Conditionally Specified Space-Time Models for Multivariate Processes

Michael J. DANIELS, Zhigang ZHOU, and Hui ZOU

This article proposes a class of conditionally specified models for the analysis of multivariate space-time processes. Such models are useful in situations where there is sparse spatial coverage of one of the processes and much more dense coverage of the other process(es). The dependence structure across processes and over space, and time is completely specified through a neighborhood structure. These models are applicable to *both* point and block sources; for example, multiple pollutant monitors (point sources) or several county-level exposures (block sources). We introduce several computational tricks that are integral for model fitting, give some simple sufficient and necessary conditions for the space-time covariance matrix to be positive definite, and implement a Gibbs sampler, using Hybrid MC steps, to sample from the posterior distribution of the parameters. Model fit is assessed via the DIC. Predictive accuracy, over both time and space, is assessed both relatively and absolutely via mean squared prediction error and coverage probabilities. As an illustration of these models, we fit them to particulate matter and ozone data collected in the Los Angeles, CA, area in 1995 over a three-month period. In these data, the spatial coverage of particulate matter was sparse relative to that of ozone.

Key Words: Bayesian inference; CAR models; Markov random field model.

1. INTRODUCTION

Conditionally specified, or Markov random field, models can often be easily formulated for processes over space and time, but computations are intractable. We propose a particular class of conditionally specified space-time models for a multivariate process here that will be computationally tractable and that allows a neighborhood structure to be explicitly specified over both space and time. These models will also accommodate nonstationary spatial and temporal structures and are applicable to both point and block source data, whether as a

Michael J. Daniels is an Associate Professor, Department of Statistics, University of Florida, Gainesville, FL 32611 (E-mail: mdaniels@stat.ufl.edu). Zhigang Zhou is a Ph.D. Student, Department of Statistics, Iowa State University, Ames, IA 50011 (E-mail: zgzhou@iastate.edu). Hui Zou is an Assistant Professor, Department of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: hzou@stat.umn.edu).

model for the data itself or as a model for random effects embedded within a regression model for the data.

Some recent work on conditional specification of models for spatial data include Cressie and Chan (1989) and Cressie et al. (1999) who used "distance"-based neighbors for block and point sources, respectively. Bayesian conditional autoregressive (CAR) models using distance-based neighbors have been applied in Conlon and Waller (1998) and Best et al. (1999). Kaiser, Daniels, Furakawa, and Dixon (2002) specified anisotropic distance-based neighborhood structures which vary over time. Conditions under which conditionally specified models induce a valid joint distribution were originally given by Besag (1974) and a weaker set of conditions was given by Kaiser and Cressie (2000). Hrafnkelsson and Cressie (2003) compared geostatistics and conditionally specified models for predicting at many locations using a fine grid; they concluded that the conditional approach is a viable alternative for spatial environmental data. Recent Bayesian work on specifying priors for particular conditionally specified models includes a series of papers by Sun, Tsutakawa, and colleagues (Sun, Tsutakawa, and Speckman 1999; Kim, Sun, and Tsutakawa 2001); these articles and many articles using CAR models focus on placing these conditional priors on random effects in the setting of block (areal) sources.

There has also been work on conditionally specified models for *multivariate* spatial processes. Mardia (1988) generalized CAR models to the multivariate setting, referring to them as MCAR models. These models are somewhat restrictive as the same spatial dependence parameters are assumed for each of the multivariate processes. Gelfand and Vounatsou (2003) generalized this work to allow a separate spatial dependence parameter for each multivariate process. Carlin and Banerjee (2003) proposed separable models with a CAR model for spatial dependence and direct specification of the multivariate dependence. Knorr-Held and Rue (2002) developed intrinsic CAR models for bivariate processes.

In addition, there has also been considerable work on space-time modeling in recent years, in some cases with multivariate processes. Carroll et al. (1997) specified a space-time covariance function for a Gaussian random field on ozone (for a discussion of their choice, see the accompanying discussion of this article by Cressie). Wikle, Berliner, and Cressie (1998) built complex Bayesian space-time models that capture spatial dependence using Gaussian Markov random fields (GMRFs) and temporal dependence using vector autoregressive (VAR) models. Wikle and Cressie (1999) developed a space-time Kalman filter that is more general than the (kriged) Kalman filter previously proposed by Mardia, Goodall, Redfern, and Alonso (1998). Le, Sun, and Zidek (1997), and Kibria et al. (2002) used multivariate t models for daily multivariate spatial prediction of pollutants while assuming independence over time using the concept of gauged (sometimes observed) and ungauged (never observed) sites. To deal with the space-time problem, Zidek et al. (2002) extended this approach by taking temporal correlation into account through prewhitening the time series (as a block) which results in minimal loss of the spatial correlation of the residuals, which are again modeled using multivariate t models. Haas (1995, 2002) used a moving "cylinder" to do local prediction in a space time process. Niu, McKeague, and Elsner (2003) proposed a general model for spatio-temporal data on a regular lattice; af-

ter "differencing" to remove seasonal patterns and/or trends, they regressed the response at each location on the grid at each time using ARMA models to account for temporal dependence and order r spatial neighborhoods. Stein (2003) recently proposed a class of nonseparable space-time covariance functions which require pseudo-likelihood approaches for estimation. Shaddick and Wakefield (2002) constructed models for multivariate space-time models using geostatistics models for the spatial component and vector autoregressive models for the temporal and multivariate dependence. Other work in the development of space-time models includes Handcock and Wallis (1994), Stroud, Muller, and Sanso (2001), and Brown, Diggle, Lord, and Young (2001).

In general, computations in space-time models are often problematic, requiring the inversion of large covariance matrices and as such, separable (over space and time) models are often considered (Gelfand, Zhu, and Carlin 2001). In addition, when both spatial and temporal dependence are considered the literature is lacking on models with such dependence represented explicitly via spatial *and* temporal neighbors. Some work in this area can be found in Knorr-Held (2000), who considered neighborhoods among spatial and temporal random effects and with space-time interaction terms.

A general issue these models will address is the use of information on a "densely" observed space-time processes to help "fill-in" a correlated, but sparsely observed, space-time process, whether the process is observed at a point or block. As an illustration of this issue and the methodology, we analyze data from a joint pollutant field of particulate matter and ozone in the Los Angeles (LA) area (van den Eeden 2001), where there were many monitors for ozone, but only a few for particulate matter. Models such as those proposed by Le et al. (1997) and Zidek, Sun, Le, and Ozkaynak (2002) would be well suited to this application, but would not generalize to block source data for which conditionally specified models such as the one proposed here are better suited.

Section 2 introduces the model and discusses specification of the components of the covariance structure, computations, and model fitting. Section 3 presents the results from analyzing the LA pollutant data. Section 4 proposes some extensions of the model. Section 5 contains conclusions and discussion.

2. MODEL

Let $Y_k(s_{ki}, t) : k = 1, \ldots, K; i = 1, \ldots, N_k; t = 1, \ldots, T$ denote the possibly transformed measured value for the kth process (pollutant) at its ith location at time t, with $p = \sum_k N_k$ the total number of subsites. We refer to the s_{ki} as subsites because more than one of the K processes may be measured at the same location. Now, consider models with the following form:

$$\mathbf{Y} \sim N(\mathbf{W}\boldsymbol{\beta}, [\mathbf{M}^{-1}(\boldsymbol{\tau}^2)(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))]^{-1}), \tag{2.1}$$

where $\mathbf{Y} = (\mathbf{Y}^{\star}(1)', \mathbf{Y}^{\star}(2)', \ldots, \mathbf{Y}^{\star}(T)')')'$ is a $pT \times 1$ vector, with $\mathbf{Y}^{\star}(t) = (Y_1(s_{11}, t), Y_1(s_{12}, t), \ldots, Y_1(s_{1,N_1}, t), Y_2(s_{21}, t), \ldots, Y_2(s_{2,N_2}, t), \ldots, Y_K(s_{K,N_K}, t))$ a $p \times 1$ vector, $\mathbf{M}$ is a block diagonal matrix of the form, $\mathbf{M} = \mathbf{I}_T \otimes \mathbf{M}^{\star}$, where $\mathbf{M}^{\star}$ is a $p \times p$ matrix,

and

$$\mathbf{C}(\boldsymbol{\eta}) = \begin{pmatrix} \mathbf{V}_s(\boldsymbol{\eta}_s) & \mathbf{V}_t(\boldsymbol{\eta}_t) & 0 & \cdots \\ \mathbf{V}_t(\boldsymbol{\eta}_t) & \mathbf{V}_s(\boldsymbol{\eta}_s) & \mathbf{V}_t(\boldsymbol{\eta}_t) & \cdots \\ 0 & \mathbf{V}_t(\boldsymbol{\eta}_t) & \mathbf{V}_s(\boldsymbol{\eta}_s) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{2.2}$$

$\mathbf{V}_s(\boldsymbol{\eta}_s)$ is a $p \times p$ matrix characterizing the spatial neighborhoods with $\mathrm{diag}(\mathbf{V}_s) = \mathbf{0}_p$ and parameters $\boldsymbol{\eta}_s$, and $\mathbf{V}_t(\boldsymbol{\eta}_t)$ is a $p \times p$ matrix, characterizing the temporal neighborhoods, parameterized by $\boldsymbol{\eta}_t$.

This covariance structure implies the following two conditional specifications, the first in terms of an individual element of the $\mathbf{Y}$ vector when $\mathbf{M}^\star$ is assumed to be diagonal, and the second in its partitioned form, $\mathbf{Y}^\star(t)$, where $\mathbf{M}^\star$ is not assumed diagonal,

$$Y_k(s_{ki}, t)|Y_{k'}(s_{k'i'}, t') \neq Y_k(s_{ki}, t)$$

$$\sim N\left(\mathbf{w}_{ikt}\boldsymbol{\beta} + \sum_{i',k',t'} c_{(i,i'),(t,t')}^{(k,k')}(y_{k'}(s_{k'i'}, t') - \mathbf{w}_{i'k't'}\boldsymbol{\beta}), \tau_{ik}^2\right), \tag{2.3}$$

$$\mathbf{Y}^\star(t)|\mathbf{Y}^\star(j), j \neq t \sim N(\mathbf{W}_{..t}\boldsymbol{\beta} + \mathbf{V}_t(\boldsymbol{\eta}_t)[(\mathbf{Y}_{t-1}^\star - \mathbf{W}_{..t-1}) + (\mathbf{Y}_{t+1}^\star - \mathbf{W}_{..t+1})], \mathbf{M}^\star),$$
$$\tag{2.4}$$

where $c_{(i,i'),(t,t')}^{(k,k')}$ denotes the element of the $\mathbf{C}(\boldsymbol{\eta})$ matrix corresponding to the pair of values $(Y_k(s_{ki}, t), Y_{k'}(s_{k'i'}, t'))$, $\mathbf{w}_{ikt}$ is the row of the design matrix $\mathbf{W}$ corresponding to the ith location for the kth process at time t, $\mathbf{W}_{..t}$ are the rows of the design matrix corresponding to time t, and τ_{ik}^2 is the diagonal element from the $\mathbf{M}$ matrix corresponding to $Y_k(s_{ki}, t)$. Note that these models are often constructed starting with conditional specification (2.3) or (2.4).

Some features and remarks on the model:

- These models assume a constant spatial neighborhood and parameters over time. However, there are no restrictions on the spatial neighborhood matrix, $\mathbf{V}_s$ (other than $\mathbf{M}^{\star -1}(\mathbf{I}_p - \mathbf{V}_s)$ being positive definite).
- These models assume an AR(1)-type temporal neighborhood structure through the matrix $\mathbf{V}_t(\boldsymbol{\eta}_t)$. However, nonstationarity is allowed by varying the AR(1) parameters over space, and different sites can be lag 1 neighbors. The elements of this matrix are related to the coefficients of vector autoregressive (VAR) models (they are, in fact, just appropriately re-scaled versions of these parameters).
- Proper spatial CAR (e.g., Best et al. 1999) and MCAR (Mardia 1988) models are special cases when $\mathbf{V}_t = \mathbf{0}_p$ and the appropriate structure is placed on $\mathbf{M}^\star$ and $\mathbf{V}_s$. The $\boldsymbol{\eta}_s$ parameters play the role of the spatial autocorrelation parameters in proper spatial (M)CAR models.
- The conditional specifications of this model (2.3) or (2.4) demonstrate how straightforward predictions using both spatial and temporal neighbors can be done.

- Although the models have a similar form to separable geostatistics models (see, e.g., Gelfand et al. 2001), the resulting covariance matrix in these models is *not* separable.
- Asymptotics in these models when modeling point sources are not well understood. In Kaiser et al. (2002), no measures of uncertainty were reported for the parameter estimates. We fit fully Bayesian models here to conduct exact Bayesian inference. This facilitates taking into account all sources of variability when characterizing the variability of the predictions and assessing the variability of the parameters, in general.

2.1 NEIGHBORHOOD STRUCTURE, DEPENDENCE PARAMETERS, AND CONDITIONAL (CO-)VARIANCES

The neighborhood/dependence structure given in the spatial matrix, $\mathbf{V}_s$, and the temporal matrix, $\mathbf{V}_t$, can be specified with great flexibility subject to the constraint that $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ is positive definite. For example, if a natural barrier exists between two sites, the corresponding component of $\mathbf{C}(\boldsymbol{\eta})$ can be set to 0. Such issues can be difficult to handle with direct specification of covariances, as in geostatistics models. Choices for the $\mathbf{M}^\star$ matrix and the spatial and temporal neighborhood matrices, $\mathbf{V}_s$ and $\mathbf{V}_t$, respectively, can be made by equating these matrices to their counterparts in common (multivariate) spatial models and temporal models. For example, if $N_k = N$ for $k = 1, \ldots, K$ and $s_{ki} = s_{k'i}$ for all i, then the form of the $\mathbf{M}^\star$ and $\mathbf{V}_s$ can be specified as in MCAR models (Mardia 1988). Section 4.3 discusses extending this to the more flexible MCAR models of Gelfand and Vounatsou (2003). Any type of spatial dependence/parameters can be used in $\mathbf{V}_s(\boldsymbol{\eta})$ as long as they are constant over time. The form of the $\mathbf{V}_t$ matrix can be made similarly to the choices commonly made in vector autoregressive (VAR) time series models for multivariate responses. It might be chosen to be diagonal with a separate parameter for each process/site or to be a full matrix with the temporal autoregressive parameters decreasing with distance from the site. Simplified (diagonal) forms for the $\mathbf{M}^\star$ matrix include $\mathbf{M}^\star = \tau^2\mathbf{I}_p$, a constant conditional variance across sites and time and process or $\mathbf{M}^\star = \mathrm{diag}_p(\tau_k^2)$, that is, a separate conditional variance for each process. Additional discussion of specification of the $\mathbf{M}^\star$ matrix can be found in Section 4.4. We introduce the neighborhood structure, $\boldsymbol{\eta}$ parameters, and the parameterization of the $\mathbf{M}^\star$ matrix used in our example in Section 3.

2.2 PRIOR DISTRIBUTIONS

A diffuse normal prior was assumed on the regression parameters, $\boldsymbol{\beta}$, and inverse Wishart/gamma priors on $\mathbf{M}^\star$ matrix and/or its components as appropriate. Prior specification for the dependence parameters, $\boldsymbol{\eta}$, is tricky and an open problem. Here, for simplicity, we consider uniform priors on the space over which $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ is positive definite. For some models, this region can be determined explicitly. For details, see Section 3.3. Prior

specification on simpler CAR models within a multilevel setup can be found in Sun et al. (1999).

2.3 Computational Difficulties and Solutions

Space-time models often create computational challenges. This section discusses how this class of models can be fit with minimal computational difficulty. The main challenge involves computation of the determinant, inverse, and eigenvalues of a matrix which is $Tp \times Tp$ where T is the number of time points and $p = \sum_k N_k$ is the total number of spatial locations over each process. To accomplish these tasks, clever, but simple, computational approaches are required. We consider the case where $\mathbf{M} = \mathbf{I}_T \otimes \mathbf{M}^\star$.

Result 1:

If $\mathbf{M}^{\star\,-1}(\mathbf{I}_p - \mathbf{V}_s)$ is not positive definite, then $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ is not positive definite.

Proof: It is a leading submatrix of $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$. $\square$

Result 2:

Given the form of $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$, we can rewrite it as

$$\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta})) = [\mathbf{I}_T \otimes \mathbf{M}^{\star\,-1}(\mathbf{I}_p - \mathbf{V}_s(\boldsymbol{\eta}_s))] - \boldsymbol{\rho} \otimes (\mathbf{M}^{\star\,-1}\mathbf{V}_t(\boldsymbol{\eta}_t)), \qquad (2.5)$$

where $\boldsymbol{\rho}$ is the $T \times T$ matrix defined below.

$$\boldsymbol{\rho} = \begin{pmatrix} 0 & 1 & 0 & \dots \\ 1 & 0 & 1 & \dots \\ 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Result 3: Using some Kronecker product identities,

$$\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta})) = (\mathbf{I}_T \otimes \mathbf{A}^{-1})(\mathbf{I}_{pT} - \boldsymbol{\rho} \otimes \mathbf{B})(\mathbf{I}_T \otimes \mathbf{A}^{-1}),$$

where $\mathbf{A} = (\mathbf{I}_p - \mathbf{V}_s)^{-1/2}\mathbf{M}^{\star\,1/2}$ and $\mathbf{B} = \mathbf{A}\mathbf{M}^{\star\,-1}\mathbf{V}_t\mathbf{A}$. This result will hold given Result 1.

Result 4: $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ is positive definite if and only if $\mathbf{A}$ is full rank and $(\mathbf{I}_{pT} - \boldsymbol{\rho} \otimes \mathbf{B})$ is positive definite. (follows from an adaptation of Theorem A.1.1 in Anderson (1984, p. 583)).

Theorem 1. *Define γ_i to be the eigenvalues of $\mathbf{M}^{\star\,-1}(\mathbf{I}_p - \mathbf{V}_s)$, λ_i to be the eigenvalues of $\boldsymbol{\rho}$, and μ_j to be the eigenvalues of $\mathbf{B}$. Then $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ will be positive definite iff $\gamma_i > 0$ and $\lambda_i \mu_j < 1$ for all i and j.*

Proof: Follows directly from Results 1–4. $\square$

So to check the positive definiteness of the joint covariance matrix $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$, we only need to compute the eigenvalues of two $\sum_k N_k$-dimensional matrices, $\mathbf{A}$ and $\mathbf{B}$, and a

T-dimensional matrix ρ, which has a simple form and is fixed; this implies that even within a sampling based algorithm, the eigenvalues of ρ only need to be computed initially. Given the eigenvalues, it is easy to compute the determinant. To invert $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$, we just need the eigenvectors of these same matrices. Thus, all the computations reduce to dealing with T and $p = \sum_k N_k$ dimensional matrices. Based on specific neighborhood structure, calculations can be simplified even further (see Section 3.3).

The form of the $\mathbf{C}(\boldsymbol{\eta})$ matrix given in Equation (2.2) implies a "balanced" $\mathbf{C}(\boldsymbol{\eta})$ matrix, that is, the same sites over time and the same times for all sites. The computational tricks above require such balance. If the actual data are unbalanced, balance can be achieved by filling in the intermittent "missing" data using data augmentation within an MCMC algorithm. The details of this will be discussed in Section 2.5.

So, similar to separable space-time geostatistics models, the "limiting" factor in computations is the size of T and $p = \sum_{k=1}^{K} N_k$. Manipulating other T, p, or N_k dimensional matrices can sometimes facilitate posterior computation (see Section 3.3).

2.4 POSTERIOR SAMPLING

We use a Gibbs sampler to obtain a sample from the posterior distribution of the parameters. The full conditional distributions for β is normal. Sampling $\mathbf{M}^\star$ will depend on the form of this matrix; for example, if $\mathbf{M}^\star = \tau^2 \mathbf{I}_p$, then the full conditional distribution of τ^2 will be inverse gamma. Otherwise, we propose to use an appropriate version of the Metropolis-Hastings algorithm. To sample the $\boldsymbol{\eta}$, we used a Hybrid MC algorithm (Neal 1994). This requires evaluation of the first derivative of the log full conditional with respect to $\boldsymbol{\eta}$, which is simple to compute given the form of the likelihood,

$$\log(L(\boldsymbol{\eta})) \propto (1/2) \log |\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))|$$

$$-(1/2)\mathrm{tr}[\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))(\mathbf{Y} - \mathbf{W}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{W}\boldsymbol{\beta})^T] \quad (2.6)$$

as $\boldsymbol{\eta}$ appear directly in the components of the $\mathbf{C}(\boldsymbol{\eta})$ matrix. This algorithm facilitates efficient movement through the multivariate $\boldsymbol{\eta}$ space.

When intervals over which the $\boldsymbol{\eta}$ resulted in a positive definite $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ matrix could be determined explicitly, the parameters were transformed using a logit type transformation on that interval, $\log(\frac{\eta - \eta_{\min}}{\eta_{\max} - \eta})$, where $(\eta_{\min}, \eta_{\max})$ is the interval; this was done in order to transform η from a closed interval for the Hybrid MC algorithm. If an interval cannot be determined, a check of positive definiteness (pd) can be conducted using Theorem 1, and if the values corresponds to a nonpositive definite covariance matrix, the value is thrown out (since it has prior probability 0). Details using the structure of the $\mathbf{C}$ and $\mathbf{M}$ matrix in our example will be given in Section 3.3.

2.5 Model Comparison

As discussed by Kaiser et al. (2002), evaluation of model fit and predictive ability are two different questions and can result in differing conclusions. To evaluate model fit and compare models, we use the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, and van der Linde 2002). We will give some details in the following. Define the parameter vector to be $\boldsymbol{\theta}$. The DIC is defined to be $\text{Dev}(\hat{\boldsymbol{\theta}}) + 2p_D$, where $\text{Dev}(\boldsymbol{\theta}) = -2\text{loglik}(\boldsymbol{\theta})$, $\hat{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$, and $p_D = \hat{\text{Dev}} - \text{Dev}(\hat{\boldsymbol{\theta}})$ with $\hat{\text{Dev}}$ the posterior mean of the Deviance. Here, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{M}(\boldsymbol{\tau}^2), \boldsymbol{\eta})$ and the likelihood is given in (2.6). Smaller values of the DIC are preferred.

To compare the predictive ability of these models, we compute both the prediction error and coverage probabilities. We use a cross-validation approach. To assess spatial prediction, we drop out the data on a subsite at all times, refit the model, and then predict the process values at the dropped subsite. For temporal prediction, we drop out all subsites at the first and last times, refit the model, and then predict the process values at the dropped times.

Some details on prediction follow (under the restriction that $\mathbf{M}^\star$ is diagonal; extension to nondiagonal $\mathbf{M}^\star$ is straightforward). Similar to the approach in Kaiser et al. (2002), we drop a subsite (time) and fit the model to the remaining sites to obtain the posterior distributions of the parameters conditional on these subsites. Then, we assume the pollution process at the dropped subsite follows the same distributional form as (2.3), that is, the conditional distribution of $\mathbf{Y}_0$, the vector of responses at the subsites/times to be predicted (i.e., the dropped subsites/times), given the remaining subsites/times, $\mathbf{Y}_{\text{obs}}$, follows a normal distribution with mean,

$$\mathbf{Y}_{0,\text{pred}} = \mathbf{W}_0\boldsymbol{\beta} + \sum_{k=1}^{K}\sum_{i=1}^{N_k}\sum_{t=1}^{T}\mathbf{c}_{(0,i),(0,t)}^{(0,k)}(Y_k(s_{ki},t) - \mathbf{W}_{ikt}\boldsymbol{\beta}) \tag{2.7}$$

and variance

$$\mathbf{V}_y^{-1} = \mathbf{M}_0^{-1}(\mathbf{I}_{\dim(\mathbf{Y}_{0,\text{pred}})} - \mathbf{C}_0), \tag{2.8}$$

where $\mathbf{C}_0$ is the neighborhood matrix for the subsites to be predicted, $\mathbf{W}_0$ is the design matrix for the subsites to be predicted, $\mathbf{c}_{(0,i),(0,t)}^{(0,k)}$ is the vector of elements of the $\mathbf{C}(\boldsymbol{\eta})$ matrix corresponding to the regression of the subsites to be predicted on the remaining sub-sites (see (2.3)), and $\mathbf{M}_0$ is a diagonal matrix containing the appropriate conditional variances; these are formed using (2.3). So, for prediction, the dimension of the matrix that needs to be inverted will be the dimension of $\mathbf{Y}_0$.

The $\mathbf{C}(\boldsymbol{\eta})$ matrix is constructed to include intermittent missing values. At each iteration of the sampler, these values are filled in via data augmentation using a normal distribution with mean and variance analogues of (2.7) and (2.8); this is necessary for the computational results outlined in Section 2.3 to hold. We emphasize that by using this approach, we are implicitly using Monte Carlo integration to account for the uncertainty of these missing values.

We will evaluate prediction using mean squared prediction error (MSPE) and coverage probabilities. Mean square prediction error is defined as

$$\text{MSPE} = \sum (Y_{0,\text{pred}} - Y_{0,\text{obs}})^2,$$

where $Y_{0,\text{pred}}$ is the conditional posterior mean given in (2.7) and $Y_{0,\text{obs}}$ is the observed value of the dropped sub-site/time. To help calibrate MSPE, we can compare its value to the conditional variances (τ_{ik}^2), estimated from fitting the model with all subsites/times.

To evaluate coverage, we compute 95% prediction intervals using the following approach. We will illustrate the approach under the setting of dropping out a subsite at all times (dropping out all subsites at one time follows similarly). At iteration m (of M) of the Gibbs sampler, we create 95% credible regions using $(\mathbf{Y}_{0,\text{obs}}(s_{ki}, \cdot) - \mathbf{Y}_{0,\text{pred}}^m(s_{ki}, \cdot))^T (\mathbf{V}_y^m)^{-1}$ $(\mathbf{Y}_{0,\text{obs}}(s_{ki}, \cdot) - \mathbf{Y}_{0,\text{pred}}^m(s_{ki}, \cdot))$ where $\mathbf{Y}_{0,\text{pred}}^m(s_{ki}, \cdot)$ is given in (2.7) and $\mathbf{V}_y^m$ is given in (2.8), using the values of (τ^2, η, β) at iteration m. Because the conditional distribution is normal, this should follow a chi-squared distribution on T degrees of freedom. We can count for how many iterations these regions cover the true value of the dropped subsite; ideally, close to 95% for each subsite. A referee has pointed out that these regions may be too small, especially if there is considerable uncertainty/variability reflected in the posterior distribution of (β, η, τ^2). As an alternative, we also examine marginalized credible regions that integrate over the uncertainty in (β, η, τ^2). These are constructed as $(\mathbf{Y}_{0,\text{obs}}(s_{ki}, \cdot) - \mathbf{Y}_{0,\text{pred}}(s_{ki}, \cdot))^T (\mathbf{V}_y^\star)^{-1}(\mathbf{Y}_{0,\text{obs}}(s_{ki}, \cdot) - \mathbf{Y}_{0,\text{pred}}^m(s_{ki}, \cdot))$, where $\mathbf{Y}_{0,\text{pred}}(s_{ki}, \cdot)$ and $\mathbf{V}_y^\star$ are computed using Rao-Blackwellization (Gelfand and Smith 1990) as follows, $\mathbf{Y}_{0,\text{pred}}(s_{ki}, \cdot) = \frac{1}{m} \sum \mathbf{Y}_{0,\text{pred}}^m(s_{ki}, \cdot)$ and $\mathbf{V}_y^\star = \frac{1}{m}[\mathbf{V}_y^m + \mathbf{Y}_{0,\text{pred}}^m(s_{ki}, \cdot)$ $\mathbf{Y}_{0,\text{pred}}^m(s_{ki}, \cdot)^T] - \mathbf{Y}_{0,\text{pred}}(s_{ki}, \cdot)\,\mathbf{Y}_{0,\text{pred}}(s_{ki}, \cdot)^T$. For this region, we again use a chi-squared distribution with T degrees of freedom to create a cutoff; however, here, this will only be approximate, with the approximation improving as the posterior of (β, η, τ^2) concentrates around the true values.

If the original data were transformed, the predictive variance will be needed to obtain the posterior predictive mean on the original scale. For example, in the application, we log transformed the responses. So, the prediction on the original scale, X_{pred}, is given as

$$X_{\text{pred}} = \exp\{Y_{\text{pred}} + V_y/2\}.$$

We will compute MSPE on both the original scale and the transformed scale of the data .

In the following example, we will examine MSPE and coverage overall, and for each process separately.

3. DATA EXAMPLE

We illustrate this methodology on pollutant data from the LA area. There were a total of 35 sites with pollutant monitors, 5 of which monitored both particulate matter (PM) and ozone and 30 which only monitored ozone; a map of the sites is given in Figure 1. We analyze data collected during a 54-day period during the summer in 1995, July 10 through
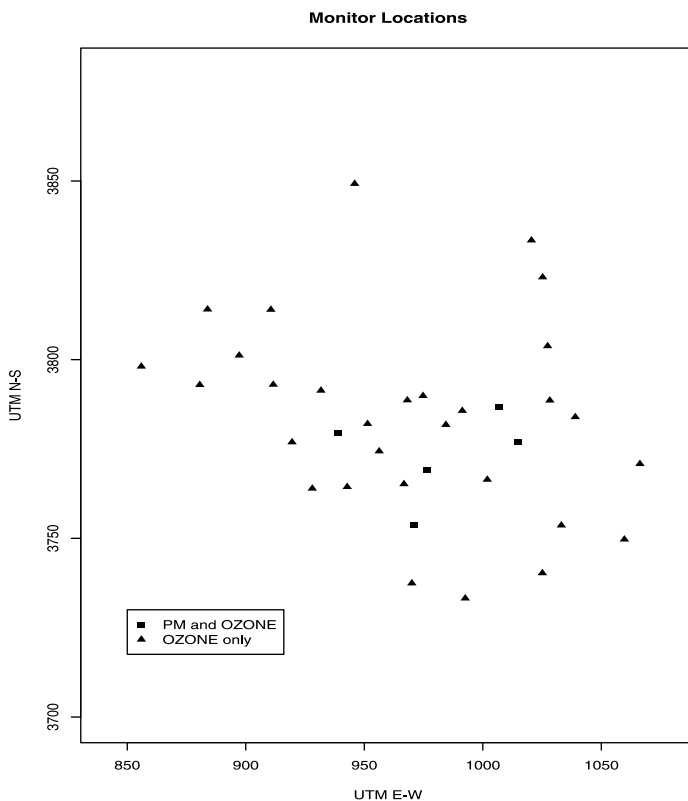
Figure 1.   PM and ozone monitor locations.

September 1 (this period was chosen since it contained no missing data); however, we point out that that the time window was chosen for ease of demonstration of these models. These models easily accommodate intermittent missing data (even if missing by design) as discussed in Section 2.5. We also mention that we undertake a simple analysis of this data here. A more definitive analysis would include many extensions which we discuss more thoroughly in Sections 3.7 and 4. One objective of this modeling was to use ozone to help predict PM given the sparsity of the PM sites.

A log transformation of both ozone and PM seemed to provide a good approximation to normality and constant variance. This agrees with previous work by Carroll et al. (1997) on ozone and Cressie et al. (1999) on PM. The design matrix ultimately chosen, $\mathbf{W}$, consisted of 9×1 design vectors which included a set of ozone-specific components: an intercept, terms for both latitude and longitude and an interaction, and a quadratic in time; and a set of PM-specific components: an intercept and a quadratic in time. We also considered some additional design vectors that included/excluded spatial components for the mean structure for PM/ozone. Some weather covariates were also available, but there was so much missingness here that they could not be used without explicitly building models for them. Ultimately, the most complex mean structure "fit" the data best (by DIC), but resulted in poorer cross-validation predictive accuracy.

A summary of the data by site appears in Table 1.

3.1 NEIGHBORHOOD STRUCTURE

We consider the following possible neighbors of $Y_k(s_{ki}, t)$

1. $Y_k(s_{ki}, t+1)$. The pollutant value at the same location on the subsequent day.
2. $Y_k(s_{ki}, t-1)$. The pollutant value at the same location on the previous day.
3. $\{Y_k(s_{kl}, t) : l \neq i\}$. The pollutant values on the same day at other locations.
4. $\{Y_j(s_{ji}, t) : j \neq k\}$. The pollutant values of other pollutants on the same day at the same location.

Table 1. Summary Statistics for Each Site Over Time on the Log Scale. Sites 1–5 are measured for both PM and ozone.

Pollutant	Site no.	Mean	Variance	Min	Max
PM	1	3.93	0.077	3.42	5.27
	2	3.70	0.056	3.13	4.76
	3	4.38	0.059	3.75	4.81
	4	4.21	0.112	3.36	5.35
	5	3.79	0.040	3.47	4.34
ozone	1	3.22	0.042	2.73	3.61
	2	3.15	0.073	2.65	3.65
	3	3.78	0.059	3.35	4.25
	4	3.68	0.082	3.21	4.27
	5	3.38	0.082	2.73	3.85
	6	2.67	0.094	1.72	3.24
	7	3.66	0.046	3.30	4.10
	8	3.73	0.121	2.83	4.55
	9	3.46	0.051	2.96	4.05
	10	3.57	0.118	3.14	4.60
	11	4.00	0.282	1.97	4.59
	12	3.28	0.089	2.48	3.88
	13	3.13	0.092	2.45	3.71
	14	3.83	0.061	3.41	4.38
	15	3.30	0.044	2.59	3.71
	16	3.79	0.043	3.23	4.19
	17	3.96	0.054	3.39	4.45
	18	3.24	0.067	2.64	3.73
	19	3.81	0.058	3.17	4.32
	20	3.92	0.075	3.36	4.35
	21	3.63	0.095	3.05	4.28
	22	3.66	0.053	3.29	4.11
	23	3.54	0.047	3.04	4.01
	24	3.86	0.058	3.34	4.41
	25	3.43	0.038	3.02	3.84
	26	3.53	0.040	3.13	3.96
	27	3.53	0.068	3.08	4.03
	28	4.04	0.083	3.37	4.54
	29	3.48	0.062	2.82	4.04
	30	3.95	0.065	3.53	4.41
	31	3.80	0.062	3.35	4.60
	32	3.59	0.051	3.03	4.60
	33	3.75	0.074	3.33	4.28
	34	3.84	0.139	2.57	4.60
	35	3.31	0.046	2.62	3.73

Let $c^{(k,k')}_{(i,i'),(t,t')}$ denote the element of the $\mathbf{C}(\boldsymbol{\eta})$ matrix corresponding to the pair of values $(Y_k(s_{ki},t), Y_{k'}(s_{k'i'},t'))$. Using the neighborhood structure given above, we specify the $\mathbf{C}(\boldsymbol{\eta})$ matrix to have the following structure,

1. $c^{(k,k)}_{(i,i),(t,t+1)} = \eta^{(k)}_t$,
2. $c^{(k,k)}_{(i,i),(t,t-1)} = \eta^{(k)}_t$,
3. $c^{(k,k)}_{(i,l),(t,t)} = \eta^{(k)}_s f(s_i, s_l)$,
4. $c^{(k,j)}_{(i,i),(t,t)} = \eta^{(k,j)}_s$,

and 0 otherwise. In this application, we set $f(s_i, s_l) = \min_{i',l'}(d_{i',l'})/d_{i,l}$, where $d_{i,l}$ is the Euclidean distance between the two sites. This is similar to the choice of Cressie and Chan (1989), though they set the dependence to 0 when the distance between locations exceeded some data determined threshold. For a good approach for determining distance based neighborhood structures using geostatistical techniques, we refer the reader to Hrafnkellson and Cressie (2003), which we will discuss further in Section 4. For block source data, $f(s_i, s_l)$ might be chosen to be an indicator of whether sites s_i and s_l are neighbors or some measure of the "quality" of the neighbors (see Section 4.4).

3.2 CHOICE OF THE CONDITIONAL COVARIANCE MATRIX, $M^\star$

For the $\mathbf{M}^\star$ matrix, we consider two specifications. For the first, we assume a common conditional variance over space, that is, $\mathbf{M}^\star = \tau^2 \mathbf{I}_p$. For the second, we assume a separate conditional variance for each process (pollutant); in our application, τ^2_1 and τ^2_2. The elements of the $\mathbf{C}(\boldsymbol{\eta})$ matrix then need to be adjusted for the matrix $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ to be symmetric. The elements of the $\mathbf{C}(\boldsymbol{\eta})$ matrix will remain as specified above except for the elements of the $\mathbf{C}(\boldsymbol{\eta})$ matrix corresponding to cross-pollutant neighbors. These will have the following structure,

1. $c^{(k,j)}_{(i,i),(t,t)} = \eta^{(j,k)}_s \tau_j/\tau_k$,
2. $c^{(j,k)}_{(i,i),(t,t)} = \eta^{(j,k)}_s \tau_k/\tau_j$,

where the first term is in the upper triangle and the second term is in the lower triangle of the $\mathbf{C}(\boldsymbol{\eta})$ matrix.

3.3 POSTERIOR SAMPLING DETAILS

The full conditional for τ^2 in the model with $\mathbf{M}^\star = \tau^2 \mathbf{I}_p$ will be inverse gamma. For the more general specification of $\mathbf{M}^\star$, the $\mathbf{C}(\boldsymbol{\eta})$ matrix will contain the diagonal elements of the $\mathbf{M}^\star$ matrix. For this choice of $\mathbf{M}^\star$, we use a random walk Metropolis-Hastings algorithm. For the neighborhood structure, given in Section 3.1, we can compute "necessary" intervals for positive definiteness (i.e., values outside the interval, correspond to a nonpositive definite matrix). For $\mathbf{M} = \tau^2 \mathbf{I}$ and the specific $\mathbf{M}^\star$ described in Section 3.2, necessary intervals, given our neighborhood structure, for $\eta^{(k)}_s$ are $(1/\psi_{k,\max}, 1/\psi_{k,\min})$, where $\psi_{k,\min}$ and $\psi_{k,\max}$ are the smallest and largest eigenvalues, respectively, of the distance matrices for each pollutant (k), that is, the $N_k \times N_k$ matrix of pairwise distances (as specified in Section 3.1

Table 2.   Models Fit

| Model | Parameters | |
	Spatial	Temporal
I	—	—
II	—	$(\eta_t^{(1)}, \eta_t^{(2)})$
III	$(\eta_s^{(1)}, \eta_s^{(2)})$	—
IV	$(\eta_s^{(1)}, \eta_s^{(2)})$	$(\eta_t^{(1)}, \eta_t^{(2)})$
V	$(\eta_s^{(1)}, \eta_s^{(2)}, \eta_s^{(1,2)})$	$(\eta_t^{(1)}, \eta_t^{(2)})$

with 0's on the main diagonal). In addition, the following bound must be satisfied which provides additional restrictions on $\eta_t^{(k)}$ given $\eta_s^{(k)}$ and vice versa, $|\eta_t^{(k)}| < \frac{1}{\max_l \lambda_l}(1 - \eta_s^{(k)}\psi_{k,\max})$. Here, we have assumed that $\eta_s^{(j,k)} = 0$; the intervals are further restricted if this is not the case. In Section 3.4, we use these bounds to help interpet the magnitude of the estimated dependence parameters ($\boldsymbol{\eta}$). However, we cannot use a logit transformation based on this last interval as the bounds for $\eta_t^{(k)}$ depend on the current value of $\eta_s^{(k)}$ and vice versa. However, we can use the first set of bounds for $\eta_s^{(k)}$ and simpler bounds on $\eta_t^{(k)}$, $(-.5, .5)$, based on the eigenvalues of the $\boldsymbol{\rho}$ matrix, $\boldsymbol{\lambda}$.

The Gibbs sampling algorithm converged quickly using several starting values. Fifty-thousand iterations were run, with the first 1,000 thrown out as burn-in; computing time for complex models like Models IV and V was four hours on a PC with a Pentium 4, 1.8 GHz processor. For Model V, we broke the dependence parameters, $\boldsymbol{\eta}$'s into two blocks: (1) $(\eta_s^{(1)}, \eta_s^{(2)}, \eta_t^{(1)}, \eta_t^{(2)})$, and (2) $\eta_s^{(1,2)}$. The dependence parameters mixed quite well using the Hybrid MC algorithm though the cross-pollutant parameter, $\eta_s^{(1,2)}$ mixed more slowly than the others. Convergence was assessed by examining trace plots from four chains.

3.4   Model Fit and Inference on η Parameters

We fit five dependence models as given in Table 2. Model fit, as discussed in Section 2.5, was assessed via the DIC. DIC values for the given models, are given in Table 3. Clearly, the model with all the spatial and temporal dependence parameters fit best.

The estimates of $\boldsymbol{\eta}$, the dependence parameters, and τ_k^2, $k = 1, 2$, the variance parameters, for the best fitting model, Model V, appear in Table 4. Using the information on the necessary intervals/bounds in Section 3.3, all of the spatial and temporal dependence parameters are quite close to their upper bounds. As the intervals in Section 3.3 do not

Table 3.   DIC and Effective Number of Parameters (p_D) for Models Fit

Model	DIC	p_D
I	1259	11.7
II	269	13.8
III	755	14.7
IV	99.7	14.9
V	78.1	14.3

Table 4. Posterior Means and 95% Credible Intervals for the Spatial Dependence Parameters. (1) corresponds to PM and (2) to ozone.

Parameter	Posterior mean (CI)
$\eta_{\text{space}}^{(1)}$	0.046 (0.005,0.099)
$\eta_{\text{space}}^{(2)}$	0.043 (0.035,0.051)
$\eta_{\text{space}}^{(1,2)}$	0.065 (0.033,0.094)
$\eta_{\text{time}}^{(1)}$	0.42 (0.37,0.45)
$\eta_{\text{time}}^{(2)}$	0.39 (0.37,0.41)
τ_1^2	0.055 (0.046,0.066)
τ_2^2	0.049 (0.045,0.052)

include $\eta_s^{(1,2)}$, we did some simulations conditional on the other η parameters and found this was also, in fact, near its upper bound as well. So, clearly, both the spatial and temporal dependence are strong in this data.

3.5 Prediction Error

For prediction, for each subsite, 5,100 iterations were run, with a burn-in of 100; 22 hours were required to run the prediction for all the subsites in the model. Predictive ability as measured by MSPE is given in Tables 5–8 for prediction on the log and original scale both spatially and temporally. We use the posterior means of τ_1^2 and τ_2^2 and Table 1 to help calibrate the magnitude of the prediction errors for PM and ozone below.

The overall gains in spatial prediction when comparing the independence model to the best dependence model, Model V, was to decrease MSPE by about 30%. The magnitude of the prediction error was about twice the value of the respective posterior means of τ^2 (see Table 4). To assess the gains in PM prediction by modeling it jointly with ozone, we compare models IV and V. The gains in predicting PM here are not large, on the order of a 2–4% reduction in MSPE. The bias in general was quite small (see Table 1).

The superior temporal prediction for PM versus ozone is mainly due to the fact that an AR(1) type model fit PM very well, but not ozone (not shown). However, we do see a decrease in MSPE of over 30% for ozone when including temporal dependence in the model. The huge differences between the MSPE for PM was due to very poor prediction for the independence model on the first day. We can see that in Model V, the magnitude of

Table 5. MSPE and Bias Overall and Broken Down by Pollutant for Predicting One Site at All Times on Log Scale

	Overall		PM		Ozone	
Model	MSPE	Bias	MSPE	Bias	MSPE	Bias
I	0.119	0.003	0.169	0.000	0.112	0.003
IV	0.094	0.004	0.107	−0.003	0.092	0.005
V	0.093	0.005	0.103	−0.004	0.091	0.007

Table 6. MSPE and Bias Overall and Broken Down by Pollutant for Predicting One Site at All Times on Original Scale

	Overall		PM		Ozone	
Model	MSPE	Bias	MSPE	Bias	MSPE	Bias
I	239	0.4	735	0.15	168	0.41
IV	180	0.0	508	−0.6	133	0.1
V	178	0.0	498	−0.9	133	0.1

Table 7. MSPE and Bias Overall and Broken Down by Pollutant for Predicting all Sites at the First and Last Times on Log Scale

	Overall		PM		Ozone	
Model	MSPE	Bias	MSPE	Bias	MSPE	Bias
I	1.26	−0.30	9.1	−2.1	0.133	−0.04
V	0.092	−0.09	0.030	−0.08	0.100	−0.10

the prediction error for PM was less than the posterior mean for τ_1^2 while it was about twice the value of τ_2^2 for ozone. This indicates solid temporal prediction for PM.

3.6 COVERAGE RESULTS

To obtain some absolute measure of model fit, we examined the coverage probabilities (based on 95% confidence regions) for Model V using the cross-validation approach. For the PM spatial predictive coverage, we observed coverages between 0.94 and 1.0 for four of the five sites, with one site having coverage of 0.0; for ozone, we observed coverage between 0.94 and 1.0 for 28 of the 35 sites, with 7 sites having less than 0.04 coverage. Thus, for both PM and ozone, the model appears to be fitting about 80% of the subsites well. This agreed with the marginal credible regions which covered 80% of the sub-sites. For PM temporal predictive coverage, coverage was over 90% for both days while for ozone, coverage was about 0.98 for the last day, but 0.00 for the first day.

We did some further exploration of the subsites with poor coverage. These tended to be sites with the fewest "quality" neighbors (for a discussion of "quality" neighbors, see Section 4.4) and with considerable variability over time (relative to the other subsites; see Table 1). Poorer prediction at sub-sites with not too many "quality" neighbors is not surprising.

Table 8. MSPE and Bias Overall and Broken Down by Pollutant for Predicting all Sites at the First and Last Times on Original Ccale

	Overall		PM		Ozone	
Model	MSPE	Bias	MSPE	Bias	MSPE	Bias
I	547	−5.5	2730	−35	235	−1.3
V	190	−4.0	87	−4.0	204	−3.9

3.7   Model Expansions

These results suggest further model expansion from several perspectives. In terms of spatial dependence, nonstationary spatial structures should be considered, especially given the topography in this region of the United States, as well as fine tuning the distance-based neighborhood structure using the approach of Hrafnkellson and Cressie (2003) (see Section 4.5). Other extensions might include adjusting the $\mathbf{M}$ matrix to account for the "quality" of neighbors (see Section 4.4) and taking into account the measurement error in the monitors (Section 4.2). We will discuss some details about these expansions and others in the next section.

4.   MODEL GENERALIZATIONS AND EXTENSIONS

This section discusses several generalizations and extensions to the model, both with respect to the neighborhood structure and incorporating additional sources of variability.

4.1   More Than Lag 1 Temporal Dependence

Only lag 1 temporal dependence has been considered here. Increasing the temporal neighborhood structure beyond the single nearest neighbor creates major difficulties with computations. Such matrices will have a block Toeplitz form for which the inverse can be computed efficiently (Dietrich 1991), but checking positive definiteness and computing the determinant are not computationally tractable.

4.2   Measurement Error

Another extension would be a to build a hierarchical (mixture) model to allow for measurement error. These models are most appropriate when the variance of the measurement error is known or there are replications. In our setting, we did not know the measurement error variances and we did not have replications. However, the measurement error variance still could be identified by assuming independence of the measurement errors, conditional on the true values of the process; the measurement error variance estimated in this way would likely be confounded by residual error not explained in the spatial dependence model for the true process (see below). An alternative approach would be to put an informative prior on σ^2.

The extension to our model is as follows,

$$\mathbf{Y} \sim N(\boldsymbol{\theta}, \mathbf{D}), \tag{4.1}$$

$$\boldsymbol{\theta} \sim N(\mathbf{W}\boldsymbol{\beta}, [\mathbf{M}^{-1}(\boldsymbol{\tau}^2)(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))]^{-1}), \tag{4.2}$$

with $\mathbf{D}$ a diagonal matrix often specified as $\sigma^2\mathbf{I}$. We have added a measurement error component into the model and the true process of interest is now represented by $\boldsymbol{\theta}$, not $\mathbf{y}$. How does this impact computing? Sampling from the full conditional distributions of $\boldsymbol{\beta}, \boldsymbol{\eta}, \tau^2$ will be the same, but with $\mathbf{Y}$ replaced by $\boldsymbol{\theta}$. The full conditional for σ^2 will be inverse gamma (assuming a conjugate prior). The full conditional distribution of $\boldsymbol{\theta}$ will be

normal and the tools used in this article to deal with $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ can be used for computing this distribution. Finally, intermittent missing data can easily be handled more easily within this model as the missing $\mathbf{Y}$'s can be sampled from a $N(\boldsymbol{\theta}, \mathbf{D})$ distribution.

For count or binary data, (4.1) could be replaced by a Poisson or logistic regression model with $\boldsymbol{\theta}$ now being a random effect.

4.3 EXTENSION TO FLEXIBLE MCAR MODELS

Gelfand and Vounatsou (2003) considered linear transformations of the data/random effects to extend the MCAR models of Mardia (1988) to allow a separate spatial parameter for each of the multivariate processes. We will give some details on incorporating these ideas into our model. Consider a $p \times p$ linear transformation matrix $\mathbf{G}$, which is block diagonal with blocks of size K. We can modify the model proposed in this article by replacing the covariance matrix $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ with the matrix $(\mathbf{G} \otimes \mathbf{I}_T)\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))(\mathbf{G} \otimes \mathbf{I}_T)'$. The benefit of extending the model using this approach is a more flexible specification for multivariate spatial dependence with simple constraints on the spatial dependence parameters for positive definiteness; thus, no check would need to be made on the eigenvalues, γ_i, as in Theorem 1. In addition, Theorem 1 will hold as before with appropriate modification to the matrix $\mathbf{B}$. Care, however, must be taken to account for the impact of $\mathbf{G}$ on the temporal neighborhood matrix, $\mathbf{V}_t$.

4.4 ADJUSTMENT FOR ''QUALITY'' OF NEIGHBORS

In the air pollution dataset here, and in many other settings with an irregular lattice, there is an asymmetry in terms of the number of neighbors (for block sources) or the number and quality of neighbors (with point sources). The "adjustment" often made to the conditional variance in CAR models with block sources is to divide a constant unknown variance by the number of neighbors, τ^2/n_i (Sun et al. 1999). The motivation for such an adjustment is that we would expect the conditional variance of sites with more neighbors to be smaller than sites with fewer neighbors. Adjustments are also frequently made for the quality of neighbors as well; for example, how much of the total boundary do they share. In point source data, often all sites are neighbors to each other. We might construct a similar adjustment to the standard CAR models in this case by dividing τ^2 by a measure of the "quality" of the neighbors. We might define this quantity, $n^\star$, by the inverse of the harmonic average of the distances from the current site to all other sites. Because the neighborhood structure in this model is assumed static over time, $n^\star$ will be constant over time and computations will proceed under the $\mathbf{M} = \mathbf{I}_T \otimes \mathbf{M}^\star$ case. Appropriate adjustments then need to be made to the conditional mean (through the $\mathbf{C}$ matrix) to ensure $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$ remains symmetric. This idea will be explored in future work and might improve the prediction for the pollution example as discussed in Section 3.6.

4.5 ALTERNATIVE NEIGHBORHOOD STRUCTURES AND NONSTATIONARITY

The temporal neighborhood structure used in the example could be supplemented by including the following additional neighbors for $Y_k(s_{ki}, t)$; $Y_k(s_{kl}, t+1), l \neq i$ (lag 1 temporal correlation across sites) and/or $Y_j(s_{jl}, t+1), l \neq i, j \neq k$ (lag 1 temporal correlation across sites and pollutant). In addition, to better calibrate distance-based neighbors, the techniques in Hrafnkellson and Cressie (2003) could be implemented which involves, among other things, finding a range of dependence for the neighbors by "equating" the neighborhood structure to the results from fitting variograms using geostatistical techniques. In this article, the distance-based neighborhood structure was specified without using such techniques as in Kaiser et al. (2002) among others.

The neighborhood structure proposed for our models might be characterized as stationary. To extend to a nonstationary dependence, we can allow the neighborhood dependence parameters only to operate locally and/or allow them to vary over space. The deformation of distance approach (Sampson and Guttorp 1992) might be a reasonable methodology to try to integrate into this model. Care must be taken in order to ensure the conditional variance matrix $\mathbf{M}^\star$ and the neighborhood matrix, $\mathbf{C}(\boldsymbol{\eta})$ are specified appropriately so as to make $\mathbf{M}^{-1}(\mathbf{I}_{pT} - \mathbf{C}(\boldsymbol{\eta}))$, symmetric.

5. DISCUSSION

We have proposed a conditionally specified space-time model which provides both tractable computations and flexibility in specifying the spatial dependence structure while allowing for simple temporal dependence. This provides another class of models that can be considered for space-time data with either point or block sources.

The multi-pollutant example illustrated the potential gains from modeling multiple space-time processes jointly via the slightly increased predictive accuracy and greatly improved model fit (DIC). Larger gains would likely have been seen in the example by modeling pollutants that are more related than PM and ozone and by introducing some of the model expansions discussion in Sections 3.7 and 4.

In addition to "modeling" the pollutant field as an end in and of itself, the "output" from these models has other uses. For example, these multi-pollutant space-time models might be used as input to health effects models. However, the health effects data are usually at a different level of aggregation than the pollutant data. In particular, the pollutant data are point sources and the health effects data tend to be over areas. So, in some sense, an average exposure (pollutant level) for the area would be desired. This could be accomplished by creating a fine grid over the areas, constructing the appropriate $\mathbf{C}(\boldsymbol{\eta})$ matrix on this grid, predicting the pollutants at each point on the grid (Hrafnkelsson and Cressie 2003), and then averaging. Such an approach would avoid the change of support problem discussed recently by Gelfand et al. (2001). In addition, the posterior draws from the Gibbs sampler can be used to address the uncertainty of the pollutant exposure in the health effects models. Recent work on such modeling has been done by Van den Eeden (2001), and Zhu, Gelfand,

and Carlin (2003). For block sources, we might use these models to fill in spatial covariates for a regression model where some of the covariates (exposures) are missing in certain blocks.

In the LA data, weather variables were also collected. However, there was considerable missingness. These might be used in both the spatial model for the pollutants and the health effects models by fitting additional spatial models to the weather data to obtain a predicted weather field that could be used as covariates in both the pollutant and health effects models.

A criticism of conditionally specified models is that the marginal distribution of the original sites changes when we add in additional sites; that is, lack of marginal invariance. For block sources this is less of an issue than with point sources. However, the ultimate goal of the modeling may either be a good fit of the model to the data and/or good prediction; the model should be judged more on these criteria than the lack of marginal invariance property.

An alternative approach for prediction (see Section 2.5) would be to augment the $\mathbf{C}(\boldsymbol{\eta})$ matrix with the site(s) to be predicted, Y_0, and then to sample from the posterior distribution with the additional site(s) in the model, similar to the data augmentation approach used to "fill-in" intermittent missing data. This approach may potentially result in slightly different predictions and posterior distributions of the parameters since the marginal distribution of the observed sites is altered when adding in the site to be predicted. This was the approach used by Hrafnkelsson and Cressie (2003).

The fit of the covariance structure could be assessed via posterior predictive checks (Gelman et al. 2003). Such checks have been used recently to assess the fit of the correlation structure in complex hierarchical models (see, e.g., Ilk and Daniels 2005).

Future work will include doing many of the model expansions discussed here for a more definitive analysis of the pollutant data and making comparisons of such models with geostatistics models with a separable space-time structure. In addition, we will report on the propriety of the posterior under improper priors on the regression coefficients, $\boldsymbol{\beta}$, and the diagonal matrix of variances, $\mathbf{M}$, including derivation of default reference priors for the dependence parameters. We will also explore efficient sampling algorithms as the neighborhood/dependence structure gets more complex.

REFERENCES

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

Besag, J. E. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society*, Series B, 36, 192–225.

Best, N. G., Arnold, R. A., Thomas, A., Waller, L. A., and Conlon, E. M. (1999), "Bayesian Models for Spatially Correlated Disease and Exposure Data," in *Bayesian Statistics 6*, Oxford: Oxford University Press, pp. 131–156.

Brown, P. E., Diggle, P. J., Lord, M. E., and Young, P. C. (2001), "Space-Time Calibration of Radar Rainfall Data," *Applied Statistics*, 50, 221–241.

Carlin, B. P., and Banerjee, S. (2003), "Hierarchical Multivariate CAR Models for Spatially Correlated Survival Data," in *Bayesian Statistics 7*, Oxford: Oxford University Press, pp. 45–64.

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" (with discussion), *Journal of the American Statistical Association*, 92, 392–415.

Conlon, E., and Waller, L. (1998), "Flexible Neighborhood Structures in Hierarchical Models for Disease Mapping," University of Minnesota Biostatistics Research Report, No. rr98-018.

Cressie, N. A. C., and Chan, N. H. (1989), "Spatial Modeling of Regional Variables," *Journal of the American Statistical Association*, 84, 393–401.

Cressie, N., Kaiser, M. S., Daniels, M. J., Aldworth, J., Lee, J., Lahiri, S. N., and Cox, L. H. (1999), "Spatial Analysis of Particulate Matter in an Urban Environment," in *geoENV II—Geostatistics for Environmental Applications*, eds. J. Gomez-Hernandez, A. Soares, and R. Froidevaux, Dordrecht: Kluwer, pp. 41–52.

Dietrich, C. R. (1991), "Computationally Efficient Cholesky Factorization of a Covariance Matrix with Block Toeplitz Structure," *Journal of Statistical Computation and Simulation*, 45, 210–218.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Gelfand, A. E., and Vounatsou, P. (2003), "Proper Multivariate Conditional Autoregressive Models for Spatial Data Analysis," *Biostatistics*, 4, 11–25.

Gelfand, A. E., Zhu, L., and Carlin, B. P. (2001), "On the Change of Support Problem for Spatio-Temporal Data," *Biostatistics*, 2, 31–45.

Gelman, A., Carlin, J. B, Stern, H., and Rubin, D. (2003), *Bayesian Data Analysis* (2nd ed.), Boca Raton: Chapman and Hall/CRC.

Handcock, M. S., and Wallis, J. R. (1994), "An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields," *Journal of the American Statistical Association*, 89, 368–378.

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.

——— (2002), "New Systems for Estimating, and Predicting a Multivariate Spatio-Temporal Process," *Environmetrics*, 13, 311–332.

Hrafnkelsson, B., and Cressie, N. (2003), "Hierarchical Modeling of Count Data with Application to Nuclear Fall-Out," *Environmental and Ecological Statistics*, 10, 179–200.

Ilk, O., and Daniels, M. J. (2005), "Marginalized Transition Random Effects Models for Multivariate Longitudinal Binary Data," submitted.

Kaiser, M. S., Daniels, M. J., Furakawa, K., and Dixon, P. (2002), "Analysis of Particulate Matter Air Pollution Using Markov Random Field Models of Spatial Dependence," *Environmetrics*, 13, 615–628.

Kaiser, M. S., and Cressie, N. (2000), "The Construction of Multivariate Distributions from Markov Random Fields," *Journal of Multivariate Analysis*, 73, 199–220.

Kibria, B., Golam, M., Sun, L., Zidek, J.V., and 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.

Kim, H., Sun, D., and Tsutakawa, R. K. (2001), "A Bivariate Bayes Method for Improving the Estimates of Mortality Rates with a Two-Fold Conditional Autoregressive Model," *Journal of the American Statistical Association*, 96, 1506–1521.

Knorr-Held, L. (2000), "Bayesian Modeling of Inseparable Space-Time Variation in Disease Risk," *Statistics in Medicine*, 19, 2555–2567.

Knorr-Held, L., and Rue, H. (2002), "On Block Updating of Markov Random Field Models for Disease Mapping," *Scandinavian Journal of Statistics*, 29, 597–614.

Le, N. D., Sun, W., Zidek, J. V. (1997), "Bayesian Multivariate Spatial Interpolation with Data Missing by Design," 59, 501–510.

Mardia, K. V. (1988), "Multi-Dimensional Multivariate Gaussian Markov Random Fields with Application to Image Processing," *Journal of Multivariate Analysis*, 24, 265–284.

Mardia, K. V., Goodall, C., Redfern, E. J., and Alonso, F. J. (1998), "The Kriged Kalman Filter" (with discussion), *TEST*, 7, 217–285.

Neal, R. M. (1994), "An Improved Acceptance Procedure for the Hybrid Monte-Carlo Algorithm," *Journal of Computational Physics*, 111, 194–203.

Niu, X.-F., McKeague, I. W., and Elsner, J. B. (2003), "Seasonal Space-Time Models for Climate Systems," *Statistical Inference for Stochastic Processes*, 6, 111–133.

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119.

Shaddick, G., and Wakefield, J. (2002), "Modelling Daily Multivariate Pollutant Data at Multiple Sites," *Applied Statistics*, 51, 351–372.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society*, Series B, 64, 583–616.

Stein, M. L. (2003), "Space Time Covariance Functions," The University of Chicago, Center for Integrating Statistical and Environmental Science Technical Report, No. 4.

Stroud, J. R., Muller, P., and Sanso, B. (2001), "Dynamic Models for Spatiotemporal Data," *Journal of the Royal Statistical Society*, Series B, 63, 673–689.

Sun, D., Tsutakawa, R. K., and Speckman, P. L. (1999), "Posterior Distribution of Hierarchical Models using CAR(1) Distributions," *Biometrika*, 86, 341–350.

Van den Eeden, S. (2001), "Air Pollution and Morbidity and Mortality Among California Kaiser Permanente members," Kaiser Permanente Research Report.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998), "Hierarchical Bayesian Space-Time Models," *Environmental and Ecological Statistics*, 5, 117–154.

Wikle, C. K., and Cressie, N. (1999), "A Dimension-Reduced Approach to Space-Time Kalman Filtering," *Biometrika*, 86, 815–829.

Zhu, L., Gelfand, A. E., and Carlin, B. P. (2003), "Hierarchical Regression with Misaligned Spatial Data: Relating Ambient Ozone and Pediatric Asthma ER Visits in Atlanta," *Environmetrics*, 14, 537–557.

Zidek, J. V., Sun, W., and Le, N. D. (2001), "Designing and Integrating Composite Networks for Monitoring Multivariate Gaussian Pollution Fields," *Applied Statistics*, 49, 63–79.

Zidek, J. V., Sun, W., Le, N. D., and Ozkaynak, H. (2002), "Contending with Space-Time Interaction in the Spatial Prediction of Pollutions: Vancouver's Hourly ambient PM_{10} Field," *Environmetrics*, 13, 595–613.