'\boldsymbol{\beta} + w(\mathbf{s}_0, t).
$$

It is easy to see that $Y(\mathbf{s}_0, t)$ can only be sequentially determined using all the previous $Y(\mathbf{s}_0, t)$, including $Y(\mathbf{s}_0, 0)$, up to time t. Hence, we introduce the notation $\mathbf{Y}(\mathbf{s}, [t])$ to denote the vector $(Y(\mathbf{s}, 1), \ldots, Y(\mathbf{s}, t))'$ for $t \geq 1$. Note that a value of $Y(\mathbf{s}_0, 0)$ is required for this prediction problem. This value should be taken according to the prior distribution assumed for the initial condition on $\mathbf{Y}_0$. If, however, $\mathbf{Y}_0$ has been taken to be a fixed constant, then $Y(\mathbf{s}_0, 0)$ can also be taken as that same constant, as has been done in our illustration here.

The posterior predictive distribution of $Z(\mathbf{s}_0, t)$ is obtained by integrating over the unknown quantities in (6) with respect to the joint posterior distribution, i.e.,

$$
\pi\left(Z(\mathbf{s}_0, t)|\mathbf{z}\right) = \int \pi\left(Z(\mathbf{s}_0, t)|Y(\mathbf{s}_0, [t]), \sigma^2_\epsilon\right)\, \pi\left(Y(\mathbf{s}_0, [t])|\boldsymbol{\theta}, \mathbf{v}\right) \pi(\boldsymbol{\theta}, \mathbf{v}|\mathbf{z})\, dY(\mathbf{s}_0, [t])\, d\boldsymbol{\theta}\, d\mathbf{v}. \tag{7}
$$

When using MCMC methods to draw samples from the posterior, the predictive distribution (7) is sampled by composition. Draws from the posterior distribution $\pi(\boldsymbol{\theta}, \mathbf{v}|\mathbf{z})$ facilitates evaluation of the above integral, details are provided below.

We draw $Y(\mathbf{s}_0, t)$ from its conditional distribution given $\boldsymbol{\theta}, \mathbf{v}$ and $Y(\mathbf{s}_0, [t-1])$. Analogous to (5), we obtain for $t \geq 0$

$$\begin{pmatrix} Y(\mathbf{s}_0, t) \\ \mathbf{Y}_t \end{pmatrix} \sim N\left[\begin{pmatrix} \rho\, Y(\mathbf{s}_0, t-1) + x(\mathbf{s}_0, t)'\boldsymbol{\beta} \\ \rho\, \mathbf{Y}_{t-1} + X_t\boldsymbol{\beta} \end{pmatrix},\ \sigma_w^2 \begin{pmatrix} 1 & S_{w,12} \\ S_{w,21} & S_w \end{pmatrix} \right]$$

where $S_{w,12}$ is $1 \times n$ with the ith entry given by $\exp(-\phi_w d(\mathbf{s}_i, \mathbf{s}_0))$ and $S_{w,21} = S'_{w,12}$. Hence,

$$Y(\mathbf{s}_0, t)|\mathbf{Y}_t, \boldsymbol{\theta}, \mathbf{v} \sim N(\chi,\ \Lambda) \tag{8}$$

where $\Lambda = \sigma_w^2 \left(1 - S_{w,12} S_w^{-1} S_{w,21}\right)$ and

$$\chi = \rho\, Y(\mathbf{s}_0, t-1) + x(\mathbf{s}_0, t)'\boldsymbol{\beta} + S_{w,12} S_w^{-1} \left(\mathbf{Y}_t - \rho\, \mathbf{Y}_{t-1} - X_t\boldsymbol{\beta}\right).$$

In summary, we implement the following algorithm to predict $Z(\mathbf{s}_0, t), t = 1, \ldots, T$.

1. Draw a sample $\boldsymbol{\theta}^{(j)}, \mathbf{v}^{(j)}, j \geq 1$ from the posterior distribution.
2. Draw $\mathbf{Y}^{(j)}(\mathbf{s}_0, [t])$ sequentially using (8). Note that the initial value $Y^{(j)}(\mathbf{s}_0, 0)$ is a constant for all $\mathbf{s}_0$ in our implementation.
3. Finally draw $Z^{(j)}(\mathbf{s}_0, t)$ from $N\left(Y^{(j)}(\mathbf{s}_0, t),\ \sigma_\epsilon^{2(j)}\right)$.

The air pollution concentration on the original scale is the square of $Z^{(j)}(\mathbf{s}_0, t)$. If we want the predictions of the smooth pollution process without the nugget term we simply omit the last step in the above algorithm and square the realizations $\mathbf{Y}^{(j)}(\mathbf{s}_0, t)$. We use the median of the MCMC samples and the lengths of the 95% intervals to summarize the predictions. The median as a summary measure preserves the one-to-one relationships between summaries for Y and Z, and for Y^2 and Z^2.

3.1 Calculating Summaries

We now develop methodology for obtaining temporal summaries of air pollution. We illustrate by detailing methodologies for calculating the annual $4th$ highest ozone concentration at any site $\mathbf{s}_0$.

The true annual $4th$ highest daily maximum 8-hour average ozone concentration, denoted by $f(\mathbf{s}_0)$, is given by the $4th$ highest value of the series $Y^2(\mathbf{s}_0, 1), \ldots, Y^2(\mathbf{s}_0, T)$. (Note that we model ozone on the square-root scale.) At each MCMC iteration, j, we calculate $f^{(j)}(\mathbf{s}_0)$ and then the summaries of these posterior predictive realizations $f^{(j)}(\mathbf{s}_0)$ are used for predictions of the annual $4th$ highest daily maximum 8-hour average ozone concentration (and to obtain their uncertainties).

3.2 Forecasting

The one-step ahead Bayesian forecast at a location $\mathbf{s}_0$ is given by the posterior predictive distribution of $Z(\mathbf{s}_0, T+1)$ which is determined by $Y(\mathbf{s}_0, T+1)$. Note that using (8), we already have the conditional distribution of $Y(\mathbf{s}_0, T)$ given $\mathbf{Y}_t, \boldsymbol{\theta}$ and $\mathbf{v}$. We use model equation (2) to advance this conditional distribution one unit of time in future. The mean of the one step-ahead forecast distribution is given by $\rho\, Y(\mathbf{s}_0, T) + x(\mathbf{s}_0, T+1)'\boldsymbol{\beta}$, according to (2), and $Y(\mathbf{s}_0, T+1)$ should be equal to this if we are interested in forecasting the mean. If, however, we want to forecast an observation at location $\mathbf{s}_0$ we simulate $Y(\mathbf{s}_0, T+1)$ from the marginal distribution which has the above mean and variance σ_w^2. We work with this marginal distribution rather than the conditional distribution like (8) above since conditioning with respect to the observed information (i.e. kriging) upto time T at the observation locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ has already been done to obtain $Y(\mathbf{s}_0, T)$, and at the future time $T+1$ there is no new available information to condition on except for the new values of the regressor, $x(\mathbf{s}_0, T+1)$. Then we follow the above algorithm to obtain the forecasts and their uncertainty estimates using ergodic averages of MCMC output.

4 An Example

We illustrate with the daily maximum 8-hour average ozone concentration levels in 2006 observed in 117 monitoring sites in the three mid-western states, Illinois, Indiana and Ohio. This study region, see Figure 1, provides a good mix of developed industrial areas in Ohio and large cities like Chicago separated by vast rural areas. We use data from 12 randomly selected sites for validation purposes. The data from remaining 105 sites are used for model fitting.

Our analysis uses daily data for the $T = 153$ days in the high ozone season between May to September. This is a moderately large data set rich in both space and time with 16,065 observations ($= 105 \times 153$); 291 ($= 1.8\%$) of these are missing. The mean value is 47.62 parts per billion (ppb) and the range is 6.75 to 131.38ppb. A site-wise box-plot, see Figure 3, shows much spatial variation in the average levels between the sites. However, the variability within the sites is seen to be roughly constant which can be explained by the fact that the daily observations are all based on 8 hourly averages. A time series plot of the data for two randomly selected sites, provided in Figure 2, shows high ozone values during the three hottest months of June, july and August. The plot also shows the presence of moderate temporal dependence.

Following Sahu *et al.* (2009b), we include the output of a computer simulation model known as the CMAQ (Community Multiscale Air Quality), see *http://www.cmaq-model.org/*, as the single co-variate in the model. The CMAQ model is based on emission inventories, meteorological information, and land use, and it produces average ozone concentration levels at each cell of a 12 square-kilometer grid covering the whole of the continental US retrospectively, although there is a version of the model known as Eta-CMAQ which produces forecasts upto two days in advance. In this chapter we use the retrospective daily maximum eight-hour average CMAQ ozone concentration for the grid cell covering the monitoring site. We provide a scatter plot of the ozone concentration values and the corresponding CMAQ values in Figure 4. The plot shows a strong linear relationship between the two, but clearly CMAQ values are upwardly biased. This points to the need for a more accurate empirical model as is done here. Note that for this plot and also for modeling we have adopted the square-root scale. The spatial predictions at the unmonitored sites are performed using the CMAQ output at the corresponding grid cells. We have also attempted to include other meteorological co-variates such as the daily maximum temperature, but none of those turned out to be significant in the presence of the CMAQ output.

In addition to the CMAQ output we include an overall intercept β_0 in the model. Thus the mean of the true process $Y(\mathbf{s}, t)$ is given by $\rho Y(\mathbf{s}, t-1) + \beta_0 + \beta x(\mathbf{s}, t)$ where $x(\mathbf{s}, t)$ denotes the CMAQ output at the grid-cell that includes the location $\mathbf{s}$. The model also contains the two variance components σ^2_ϵ and σ^2_w, and the spatial decay parameter, ϕ_w. Of-course, all the $Y(\mathbf{s}_i, t)$ and the missing $Z(\mathbf{s}_i, t)$ are also need to be simultaneously estimated. We implement the Gibbs sampler with a Metropolis step for ϕ_w to simulate these parameters from their conditional distributions provided in the Appendix.

We tune the variance of the proposal distribution in the Metropolis step for the decay parameter ϕ_w to have a reasonable acceptance rate in the range $(0.15, 0.40)$. The tuning parameter finally adopted gave us an acceptance rate of 27.35% from 25,000 iterations. As is usual in MCMC computation we have run the chains with many different starting values and monitored convergence by plotting traces of the parameters $\rho, \beta_0, \beta, \sigma^2_\epsilon, \sigma^2_w$ and ϕ_w. We have also examined the auto-correlation plots of these parameters and found those to be reasonable, i.e. the auto-correlations die down for moderate values of the lag parameter. There is, however, high cross-correlation between the parameters σ^2_w and ϕ_w and the MCMC chain mixes somewhat slowly because of this. This mixing problem occurs due to weak identifiability of the parameters and has been noted by many authors, see e.g. Zhang (2004) and Stein (1999). The problem disappears if ϕ_w is not estimated and kept at a fixed value chosen by out of sample validations, see e.g. Sahu *et al.* (2007). Here, we decide to sample ϕ_w for making inference using a large number of MCMC iterations, 20,000 after discarding first 5,000.

We use out of sample data from the 12 sites to validate the model. Out of the 1836 (12×153) validation

ozone values 32 are missing in our data. Figure 5 provides a plot of the 1804 out of sample predictions against the corresponding observations. The 95% prediction intervals are super-imposed along with the $y = x$ line. The figure shows a very slight over and under prediction at the two ends of the ozone scale, otherwise, there is a good agreement between the observations and predictions. The nominal coverage of the 95% prediction intervals is 96.2% which confirms the adequacy of the model. The validation mean square errors (VMSE) calculated for the 12 sites are between 8.91 to 114.68; the overall VMSE calculated using all the 1804 observations and their predictions is 38.02. These compare very favorably against CMAQ since the overall VMSE for CMAQ output is 144.18 and the site-wise CMAQ VMSEs range between 58.39 and 451.88. The overall VMSE value, 38.02, for the model based predictions is also smaller than the same, 48.5, for a down-scaler model recently proposed by Berrocal *et al.* (2011) for a similar data set.

The point and interval estimates of the model parameters are given in Table 1. We found moderate temporal dependence among successive day ozone concentrations (estimate of $\rho = 0.2687$). There is also strong spatial correlation since the point estimate of ϕ_w is 0.0027 implying an approximate range of 1109 kilometers. In addition to these strong spatial and temporal dependencies, the ozone concentrations also entertain the CMAQ output as a significant predictor since the point estimate is 0.4976 and the 95% credible interval does not include zero. A direct interpretation of this estimate is difficult due to the square-root transformation used in modeling. Note that both auto-regressive and the regression terms are significant predictors and hence their inclusion will provide better model fitting and prediction and hence are retained in the model. Finally, the estimates of the variance components σ^2_ϵ and σ^2_w show that more variation is explained by the spatio-temporal effects than the pure error process $\epsilon(\mathbf{s}, t)$.

We now plot the annual $4th$ highest daily maximum 8-hour average true ozone values by linearly interpolating the predictions at the centers of the 900 randomly selected CMAQ grid cells in the study region (see Figure 6). We find very good agreement among the predicted and observed maximum values. In fact, to quantify this with set aside data from the 12 validation sites we provide the observed, the model predicted and the CMAQ output for the annual $4th$ highest daily maximum 8-hour average ozone concentration values in Table 2. The mean square error for the model based predictions for these 12 validation sites is 10.4 while the same for the CMAQ output is 17.3. Thus, the model provides even more accurate predictions than the very accurate CMAQ output and the model is predicting the annual $4th$ highest value within a range of 3.2 ($= \sqrt{10.4}$) ppb on average. Figure 7 shows the uncertainties in the model predictions by providing a map of the lengths of the 95% prediction intervals. As expected, these intervals are larger in non-monitored areas compared with monitored areas. No such uncertainty map is possible for the deterministic CMAQ output.

5 Further discussion

The modeling methods discussed in this chapter are suited for monitoring compliance with respect to air regulatory standards. High resolution spatial and fine scale temporal modeling allows inference on aggregated spatial (e.g. regional) and temporal summaries (e.g. annual). The Bayesian computation methods also enable accurate assessment of uncertainties in the aggregated summaries.

There are several other important areas of research in air pollution modeling. A number of papers are devoted to assessing exposure to air pollution and to fuse monitoring data with computer model output, see e.g. Gelfand and Sahu (2010) for a recent review. Important modeling developments are also taking place in analyzing other pollutants such as sulfate and nitrate oxides. Deposition of these through precipitation is also of very much interest to researchers, see e.g. Sahu *et al.* (2010) and the references therein. Forecasting of air pollution both for short-term and long-term periods also provide challenging statistical problems to the modeling community. Sahu *et al.* (2009a&b) develop methods for instantaneous forecasting of hourly and daily ozone levels.

Acknowledgment

The authors thank David Holland of the USEPA for many helpful comments and also for providing the monitoring data and CMAQ model output used in this chapter.

Appendix: Conditional Distributions for Gibbs sampling

1. **Sampling Missing Data** Any missing value, $Z(\mathbf{s},t)$ is to be sampled from $N(Y(\mathbf{s},t),\sigma^2_\epsilon)$, $t=1,\ldots,T$.

2. **Sampling σ^2_ϵ and σ^2_w.** Straightforward calculation yields the following complete conditional distributions:
$$\frac{1}{\sigma^2_\epsilon} \sim G\left(\frac{nT}{2}+a, b+\frac{1}{2}\sum_{t=1}^{T}(\mathbf{Z}_t-\mathbf{Y}_t)'(\mathbf{Z}_t-\mathbf{Y}_t)\right),$$
$$\frac{1}{\sigma^2_w} \sim G\left(\frac{nT}{2}+a, b+\frac{1}{2}\sum_{t=1}^{T}(\mathbf{Y}_t-\boldsymbol{\vartheta}_t)'S_w^{-1}(\mathbf{Y}_t-\boldsymbol{\vartheta}_t)\right).$$

3. **Sampling $\mathbf{Y}_t$.** Let $Q_w=\Sigma_w^{-1}$. The full conditional distribution of $\mathbf{Y}_t$ is $N(\Lambda_t\boldsymbol{\chi}_t,\,\Lambda_t)$ where

   Case 1: For $1\le t<T-1$:
$$\Lambda_t^{-1}=\frac{I_n}{\sigma^2_\epsilon}+(1+\rho^2)Q_w,$$
$$\boldsymbol{\chi}_t=\frac{\mathbf{Z}_t}{\sigma^2_\epsilon}+Q_w\left\{\rho\mathbf{Y}_{t-1}+X_t\boldsymbol{\beta}+\rho\left(\mathbf{Y}_{t+1}-X_{t+1}\boldsymbol{\beta}\right)\right\}.$$

   Case 2: For $t=T$
$$\Lambda_t^{-1}=\frac{I_n}{\sigma^2_\epsilon}+Q_w,$$
$$\boldsymbol{\chi}_t=\frac{\mathbf{Z}_t}{\sigma^2_\epsilon}+Q_w\left(\rho\mathbf{Y}_{t-1}+X_t\boldsymbol{\beta}\right).$$

4. **Sampling ρ.** The full conditional distribution of ρ is $N(\Lambda\chi,\ \Lambda)$ where
$$\Lambda^{-1}=\sum_{t=1}^{T}\mathbf{Y}'_{t-1}Q_w\mathbf{Y}_{t-1}+10^{-4},\ \chi=\sum_{t=1}^{T}\mathbf{Y}'_{t-1}Q_w(\mathbf{Y}_t-X_t\boldsymbol{\beta}),$$
   restricted in the interval (0, 1).

5. **Sampling $\boldsymbol{\beta}$**. The full conditional distribution of $\boldsymbol{\beta}$ is $N(\Lambda\boldsymbol{\chi},\ \Lambda)$ where
$$\Lambda^{-1}=\sum_{t=1}^{T}X'_tQ_wX_t+\Sigma_\beta^{-1},\ \text{and}$$
$$\boldsymbol{\chi}=\sum_{t=1}^{T}X'_tQ_w(\mathbf{Y}_t-\rho\mathbf{Y}_{t-1}).$$

6. **Sampling** ϕ_w. The full conditional distribution of ϕ_w is non-standard and must be calculated from the prior and likelihood terms involving ϕ_w and is given by:

$$\log(\pi(\phi_w|\cdots) = -\frac{1}{2}|S_w| - \frac{1}{2\sigma_w^2}\sum_{t=1}^{T}(\mathbf{Y}_t - \boldsymbol{\vartheta}_t)'S_w^{-1}(\mathbf{Y}_t - \boldsymbol{\vartheta}_t) + \log(\pi(\phi_w))$$

upto a normalizing constant, where $\cdots$ denotes all the data and parameters except for ϕ_w. We adopt a Metropolis-Hastings step to obtain sample from this full conditional distribution as follows. Let $\phi_w^{(p)}$ denote a sample from a proposal distribution $q\left(\phi_w^{(p)}|\phi_w^{(c)}\right)$ where the current value is $\phi_w^{(c)}$. The sampled value, $\phi_w^{(p)}$ is accepted with probability

$$\alpha\left(\phi_w^{(p)},\, \phi_w^{(c)}\right) = \min\left\{1, \frac{\pi(\phi_w^{(p)}|\cdots)\, q\left(\phi_w^{(c)}|\phi_w^{(p)}\right)}{\pi(\phi_w^{(c)}|\cdots)\, q\left(\phi_w^{(p)}|\phi_w^{(c)}\right)}\right\}.$$

This acceptance probability simplifies considerably when $q\left(\phi_w^{(p)}|\phi_w^{(c)}\right)$ is taken to be symmetric in its arguments $\phi_w^{(p)}$ and $\phi_w^{(c)}$, i.e. when $q\left(\phi_w^{(p)}|\phi_w^{(c)}\right) = q\left(\phi_w^{(c)}|\phi_w^{(p)}\right)$. In this case, the ratio of densities in the acceptance probability is simply calculated by the ratio $\pi(\phi_w^{(p)}|\cdots)/\pi(\phi_w^{(c)}|\cdots)$. The resulting algorithm is known as the Metropolis algorithm.

We implement the Metropolis algorithm by taking the proposal distribution as the normal distribution with the mean at the current value and the variance σ_p^2 which we tune to have an acceptance rate between 15-40%, see Gelman *et al.* (1996) for theoretical justifications. Moreover, we implement the Metropolis algorithm on the log-scale for ϕ_w, i.e. we work with the density of $\log(\phi_w)$ instead of ϕ_w since the support of the normal proposal distribution is the real line. Keeping ϕ_w within a range is trivial since any proposal value outside the range is rejected forthwith.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions.* New York:Dover.

Banerjee, S., Carlin, B.P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall/CRC.

Berrocal, V. J. Gelfand, A. E. and Holland, D. M. (2010). A Spatio-Temporal Downscaler for Output From Numerical Models. *Journal of Agricultural, Biological and Environmental Statistics,* **15**, 176-197.

Berrocal, V. J. Gelfand, A. E. and Holland, D. M. (2011). Space-time data fusion under error in computer model output: an application to modeling air quality. *Technical Report.* Duke University.

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.

Cocchi, D., Fabrizi, E., and Trivisano, C. (2005). A stratified model for the assessment of meteorologically adjusted trends of surface ozone. *Environmental and Ecological Statistics* **12**, 1195-1208.

Cocchi, D., Greco, F. and Trivisano, C. (2007). Hierarchical space-time modelling of PM10 pollution. *Atmospheric Environment*, **41**, 532-542.

Cox, W. M. and Chu, S. H. (1992). Meteorologically adjusted trends in urban areas, a probabilistic approach. *Atmospheric Environment*, **27**, 425-434.

Cressie, N. (1994) Comment on "An approach to statistical spatial-temporal modeling of meteorological fields" by M. S. Handcock and J. R. Wallis. *Journal of the American Statistical Association*, **89**, 379–382.

Cressie, N, Shi, T., and Kang, E. L. (2010) Fixed Rank Filtering for Spatio-Temporal Data. *Journal of Computational and Graphical Statistics*, **19**, 724–745.

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.

Cressie, N. and Wikle, C. K. (2011) *Statistics for Spatio-Temporal Data.* John Wiley & Sons.

Dou, Y., Le, N. D. and Zidek, J. V. (2010). Modeling Hourly Ozone Concentration Fields. *Annals of Applied Statistics*, **4**, 1183-1213.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient Metropolis Jumping Rules. In *Bayesian Statistics 5*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith. Oxford University Press, pp 599-607.

Gelfand, A. E., Banerjee, S. and Gamerman, D. (2005). Spatial process modeling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**, 465-479.

Gelfand, A. E. and Sahu, S. K. (2010) Combining Monitoring Data and Computer model Output in Assessing Environmental Exposure. In *Handbook of Applied Bayesian Analysis*, OHagan, A. and West, M. (eds) Oxford: Oxford University Press, pp 482-510.

Goodall, C. and Mardia, K. V. (1994) Challenges in multivariate spatio-temporal modeling. In *Proceedings of the XVIIth International Biometric Conference*, Hamilton, Ontario, Canada, 8–12 August 1994, pp 1–17.

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

Kibria, B. M. G., Sun, L., Zidek, J. V., and Le, N. D. (2002). Bayesian spatial prediction of random space-time fields with application to mapping PM2.5 exposure. *Journal of the American Statistical Association*, **97**, 112-124.

Kyriakidis, P. C. and Journal, A. G. (1999) Geostatistical space-time models: A review. *Mathematical Geology*, **31**, 651–684.

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., Goodall C., Redfern E.J., and Alonso F.J. (1998) The Kriged Kalman filter (with discussion). *Test*, **7**, 217–252.

McMillan, N., Bortnick, S. M., Irwin, M. E. and Berliner, M. (2005). A hierarchical Bayesian model to estimate and forecast ozone through space and time. *Atmospheric Environment*, **39**, 1373–1382.

Pollice, A. and Lasinio, G. J. (2010). Spatiotemporal analysis of PM10 concentration over the Taranto area. *Environmental Monitoring and Assessment*, **162**, 177-190.

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.

Sahu, S. K., Gelfand, A. E. and Holland, D. M. (2006). Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 61-86.

Sahu, S. K., Gelfand, A. E. and Holland, D. M. (2007). High Resolution Space-Time Ozone Modeling for Assessing Trends. *Journal of the American Statistical Association*, **102**, 1221-1234.

Sahu, S. K., Gelfand, A. E. and Holland, D. M. (2010). Fusing point and areal space-time data with application to wet deposition. *Journal of the Royal Statistical Society*, Series C, **59**, 77-103.

Sahu, S. K., Yip, S. and Holland, D. M. (2009a). A fast Bayesian method for updating and forecasting hourly ozone levels. *Environmental and Ecological Statistics*, doi:10.1007/s10651-009-0127-y.

Sahu, S. K., Yip, S. and Holland, D. M. (2009b). Improved space-time forecasting of next day ozone concentrations in the eastern US. *Atmospheric Environment*, doi:10.1016/j.atmosenv.2008.10.028.

Stein, M. 1999. *Interpolation of Spatial Data: Some Theory for Kriging:* Springer Verlag.

Stroud, J. R., Müller, P. and Sansó, B. (2001) Dynamic models for Spatio-temporal data. *Journal of the Royal Statistical Society*, B, **63**, 673–689.

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*, 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.

Wikle, C. K. (2003). Hierarchical models in environmental science. *International Statistical Review*, **71**, 181-199.

Wikle, C. K. and Cressie, N. (1999) A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, **86**, 815-829.

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 1: Estimation of the parameters. CI stands for equal tailed credible intervals.

	mean	sd	95%CI
ρ	0.2687	0.0108	(0.2469, 0.2890)
β_0	1.4152	0.0667	(1.2885, 1.5485)
β	0.4976	0.0081	(0.4820, 0.5136)
σ^2_ϵ	0.2165	0.0042	(0.2085, 0.2248)
σ^2_w	0.4246	0.0229	(0.3848, 0.4738)
ϕ_w	0.0027	0.0002	(0.0024, 0.0031)

Table 2: Annual $4th$ highest daily maximum 8-hour average ozone concentrations in ppb units.

Validation Site	Observed	Predicted	CMAQ
1	71.13	64.38	67.86
2	60.25	62.08	70.59
3	72.75	70.03	72.16
4	76.63	76.64	70.98
5	70.75	71.02	70.55
6	75.50	79.36	78.59
7	81.00	78.68	83.96
8	72.25	74.90	76.43
9	70.80	75.61	72.51
10	70.25	73.81	72.78

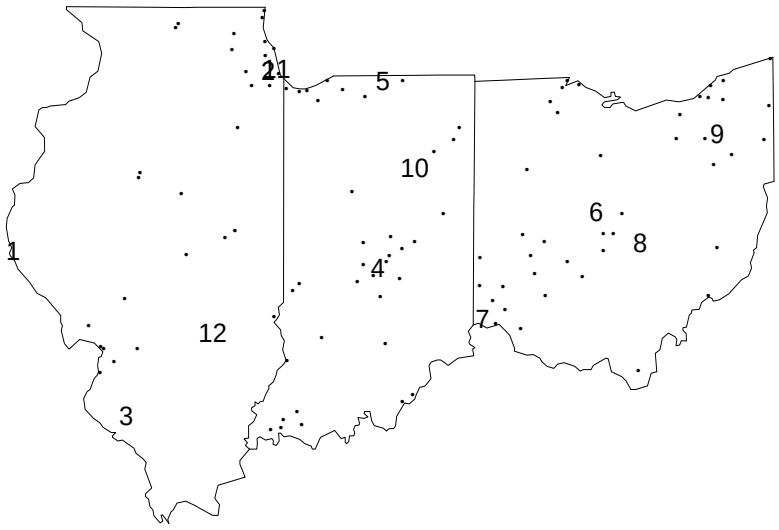

Figure 1: A map of Illinois, Indiana and Ohio with the 105 ozone monitoring sites plotted as points. The validation sites are labeled $1, \ldots, 12$.

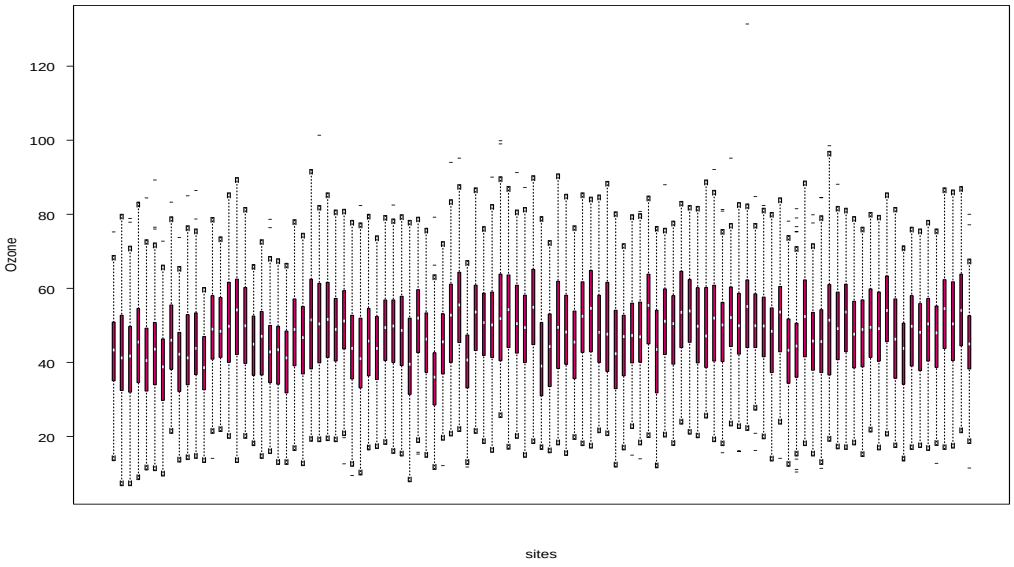

Figure 2: Time series plot of daily ozone values from two randomly selected sites.

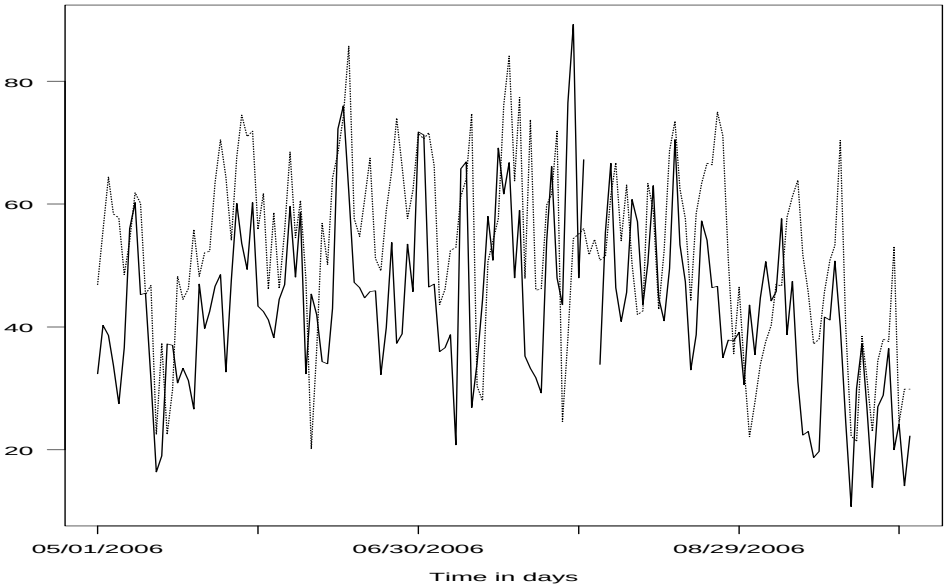

Figure 3: Boxplots of the 153 daily maximum 8-hour ozone concentration levels in 2006 for each of the 105 sites located in Illinois, Indiana.

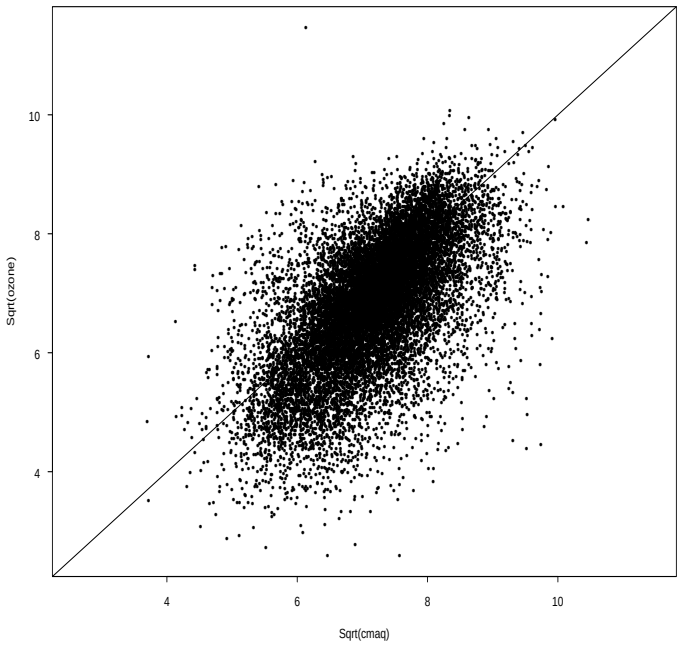

Figure 4: A scatter plot of daily maximum 8-hour ozone concentration levels for the 105 sites in 2006 against the corresponding CMAQ values on the square-root scale.

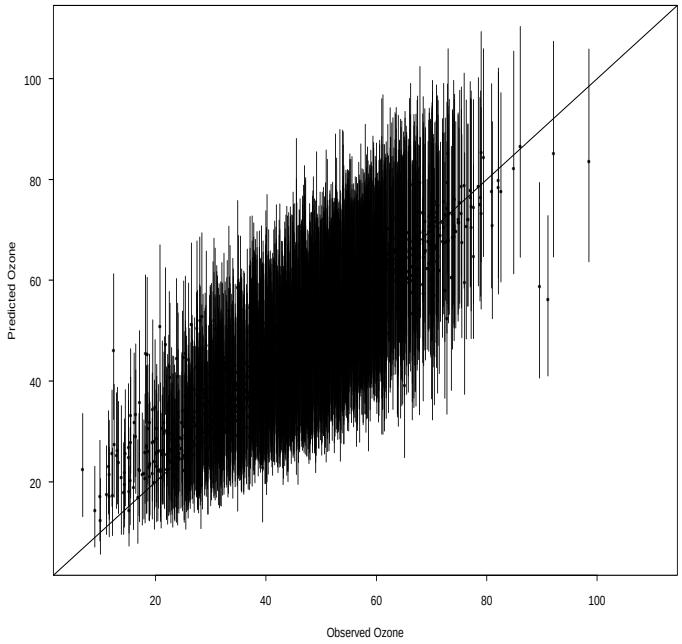

Figure 5: A plot of the validation predictions against the observations along with 95% prediction intervals. The $y = x$ line is super-imp

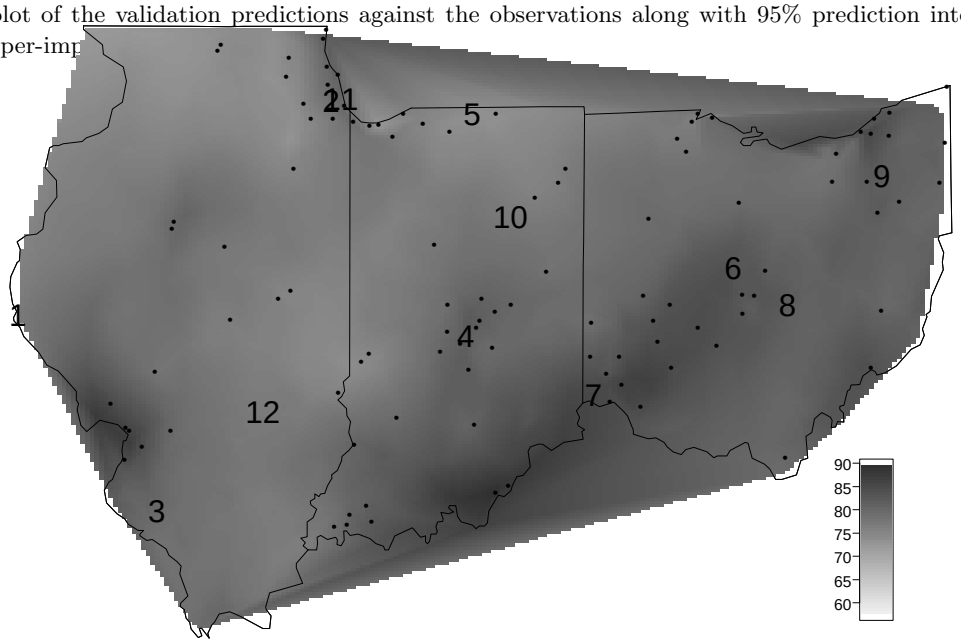

Figure 6: Model based predictions of the true annual $4th$ highest daily maximum 8-hour average ozone levels in 2006. The 105 fitting sites are plotted as points and the validation sites are labeled $1, \ldots, 12$ as in Figure 1.

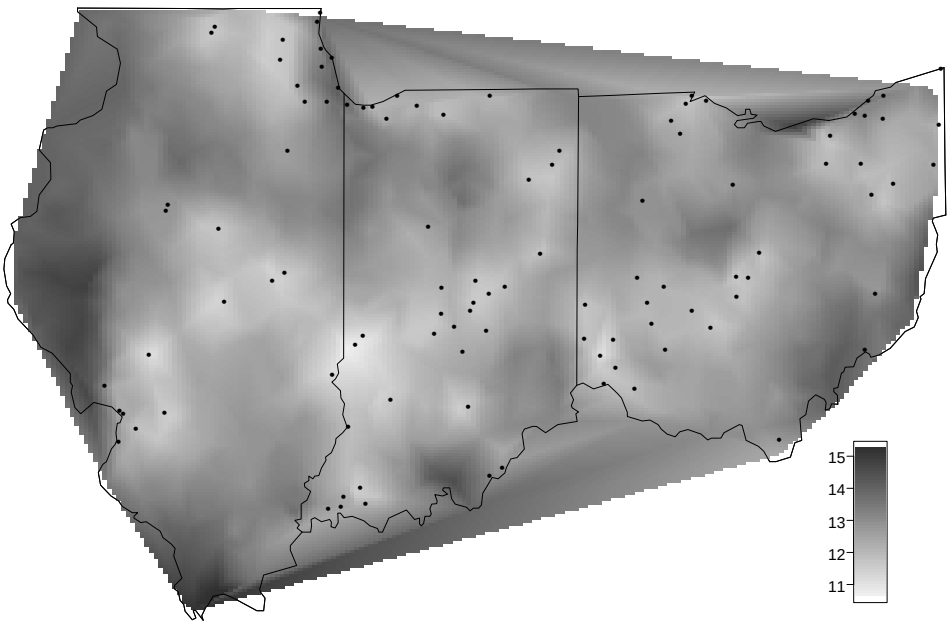

Figure 7: Lengths of 95% prediction intervals for the true annual $4th$ highest daily maximum 8-hour ozone ozone levels in 2006.