Dynamic Factor Process Convolution Models for Multivariate Space-Time Data with Application to Air Quality Assessment

Catherine A. Calder
Department of Statistics
The Ohio State University
1958 Neil Ave.
Columbus, OH 43221

calder@stat.ohio-state.edu

Department of Statistics Preprint No. 750, The Ohio State University

March 1, 2005

Abstract

We propose a Bayesian dynamic factor process convolution model for multivariate spatial temporal processes and illustrate the utility of this approach in modeling large air quality monitoring data. Key advantages of this modeling framework are a descriptive parametrization of the cross-covariance structure of the space-time processes and dimension reduction features that allow full Bayesian inference procedures to remain computationally tractable for large data sets. These features result from modeling space-time data as realizations of linear combinations of underlying space-time fields. The underlying latent components are constructed by convolving temporally-evolving processes defined on a grid covering the spatial domain and include both trend and cyclical components. We argue that mixtures of such components can realistically describe a variety of space-time environmental processes and are especially applicable to air pollution processes that have complex space-time dependencies. In addition to computational benefits that arise from the dimension reduction features of the model, the process convolution structure permits misaligned and missing data without the need for imputation when fitting the model. This advantage is especially useful when constructing models for data collected at monitoring stations that have misaligned sampling schedules and that are frequently out of service for long stretches of time. We illustrate the modeling approach using a multivariate pollution dataset taken from the EPA's CASTNet database.

Key Words: Bayesian modeling, dimension reduction, factor analysis, misaligned space-time data, missing data

1 Introduction

Large multivariate space-time datasets are often collected to study or monitor environmental phenomena. For example, in the air quality field, monitoring networks are used to collect time-series data of atmospheric concentrations of multiple pollutants across a spatial region. These data are used for a variety of purposes such as determining compliance with air quality regulations, studying the acute and long-term effects of air pollution exposure, and evaluating environmental justice concerns. In some situations, the monitoring data may be spatially interpolated to estimate the pollutant levels at locations or time points where the atmospheric concentrations are not measured. In other situations, network data may be collected primarily for assessing global or regional trends in pollutant levels over time. The latter situation is typically the goal of analyses of data taken from networks with a broad spatial coverage but limited spatial resolution. These networks are built to provide representative information on pollutant levels over large spatial regions, as opposed to being designed with a fine enough resolution to permit accurate spatial interpolation. In addition, such networks usually collect readings at regular intervals of a length appropriate for assessing long-term trends ($i.e.$, weekly or quarterly) but are insufficient for accurate temporal interpolation.

In analyzing the type of network data described above, there are a number of complexities inherent in the data that prevent the use of standard exploratory statistical and graphical analyses as well as formal statistical modeling. For example, an initial exploratory graphical approach might produce time-series plots of the concentrations of the multiple pollutants at a monitoring site in the network. However, based on this type of graphical summary, it may be difficult to ascertain the complex multivariate/space/time relationships in network data. This type of graphical representation of network data is limited by the fact that these data are often *misaligned* (not collected at the same spatial locations over time or collected at different temporal resolutions). In addition, individual monitors often malfunction or otherwise cease to take readings for long stretches of time resulting in large amounts of *missing* data. These two issues also impair the use of standard model-based statistical analyses, which are further limited by the computational demands that result from the large quantity of data available for analysis. In terms of assessing spatial variation in long-term trends in pollutant levels, the non-uniformity of the spatial locations

of the monitoring sites may inhibit the ability of traditional space-time models to provide a spatially representative summary of temporal trends. Motivated by the need to address these issues, we present a model-based technique for analyzing large multivariate spatial temporal datasets that is amenable to the concerns of misaligned and missing data collected on a non-uniform monitoring network. Our proposed modeling approach incorporates a latent variable structure of reduced dimension with respect to the original data. This feature eases the computational demands of model-fitting, while simultaneously addressing concerns about the ability of data collected on a spatially non-uniform network to provide representative summaries of spatial and temporal trends.

Our latent variable model for multivariate space-time data is based on the discrete process convolution approach to modeling spatial data (Higdon, 1998). This framework underlies the dynamic process convolution model for space-time data (Calder et al., 2001; Higdon, 2002). We review these classes of models in Section 2 and highlight the features that motivate our multivariate extension, which is designed to provide efficient summaries of the relationship between multiple types of space-time data in terms of global or regional trends. Our multivariate extension, which we we term a dynamic factor process convolution model, is introduced in Section 3 and is compared to other standard approaches used to model multivariate space-time data. To illustrate the utility of the proposed model, we present a simulation study designed to highlight its strengths in Section 4. In addition, Section 5 discusses an analysis of a subset of data taken from the EPA's Clean Air Status and Trend Network (CASTNet) using our model where our goal is to efficiently report general temporal trends in atmospheric concentrations of five pollutants over a period of an 11 years across the eastern United States resulting from reductions in pollutant emissions required by the 1990 Clean Air Act Amendment.

2 Background

2.1 Gaussian Process Models for Spatial Data

Gaussian processes (GPs) are commonly used to model point-referenced spatial data. If a spatial process ψ defined over a domain $\mathcal{D}$ (typically a subset of $\mathcal{R}^2$) is a GP, then the distribution of all

finite subsets $\{\psi(\boldsymbol{s})\}$ for all $\boldsymbol{s} = (s_1, s_2, \ldots, s_N) \subset \mathcal{D}$ is multivariate normal. Consequently, a necessary condition is that the covariance matrix Σ associated with the process at any finite set of locations in $\mathcal{D}$ must be non-negative definite. When fitting GP models to data, this restriction makes the choice of the covariance function of the process cumbersome. Usually covariance functions for GP models of spatial data are chosen from a set of standard parametric families that ensure that the non-negative definite restriction holds (see Stein, 1999, for examples of such families of covariance functions). While the use of these standard parametric families allows the covariance structure of a GP to be specified directly, often the covariance functions available from these families may not be appropriate for modeling spatial data with non-standard, $i.e.$, non-stationary or anisotropic, covariance structures.

Process convolutions (PCs) are an alternative, constructive approach for specifying GPs that allow more flexibility in modeling the covariance structure than the direct specification (Thiébaux and Pedder, 1987; Barry and Ver Hoef, 1996). The constructive, as opposed to direct, nature of the covariance function modeling has led to the increased popularity of PC models for spatial modeling in recent years (see the examples in Higdon, 2002). In the PC approach, rather than defining the covariance function of the process directly, a GP is created by convolving a continuous white noise process x defined over $\mathcal{D}$ with a smoothing kernel κ so that

$$\psi(s) = \int_D \kappa_s(\omega) x(\omega) d\omega, \text{ for } s \in D, \tag{1}$$

where $\kappa_s(\omega)$ denotes the value at locations ω of the smoothing kernel centered at site s. When κ is taken to be stationary ($i.e.$, $\kappa_s(u) = \kappa(s-u)$ for all $s, u \in \mathcal{D}$), then resulting covariance function for ψ is given by

$$c(\Delta s) = Cov(\psi(s_1), \psi(s_2)) = \int_D \kappa(\omega - s_1)\kappa(\omega - s_2) d\omega = \int_D \kappa(\omega - \Delta s)\kappa(\omega) d\omega, \tag{2}$$

where Δs is the vector connecting locations s_1 and s_2. Further, it can be shown that if κ is isotropic ($i.e.$, $\kappa(\Delta s) = \kappa(\mid \Delta s \mid)$) and either κ is both integrable and square integrable or $c(\Delta s)$ is integrable and positive definite, then there is a one-to-one relationship between the covariance function $c(\cdot, \cdot)$ and the smoothing kernel κ implied by the convolution theorem for Fourier transforms (see Kern, 2000, for a proof).

The moving average construction of ψ guarantees that the implied covariance function of the GP is positive definite. As a result, this constructive specification leads to increased modeling flexibility as illustrated by Barry and Ver Hoef (1996), Ver Hoef and Barry (1998), and Kern (2000). However, a disadvantage of the constructive approach is the computational burden of performing the required numerical integration when fitting the a GP with a covariance structure specified using the PC representation to spatial data. However, Ver Hoef et al. (2004) showed that the Fast Fourier Transform (FFT) can be used to simplify implementation of the PC model.

In this remaining section, we do not emphasize the ability of PCs to allow flexibility in modeling spatial covariance functions. Instead, we focus on their utility in capturing temporal features of space-time processes. In our simulation study and application, we rely on simple circular Gaussian kernels whose parameters are assumed constant across space. This assumption is reflected in the notation used in the following sections. While our models could easily incorporate different types of kernels, we have chosen to focus our efforts primarily in the temporal domain rather than on exploring kernel choice.

2.2 The Discrete Process Convolution Model

Higdon (1998) proposed a model for spatial data based on a discretized version of Equation 1, as opposed to using the discretization for numerical integration purposes. In this approach, a discrete white noise process is defined at a finite set of locations on a coarse grid $\{\omega_1, \omega_2, ..., \omega_M\}$ that covers $\mathcal{D}$, but that does not fill $\mathcal{D}$. This discrete white noise process is smoothed by the kernel κ to create a continuous spatial process, which we refer to as a discrete PC. For any $s \in \mathcal{D}$, this process $\tilde{\psi}$ can be written so that

$$\tilde{\psi}(s) = \sum_{i=1}^{M} \kappa(\omega_i - s)x(\omega_i), \tag{3}$$

where $x(\omega_i) \overset{\text{i.i.d.}}{\sim} N(0, \lambda_x)$ and M is the number of locations where the underlying process is defined. Then for any set of locations $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)$ in $\mathcal{D}$,

$$\tilde{\psi}(\boldsymbol{s}) = \mathbf{K}\boldsymbol{x}, \tag{4}$$

where $\boldsymbol{x} = [x(\omega_1), x(\omega_2), ..., x(\omega_M)]'$ and $\mathbf{K}$ is a matrix with rows $\mathbf{K}(n) = [\kappa(\omega_1 - s_n), \kappa(\omega_2 - s_n), ..., \kappa(\omega_M - s_n)]$.

Typically, point-referenced spatial data are modeled as the sum of non-spatial mean term, an underlying spatial process (typically a GP), and a measurement error process that is taken to be spatially independent. For simplicity, we assume that the mean is equal to zero across all spatial locations, although this assumption will be relaxed later. When the underlying spatial process component is defined using a discrete PC, for data observed at N spatial locations, the resulting model can be written as

$$\begin{aligned} y(s_n) &= \tilde{\psi}(s_n) + \epsilon(s_n) \\ &= \sum_{i=1}^{M} \kappa(\omega_i - s_n) x(\omega_i) + \epsilon(s_n), \end{aligned} \tag{5}$$

where $y(s_n)$ is the observed value at location s_n. As above, we assume that the $x(\omega_i)$s are independent and have a $N(0, \lambda_x)$ distribution. The measurement error terms are assumed to be normally distributed with mean 0 and variance λ_ϵ independently across the spatial locations. For a more detailed discussion of implementing spatial models based on discrete PCs, including issues such as determining the appropriate resolution of the coarse grid and the form of the kernel, see Kern (2000). An advantage of using the discrete PC specification of an underlying spatial process over using a standard GP is that the number of spatial locations where the underlying process is defined does not depend on the number of locations where data are observed. This leads to computational advantages in model-fitting in that the dimension of the spatial component of the model is reduced from N to M. Discrete PC models have been successfully applied to data arising in a variety of fields of application (see for example Higdon, 1998; Higdon et al., 1998; Higdon, 2002; Lee et al., 2005).

2.3 Dynamic Process Convolution Models

The discrete PC approach to modeling spatial data can be extended to the space-time arena in two ways. One approach used by Higdon (1998), is to convolve a three-dimensional smoothing kernel with a three-dimensional discrete underlying process $x(\omega_{i,t})$ defined at M two-dimensional spatial

locations at T time points. In this approach, the $x(\omega_{i,t})$s are assumed to be normally distributed with mean 0 and variance λ_x and to be independent across space and time. A disadvantage of this natural extension to the spatial discrete PC model is that inference procedures are computationally challenging for large values of T. To get around this problem, simplifying assumptions can be made about the form of the smoothing kernel. For example, κ can be chosen to have a separable form, $i.e.$,

$$\kappa(\Delta s, \Delta t) = \kappa_s(\Delta s)\kappa_t(\Delta t).$$

However, this restriction greatly reduces the flexibility of the space-time covariance structure in that the covariance function of the resulting space-time process is also separable (Higdon, 1998).

Alternatively, a temporal component can be incorporated into the discrete PC model by allowing the latent x process defined on a three-dimensional lattice to be temporally dependent. A space-time process that is continuous in space, but discrete in time, can be constructed by convolving the x process with a two-dimensional smoothing kernel at each time point. We use the term dynamic process convolution model (DPC) to describe this type of space-time model and argue that this temporal extension of the spatial discrete PC modeling approach is often sufficient in practice since space-time data are typically observed at fixed time steps and rarely is temporal interpolation the goal of a statistical analysis of such data. If the covariance of the x process in a DPC is not space-time separable, then it is clear that the covariance structure of the continuous spatial process created by convolving the latent x process at each time point is also not separable. In most situations where a non-separable space-time covariance structure is required, modeling the space-time dependence structure of a temporal process on a two-dimensional lattice is much more straightforward than specifying an appropriate three-dimensional kernel to convolve an independent, three-dimensional x process or than specifying the space-time covariance function directly.

Another motivation for the use of DPC models is that they can be specified as a multivariate state-space model, which can capture a variety of types of temporal dependence structures and can allow Kalman filtering-type algorithms to facilitate model-fitting when the T is large (Kalman, 1960). In introducing the state-space version of a dynamic process convolution model, we assume the space-time process is observed at N spatial locations at each time point, however, this is by no

means a necessary assumption. In this approach, the observation at location s_n at time t is modeled as

$$y(s_n, t) \quad = \mathbf{K}(n)\boldsymbol{x}_t + \epsilon_{n,t}, \tag{6}$$

such that the underlying process evolves as

$$\boldsymbol{x}_t \quad = G(D_{t-1}) + \boldsymbol{\nu}_t, \tag{7}$$

where $\boldsymbol{x}_t$ is the vector of the M values of the latent space-time process at time t and $\mathbf{K}(n)$ is the n^{th} row of the matrix $\mathbf{K}$. The function $G(\cdot)$ controls the evolution of the x process; D_t is defined to be the set of values of the x process at all time points up to and including time t, $i.e.$, $D_t = \{x(\omega_{1,\tau}), x(\omega_{2,\tau}), \ldots, x(\omega_{M,\tau}) : \tau \leq t\}$; the distribution of the measurement error term $\epsilon_{s,t}$ is assumed to be $N(0, \lambda_\epsilon)$; and the distribution of the process or evolution error term $\boldsymbol{\nu}_t$ is taken to be $N(0, \boldsymbol{I}_M \otimes \lambda_\nu)$, where $\boldsymbol{I}_M$ is an $M \times M$ identity matrix. To complete the state-space model specification, the distribution of $\boldsymbol{x}_0$ is assumed to be $N(\boldsymbol{m}_0, \boldsymbol{C}_0)$. Figure 1 illustrates the conditional independence structure of the DPC model. The spatial process at time t is constructed by convolving the underlying process on a lattice. Moving from time t to time $t+1$, the underlying process evolves according to the evolution equation (Equation 7). Then the new underlying process at time $t+1$ is smoothed to create the continuous spatial process at time $t+1$. Therefore, conditional on the values of the underlying x process, the smooth spatial fields are independent over time.

The state-space dynamic PC model defined by Equations 6 and 7 falls into the general class of space-time models known as reduced-dimension space-time Kalman filters as defined by Cressie and Wikle (2002). The reduced-dimension feature of this class of models refers to the reduction in the spatial dimension of the latent space-time component of the model. In our case, the dimension is reduced from N to M. An advantage of the DPC model over other reduced-dimension space-time Kalman filters, such as those based on empirical orthogonal functions ($e.g.$, Wikle et al., 2001), is that they allow a spatially-varying temporal dependence structure. Besides the computational advantages that result from the reduction in the dimension of the underlying process and the use of the Kalman filtering-type algorithms, these models also have the benefit that they permit descriptive features to be included in the evolution of the underlying x process. For example,

the temporal evolution can be based on physical assumptions about the temporal evolution of the process as advocated by Wikle et al. (1998). An example of a physically-based evolution structure in a state-space DPC model can be found in Xu et al. (2003) where a partial differential equation motivates the form of the evolution of the latent process. Another descriptive approach to modeling the temporal evolution of the xs is to allow the evolution function $G(\cdot)$ to depend on unknown parameters. By using a parametric evolution function, estimates of the evolution parameters can efficiently summarize the space-time dependence structure of the observed process. This idea was used by Calder et al. (2001) in a Bayesian state-space DPC model with an evolution form composed of parameters that describe the directional space-time dependence of ozone over the eastern United States.

Bayesian versions of state-space DPC models can readily incorporate dynamic linear modeling (DLM) tools used in time-series analysis (West and Harrison, 1997) into models for spatio-temporal data. DLM-based space-time models have been used to analyze processes where the spatial covariance structure is specified directly. For example, Huerta et al. (2004) used DLM tools, including autoregressive regression coefficients and variance discounting methods for specifying error variances, in a model for relating temperature readings and ozone levels across Mexico City. The multivariate state-space DPC model proposed in the next section for exploring the cross-covariance structure of multiple space-time processes is based on the DLM approach to constructing time-series models as sums of latent polynomial trend and cyclical components described in West (1995).

3 Dynamic Factor Process Convolution Models

3.1 General Framework

When analyzing multivariate space-time data collected to monitor environmental phenomena, often the goal is provide insight into the dependence between the multiple types of data and reveal global trends in the space-time processes that generated the data. The goal in this situation differs from the motivation underlying most spatial models where accurate spatial prediction or interpola-

tion is typically the main priority. Multivariate spatial models that are known to have good predictive abilities such as linear models of coregionalization (Wackernagel, 1998; Gelfand et al., 2002) and shift-factor models (Christensen and Amemiya, 2001, 2002) have difficulty accommodating misaligned or large amounts of missing space-time data and can be computationally prohibitive for large datasets. In addition, these models are often not able to reveal or summarize important global features of multivariate space-time processes through inference on unknown model parameters. For example, when the goal of a statistical analysis is to assess the overall effect of legislation limiting pollutant emissions on atmospheric concentrations of primary and secondary pollutant compounds, alternative models are needed.

The multivariate models developed in this section are based on the Bayesian state-space DPC framework. In constructing these models, we assume that they will be used to assess global or regional (as opposed to local) trends in multiple space-time processes and to provide an explanation for the observed cross-type dependences. Since accurate spatial prediction is not a priority, we can take the number of locations where the latent space-time processes are defined to be M where $M << N$. The resolution of the grid where the underlying process is defined and the width of smoothing kernels can be chosen as a compromise between the spatial resolution needed to assess spatial variation in temporal trends and the computational limitations associated with implementation of the model-fitting algorithms.

In light of our modeling goals, we define $J = J_1 + J_2$ independent underlying processes on the M grid locations. The first J_1 underlying processes are assumed to evolve as spatially independent random walks. This implies that for $j = 1, 2, \ldots, J_1$,

$$\boldsymbol{x}_t^{(j)} = \boldsymbol{x}_{t-1}^{(j)} + \boldsymbol{\nu}_t^{(j)}, \tag{8}$$

where $\boldsymbol{\nu}_t^{(j)} \sim N(\mathbf{0}, \boldsymbol{I}_M \otimes \lambda_\nu^{(j)})$. The next J_2 underlying processes are taken to be spatially independent cyclic second-order autoregressions. For $j = J_1 + 1, J_1 + 2, \ldots, J$,

$$\boldsymbol{x}_t^{(j)} = (\boldsymbol{I}_M \otimes \beta^{(j)})\boldsymbol{x}_{t-1}^{(j)} - \boldsymbol{x}_{t-2}^{(j)} + \boldsymbol{\nu}_t^{(j)}, \tag{9}$$

where $\boldsymbol{\nu}_t^{(j)} \sim N(\mathbf{0}, \boldsymbol{I}_M \otimes \lambda_\nu^{(j)})$. Due to the second-order dependence of the cyclic components in Equation 9, $\boldsymbol{x}_{-1}^{(j)}$ in addition to $\boldsymbol{x}_0^{(j)}$ (the values of the cyclical underlying processes at times 0

and -1) must be given a prior distribution; as with the prior assumption on the latent process at $t = 0$, we assume that all of the $\boldsymbol{x}_{-1}$s are *a priori* $N(\mu_{-1}, C_{-1})$. Therefore, the first J_1 underlying components are first-order polynomials which can describe a variety of acyclic, smooth temporal trends. The next J_2 components have a cyclic structure with a period determined by $\beta^{(j)}$. As discussed in West (1995), for $\beta^{(j)} = \beta$, the cyclic component has a period of $2\pi/\cos^{-1}(\beta/2)$ and the $\beta^{(j)}$s can be treated as unknown parameters. However, in most environmental applications, yearly or seasonal cycles are most prevalent, so we can often assume these parameters to be known.

Underlying space-time processes with evolution specifications given by Equations 8 and 9 can be embedded in a multivariate state-space framework. If we define $\boldsymbol{y}_t$ to be an $N \times I$ matrix corresponding to the I types of data observed at N locations (where again the assumption that each of the I data types are observed at each spatial location at each time point can easily be relaxed) and the $M \times J$ matrix $\boldsymbol{x}_t = [\boldsymbol{x}_t^{(1)}, \boldsymbol{x}_t^{(2)}, \cdots, \boldsymbol{x}_t^{(J)}]$, we can extend the model given by Equation 6 as

$$\boldsymbol{y}_t = \mathbf{K}\boldsymbol{x}_t\boldsymbol{F}' + \boldsymbol{\epsilon}_t, \tag{10}$$

where $Vec(\boldsymbol{\epsilon}_t) \sim N(\mathbf{0}, \boldsymbol{I}_N \otimes diag(\boldsymbol{\lambda}_\epsilon))$ such that $\boldsymbol{\lambda}_\epsilon = (\lambda_\epsilon^{(1)}, \lambda_\epsilon^{(2)}, \ldots, \lambda_\epsilon^{(I)})$ and where the operator $Vec(\cdot)$ creates a vector by stacking the columns of its argument. This model for multivariate space-time data can be written in the form of a multivariate state-space model by converting the matrices in Equation 10 to vectors so that

$$Vec(\boldsymbol{y}_t) \;\; = (\mathbf{K} \otimes \boldsymbol{F})Vec(\boldsymbol{x}_t) + Vec(\boldsymbol{\epsilon}_t), \tag{11}$$

and converting the matrices in Equations 8 and 9 to vectors as well as combining the evolution equations for the two types of underlying processes allows Equations 8 and 9 to be written as

$$Vec(\boldsymbol{x}_t) = (\boldsymbol{I}_M \otimes \boldsymbol{G})Vec([\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}]) + Vec(\boldsymbol{\nu}_t). \tag{12}$$

In this joint specification of the evolution of both types of underlying components, the distribution of the evolution error term $Vec(\boldsymbol{\nu}_t)$ is $N(\mathbf{0}, \boldsymbol{I}_M \otimes diag(\boldsymbol{\lambda}_\nu))$ where $\boldsymbol{\lambda}_\nu = (\lambda_\nu^{(1)}, \lambda_\nu^{(2)}, \ldots, \lambda_\nu^{(J)})$. The matrix $\boldsymbol{G}$ is defined based on the evolution structure of the J underlying processes given by Equations 8 and 9. Combining these two evolution models implies that $\boldsymbol{G} = [\boldsymbol{G}^{(1)}, \boldsymbol{G}^{(2)}]$ where

$\boldsymbol{G}^{(1)}$ is a $J \times J$ diagonal matrix with diagonal elements $(1, 1, \ldots, 1, \beta^{(J_1+1)}, \beta^{(J_1+2)}, \ldots, \beta^{(J)})$ and $\boldsymbol{G}^{(2)}$ is also a $J \times J$ diagonal matrix with the first J_1 diagonal elements equal to zero and the last J_2 equal to -1.

In order for the model defined by Equations 11 and 12 to be identifiable, constraints must be placed on the matrix $\boldsymbol{F}$, which links the J types of underlying processes to the I data types. While a number of possible constraints are available, we choose to use a set of restrictions based on those used by Aguilar and West (2000) in their dynamic factor models for multiple time-series. In our dynamic factor PC model, the matrix $\boldsymbol{F}$ in Equation 11 will be of the form $\boldsymbol{F} = [\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}]$ where

$$\boldsymbol{F}^{(1)} = \begin{bmatrix} 1 & 0 & \ldots & & 0 \\ f_{2,1}^{(1)} & 1 & 0 & \ldots & 0 \\ \vdots & & & & \vdots \\ f_{J_1,1}^{(1)} & \ldots & & & f_{J_1,J_1}^{(1)} \\ \vdots & & & & \vdots \\ f_{I,1}^{(1)} & \ldots & & & f_{I,J_1}^{(1)} \end{bmatrix} \text{ and } \boldsymbol{F}^{(2)} = \begin{bmatrix} 1 & 0 & \ldots & & 0 \\ f_{2,1}^{(2)} & 1 & 0 & \ldots & 0 \\ \vdots & & & & \vdots \\ f_{J_2,1}^{(2)} & \ldots & & & f_{J_2,J_2}^{(2)} \\ \vdots & & & & \vdots \\ f_{I,1}^{(2)} & \ldots & & & f_{I,J_2}^{(2)}. \end{bmatrix}.$$

The elements of $\boldsymbol{F}^{(1)}$ and $\boldsymbol{F}^{(2)}$ denoted by f are free parameters in the model, and the other components of $\boldsymbol{F}$ are restricted to be either one or zero. Based on the results in Aguilar (1998), this restriction on $\boldsymbol{F}$ insures that the likelihood for the data implied by Equation 11 is invariant under invertible linear transformations of the underlying factor processes and that $\boldsymbol{F}$ is of full rank. An artifact of assuming this form for $\boldsymbol{F}$ is that an interpretation of the latent space-time processes is implied since the first data type is associated with only first trend and the first cyclical component of the model, the second data type is modeled as a linear combination of the first trend and the first cyclical component plus the second trend and second cyclical component, etc. As a result of this artifact, the order of the data types influences the underlying factors, and therefore, the data types should be ordered in such a way that the first J data types explain as much of the variation in the I total data types as possible. Arranging the I data types can be done by performing principal components decompositions on subsets of the original data or can be chosen based on scientific understanding of the relationship between the I space-time processes. Since our modeling goals are primarily exploratory in nature, fitting various models using different orderings of the data types may reveal additional information about the dependence structure of the processes. This strategy

of fitting several models is often feasible even with large datasets because of the dimension reduction features of the model resulting from the fact that $M << N$ and that J is usually chosen to be less than I. Similar to the issue of ordering the data types, the choice of the appropriate number of factors J can be based on initial exploratory analyses or on scientific understanding, and we believe that fitting models with different numbers of factors can be illuminating. Alternatively, the appropriate number of factors can be chosen using an information criteria approach or by allowing $J^{(1)}$ and $J^{(2)}$ to be unknown model parameters (see Lopes and West, 2004, for a discussion of various approaches for determining the number of factors in Bayesian factor analysis). However, since predictive ability is not of highest priority and due to the exceptional computational challenges associated with fitting the number of factors that would likely be prohibitive in our situation, we argue that formally determining the best value for J is not a necessity.

A mean component can be added to the observation equation (Equation 11) to allow covariate information to be incorporated into the model or to allow the data of type i, where $i > \max(J_1, J_2)$, to be modeled as linear combinations of the latent processes plus an overall mean level specific to type i, which may be appropriate in certain applications. In the latter situation, the data types i where $i \leq \max(J_1, J_2)$ do not need to have a unique mean parameter since the underlying processes fit their mean. If the mean component for the other data types are modeled, Equation 11 can be altered so that

$$Vec(\boldsymbol{y}_t) \quad = (\mathbf{K} \otimes \boldsymbol{F}) Vec(\boldsymbol{x}_t) + Vec(\boldsymbol{I}_N \otimes \boldsymbol{\mu}) + Vec(\boldsymbol{\epsilon}_t) \tag{13}$$

where $\boldsymbol{\mu} = (0, 0, \ldots, 0, \mu_{\max(J_1,J_2)+1}, \mu_{\max(J_1,J_2)+2}, \ldots, \mu_I)'$.

Finally, as we mentioned before, the restriction of a constant smoothing kernel matrix $\mathbf{K}$ is not necessary. In fact, the ability to allow $\mathbf{K}$ to vary with time is an important feature of the DPC framework. If $\mathbf{K}$ depends on time, missing and misaligned data can be handled in a straightforward manner, without a need for impution. The underlying process at time t can be related to only the data that are actually observed at time t. If, for example, there was one location where an observation was not taken at a particular time point, we can simply adjust the $\mathbf{K}$ matrix by deleting the corresponding row. In the same way, the coefficent $(\mathbf{K} \otimes \boldsymbol{F})$ relating the underlying processes to the observed values in Equation 11 can be temporally varying. This extension would allow the

model to handle situations when some types of data at a particular location are observed, but others are not. Additionally, by adjusting this coeffficient, we can allow the form of the smoothing kernel to depend on time or data type. When spatial prediction is a priority, such careful modeling of this term can be especially fruitful. However, since we are predominently interested in capturing global trends in space-time processes in our situation, we only adjust the $(\mathbf{K} \otimes \boldsymbol{F})$ matrix to accomodate missing observations.

3.2 Posterior Structure and Simulation-Based Model Fitting

The Bayesian DPC model requires the specification of priors on the unknown parameters. We assume diffuse, conjugate priors for the static model parameters. This assumption corresponds to independent normal distributions for the free parameters in the $\boldsymbol{F}$ matrix, which we denote $\boldsymbol{F}^{free}$, and the free parameters of $\boldsymbol{\mu}$, which we denote $\boldsymbol{\mu}^{free}$. The components of the observation and error variances, $\boldsymbol{\lambda}_\epsilon$ and $\boldsymbol{\lambda}_\nu$, respectively, are given inverse gamma prior distributions. For the the dynamic parameters, the $\boldsymbol{x}_t$s, prior distributions only need to be specified for the initial states, since the prior distributions on the other states are implied by the evolution structure given in Equation 12. We assume vague but proper independent normal priors for the components of $\boldsymbol{x}_0^{(j)}$ for $j = 1, 2, \ldots, J$ and for the components of $\boldsymbol{x}_{-1}^{(j)}$ for $j = J_1 + 1, J_1 + 2, \ldots, J$. Thus, the joint prior distribution of all unknown parameters can be written as

$$\begin{aligned} p(\boldsymbol{F}^{free}, \boldsymbol{\mu}^{free}, \boldsymbol{\lambda}_\epsilon, \boldsymbol{\lambda}_\nu, \boldsymbol{x}_0, \boldsymbol{x}_{-1}) =& p(\boldsymbol{F}^{free}) p(\boldsymbol{\mu}^{free}) \left(\prod_{i=1}^{I} p(\boldsymbol{\lambda}_\epsilon^{(i)}) \right) \\ & \left(\prod_{j=1}^{J} p(\boldsymbol{\lambda}_\epsilon^{(j)}) p(\boldsymbol{x}_0^{(j)}) \right) \left(\prod_{j=J_1+1}^{J} p(\boldsymbol{x}_{-1}^{(j)}) \right), \end{aligned} \tag{14}$$

where $\boldsymbol{x}_0 \doteq \{\boldsymbol{x}_0^{(j)} : j = 1, 2, \ldots, J\}$ and $\boldsymbol{x}_{-1} \doteq \{\boldsymbol{x}_{-1}^{(j)} : j = J_1 + 1, J_1 + 2, \ldots, J\}$.

The joint posterior distribution of parameters in the DPC is equal

$$p(\boldsymbol{F}^{free}, \boldsymbol{\mu}^{free}, \boldsymbol{\lambda}_\epsilon^*, \boldsymbol{\lambda}_\nu^*, \boldsymbol{x}_0^*, \boldsymbol{x}_{-1}^*, \boldsymbol{x}^* | \boldsymbol{y}^*)$$

where $\boldsymbol{x}^* \doteq \{\boldsymbol{x}_t : t = 1, 2, \ldots, T\}$ and $\boldsymbol{y}^* \doteq \{\boldsymbol{y}_t : t = 1, 2, \ldots, T\}$. We use a simulation-based approach to analyzing this posterior distribution. In particular, we employ a Gibbs sampling algorithm (Gelfand and Smith, 1990) to iteratively sample from the full conditional distributions of the

model parameters, and we base our inferences of the marginal posterior distribution represented by the simulated values of the parameters. When implementing our Gibbs sampler, block sampling is employed to facilitate mixing of our Markov chain. For each of the static parameters, we sample from the joint full conditional distribution of all components of the parameters. The dynamic parameters ($i.e.$, $\boldsymbol{x}_{-1}$, $\boldsymbol{x}_0$, and $\boldsymbol{x}^*$) can also be sampled jointly. While this step involves sampling from a high-dimensional ($M \times K \times T$) multivariate normal distribution, the covariance matrix of this distribution is quite sparse and numerical techniques such as the conjugate gradient algorithm for matrix inversion and sparse matrix routines can greatly facilitate the computational effort required to sample from this distribution. We also note that this step of our model fitting strategies is where the benefits of dimension reduction come into play. If our model had an underlying structure based on traditional multivariate space-time models that do not have dimension reduction features, then corresponding dimension of the $\boldsymbol{x}$ components would be $N \times I \times T$. Thus, the dimension reduction features of our model have the potential to reduce the order of the model-fitting algorithm by several orders of magnitude compared to competing approaches.

Despite the computational advantages afforded by dimension reduction techniques, sampling from the joint conditional distribution of the $\boldsymbol{x}$s may not always be possible. This situation arises primarily when T is large, $i.e.$, when T is on the order of several hundreds or thousands. When using the dynamic factor PC models to analyze multivariate space-time process with large T, sampling from the full conditional distribution of $\boldsymbol{x}$ can be performed using the Forward Filtering Backward Smoothing (FFBS) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). The FFBS algorithm makes use of the Kalman filtering recursions to facilitate sampling from the joint distribution of the dynamic, or state, parameters in a Bayesian state-space models. The original version of the FFBS algorithm was developed for state-space models with a first-order evolution structure, $i.e.$, the prior distribution of the state parameter at time t, depends only on the state parameters at time $t-1$. However, the algorithm can be extended to include higher-order lags in the evolution equation as illustrated by West (1995). Finally, we note that multivariate FFBS algorithms are amenable to situations where the dimension of the data changes with t, so they are applicable when using the dynamic factor PC to analyze multivariate space-time data with missing

observations.

4 Simulation Study

In the process of developing and testing our Gibbs sampling algorithm, several simulation studies were performed in which data was simulated based on assumed 'true' values of the model parameters, and the posterior distribution of the model parameters resulting from fitting the model to the simulated data were compared to the 'true' parameter values. These simulations were used to assess the ability of our model to find global temporal trends in multivariate space-time processes. In addition, they helped assess convergence rates of our Gibbs sampling algorithm, sensitivity to prior assumptions, as well as consequences of model misspecification ($e.g.$, incorrect number of factors or ordering of the data types). With regard to this last concern, our model appears to be fairly robust to the ordering of the data types so long as types that account for most of the variation in the original data are near the top of the ordering, but not necessarily in the order in which they were simulated. If the number of factors is taken to be less than the number of factors used to generate the data, in most situations our model is still able to pick out some of the factors. In addition, we found that the fitted factor processes could be used to diagnose model misspecification. If fitted trend components showed some cyclic behavior or the cyclic components showed a trend, then the model benefited from being refit with more factors. Similarly, if the number of factors was chosen to be greater than the 'true' number of factors, some of the fitted factor processes either appear to be the same as other factors or they appeared to be noise. Thus, given the exploratory nature of the dynamic factor PC framework, we feel that it can be used to successfully gain insight into the dependence structure of large multivariate space-time processes.

In Figure 2, we present the results from one of our simulation analyses. The plot in the top left corner shows the locations of the simulated data points along with the locations of the underlying process (represented as '+'s) and one standard deviation ellipses of the smoothing kernels. The data at each location were taken to be spatially varying linear combinations of three time-series (one linear trend, one trend that linearly decreased for the first 30 timesteps and linearly increased for the remaining 30 timesteps, and a sinusoidal curve with a period of 10). The time-series at the data

locations represented as squares in the top left plot are shown in the top right plot. While the data types show some dependence across the different locations and types, it is not readily apparent how these different series are related. The bottom plot shows the posterior distributions of the three latent factor processes. The solid lines correspond to the mean (across the 9 locations of the underlying process) of the posterior means of each factor process at each time point. The dashed lines correspond to the 25^{th} and 75^{th} percentiles of the posterior means of the factors across the 9 underlying locations. Reordering the data types did not dramatically affect these plots, except for the reordering of the first two trend factors. As expected, the fitted values of the free parameters in $\boldsymbol{F}$ were sensitive to the orderings of the original data types. However, given both the fitted factor processes and the fitted $\boldsymbol{F}$ matrix, similar conclusions could be drawn about the dependencies between the different types of space-time data.

5 Air Quality Assessment Example

In this section, we use the dynamic factor PC model to analyze atmospheric concentrations of five pollutants: sulfur dioxide (SO_2), nitric acid (HNO_3), particulate sulfate (SO_4^{2-}), particulate nitrate (NO_3^-), and particulate ammonium (NH_4^+). These concentration measurements were taken from a subset of data collected by the Environmental Protection Agency's (EPA) Clean Air Status and Trends Network (CASTNet). This network was established to monitor the effectiveness of emissions reductions required by the 1990 Clean Air Act, and according Lavery et al. (2002) one of the primary roles of the network was to detect and quantify trends in pollutants.

We focus our analysis on the levels of the five pollutants collected at 55 sites across the eastern United States shown in Figure 3. We model the $\log$ of the measured concentrations collected on a weekly basis from January 1990 through December 2000 (572 time points). An example of the data collected at a single location is shown in Figure 4. These observations were taken from a monitor located on the border of Indiana and Ohio (denoted by a triangle in the map shown in Figure 3). Each of the pollutant series exhibits a yearly cycle and some of the pollutants show fairly minor decreasing trends. This monitor site differs from many of the other sites in CASTNet in that it has few missing observations. Other monitors have much less temporal coverage due to the fact

that they were added to the network sometime after 1990 or taken out of the network sometime before 2000. In addition, monitor malfunctions caused data to not be collected for periods of time at several sites. Examples of sulfur dioxide concentration levels collected at two sites with less temporal coverage are shown in Figure 5. The location of the site corresponding to the top graph is depicted as a diamond in Figure 3 and a neighboring site corresponding to the the bottom graph is represented by a square. In terms of the top graph, assessing temporal trends is impeded by the large amounts of missing data. Our hope is that by borrowing strength from neighboring sites that we will be able to use the information in the CASTNet data to gain a better understanding of pollutant trends.

In using the dynamic factor PC model to analyze the CASTNet data, we focus our attention on models with up to three trend components and one cyclic component, based on scientific knowledge about the relationship between the five pollutants. Atmospheric concentrations of the pollutants are primarily driven by SO_2, NO_X, and NH_3 emissions and weather variables as depicted in Figure 6. Based on this understanding, we arranged the series so that: type 1 - SO_2, type 2 - HNO_3, type 3 - NH_4^+, type 4 - SO_4^{2-}, and type 5 - NO_3^-.

Prior distributions were assumed to be of the forms discussed in Section 3.2. The vague normal priors were taken to be $N(0, 10^6)$, and the inverse gamma priors were assumed to be $IG(10^{-3}, 10^{-3})$, where the inverse gamma distribution is parameterized so that a random variable that is distributed as $IG(a, b)$ has mean a/b. As a consequence of the large amounts of missing data, we chose to fix the components of the evolution error variances, the $\boldsymbol{\lambda}_\nu^{(j)}$s, to 0.01. We determined that our results were not very sensitive to this assumption based on comparisons with the results given with these parameters set to 0.1 and 0.001. Despite the dimension reduction resulting from modeling the underlying process at 34 locations as opposed to the 55 locations where the data were observed, we fit the modeling using the FFBS algorithm discussed in Section 3.2 to sample from the full conditional distributions of the $\boldsymbol{x}$s rather than sampling from their joint distribution directly. Finally, the smoothing kernels were chosen to be circular normal kernels with a standard deviation of $3\sqrt{2} \approx 4.25$ degrees.

After examining models with up to three trend components and one cyclical component, we

found that one trend and one cyclic component were sufficient for explaining most of the variation in pollutant data, $i.e.$, adding additional trend factors appeared to be fitting noise as opposed to signal in the data. As in the simulation study, Figure 7 summarizes the posterior distributions of the two factors by presenting the posterior means and 25^{th} and 75^{th} percentiles (averaged across the 34 locations of the underlying process) for the two underlying factors. The posterior means of the free parameters in the $\boldsymbol{F}$ matrix and the components of the observation error variance are

$$\hat{\boldsymbol{F}}^{(1)} = \begin{bmatrix} 1 \\ 0.311\ (0.300, 0.324) \\ 0.149\ (0.135, 0.163) \\ 0.663\ (0.642, 0.680) \\ -0.107\ (-0.136, -0.080) \end{bmatrix} \quad \hat{\boldsymbol{F}}^{(2)} = \begin{bmatrix} 1 \\ -0.642\ (-0.674, -0.610) \\ -0.854\ (-0.887, -0.821) \\ -1.352\ (-1.392, -1.314) \\ 1.747\ (1.681, 1.815) \end{bmatrix}$$

and

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_\epsilon =& (0.290\ (0.285, 0.296), 0.397\ (0.387, 0.407), 0.338\ (0.328, 0.349), 0.183\ (0.177, 0.190), \\ & 1.61\ (1.566, 1.66)). \end{aligned}$$

These results indicate that over the 11-year period, sulfur dioxide levels have decreased as captured by the first factor. In each of the plots of the CASTNet data we include a vertical line representing the beginning of 1995, which is the time when emission reduction mandates went into effect. It is evident that these mandates had an effect on the atmospheric concentrations of the pollutants as the time interval corresponding to the most dramatic decline in the trend factor is centered at 1995. As we would expect, SO_2 and SO_4^{2-} are most heavily related to the underlying trend factor. HNO_3 and NH_4^+ show a weaker decline, and NO_3^- levels remained fairly constant over this time period (since the 95 percent credible interval for $f_{5,1}^{(1)}$ nearly covers 0). In terms of the cyclic nature of the pollutants, each type of pollutant showed either a positive or negative relationship with the cyclic factor as a result of their opposite yearly cycles.

6 Discussion

In our analysis of the multivariate space-time data taken from the CASTNet, we have demonstrated the ability of the dynamic process convolution framework to extract global trends in spatially dependent data. Since our analysis was driven by a desire to provide understanding and summarize the data, we have chosen to present our results by highlighting the latent factor processes and their levels of uncertainty. We readily admit that our model is not geared toward spatial interpolation or capturing interesting spatial features of the data such as non-stationarity or anisotropy. While the framework could in fact allow more precise spatial modeling, this would come with the expense of increased computational effort. Since CASTNet is designed to inform about general trends in atmospheric concentrations of pollutants resulting for widespread emissions reduction, we believe that the dynamic process convolution modeling framework is appropriate.

Besides the ease of handling missing and misaligned data, another benefit of the modeling framework that stems from the underlying process being defined on a regular grid is that the sampling scheme of the data has less of an impact on inferences than would be the case were the underlying spatial process modeled at the observation locations. The number of monitoring sites near the locations of the underlying factor processes will affect the uncertainty associated with the factor processes. As a result, by summarizing our factors on a regular spatial grid, we can explore the variation in information about the pollutant levels over space.

Finally, we point out that a priority of future work should be devising numerical and graphical summaries for reporting the output a dynamic factor PC analyses. Summarizing only the average of a few summary statistics of the posterior distributions of the underlying factors does not do justice to the information available from such an analysis. By developing animated or interactive visualization tools, the output could be better explored for features such as space-time variation in factor processes and their associated levels of uncertainty.

References

Aguilar, O. (1998). "Latent structure in Bayesian multivariate time series models." Ph.D. thesis, Duke University, Durham, NC 27708.

Aguilar, O. and West, M. (2000). "Bayesian dynamic factor models and variance matrix discounting for portfolio allocation." *Journal of Business & Economic Statistics*, 18, 338–357.

Barry, R. and Ver Hoef, J. (1996). "Blackbox kriging: Spatial prediction without specifying variogram models." *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 297–322.

Calder, C. A., Holloman, C., and Higdon, D. (2001). "Exploring space-time structure in ozone concentration using a dynamic process convolution model." In *Case Studies in Bayesian Statistics 6*, 165–176. Springer-Verlag.

Carter, C. and Kohn, R. (1994). "On Gibbs sampling for state space models." *Biometrika*, 81, 3, 541–553.

Christensen, W. and Amemiya, Y. (2001). "Generalized shift-factor analysis method for multivariate geo-referenced data." *Mathematical Geology*, 33, 7, 801–824.

— (2002). "Latent variable analysis of multivariate spatial data." *Journal of the American Statistical Association*, 97, 457, 302–317.

Cressie, N. and Wikle, C. (2002). *Encyclopedia of Environmetrics*, chap. Space-time Kalman Filtering. John Wiley & Sons.

Frühwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models." *Journal of Time Series Analysis*, 15, 183–202.

Gelfand, A., Schmidt, A., and Sirmans, C. (2002). "Multivariate spatial process models: Conditional and unconditional Bayesian approaches using coregionalization." Tech. rep., University of Connecticut.

Gelfand, A. and Smith, A. (1990). "Sampling-based approaches to calculating marginal densities." *Journal of the American Statistical Association*, 85, 410, 398–409.

Higdon, D. (1998). "A process-convolution approach to modeling temperatures in the North Atlantic Ocean." *Journal of Environmental and Ecological Statistics*, 5, 173–190.

— (2002). "Space and space-time modeling using process convolutions." In *Quantitative Methods for Current Environmental Issues*, eds. C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi, 37–56. Springer Verlag.

Higdon, D., Swall, J., and Kern, J. (1998). "Non-stationary spatial modeling." In *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 761–768. Oxford University Press.

Huerta, G., Sansó, B., and Stroud, J. (2004). "A spatiotemporal model for Mexico City ozone levels." *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 53, 231–248.

Kalman, R. (1960). "A new approach to linear filtering and prediction problems." *Journal of Basic Engineering (ASME)*, 82D, 35–45.

Kern, J. (2000). "Bayesian process-convolution approaches to specifying spatial dependence structure." Ph.D. thesis, Duke University, Durham, NC 27708.

Lavery, T., Rogers, C., Howell, H., Burnett, M., and Wanta, C. (2002). "Clean Air Status and Trends Network (CASTNet): 2000 Annual Report." Tech. rep., United States Environmental Protection Agency. Prepared by: Harding ESE.

Lee, H., Higdon, D., Calder, C., and Holloman, C. (2005). "Efficient models for correlated data via convolutions of intrinsic processes." *Statistical Modeling*, Forthcoming.

Lopes, H. and West, M. (2004). "Bayesian model assessment in factor analysis." *Statistica Sinica*, 14, 41–67.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag.

Thiébaux, H. J. and Pedder, M. A. (1987). *Spatial Objective Analysis with Applications in Atmospheric Science*. London: Academic Press.

Ver Hoef, J. and Barry, R. (1998). "Constructing and fitting models for cokriging and multivariable spatial prediction." *Journal of Statistical Planning and Inference*, 69, 275–294.

Ver Hoef, J., Cressie, N., and Barry, R. (2004). "Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT)." *Journal of Computational and Graphical Statistics*, 13, 265–282.

Wackernagel, H. (1998). *Multivariate Geostatistics*. Berlin: Springer.

West, M. (1995). "Bayesian inference in cyclical component dynamic linear models." *Journal of the American Statistical Association*, 90, 1301–1312.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. 2nd ed. New York: Springer-Verlag.

Wikle, C., Berliner, L., and Cressie, N. (1998). "Hierarchical Bayesian space-time models." *Environmental and Ecological Statistics*, 5, 117–154.

Wikle, C., Milliff, R., Nychka, D., and Berliner, L. (2001). "Spatio-temporal hierarchical Bayesian Modeling: Tropical Ocean Surface Winds." *Journal of the American Statistical Association*, 96, 382–397.

Xu, B., Wikle, C., and Fox, N. (2003). "A kernel-based spatio-temporal dynamical model for nowcasting radar precipitation." In review.

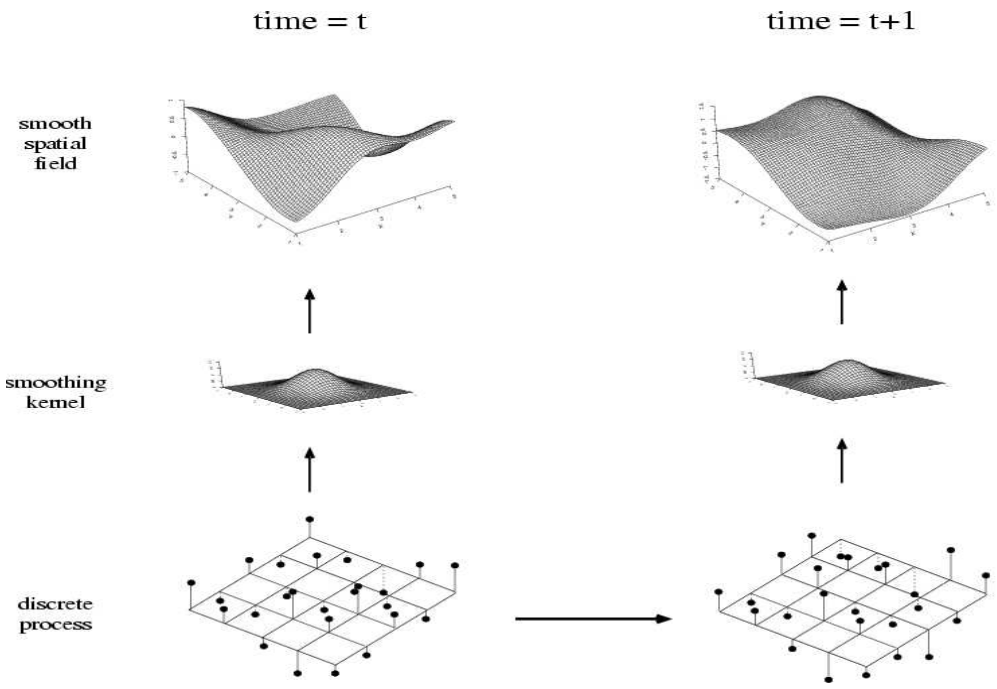

Figure 1: Conditional independence structure of the DPC model. The plot in the top left represents the smooth spatial field $\mathbf{K}\boldsymbol{x}_t$ at time t which is constructed by convolving a discrete process on a lattice (bottom left) with a smoothing kernel (center left). Going from time t to $t+1$ the underlying process evolves over time according to Equation 7 as depicted in the bottom plots. The new discrete process at time $t+1$ is then convolved to create the smooth spatial field at time $t+1$ from which the data are modeled (based on Equation 6). Thus, the smooth spatial fields at any set of time points are independent conditional on the underlying x process.

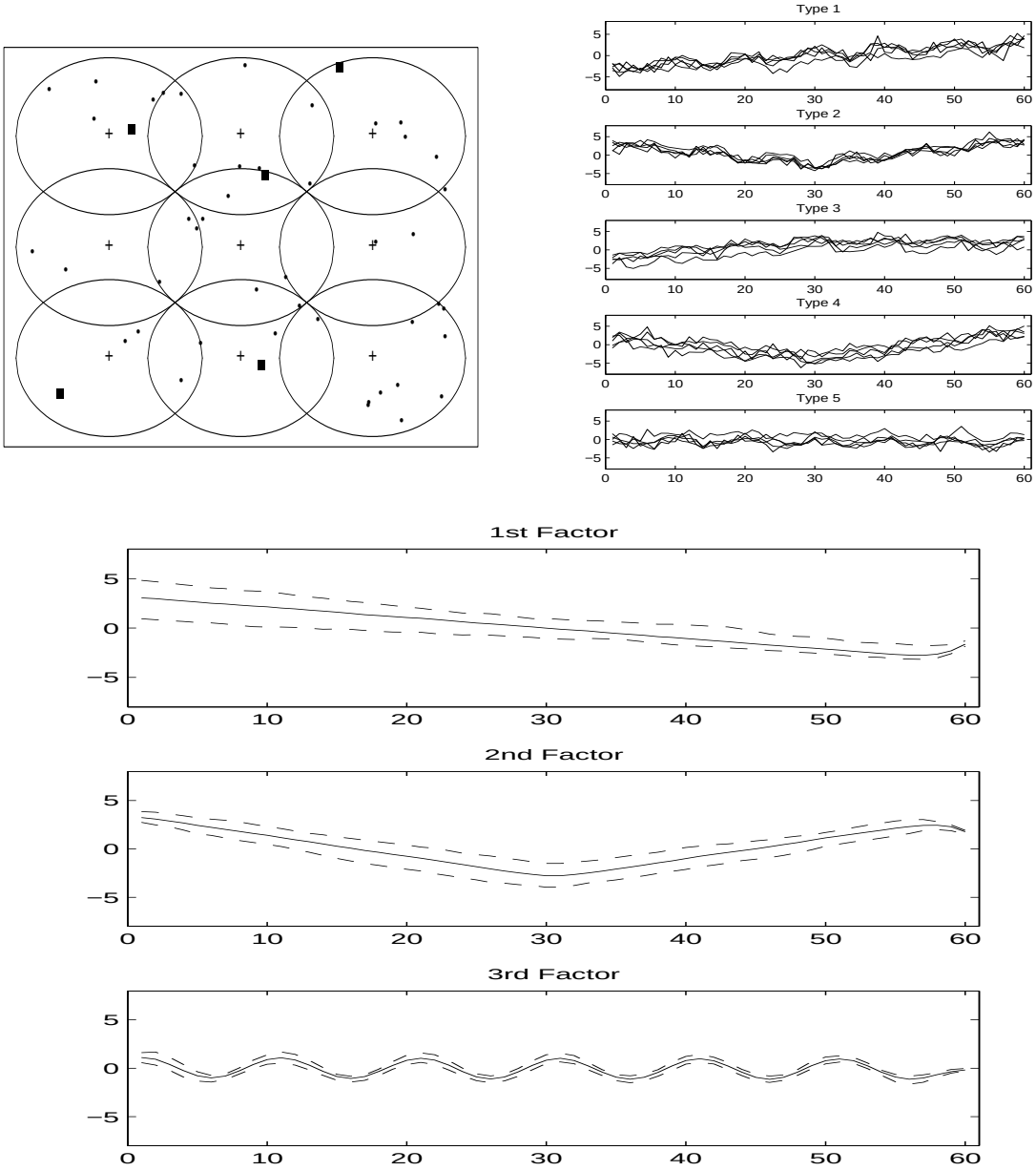

Figure 2: Plots related to the simulation study described in Section 4. Top left: Locations of the observed data are represented by points and squares. The locations of the underlying $\boldsymbol{x}$ processes are denoted by '+'s, and one standard deviation ellipses of the kernels centered at each of the underlying process locations are plotted. Top right: Time-series plots that depict the simulated data at the locations denoted by squares in the top left plot. Each graph shows one data type at these four locations. Bottom: As described in detail in the text, these plots summarize the posterior distributions of the three underlying factor processes.

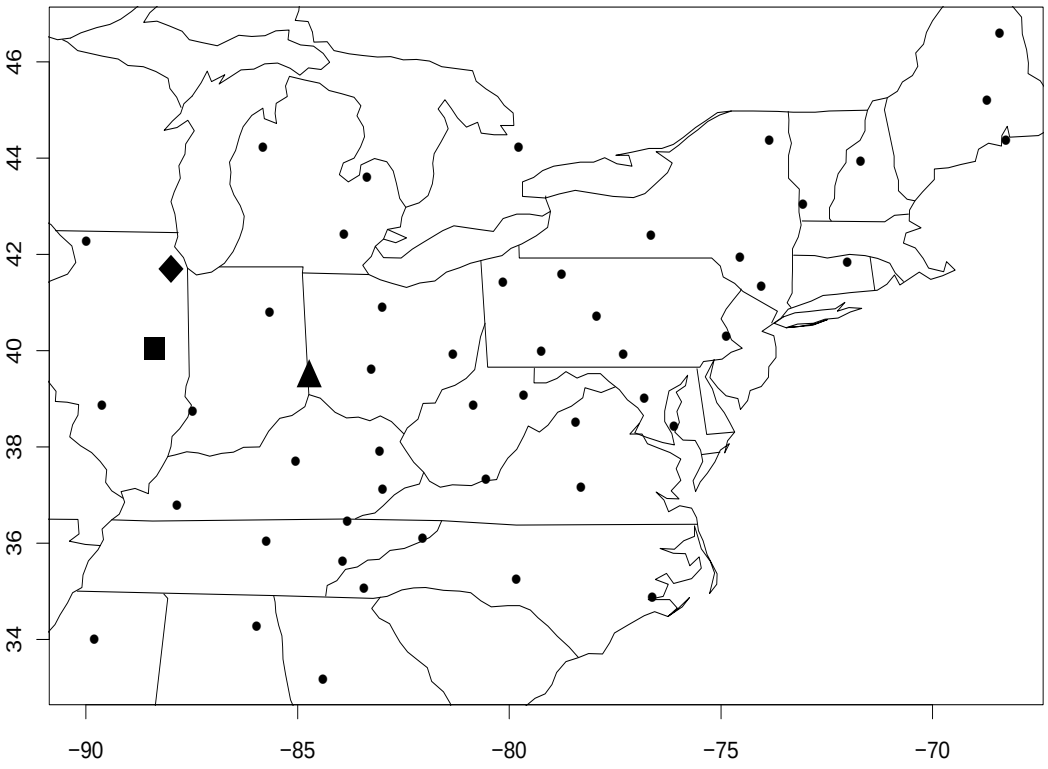

Figure 3: Locations of the 55 CASTNet monitor locations. The $\log$ concentrations of the five pollutants at the monitoring site represented as a triangle (lat $= 39.5$, lon $= -84.7$)) are shown in Figure 4. The $\log$ concentrations of sulfur dioxide registered at the locations represented as a diamond (lat $= 41.7$, lon $= -88.0$) and as a square (lat $= 38.9$, lon $= -89.6$) are shown in Figure 5.

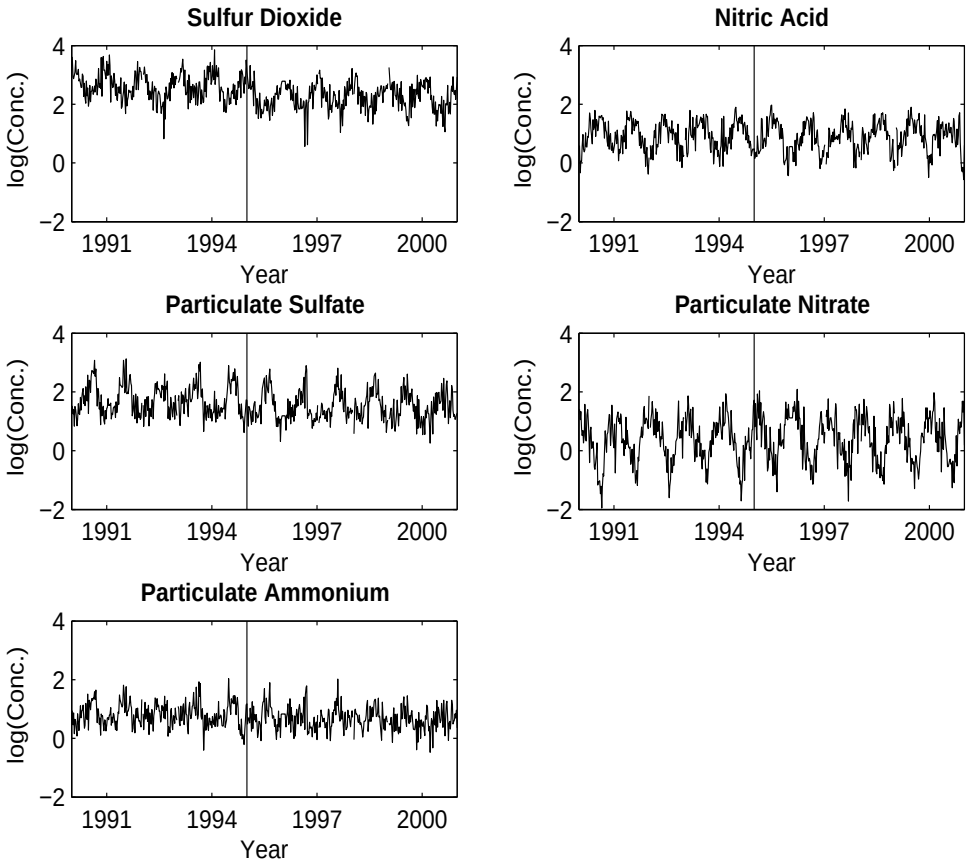

Figure 4: Weekly pollutant concentration levels from a site on the border of Indiana and Ohio (lat $= 39.5$, lon $= -84.7$) from January 1990 to December 2000

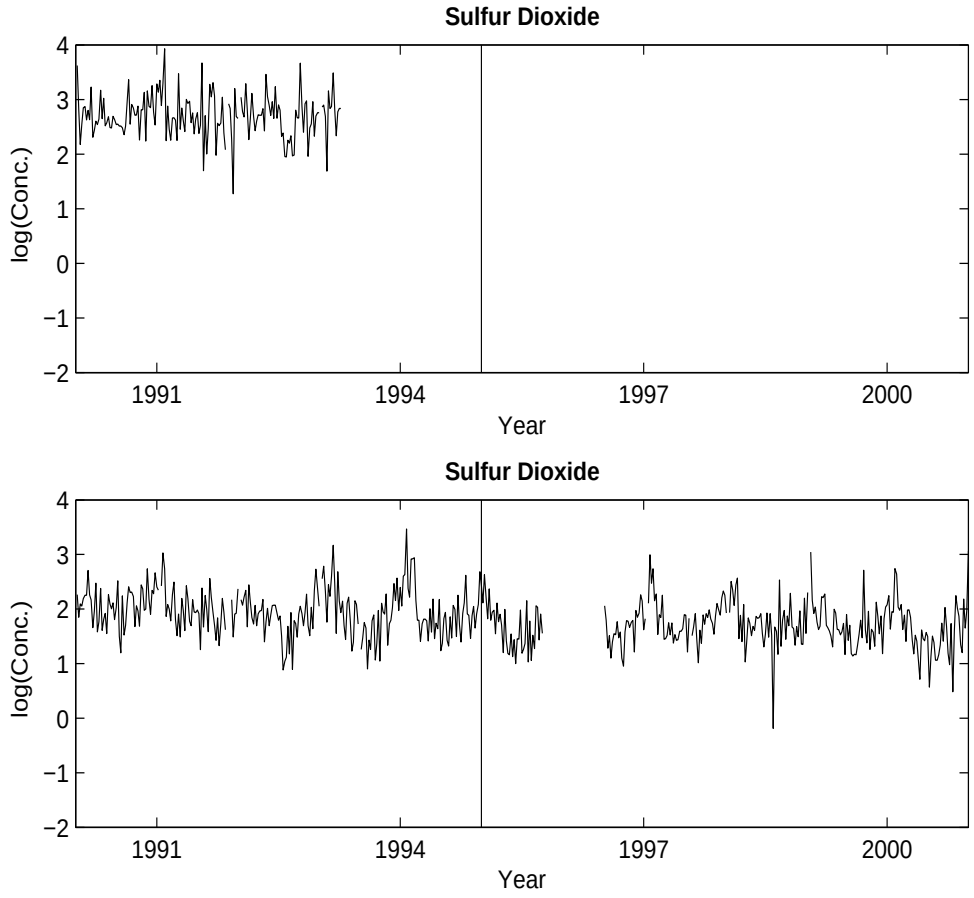

Figure 5: Top: Weekly pollution concentration levels from a site in northeastern Illinois (lat $=$ 41.7, lon $=$ -88.0) from January 1990 to December 2000. Bottom: Weekly pollutant concentration levels from a site in central Illinois (lat $=$ 38.9, lon $=$ -89.6) from January 1990 to December 2000.

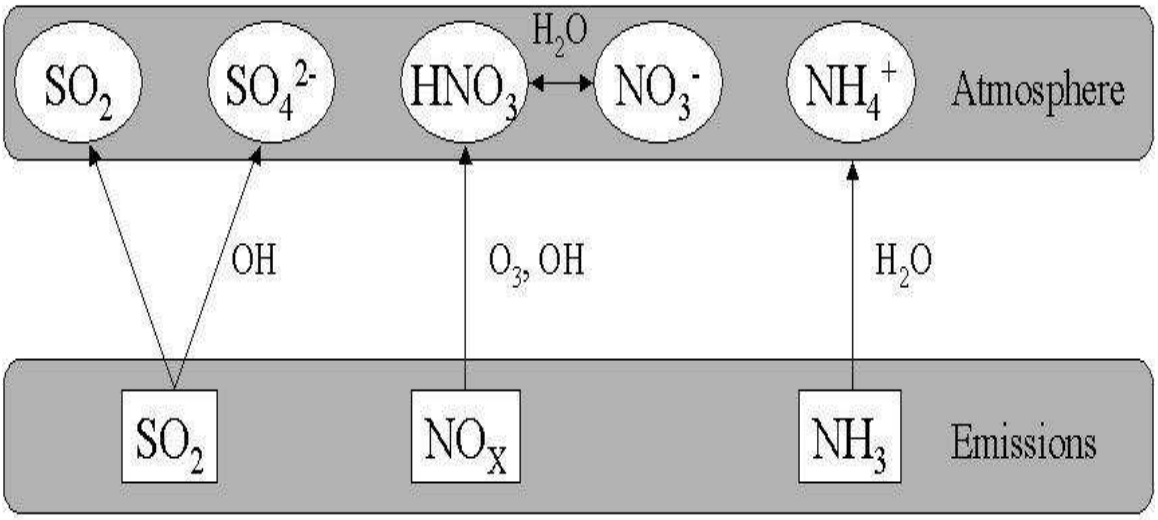

Figure 6: Relationship between the atmospheric concentrations of the five pollutants in our analysis of the CASTNet data and pollutant emissions.

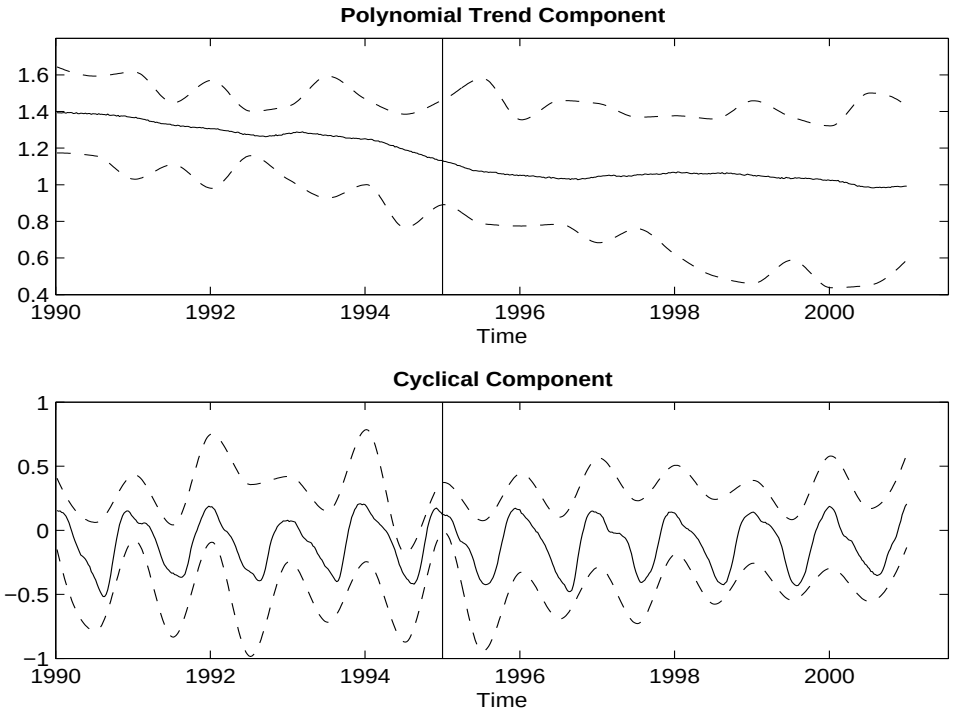

Figure 7: Average (taken over the M spatial locations) of posterior means (solid lines) and 25^{th} and 75^{th} percentiles (dashed lines) of the latent $\boldsymbol{x}$ processes.