Bayesian Analysis (2008) **3**, Number 4, pp. 759–792

Spatial Dynamic Factor Analysis

Hedibert Freitas Lopes*, Esther Salazar† and Dani Gamerman‡

Abstract. A new class of space-time models derived from standard dynamic factor models is proposed. The temporal dependence is modeled by latent factors while the spatial dependence is modeled by the factor loadings. Factor analytic arguments are used to help identify temporal components that summarize most of the spatial variation of a given region. The temporal evolution of the factors is described in a number of forms to account for different aspects of time variation such as trend and seasonality. The spatial dependence is incorporated into the factor loadings by a combination of deterministic and stochastic elements thus giving them more flexibility and generalizing previous approaches. The new structure implies nonseparable space-time variation to observables, despite its conditionally independent nature, while reducing the overall dimensionality, and hence complexity, of the problem. The number of factors is treated as another unknown parameter and fully Bayesian inference is performed via a reversible jump Markov Chain Monte Carlo algorithm. The new class of models is tested against one synthetic dataset and applied to pollution data obtained from the Clean Air Status and Trends Network (CASTNet). Our factor model exhibited better predictive performance when compared to benchmark models, while capturing important aspects of spatial and temporal behavior of the data.

Keywords: Bayesian inference, forecasting, Gaussian process, spatial interpolation, reversible jump Markov chain Monte Carlo, random fields

1 Introduction

Factor analysis and spatial statistics are two successful examples of statistical areas that have been experiencing major attention both from the research community as well as from practitioners. One could argue that the main reason for such interest is motivated by the recent increase in availability of efficient computational schemes coupled with fast, easy-to-use computers. In particular, Markov chain Monte Carlo (MCMC) simulation methods (Gamerman and Lopes, 2006, and Robert and Casella, 2004, for instance) have opened up access to fully Bayesian treatments of factor analytic and spatial models as described, for instance, in Lopes and West (2004) and Banerjee, Carlin, and Gelfand (2004) respectively, and their references.

Factor analysis has previously been used to model multivariate spatial data. Wang and Wall (2003), for instance, fitted a spatial factor model to the mortality rates for three major diseases in nearly one hundred counties of Minnesota. Christensen

*Graduate School of Business, University of Chicago, Chicago, IL, mailto:hlopes@ChicagoGSB.edu

†Graduate Program in Statistics, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, mailto:esalazar@dme.ufrj.br

‡Graduate Program in Statistics, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, dani@im.ufrj.br

 DOI:10.1214/08-BA329

and Amemiya (2002, 2003) proposed what they called the shift-factor analysis method to model multivariate spatial data with temporal behavior modeled by autoregressive components. Hogan and Tchernis (2004) fitted a one-factor spatial model and entertained several forms of spatial dependence through the single common factor. In all these applications, factor analysis is used in its original setup, i.e., the common factors are responsible for potentially reducing the overall dimension of the response vector observed at each location. In this paper, however, the observations are univariate and factor analysis is used to reduce (identify) clusters/groups of locations/regions whose temporal behavior is primarily described by a potentially small set of common dynamic latent factors. One of the key aspects of the proposed model is that flexible and spatially structured prior information regarding such clusters/groups can be directly introduced by the columns of the factor loadings matrix.

More specifically, a new class of nonseparable and nonstationary space-time models that resembles a standard dynamic factor model (Peña and Poncela 2004, for instance), is proposed

$$y_t = \mu_t^{y*} + \beta f_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \Sigma) \tag{1}$$

$$f_t = \Gamma f_{t-1} + \omega_t, \qquad \omega_t \sim N(0, \Lambda) \tag{2}$$

where $y_t = (y_{1t}, \ldots, y_{Nt})'$ is the N-dimensional vector of observations (locations $s_1, \ldots,$ s_N and times $t = 1, \ldots, T$), μ_t^{y*} is the mean level of the space-time process, f_t is an m-dimensional vector of common factors, for $m < N$ (m is potentially several orders of magnitude smaller than N) and $\beta = (\beta_{(1)}, \ldots, \beta_{(m)})$ is the $N \times m$ matrix of factor loadings. The matrix Γ characterizes the evolution dynamics of the common factors, while Σ and Λ are observational and evolutional variances. For simplicity, it is assumed throughout the paper that $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m)$. The dynamic evolution of the factors is characterized by $\Gamma = \text{diag}(\gamma_1, \ldots, \gamma_m)$, which can be easily extended to non-diagonal cases (Sections 2.3 and 2.5, for instance, deal with seasonal components and non-stationary common factors). Also, the above setting assumes that observed locations remain unchanged throughout the study period but anisotopic observation processes can be easily contemplated.

The novelty of the proposal is twofold *i)* at any given time the univariate measurements from all observed locations, either areal or point-referenced, are grouped together in what otherwise would be the vector of attributes in standard factor analysis and *ii)* spatial dependence is introduced by the columns of the factor loadings matrix. As a consequence, common dynamic factors can be thought of as describing temporal similarities amongst the time series, such as common annual cycles or (stationary or nonstationary) trends, while the importance of common factors in describing the measurements in a given location is captured by the components of the factor loadings matrix, which are modeled by regression-type Gaussian processes (see equation (3) below). The interpretability of the common factors and factor loadings matrix, typical tools in factor analysis, are of key importance under the new structure. As can be seen in the application, the dynamics of time series and their spatial correlations can be separately treated and interpreted, without forcing unrealistic simplifications. More general time series models can be entertained, through the common factors, without imposing additional constraints to the current spatial characterization of the model, and vice-versa. Standard factor analysis approach to high dimensional data sets usually trades off modeling refinement and simplicity. Therefore, this paper focuses primarily on moderate sized problems, where Bayesian hierarchical inference is a fairly natural approach.

An additional key feature of the proposed model is its ability to encompass several existing models, which are restricted in most cases to nonstochastic common dynamic factors or factor loadings matrices. More specifically, when the common factors are nonstochastic (multivariate dynamic regression or weight kernels), the well known classes of spatial priors for regression coefficients (see Gamerman, Moreira and Rue, 2003, and Nobre, Schmidt and Lopes, 2005, for instance) or dynamic models for spatiotemporal data (Stroud, Müller and Sansó, 2001) fall into the class of structured hierarchical priors introduced in Section 2. Additionally, when the factor loadings matrix is fixed, completely deterministic or a deterministic function of a small number of parameters, the proposed model can incorporate the structures that appear in Mardia, Goodall, Redfern and Alonso (1998), Wikle and Cressie (1999) and Calder (2007), to name a few. In this paper, both deterministic and stochastic forms are considered in the specification of factor loadings, which can easily incorporate external information through regression functions.

Finally, the number of common factors is treated as another parameter of the model and formal Bayesian procedures are derived to appropriately account for its estimation. Thus, sub-models with different parametric dimensions have to be considered and a customized reversible jump MCMC algorithm derived. This is an original contribution in the intersected field of spatiotemporal models and dynamic factor analysis. The number of common dynamic factors plays the usual and important role of data reduction, i.e. N time series are parsimoniously represented by a small set of m time series processes. Also, different levels of spatial smoothness and different degrees of spatial and/or temporal nonstationarity can be achieved by the inclusion/exclusion of common factors.

The remainder of the paper is organized as follows. Sections 2 and 3 specify in detail the components of the proposed model (equations (1) and (2)), along with prior specification for the model parameters, as well as forecasting and interpolation strategies. Posterior inference for a fixed or an unknown number of factors is outlined in Section 4. Simulated and real data illustrations appear in Section 5 with Section 6 listing conclusions and directions of current and future research.

2 Proposed space-time model

Equations (1) and (2) define the first level of the proposed dynamic factor model. Similar to standard factor analysis, it is assumed that the m conditionally independent common factors f_t capture all time-varying covariance structure in y_t. The conditional spatial dependencies are modeled by the columns of the factor loadings matrix β. More specifically, the j^{th} column of β, denoted by $\beta_{(j)} = (\beta_{(j)}(s_1), \ldots, \beta_{(j)}(s_N))'$, for $j = 1, \ldots, m$, is modeled as a conditionally independent, distance-based Gaussian process or a Gaussian random field (GRF), i.e.

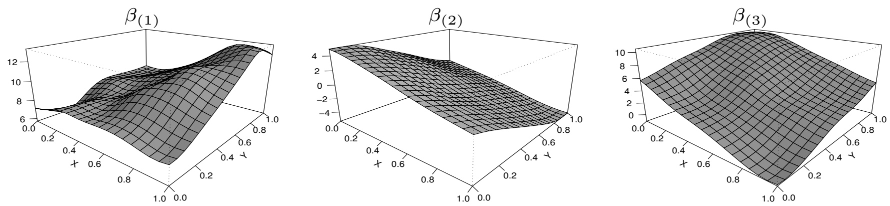
$$\beta_{(j)} \sim GRF(\mu_j^{\beta^*}, \tau_j^2 \rho_{\phi_j}(\cdot)) \equiv N(\mu_j^{\beta^*}, \tau_j^2 R_{\phi_j}), \tag{3}$$

where $\mu_j^{\beta^*}$ is a N-dimensional mean vector (see Section 2.1 for the inclusion of covariates). The (l,k)-element of R_{ϕ_j} is given by $r_{lk} = \rho_{\phi_j}(|s_l - s_k|)$, $l, k = 1, \ldots, N$, for suitably defined correlation functions $\rho_{\phi_j}(\cdot)$, $j = 1, \ldots, m$. The parameters ϕ_js are typically scalars or low dimensional vectors. For example, ϕ is univariate when the correlation function is exponential $\rho_\phi(d) = \exp\{-d/\phi\}$ or spherical $\rho_\phi(d) = (1-1.5(d/\phi)+0.5(d/\phi)^3)1_{\{d/\phi \leq 1\}}$, and bivariate when it is power exponential $\rho_\phi(d) = \exp\{-(d/\phi_1)^{\phi_2}\}$ or Matérn $\rho_\phi(d) = 2^{1-\phi_2}\Gamma(\phi_2)^{-1}(d/\phi_1)^{\phi_2}\mathcal{K}_{\phi_2}(d/\phi_1)$, where $\mathcal{K}_{\phi_2}(\cdot)$ is the modified Bessel function of the second kind and of order ϕ_2. In each one of the above families, the range parameter $\phi_1 > 0$ controls how fast the correlation decays as the distance between locations increases, while the smoothness parameter ϕ_2 controls the differentiability of the underlying process (for details, see Cressie, 1993 and Stein, 1999). The proposed model could, in principle, accommodate nonparametric formulations for the spatial dependence, such as the ones introduced by Gelfand, Kottas and MacEachern (2005), for instance.

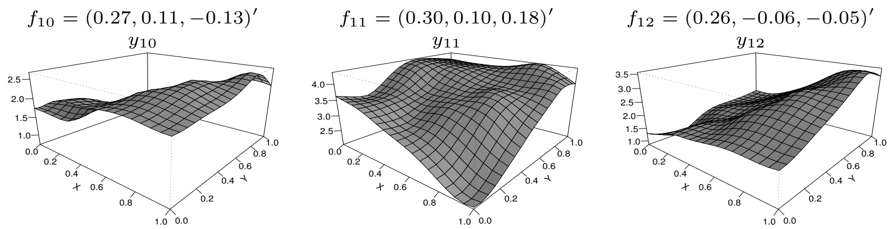
The proposed spatial dynamic factor model is defined by equations (1)–(3). Figure 1 illustrates the proposed model through a simulated spatial dynamic three-factor analysis. The first column of the factor loadings matrix, $\beta_{(1)}$, suggests that the first common factor is more important on the northeast end than on the southwest corner of the region. Similarly, the second and third common factors are more important on the northeast and northwest corners of the region. Notice, for instance, that $\beta_{(1)}$ and y_{12} behave quite similarly due to the overwhelming influence of the first common factor. Additional recent developments on Bayesian factor analysis can be found in Lopes and Migon (2002), Lopes (2003), Lopes and West (2004) and Lopes and Carvalho (2007).

Several existing alternative models can be seen as particular cases of the model proposed by letting $\tau_j^2 = 0$, for all j and by properly specifying $\mu_j^{\beta^*}$ by a pre-gridding principal component decomposition (Wikle and Cressie 1999) or by a principal kriging procedure (Sahu and Mardia, 2005, and Lasinio, Sahu and Mardia, 2005), for instance. Mardia, Goodall, Redfern and Alonso (1998), for instance, introduced the kriged Kalman filter and split the columns of β (common fields) into trend fields and principal fields, both of which are fixed functions of empirical orthogonal functions. In a related paper, Wikle and Cressie (1999) modeled the columns of β based on deterministic, complete and orthonormal basis functions. More recently, Sansó et al. (2008) and Calder (2007) used smoothed kernels to deterministically derive β. The modeling of $\mu_j^{\beta^*}$, presented in what follows, highlights some of the features of the new model. Finally, Reich et al. (2008) present a simpler, non-spatial version of our space-time model for known number of factors applied to human exposure to particulate matter.

The existence of a broad literature on general dynamic factor model for large scale problems should also be mentioned. See, among others, Bai (2003) and Forni et al. (2000). In this paper, the whole space-time dependencies are modeled through dynamic common factors and spatially structured factor loadings matrix. This is primarily moti-



(a) Factor loadings



(b) $y_t = \sum_{j=1}^{3} \beta_{(j)} f_{jt} + \epsilon_t$

Figure 1: *Simulated data:* y_t follows a spatial dynamic 3-factor model. (a) Gaussian processes for the three columns of the factor loadings matrix. (b) y_t processes at $t = 10, 11, 12$.

vated by recent interest in the factor models literature towards hierarchically structured loadings matrices (see Lopes and Migon, 2002, Lopes and Carvalho, 2007, Carvalho et al. 2008, and West, 2003, to name a few).

2.1 Covariate effects

Many specifications for the mean level of both space-time process (equation 1) and Gaussian random field (equation 3) can be entertained, with the most common ones based on time-varying as well as location-dependent covariates. For the mean level of the space-time process (μ_t^{y*} in equation (1)), a few alternatives are *i)* constant mean level model: $\mu_t^{y*} = \mu^y$, $\forall t$ (possibly $\mu^y = 0$); *ii)* regression model: $\mu_t^{y*} = X_t^y \mu^y$, where $X_t^y = (1_N, X_{1t}^y, \dots, X_{qt}^y)$ contains q time-varying covariates; *iii)* dynamic coefficient model: $\mu_t^{y*} = X_t^y \mu_t^y$ and $\mu_t^y \sim N(\mu_{t-1}^y, W)$.

Similarly, covariates can be included in the factor loadings prior specification ($\mu_j^{\beta*}$ in equation (3)) so that dependencies due to deterministic spatial variation can be entertained. Due to the static behavior of β, only spatially-varying covariates will be considered in explaining the mean level of the Gaussian random fields. The following are the specifications considered in this paper: *i)* $\mu_j^{\beta*} = 0$; *ii)* $\mu_j^{\beta*} = \mu_j^\beta 1_N$ and *iii)* $\mu_j^{\beta*} = X_j^\beta \mu_j^\beta$, where X_j^β is a $N \times p_j$ matrix of covariates. In the last case, more flexibility is brought up by allowing potentially different covariates for each Gaussian random field.

Deterministic specifications for the leadings such as smoothing kernels (Calder, 2007; Sansó and Schmidt, 2008) or orthogonal basis (Stroud, Müller and Sansó, 2001; Wikle and Cressie, 1999) can be accommodated in this formulation. The approach of this paper allows an additional stochastic component that is spatially structured and is thus potentially capable of picking more general spatial dependencies.

2.2 Spatio-temporal separability

Roughly speaking, separable covariance functions of spatiotemporal processes can be written as the product (or sum) of a purely spatial covariance function and a purely temporal covariance function. More specifically, let $Z(s,t)$ be a random process indexed by space and time. The process is separable if $\text{Cov}(Z(s_1,t_1), Z(s_2,t_2))=\text{Cov}_s(u|\theta)\text{Cov}_t(h|\theta)$ (or $\text{Cov}_s(u|\theta) + \text{Cov}_t(h|\theta)$), for model parameters $\theta \in \Theta \subset \Re^p$, $u = \|s_2 - s_1\|$ and $h = |t_2 - t_1|$. Under the proposed model, when $m = 1$ and $\mu_t^{y*} = 0$, it is easy to see that $\text{Cov}(y_{it}, y_{j,t+h}) = (\lambda\gamma^h)(1-\gamma^2)^{-1}(\tau^2\rho(u,\phi) + \mu_i^\beta\mu_j^\beta)$, so both spatial and temporal covariance functions are separately identified.

Cressie and Huang (1999) claim that separable structures are often chosen for convenience rather than for their ability to fit the data. In fact, separable covariance functions are usually severely limited because they are unable to model space-time interaction. Cressie and Huang (1999), for instance, introduced classes of nonseparable, stationary covariance functions that allow for space-time interaction and are based on closed form Fourier inversions. Gneiting (2002) extends Cressie and Huang's (1999) classes by constructions directly in the space-time domain.

It is easy to show that $\text{Cov}(y_{it}, y_{j,t+h}) = \sum_{k=1}^m(\lambda_k\gamma_k^h)(1-\gamma_k^2)^{-1}(\tau_k^2\rho(u,\phi_k)+\mu_{ik}^\beta\mu_{jk}^\beta)$, for $m > 1$. Therefore, an important property of the proposed model is that it leads to nonseparable forms for its covariance function whenever the number of common factors is greater than one.

2.3 Seasonal dynamic factors

Periodic or cyclical behavior are present in many applications and can be directly entertained by the dynamic models framework embedded in the proposed model. For example, linear combinations of trigonometric functions (Fourier form) can be used to model seasonality (West and Harrison 1997). In fact, equations (1) and (2) encompass a fairly large class of models, such as multiple dynamic linear regressions, transfer function models, autoregressive moving average models and general time series decomposition models, to name a few.

Seasonal patterns can be incorporated into the proposed model either through the common dynamic factors or through the mean level. In the former, common seasonal factors receive different weights for different columns of the factor loading matrix, so allowing different seasonal patterns for the spatial locations. In the latter, the same pattern is assumed for all locations. For example, a seasonal common factor of period p ($p = 52$ for weekly data and annual cycle) can be easily accommodated by simply letting $\beta = (\beta_{(1)}, 0, \ldots, \beta_{(h)}, 0)$ and $\Gamma = \text{diag}(\Gamma_1, \ldots, \Gamma_h)$, where

$$\Gamma_l = \begin{pmatrix} \cos(2\pi l/p) & \sin(2\pi l/p) \\ -\sin(2\pi l/p) & \cos(2\pi l/p) \end{pmatrix}, \quad l = 1 \ldots, h = p/2,$$

and $h = p/2$ is the number of harmonics needed to capture the seasonal behavior of the time series (see West and Harrison, 1997, Chapter 8, for further details). In this context the covariance matrix Λ is no longer diagonal since the seasonal factors are correlated, i.e., $\Lambda = \text{diag}(\Lambda_1, \ldots, \Lambda_h)$ and each Λ_l $(l = 1, ..., h)$ is a 2×2 covariance matrix.

It is worth noting that the seasonal factors play the role of weights for loadings that follow Gaussian processes, so implying different seasonal patterns for different stations. Inference for the seasonal model is done using the algorithm proposed below with (conceptually) simple additional steps. For instance, posterior samples for $\Lambda_l, l = 1, \ldots, h$ are obtained from inverted Wishart distributions, as opposed to the usual inverse gamma distributions. In practice, fewer harmonics are required in many applications to adequately describe the seasonal pattern of many datasets and the dimension of this component is typically small. For the sake of notation, the following sections present the inferential procedures based on the more general equations (1) and (2).

2.4 Likelihood function

It will be assumed for the remainder of the paper, without loss of generality, that $\mu_t^{y*} = 0$ and $\mu_j^{\beta*} = X_j^\beta \mu_j^\beta$. Therefore, conditional on f_t, for $t = 1, \ldots, T$, model (1) can be rewritten in matrix notation as $y = F\beta' + \epsilon$, where $y = (y_1, \ldots, y_T)'$ and $F = (f_1, \ldots, f_T)'$ are $T \times N$ and $T \times m$ matrices, respectively. The error matrix, ϵ, is of dimension $T \times N$ and follows a matric-variate normal distribution, i.e., $\epsilon \sim N(0, I_T, \Sigma)$ (Dawid, 1981 and Brown, Vannucci and Fearn, 1998), so the likelihood function of (Θ, F, β, m) is

$$p(y|\Theta, F, \beta, m) = (2\pi)^{-TN/2}|\Sigma|^{-T/2} etr\left\{-\frac{1}{2}\Sigma^{-1}(y - F\beta')'(y - F\beta')\right\}, \tag{4}$$

where $\Theta = (\sigma, \lambda, \gamma, \mu, \tau, \phi)$, $\sigma = (\sigma_1^2, \ldots, \sigma_N^2)'$, $\lambda = (\lambda_1, \ldots, \lambda_m)'$, $\gamma = (\gamma_1, \ldots, \gamma_m)'$, $\mu = (\mu_1^\beta, \ldots, \mu_m^\beta)$, $\tau = (\tau_1^2, \ldots, \tau_m^2)'$, $\phi = (\phi_1, \ldots, \phi_m)'$, $etr(X) = \exp(trace(X))$. The dependence on the number of factors m is made explicit and considered as another parameter in Section 4.1.

2.5 Prior information

For simplicity, conditionally conjugate prior distributions will be used for all parameters defining the dynamic factor model, while two different prior structures are considered for the parameters defining the spatial processes. The prior for the common dynamic factors is given in (2) and completed by $f_0 \sim N(m_0, C_0)$, for known hyperparameters m_0 and C_0. Independent prior distributions for the hyperparameters σ and λ are as follows: *i)* $\sigma_i^2 \sim IG(n_\sigma/2, n_\sigma s_\sigma/2)$, $i = 1,\dots,N$; and *ii)* $\lambda_j \sim IG(n_\lambda/2, n_\lambda s_\lambda/2)$, $j = 1,\dots,m$, where $n_\sigma, s_\sigma, n_\lambda$ and s_λ are known hyperparameters. Remaining temporal dependence on the idiosyncratic errors can also be considered (Lopes and Carvalho 2007 and Peña and Poncela 2006) but this was not pursued here.

For γ, many specifications can be considered. For example, i) $\gamma_j \sim Ntr_{(-1,1)}(m_\gamma, S_\gamma)$, where $Ntr_{(c,d)}(a,b)$ refers to the $N(a,b)$ distribution truncated to the interval $[c,d]$; and ii) $\gamma_j \sim \alpha Ntr_{(-1,1)}(m_\gamma, S_\gamma) + (1-\alpha)\delta_1(\gamma_j)$, where m_γ, S_γ and $\alpha \in (0,1]$ are known hyperparameters, $\delta_1(\gamma_j) = 1$ if $\gamma_j = 1$ and $\delta_1(\gamma_j) = 0$ if $\gamma_j \neq 1$. In the first case, all common dynamic factors are assumed stationary. In the second case, possible nonstationary factors are also entertained. Note that when $\alpha = 1$, case i) is contemplated. See Peña and Poncela (2004, 2006) for more details on nonstationary dynamic factor models and Section 5 for the application.

The parameters μ_j^β, ϕ_j and τ_j^2, for $j = 1,\dots,m$, follow one of the two prior specifications: *i)* vague but proper priors and *ii)* reference-type priors. In the first case, $\mu_j^\beta \sim N(m_\mu, S_\mu)$, $\phi_j \sim IG(2,b)$ and $\tau_j^2 \sim IG(n_\tau/2, n_\tau s_\tau/2)$, $j = 1,\dots,m$, where m_μ, S_μ, n_τ and s_τ are known hyperparameters, $b = \rho_0/(-2\ln(0.05))$ and $\rho_0 = \max_{i,j=1,\dots,N} |s_i - s_j|$ (see Banerjee, Carlin, and Gelfand (2004) and Schmidt and Gelfand (2003), for more details). In other words, $\pi(\mu_j^\beta, \tau_j^2, \phi_j) = \pi_N(\mu_j^\beta)\pi_{IG}(\tau_j^2, \phi_j)$ where

$$\pi_{IG}(\tau_j^2, \phi_j) \quad = \quad \pi_{IG}(\tau_j^2)\pi_{IG}(\phi_j) \propto \tau_j^{-(n_\tau+2)} e^{-0.5 n_\tau s_\tau/\tau_j^2} \phi_j^{-3} e^{-b/\phi_j}, \tag{5}$$

where the subscripts N and IG stand for the normal and inverted gamma densities, respectively. In the second case, the reference analysis proposed by Berger, De Oliveira and Sansó (2001) is considered, which guarantees propriety of the posterior distributions. More specifically, $\pi_R(\mu_j^\beta, \tau_j^2, \phi_j) = \pi_R(\mu_j^\beta|\tau_j^2, \phi_j)\pi_R(\tau_j^2, \phi_j)$, with $\pi_R(\mu_j^\beta|\tau_j^2, \phi_j) = 1$ and

$$\pi_R(\tau_j^2, \phi_j) \quad = \quad \pi_R(\tau_j^2)\pi_R(\phi_j) \propto \tau_j^{-2} \left\{ \mathrm{tr}(W_{\phi_j}^2) - \frac{1}{N - p_j}[\mathrm{tr}(W_{\phi_j})]^2 \right\}^{1/2}, \tag{6}$$

where $W_{\phi_j} = ((\partial/\partial\phi_j)R_{\phi_j})R_{\phi_j}^{-1}(I_N - X_j^\beta(X_j^{\beta\prime} R_{\phi_j}^{-1} X_j^\beta)^{-1} X_j^{\beta\prime} R_{\phi_j}^{-1})$. It is worth noticing that $\pi_{IG}(\tau_j^2)$ and $\pi_R(\tau_j^2)$ will be quite similar for n_τ near 0. They propose and recommend the use of the reference prior for the parameters of the correlation function. The basic justification is simply that the reference prior yields a proper posterior, in contrast to other noninformative priors. It is important to emphasize that this prior specification defines a reference prior when conditioning on the factor loadings matrix.

3 Uses of the model

3.1 Forecasting

Forecasting is theoretically straightforward. More precisely, one is usually interested in learning about the h-steps ahead predictive density $p(y_{T+h}|y)$, i.e.

$$p(y_{T+h}|y) = \int p(y_{T+h}|f_{T+h}, \beta, \Theta)p(f_{T+h}|f_T, \beta, \Theta)p(f_T, \beta, \Theta|y)df_{T+h}df_T d\beta d\Theta \tag{7}$$

where $(y_{T+h}|f_{T+h},\beta,\Theta) \sim N(\beta f_{T+h},\Sigma)$, $(f_{T+h}|f_T,\beta,\Theta) \sim N(\mu_h,V_h)$, $\mu_h = \Gamma^h f_T$ and $V_h = \sum_{k=1}^h \Gamma^{k-1}\Lambda(\Gamma^{k-1})'$, for $h>0$. Then, if $\{(\beta^{(1)},\Theta^{(1)},f_T^{(1)}),\ldots,(\beta^{(M)},\Theta^{(M)},f_T^{(M)})\}$ is a sample from $p(f_T,\beta,\Theta|y)$ (see Section 4 below), it is easy to draw $f_{T+h}^{(j)}$ from $p(f_{T+h}|f_T^{(j)},\beta^{(j)},\Theta^{(j)})$, for all $j=1,\ldots,M$, such that $\hat{p}(y_{T+h}|y) = M^{-1}\sum_{j=1}^M p(y_{T+h}|f_{T+h}^{(j)},\beta^{(j)},\Theta^{(j)})$ is a Monte Carlo approximation to $p(y_{T+h}|y)$. Analogously, a sample $\{y_{T+h}^{(1)},\ldots,y_{T+h}^{(M)}\}$ from $p(y_{T+h}|y)$ is obtained by sampling $y_{T+h}^{(j)}$ from $p(y_{T+h}|f_{T+h}^{(j)},\beta^{(j)},\Theta^{(j)})$, for $j=1,\ldots,M$.

3.2 Interpolation

The interest now is in interpolating the response for N_n locations where the response variable y has not yet been observed. More precisely, let y^o denote the vector of observations from locations in S and y^n denote the (latent) vector of measurements from locations in $S_n = \{s_{N+1},\ldots,s_{N+N_n}\}$. Also, let $\beta_{(j)} = (\beta_{(j)}^{o'},\beta_{(j)}^{n'})'$ be the j-column of the factor loadings matrix β with $\beta_{(j)}^o$ corresponding to y^o and $\beta_{(j)}^n$ corresponding to y^n, respectively. Interpolation consists of finding the posterior distribution of β^n (Bayesian kriging),

$$p(\beta^n|y^o) = \int p(\beta^n|\beta^o,\Theta)p(\beta^o,\Theta|y^o)d\beta^o d\Theta. \tag{8}$$

where $p(\beta^n|\beta^o,\Theta) = \prod_{j=1}^m p(\beta_{(j)}^n|\beta_{(j)}^o,\mu_j^\beta,\tau_j^2,\phi_j)$. Standard multivariate normal results can be used to derive, for $j=1,\ldots,m$, the distribution of $p(\beta_{(j)}^n|\beta_{(j)}^o,\mu_j^\beta,\tau_j^2,\phi_j)$. Conditionally on Θ,

$$\begin{pmatrix}\beta_{(j)}^o\\ \beta_{(j)}^n\end{pmatrix} \sim N\left[\begin{pmatrix}X_j^{\beta^o}\\ X_j^{\beta^n}\end{pmatrix}\mu_j^\beta,\tau_j^2\begin{pmatrix}R_{\phi_j}^o & R_{\phi_j}^{o,n}\\ R_{\phi_j}^{n,o} & R_{\phi_j}^n\end{pmatrix}\right]$$

where $R_{\phi_j}^n$ is the correlation matrix of dimension N_n between ungauged locations, $R_{\phi_j}^{o,n}$ is a matrix of dimension $N\times N_n$ where each element represents the correlation between gauged location i and ungauged location j, for $i=1,\ldots,N$ and $j=N+1,\ldots,N+N_n$. Therefore, $\beta_{(j)}^n|\beta_{(j)}^o,\Theta \sim N(X_j^{\beta^n}\mu_j^\beta + R_{\phi_j}^{n,o}R_{\phi_j}^{o\;-1}(\beta_{(j)}^o - X_j^{\beta^o}\mu_j^\beta);\tau_j^2(R_{\phi_j}^n - R_{\phi_j}^{n,o}R_{\phi_j}^{o\;-1}R_{\phi_j}^{o,n}))$ and the usual Monte Carlo approximation to $p(\beta^n|y^o)$ is $\hat{p}(\beta^n|y^o) = L^{-1}\sum_{l=1}^L p(\beta^n|\beta^{o(l)},\Theta^{(l)})$, where $\{(\beta^{o(1)},\Theta^{(1)}),\ldots,(\beta^{o(L)},\Theta^{(L)})\}$ is a sample from $p(\beta^o,\Theta|y)$ (see Section 4 below). If $\beta^{n(l)}$ is drawn from $p(\beta^n|\beta^{o(l)},\Theta^{(l)})$, for $l=1,\ldots,L$, then $\{\beta^{n(l)},\ldots,\beta^{n(L)}\}$ is a sample from $p(\beta^n|y^o)$. As a by-product, the expectation of non-observed measures y^n can be approximated by $\hat{E}(y^n|y^o) = L^{-1}\sum_{l=1}^L \beta^{n(l)}f^{(l)}$.

4 Posterior inference

Posterior inference for the proposed class of spatial dynamic factor models is facilitated by Markov Chain Monte Carlo algorithms designed for two cases: (1) known number m of common factors and (2) unknown m. In the first case, standard MCMC for dynamic linear models are adapted, while reversible jump MCMC algorithms are designed for when m is unknown.

Conditional on m, the joint posterior distribution of (F, β, Θ) is

$$\begin{aligned} p(F, \beta, \Theta | y) &\propto \prod_{t=1}^{T} p(y_t | f_t, \beta, \sigma) p(f_0 | m_0, C_0) \prod_{t=1}^{T} p(f_t | f_{t-1}, \lambda, \gamma) \\ &\times \prod_{j=1}^{m} p(\beta_{(j)} | \mu_j^{\beta}, \tau_j^2, \phi_j) p(\gamma_j) p(\lambda_j) p(\mu_j^{\beta}) p(\tau_j^2, \phi_j) \prod_{i=1}^{N} p(\sigma_i^2) \qquad (9) \end{aligned}$$

which is analytically intractable. Exact posterior inference is performed by a customized MCMC algorithm. The common factors are jointly sampled by means of the well known forward filtering backward sampling (FFBS) algorithm (Carter and Kohn 1994, and Frühwirth-Schnatter 1994). All other full conditional distributions are multivariate normal distributions or inverse gamma distributions, except the parameters characterizing the spatial correlations, ϕ, which are sampled based on a Metropolis-Hastings step. The full conditional distributions and proposals, where applicable, are detailed in the Appendix.

4.1 Unknown number of common factors

Inference regarding the number of common factors is obtained by computing posterior model probabilities (PMP), well-known for playing an important role in modern Bayesian model selection, comparison and averaging. On the one hand, a unique, most probable factor model can be selected and data analysis continued by focusing on interpreting both dynamic factors and spatial loadings. Such flexibility permits the identification, for instance, of spatial clusters of similar cyclical behavior. On the other hand, when interpolation and forecasting are of primary interest, and the identification/interpretation of common factors of secondary interest, PMP play the role of weighting forecasting observations at gauged or ungauged locations and many other functionals that appear in all competing models. This is done, for example, by Bayesian model averaging (Raftery, Madigan and Hoeting, 1997 and Clyde, 1999).

The selection/estimaton of the number of common factors is central in the factor analysis literature. The spatial and the temporal components of the spatial dynamic factor model can be conditionally separated and modeled by properly choosing it. In other words, location-specific residual spatial and/or temporal correlations are minimized. Bayesian model search, via RJMCMC algorithm, automatically penalizes over- and under-parametrized factor models.

Lopes and West (2004) introduced a modern fully Bayesian treatment the number of common factor in standard normal linear factor analysis by means of a customized reversible jump MCMC (RJMCMC) scheme. Their algorithm builds on a preliminary set of parallel MCMC outputs that are run over a set of pre-specified number of factors. These chains produce a set of within-model posterior samples for $\Psi_m = (F_m, \beta_m, \Theta_m)$ that approximate the posterior distributions $p(\Psi_m | m, y)$. Then, posterior moments from these samples were used to guide the choice of the proposal distributions from which candidate parameter would be drawn. This section adapts and generalizes their approach to the proposed spatial dynamic factor model. For the spatial dynamic factor model, the overall proposal distribution is

$$\begin{aligned} q_m(\Psi_m) &= \prod_{j=1}^{m} f_N(f_{(j)}; M_{f_{(j)}}, aV_{f_{(j)}}) f_N(\beta_{(j)}; M_{\beta_{(j)}}, bV_{\beta_{(j)}}) f_N(\gamma_j; M_{\gamma_j}, cV_{\gamma_j}) \\ &\times \prod_{j=1}^{m} f_{IG}(\lambda_j; d, dM_{\lambda_j}) f_N(\mu_j; M_{\mu_j}, eV_{\mu_j}) f_{IG}(\phi_j; f, fM_{\phi_j}) \\ &\times \prod_{j=1}^{m} f_{IG}(\tau_j^2; g, gM_{\tau_j}) f_{IG}(\sigma_j^2; h, hM_{\sigma_j}), \end{aligned} \tag{10}$$

where a, b, c, d, e, f, g and h are tuning parameters and M_x and V_x are sample means and sample variances based on the preliminary MCMC runs. The choice of the tuning constants depend on the form of the posterior distributions. For example, we recommend values lower than 1 for a, b, c and e used in proposal normal distributions, and values greater than 1.5 for d, f, g and g used in proposal inverse gamma distributions. By letting $p(y, m, \Psi_m) = p(y|m, \Psi_m)p(\Psi_m|m)Pr(m)$ and by giving initial values for m and Ψ_m, the reversible jump algorithm proceeds similar to a standard Metropolis-Hastings algorithm, i.e., a candidate model m' is drawn from the proposal $q(m, m')$ and then, conditional on m', $\Psi_{m'}$ is sampled from $q_{m'}(\Psi_{m'})$. The pair $(m', \Psi_{m'})$ is accepted with probability

$$\alpha = \min\left\{1, \frac{p(y, m', \Psi_{m'})}{p(y, m, \Psi_m)} \frac{q_m(\Psi_m)q(m', m)}{q_{m'}(\Psi_{m'})q(m, m')}\right\}. \tag{11}$$

A natural choice for initial values are the sample averages of Ψ_m based on the preliminary MCMC runs. Throughout this paper it is assumed that $Pr(m) = 1/M$, where M is the maximum number of common factors. It should be emphasized that the chosen proposal distributions $q_m(\Psi_m)$ are not generally expected to provide globally accurate approximations to the conditional posteriors $p(\Psi_m|m, y)$. Nonetheless, the closer $q_m(\Psi_m)$ and $p(\Psi_m|m, y)$ are, the closer acceptance probability is to $\alpha = \min\{1, p(y|m')/p(y|m) \times q(m', m)/q(m, m')\}$, which can be thought of as a stochastic model search algorithm (George and McCulloch, 1992). We argue, based on our empirical findings, that such quasi-independent proposal scheme is sound since each marginal proposal is already in the neighborhood of interest within a particular factor model. The scheme will essentially formalize the posterior model probabilities as the limits of visiting frequencies.

The above RJMCMC algorithm can be thought of as a particular case of the Metropolised Carlin and Chib method (Godsill 2001, and Dellaportas, Forster, and Ntzoufras 2002), where the proposal distributions generating both new model dimension and new parameters depend on the current state of the chain only through m. This is true here as well where proposal distributions based on initial, auxiliary MCMC runs are used. A more descriptive name is independence RJMCMC, analogous to the standard terminology for independence Metropolis-Hastings methods (see Gamerman and Lopes, 2006, Chapter 7). Finally, standard model selection criteria, formal or informal, exist and are more extensively discussed in the applications of Section 5. All our empirical findings suggest the RJMCMC as a reasonable model selection criterion and, for instance, the only one capable of sound factor model averaging.

5 Applications

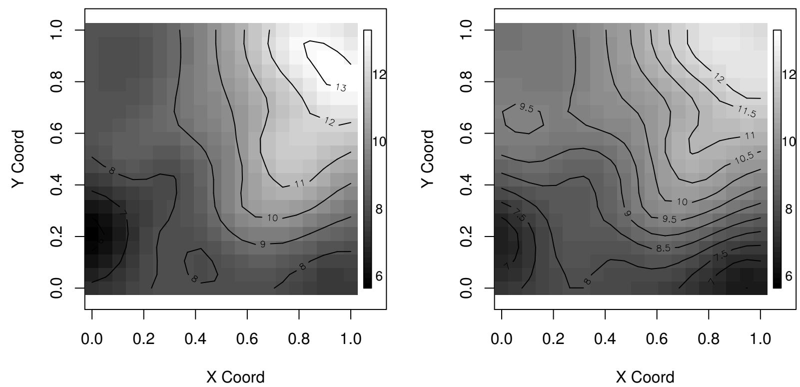
This section exemplifies the proposed spatial factor dynamic model in two situations. In the first case, space-time data is simulated from the model structure and customized MCMC and reversible jump MCMC algorithms implemented. In the second case, the levels of atmospheric concentrations of sulfur dioxide is investigated under the proposed model. The data were obtained from the Clean Air Status and Trends Network (CASTNet) and are weekly observed at 24 monitoring stations from 1998 to 2004.

5.1 Simulated study

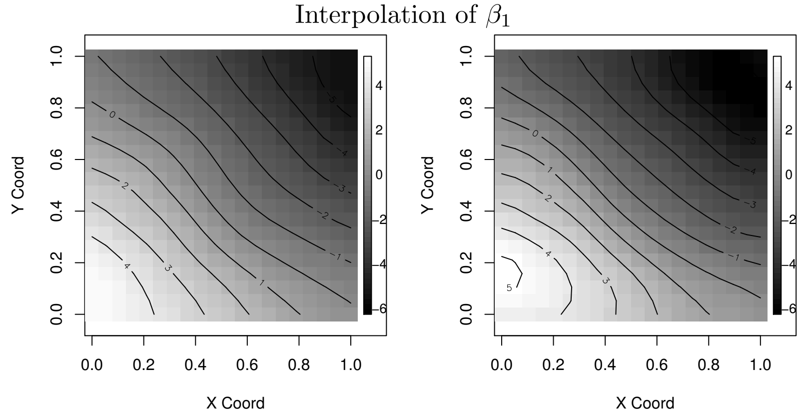
Initially, a total of 25 locations were randomly selected in the $[0, 1] \times [0, 1]$ square. Then, for $t = 1, \ldots, T = 100$, 25-dimensional vectors y_t are simulated from a dynamic 3-factor model, where *i)* $\Gamma = \mathrm{diag}(0.6, 0.4, 0.3)$ and $\Lambda = \mathrm{diag}(0.02, 0.03, 0.01)$, *ii)* the columns of the factor loadings matrix follow Gaussian processes with Matérn correlation functions with $\phi = (0.15, 0.4, 0.25)$, $\kappa = 1.5$ and $\tau = (1.00, 0.75, 0.56)$, and *iii)* $\mu_j^{\beta^*} = X^\beta \mu_j^\beta$, $\mu_1^\beta = (5, 5, 4)'$, $\mu_2^\beta = (5, -6, -7)'$, $\mu_3^\beta = (5, -8, 6)'$ and X^β contains a constant term and locations longitudes and latitudes, and *iv)* σ_i^2 were uniformly simulated in the interval $(0.01, 0.05)$, for $i = 1, \ldots, 25$. The last 10 observations were left out of the analysis for comparison purposes. Figure 1 presents the surfaces for $\beta_{(1)}, \ldots, \beta_{(3)}$ as well as some values of f_t and y_t. For $i = 1, \ldots, 25$, the prior distribution for σ_i^2 is $IG(\epsilon, \epsilon)$ with $\epsilon = 0.01$. For $j = 1, \ldots, 3$, $\lambda_j \sim IG(\epsilon, \epsilon)$, $\gamma_j \sim N(0.5, 1.0)$, $\tau_j^2 \sim IG(2, 0.75)$ and $\phi_j \sim IG(2, b)$ for $b = \max(dist)/(-2\ln(0.05))$ and $\max(dist)$ is the largest distance between locations, and μ_j^β normally distributed with mean equal to the true value and variance equal to 25.

Models with up to five common factors were entertained. Comparisons were based on posterior model probabilities (PMP), as well as commonly used information criteria, such as the AIC (Akaike 1974) and the BIC (Schwarz 1978), and goodness of fit statistics. Virtually all criteria point to the 3-factor model as the best model (see Table 1). Nonetheless, posterior model probabilities for factor models with $m = 2$ and $m = 4$ common factors are not negligible and could improve forecast and interpolation in a Bayesian model averaging set up.

For illustrative purposes, suppose that the 3-factor model is chosen for further analysis. Conditioning on $m = 3$, i.e. the true number of common factors, the MCMC algorithm outlined in Section 4 was run for a total of 50,000 iterations and posterior inference was based on the last 40,000 draws using every 10th member of the chains (a



Interpolation of β_1



Interpolation of β_2

Figure 2: *Simulated data:* Interpolation of the first two columns of the factor loadings matrix. True surfaces are the left contour plot on each panel, while interpolated ones are the right contour plot on each panel.

total of 4,000 posterior draws). Two chains were generated starting at different points of the parametric space. Convergence was analyzed through standard diagnostic tools. As an initial indication that the dynamic factor model is correctly capturing the right structure, all parameters are well estimated and all true values fall within the marginal 95% credibility intervals (Tables 2 and 3). It can be seen that most credibility intervals are not quite symmetrically around the posterior means, suggesting that (asymptotic) normal approximations would fail to properly account for model and parameter uncertainties. The model's goodness of fit is also evidenced by noticing the accuracy when estimating both the factor loadings matrix (see Figure 2) and the common dynamic factors (figure not provided).

The above results are encouraging and suggest that our procedure is able to correctly select the order of the factor model. Salazar (2008) performed several simulations under

m	AIC	BIC	MSE	MAE	MSE_P	PMP
2	1127.9	2643.3	0.11195	0.23822	2.3145	0.394
3	*-1005.5*	*1196.2*	*0.029108*	*0.13397*	2.3030	*0.437*
4	27589.7	30477.6	0.31112	0.41101	*2.2565*	0.101
5	42183.2	45757.3	0.46835	0.52360	2.2614	0.068

Table 1: *Simulated data:* Model comparison criteria. Akaike's information criterion - AIC; Schwartz's information criterion - BIC; Mean Squared Error - MSE $= N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\mu}_{it})^2$; Mean Absolute Error - MAE $= N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}|y_{it} - \hat{\mu}_{it}|$; MSE based on the last 10 predicted values - MSE_P; and Posterior Model Probability - PMP. Best models for each criterium appear in italic.

				Percentiles		
θ	True	$E(\theta)$	$\sqrt{Var(\theta)}$	2.5%	50%	97.5%
γ_1	0.60	0.504	0.091	0.325	0.504	0.687
γ_2	0.40	0.491	0.095	0.303	0.492	0.671
γ_3	0.30	0.416	0.100	0.209	0.417	0.623
λ_1	0.02	0.028	0.010	0.014	0.026	0.053
λ_2	0.03	0.019	0.004	0.011	0.018	0.028
λ_3	0.01	0.016	0.005	0.009	0.015	0.028

Table 2: *Simulated data:* Posterior summaries for the parameters characterizing the common factor's dynamics.

various spatial and temporal conditions and settings, with the great majority exhibiting good performance. In particular posterior model probabilities invariably selected the correct factor models orders.

5.2 SO_2 concentrations in eastern US

Spatial and temporal variations in the concentration levels of sulfur dioxide, SO_2, across 24 monitoring stations are examined through the proposed spatial dynamic factor model (see Figure 3). Weekly measurements in $\mu g/m^3$ are collected by the Clean Air Status and Trends Network (CASTNet), which is part of the Environmental Protection Agency (EPA) of the United States. Measurements span from the first week of 1998 to the 30th week of 2004, a total of 342 observations. The performance of the model's spatial interpolation is assessed based on stations BWR and SPD, which are left out of the analysis. Similarly, the model's forecasting performance is assessed based on the last 30 weeks, from the 1st week of 2004 to the 30th week of 2004. In summary, a total of $T = 312$ measurements on $N = 22$ stations are used in the analysis that follows.

Figure 4(a) shows the time series for the logarithm of SO_2 levels for four stations.

θ	True	$E(\theta)$	$\sqrt{Var(\theta)}$	Percentiles 2.5%	50%	97.5%
μ_{11}	5.00	4.44	0.89	2.78	4.43	6.21
μ_{21}	5.00	3.44	0.90	1.74	3.41	5.23
μ_{31}	4.00	4.10	0.85	2.44	4.09	5.76
τ_1^2	1.00	1.13	0.99	0.32	0.86	3.53
ϕ_1	0.15	0.20	0.08	0.10	0.19	0.40
μ_{12}	5.00	6.12	0.60	4.80	6.15	7.21
μ_{22}	-6.00	-6.00	0.80	-7.62	-5.99	-4.49
μ_{32}	-7.00	-7.51	0.86	-9.25	-7.45	-5.93
τ_2^2	0.75	0.51	0.91	0.11	0.30	2.29
ϕ_2	0.40	0.24	0.08	0.12	0.23	0.43
μ_{13}	5.00	4.53	0.58	3.39	4.52	5.67
μ_{23}	-8.00	-7.68	0.88	-9.43	-7.68	-6.01
μ_{33}	6.00	5.09	0.95	3.28	5.08	6.98
τ_3^2	0.56	0.44	0.40	0.15	0.34	1.36
ϕ_3	0.25	0.18	0.06	0.09	0.17	0.33

Table 3: *Simulated data:* Posterior summary for the spatial process parameters characterizing the columns of the factor loadings matrix.

Visually, the four time series exhibit seasonal behavior with an apparent annual cycle with higher values in winter. This seasonality can be explained in general terms as a result of higher rates of summertime oxidation of SO_2 to other atmospheric pollutants. Also, there seems to be a slight decrease in the time series trends over the years, probably due to implementation of EPA's Acid Rain Program in the eastern United Stated in 1995 (Phase I) and 2000 (Phase II). The logarithmic transformation normalizes the series quite reasonably (see Figure 4(b)) and will be retained hereafter. Calder (2007), for instance, also used log transformation to normalize the same SO_2 data. Nonetheless, before proceeding, a correction procedure that accounts for the effect of the curvature of the earth, commonly present when dealing with spatially distributed data, is applied to the data. Latitudes and longitudes were converted to the universal transverse Mercator (UTM) coordinates and the converted coordinates are measured in kilometers from the western-most longitude and the southern-most latitude observed in the data set. Four classes of spatial dynamic factor models are considered:

i) SSDFM(m,h): seasonal spatial dynamic model with m regular factors and h seasonal factors, $\mu_t^{y*} = 0$ and $\mu_j^{\beta*} = \mu_j^{\beta} 1_N$;

ii) SDFM(m)-cov: spatial dynamic m-factor model, common seasonal pattern, $\mu_t^{y*} = X^y \mu_t^y$ with $X^y = (1_N, lon, lat, lon^2, lon \times lat, lat^2, 1_N, 0)$ and $\mu_j^{\beta*} = \mu_j^{\beta} 1_N$;

iii) SDFM(m)-cov-GP: spatial dynamic m-factor model, common seasonal pattern,

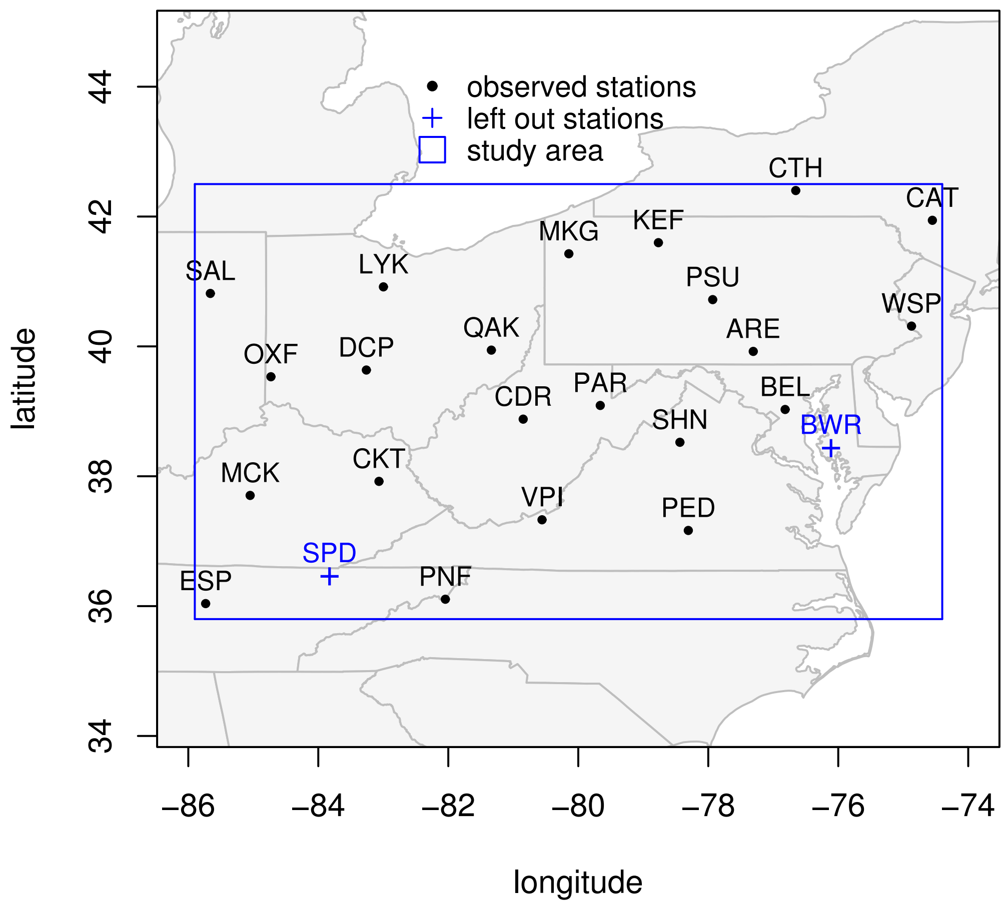


Figure 3: *CASTNet data:* Location of the monitoring stations. Stations SPD and BWR were left out for interpolation purposes.

$\mu_t^{y*} = X^y \mu_t^y$ with $X^y = (1_N, 1_N, 0)$ and $\mu_j^{\beta^*} = X^\beta \mu_j^\beta$ with $X^\beta = (1_N, lon, lat, lon^2, lon \times lat, lat^2)$;

iv) SSDFM(m, h)-cov-GP: seasonal spatial dynamic m-factor model, h seasonal factors, $\mu_t^{y*} = 0$ and $\mu_j^{\beta^*} = X^\beta \mu_j^\beta$ with $X^\beta = (1_N, lon, lat, lon^2, lon \times lat, lat^2)$.

The last two columns of X^y for models ii) and iii) correspond to the design matrix associated with the seasonal coefficients (see also Section 2.3). In those models, a common seasonal structure is considered for all monitoring stations. In model iv, this assumption is relaxed with station-specific loadings.

Each model was tested with a maximum number of factors (never larger than 6) and $h = 1$ harmonic component with cycles of 52 weeks. Models with more factors were analyzed but results are not reported when additional parameters are not statistically relevant. The correlation functions of the Gaussian processes are all Matérn, except SSDFM(m, h) which is also fitted with an exponential correlation function.

For comparison purpose, the following two benchmark spatio-temporal models were considered:

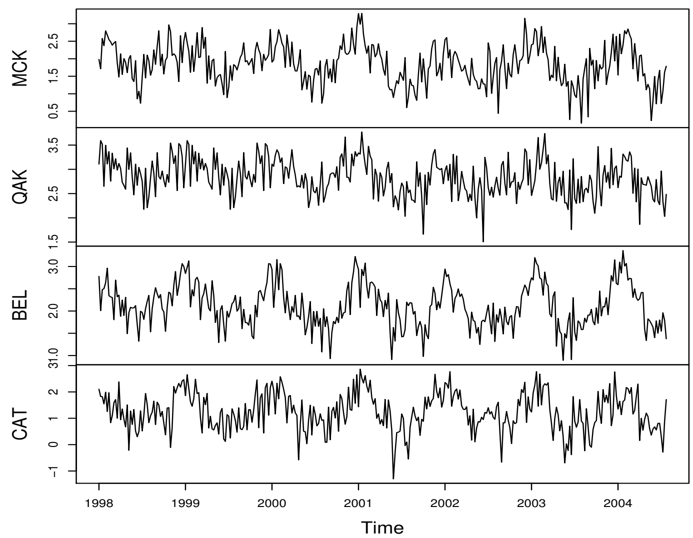
i) SGSTM: standard geostatistical spatio-temporal model, $y_t = \mu_t^{y*} + \nu_t + \epsilon_t$, $\epsilon_t \sim N(0, \sigma^2 I_N)$, $\mu_t^{y*} = X\mu_t$, $\mu_t|\mu_{t-1} \sim N(G\mu_{t-1}, W)$, $\nu_t \sim GP(0, \tau^2 R_\phi)$ with $X = (1_N, lon, lat, lon^2, lon \times lat, lat^2, 1_N, 0_N)$.

ii) SGFM(m): standard geostatistical m-factor model, $y_t = \beta f_t + \mu_t^{y*} + \nu_t + \epsilon_t$, $\beta_{j,j} = 1$, $\beta_{j,k} = 0$ $(k > j = 1, \ldots, m)$, μ_t^{y*}, ν_t and ϵ_t as in SGSTM.

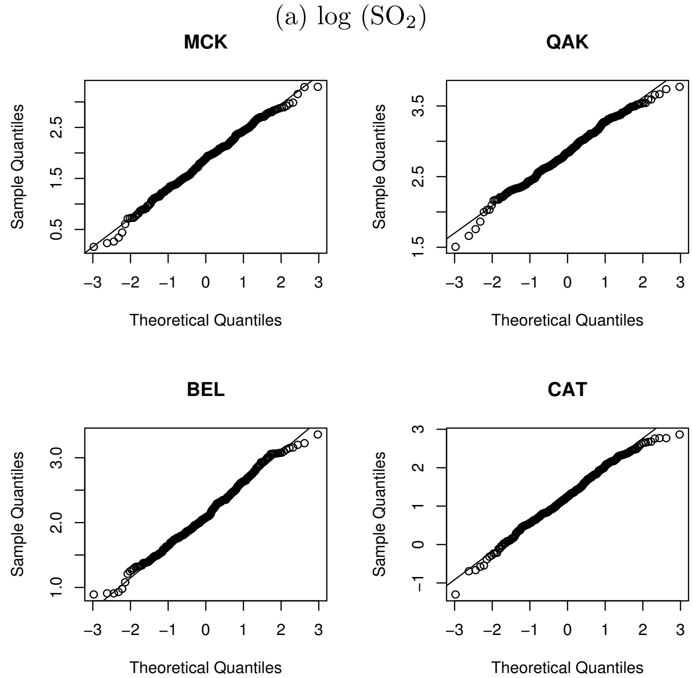
In SGSTM the temporal variation is explained, like in our proposal, through μ_t^{y*}, while the spatial variation is explained, unlike our proposal, through (temporally) independent geostatistical components ν_t. SGFM is an elaboration of SGSTM that incorporate dynamic factors. In SGFM the temporal variation is explained, like in our proposal, through μ_t^{y*}, while the spatial variation is explained, unlike our proposal, through a dynamic factor term βf_t, the difference being the absence of spatial dependence in the factor loadings β.

Relatively vague prior distributions were used for most parameters. More specifically, σ^2s are $IG(0.01, 0.01)$, λs are $IG(0.01, 0.01)$, γs are $Ntr_{(-1,1)}(0, 1)$, Λs are $IW(0.01I_2, 2)$. Additionally, the mixture prior for γ, with $\alpha = 0.5$, was implemented to allow for non-stationary common factors. Reference priors were used for the parameters of Gaussian processes with exponential correlation function. For the remainder models, the smoothness parameter, κ, of Matérn correlation functions was set as follows. First, standard normal linear m-factor models are fitted, for $m = 1, 2, 3$, and estimates of the columns of the factor loading matrices are used to model GP with reference priors. Then, the Bayes factor for a model with smoothness parameter κ equal to 7 (for κ in $\{1, \ldots, 10\}$) was selected in most cases (see Berger et al. 2001 for further details). This value is assumed for all entertained models. The prior specification for the remainder parameters of the Gaussian processes are as follows: τ^2s are $IG(2, 1)$, ϕs are $IG(2, b)$ and μ^βs are $N(a, 1)$, with $b = \max(dist)/(-2\ln(0.05))$, $\max(dist)$ is the largest distance between locations, and a is the absolute mean of observations.

The MCMC algorithm was run as in Section 5.1 but with a burn-in of 25000 draws. Competing models can be compared based on their posterior model probabilities (PMP), as well as their sum of square errors (SSE) and sum of absolute errors (SAE). Additionally, mean square errors (MSE) based on forecast and interpolated values were used for model comparison. From Table 4, it can be seen that between models with highest PMP in each class, the model SSDFM(4,1)-cov-GP has the best performance. Additionally, Table 4 shows the results for the benchmark models. It shows that the these standard models are outperformed in terms of forecasting in time and interpolation in space despite having a better fit. The results seem to indicate that the structure imposed by



(a) log (SO_2)



(b) Normal Q-Q Plot

Figure 4: *CASTNet data:* (a) Time series of weekly log (SO_2) concentrations at MCK, QAK, BEL and CAT stations. (b) Normal Q-Q plot for time series plotted in (a).

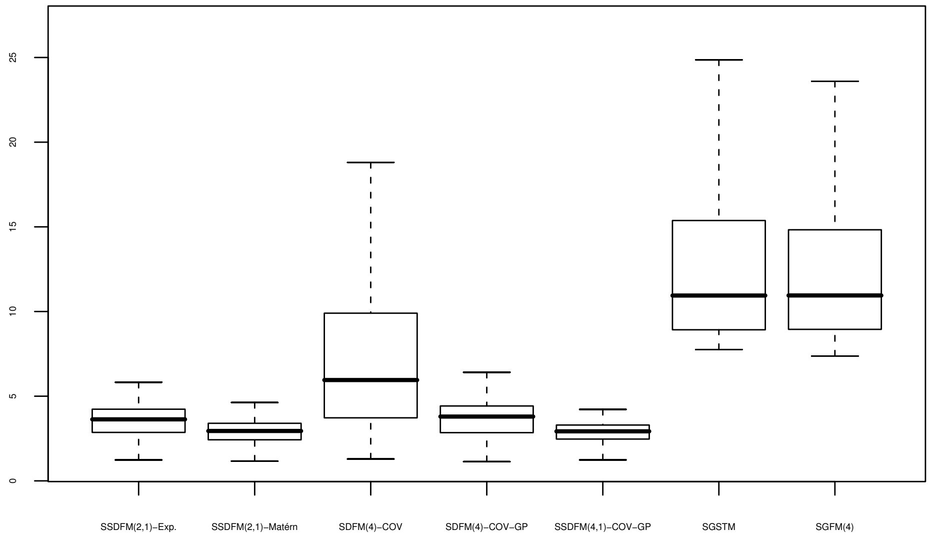


Figure 5: *CASTNet data:* Sharpness diagram for SSDFM(2,1)-Exp, SSDFM(2,1)-Matérn, SDFM(4)-cov, SDFM(4)-cov-GP, SSDFM(4,1)-cov-GP, SGSTM and SGFM(4) forecasts of weekly log(SO_2). The box plots show percentiles of the width of the 90% central prediction intervals.

our models is in fact required to improve predictions. Interpolation was not performed for SGFM because of the absence of spatial structure in the factor loadings.

Further evaluation of the predictive performance can be made using criteria proposed by Gneiting, Balabdaoui, and Raftery (2007). The main criteria suggested there are sharpness and scoring rules. Sharpness is evaluated through the width of the prediction intervals; the shorter, the sharper. Table 5 and Figure 5 show different measures of sharpness. They point to model SSDFM(4,1)-cov-GP as the sharpest. Scoring rules were also considered, as they can address calibration and sharpness simultaneously. The mean score $S(F,y) = N^{-1}H^{-1}\sum_{i=1}^{N}\sum_{h=1}^{H} S(F_{i,T+h}, y_{i,T+h})$ can be used to summarize them for any strictly proper scoring rule S. The smaller $S(F,y)$, the better. In particular, the logarithm score (LS) and the continuous ranked probability score (CRPS) are suggested by Gneiting, Balabdaoui, and Raftery (2007). LS is the negative of the logarithm of the predictive density evaluated at the observation. For each $y_{i,T+h}$, the CRPS is defined as

$$CRPS(F_{i,T+h}, y_{i,T+h}) = E_F|\hat{y}_{i,T+h} - y_{i,T+h}| - \frac{1}{2}E_F|\hat{y}_{i,T+h} - \tilde{y}_{i,T+h}|,$$

where $\hat{y}_{i,T+h}$ and $\tilde{y}_{i,T+h}$ are values from $p(y_{T+h}|y)$. This is a convenient representation because $p(y_{T+h}|y)$ is easily approximated by a sample based on MCMC output (see Gschlößl and Czado 2005 for more details). LS and CRPS values for the five best models are also presented in Table 5 and again rank SDFM(4,1)-cov-GP model as the top model, with lowest LS and CRPS.

Model	m	SSE	SAE	MSE_P	MSE_I	PMP
SSDFM$(m,1)$-Exp.	1	733.58	1651.8	0.76	0.15	0.43
	2	*594.30*	*1477.5*	*0.22*	0.16	*0.56*
	3	799.13	1738.2	0.23	*0.13*	0.01
SSDFM$(m,1)$-Matérn	1	733.66	1651.2	0.58	0.15	0.37
	2	*632.45*	*1539.6*	*0.25*	0.19	*0.54*
	3	802.20	1734.2	0.27	0.15	0.09
SDFM(m)-cov	1	544.63	1418.0	0.368	0.167	0.13
	2	473.29	1306.0	*0.237*	0.178	0.18
	3	420.12	1217.6	0.249	0.171	0.20
	4	*375.10*	*1150.0*	0.271	0.160	*0.33*
	5	621.26	1486.1	0.245	*0.130*	0.09
	6	547.83	1412.7	0.251	0.147	0.07
SDFM(m)-cov-GP	1	856.25	1811.8	0.384	*0.127*	0.04
	2	636.97	1549.0	0.638	0.167	0.13
	3	502.54	1348.5	0.308	0.148	0.26
	4	*462.57*	*1276.6*	*0.260*	0.192	*0.30*
	5	536.13	1425.6	0.498	0.213	0.16
	6	543.13	1415.9	0.304	0.216	0.11
SSDFM$(m,1)$-cov-GP	1	753.86	1673.3	0.651	0.153	0.00
	2	570.13	1450.7	0.288	0.161	0.23
	3	484.78	1320.2	0.255	*0.149*	0.31
	4	*450.95*	*1276.1*	0.229	0.158	*0.40*
	5	573.15	1446.4	*0.218*	0.165	0.06
SGSTM	-	177.26	883.6	0.341	0.172	-
SGFM(m)	4	264.18	1069.7	0.322	-	-

Table 4: *CASTNet data:* Model comparison criteria. Sum Squared Error - SSE $= \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\mu}_{it})^2$; Sum Absolute Error - SAE $= \sum_{i=1}^{N}\sum_{t=1}^{T}|y_{it} - \hat{\mu}_{it}|$; Predictive MSE (based on the last 30 weeks, 2004:01 to 2004:30) - $\text{MSE}_P = N^{-1}H^{-1}\sum_{i=1}^{N}\sum_{h=1}^{H}(y_{i,T+h} - E(y_{i,T+h}|y))^2$; Interpolation MSE (based on 312 measurements for stations BWR and SPD) - $\text{MSE}_I = N_n^{-1}T^{-1}\sum_{i=1}^{N_n}\sum_{t=1}^{T}(y_{N+i,t} - E(y_{i,t}|y))^2$, and Posterior Model Probability - PMP. Best models for each criterium appear in italic.

Model	LogS	CRPS	AW90
SSDFM(2, 1)-Exp	11.286	0.762	3.517
SSDFM(2, 1)-Matérn	11.567	0.650	2.906
SDFM(4)-cov	21.049	1.734	7.486
SDFM(4)-cov-GP	11.463	0.813	3.619
SSDFM(4, 1)-cov-GP	10.045	0.644	2.855
SGSTM	43.875	2.577	13.066
SGFM(4)	43.814	2.524	12.955

Table 5: *CASTNet data:* Forecast evaluation. Average logarithmic score (LogS), average continuous ranked probability score (CRPS) and average width of 90% central prediction intervals (AW90).

Table 5 clearly separates the models proposed in one group and the benchmark models in another group. The within-group variation, for instance, is substantially smaller than the between-group variation showing stability of our proposed methodology across a wide range of models. The improvement in the results is substantial throughout. These figures seem to jointly indicate the SDFM(4,1)-cov-GP model and so, the remainder of the analysis is conducted under this specification. Table 6 presents the results about the temporal variation of the factors. They present a wide variety of autoregressive dependence, ranging from no dependence or white noise (1st factor) to indications of non-stationarity (4th factor). The 1st factor is common across locations and should not be confused with the idiosyncratic errors that are different for every location. The other factors exhibit significant temporal dependence.

Factors may be identified according to their relative weight in the explanation of the data variability. On average the larger proportion of the data variability is associated with the 4th and the seasonal factors. They respectively account for 21% and 15% of the data variability. The 3rd factor appears after that with around 11% followed by the 1st factor with 4% and the 2nd factor with 3%.

The fourth common factor represents the grand mean as it is fairly common in factor analysis applications (see Rencher 2002 for more details). It accounts for the global time-trend variability of the series (see Figure 6). The first three factors are rather noisy but of limited variation, while the seasonal common factor is capturing the time series annual cycles. The fourth common factor, however, exhibits a rather nonstationary behavior, which is emphasized by the posterior density of γ_4 being fairly concentrated around one.

In order to verify the presence of a nonstationary factor, the mixture prior for γ was implemented with $\hat{p}(\gamma_1 = 1|y) = \hat{p}(\gamma_2 = 1|y) = \hat{p}(\gamma_3 = 1|y) = 0$ and $\hat{p}(\gamma_4 = 1|y) = 0.41$ observed. In other words, the first three factors are stationary, while the fourth common factor is stationary with roughly 60% posterior probability. Additionally, the SSDFM(4,1)-cov-GP model with $\gamma_4 = 1.0$ was analyzed and compared to its stationary counterpart and results (not shown here) show higher SSE, SAE and MSE based on

θ	$E(\theta)$	$\sqrt{Var(\theta)}$	Percentiles 2.5%	50%	97.5%
γ_1	0.009	0.069	-0.122	0.010	0.147
γ_2	0.186	0.069	0.053	0.185	0.319
γ_3	0.354	0.091	0.175	0.356	0.522
γ_4	0.997	0.002	0.992	0.997	1.000
λ_1	0.005	0.003	0.002	0.004	0.011
λ_2	0.003	0.001	0.001	0.002	0.005
λ_3	0.002	0.002	0.001	0.002	0.007
λ_4	0.002	0.001	0.001	0.002	0.003
λ_5	0.004	0.003	0.001	0.002	0.012
λ_6	0.002	0.001	0.001	0.001	0.003

Table 6: *CASTNet data:* Posterior summaries for the parameters characterizing the common factors' dynamics in the SSDFM(4,1)-cov-GP model.

forecasted values.

Table 7 present posterior summaries for the regression and remaining spatial dependence of the factor loadings. Latitude and longitude are important to describe their mean levels, but not their square and interaction terms. They imply that the posterior mean for correlation of factor loadings that are distant by 100 kilometers are 0.050, 0.063, 0.151, 0.051 and 0.061, respectively.

Figure 7 presents the mean surfaces for the columns of the factor loadings matrix β obtained by interpolation (*Bayesian kriging*, see Section 3.2). Loadings for the fourth factor are shown to be higher in the center of the interpolated area, mainly around station QAK in Ohio. Simple exploratory data analysis indicates that the highest values of SO_2 concentrations were measured at QAK, confirming the role of the fourth factor as a grand mean. Another interesting finding is that the loadings for the seasonal (fifth) factor are smaller in the highly industrialized state of Ohio and its behavior is quite the opposite to the fourth (grand mean) factor. The loadings for the first and third factors seem to be higher at the southwestern corner of the area of study and the second factor promotes a divide between east and west with higher values at the latter region. The intuitive combination of the temporal characterization of the common dynamic factors and the spatial characterization of the columns of the factor loadings matrix is one of the key features of the proposed model inherited from traditional factor analysis.

Forecasting and interpolation results are presented in Figure 8, which exhibits encouraging out-of-sample properties of the model, with data points being accurately forecast and interpolated for several steps ahead and out-of-sample monitoring stations, respectively. It is worth noting that none of the 95% credibility intervals, either based on forecasting or interpolation, are symmetric. The interpolation exercise seems to produce better fit mainly because of the presence of the neighboring structure amongst monitoring stations. The forecasting exercise relies exclusively on previous observa-

	1st column			2nd column		
	Mean	Median	95% C.I.	Mean	Median	95% C.I.
μ^{β}_{1j}	1.56	1.56	[0.61,2.49]	1.55	1.53	[0.62,2.52]
μ^{β}_{2j}	0.79	0.78	[0.00,1.71]	0.72	0.69	[-0.07,1.67]
μ^{β}_{3j}	0.92	0.91	[0.04,1.85]	1.06	1.05	[0.24,1.97]
μ^{β}_{4j}	-0.12	-0.12	[-0.27,0.01]	-0.17	-0.16	[-0.33,-0.04]
μ^{β}_{5j}	0.07	0.07	[-0.17,0.32]	0.03	0.03	[-0.18,0.24]
μ^{β}_{6j}	-0.31	-0.30	[-0.55,-0.08]	-0.14	-0.14	[-0.35,0.04]
τ^2_j	6.18	5.08	[1.52,18.20]	3.96	3.17	[0.96,11.00]
ϕ_j	32.70	32.32	[21.47,46.97]	35.51	35.06	[22.53,51.21]
	3rd column			4th column		
	Mean	Median	95% C.I.	Mean	Median	95% C.I.
μ^{β}_{1j}	1.74	1.75	[0.77,2.68]	1.32	1.32	[0.62,2.00]
μ^{β}_{2j}	1.07	1.06	[0.14,2.08]	0.20	0.18	[-0.15,0.61]
μ^{β}_{3j}	1.04	1.08	[-0.30,2.17]	0.76	0.76	[0.31,1.27]
μ^{β}_{4j}	-0.16	-0.15	[-0.35,0.00]	-0.04	-0.04	[-0.09,0.00]
μ^{β}_{5j}	0.13	0.12	[-0.15,0.45]	0.03	0.03	[-0.03,0.11]
μ^{β}_{6j}	-0.16	-0.16	[-0.41,0.08]	-0.12	-0.11	[-0.22,-0.03]
τ^2_j	14.10	8.41	[0.91,48.80]	0.43	0.36	[0.16,1.07]
ϕ_j	52.79	54.38	[24.09,77.46]	33.04	32.87	[21.17,46.82]
	5th column					
	Mean	Median	95% C.I.			
μ^{β}_{1j}	1.62	1.62	[0.85,2.36]			
μ^{β}_{2j}	0.09	0.06	[-0.25,0.59]			
μ^{β}_{3j}	0.19	0.17	[-0.26,0.75]			
μ^{β}_{4j}	0.00	0.00	[-0.06,0.03]			
μ^{β}_{5j}	0.00	0.00	[-0.07,0.07]			
μ^{β}_{6j}	-0.03	-0.02	[-0.12,0.05]			
τ^2_j	0.45	0.35	[0.14,1.29]			
ϕ_j	35.14	34.73	[21.22,51.95]			

Table 7: *CASTNet data:* Posterior summary for the spatial process parameters characterizing the columns of the factor loadings matrix in the SSDFM(4,1)-cov-GP model. C.I. stands for credibility interval.

tions, so the credibility intervals are more conservative. The predictive performance of our models is superior when compared to benchmark spatio-temporal models. Finally, Figure 9 presents the mean surfaces of SO_2 levels for nine weeks in 2003. It is clear that some parts of the map are more affected by the seasonal factor (see, for instance, the region around stations SAL in Indiana and PSU in Pennsylvania) while other parts are less affected (see, for instance, the region around station QAK). The top half of the area with time-varying portions in the east-west direction defines the region with the highest levels of SO_2 throughout the year, indicating that the proposed model is capable of accommodating both spatial and temporal nonstationarities in a nonseparable fashion.

6 Conclusions

This paper introduces the spatial dynamic factor model, which is a new class of nonseparable and nonstationary spatiotemporal models that generalizes several of the existing alternatives. It uses factor analysis ideas to frame and exploit both the spatial and the temporal dependencies of the observations. The spatial variation is brought into the modeling conditionally through the columns of the factor loadings matrix, while the time series dynamics are captured by the common dynamic factors. One of the main contributions of the paper is the exploration of factor analysis arguments in spatio-temporal models in order to explicitly model spatial and temporal components. The model takes advantage of well established literature for both spatial processes and multivariate time series processes. The matrix of factor loadings plays the important role of weighing the common factors in general factor analysis and is here incumbent of modeling spatial dependence. Similarly, the common factors follow time series decomposition processes, such as local and global trends, cycle and seasonality.

Conditional on the number of factors, posterior inference is facilitated by a customized MCMC algorithm that combines well established schemes, such as the forward filtering backward sampling algorithm, with standard normal and inverse gamma updates. Inference across models, i.e. the selection of the number of common factors, is performed by a computationally and practically feasible reversible jump MCMC algorithm that builds proposal densities based on short and preliminary MCMC runs. The true number of factors was given the highest posterior model probability and the parameters of the modal model were accurately estimated, including the dynamic common factors and the spatial loading matrix.

The applications exploited the potential of the proposed model both as an interpolation tool and as a forecasting tool in the spatiotemporal context. The factor model structure allowed direct incorporation of exogenous, predetermined variables into the analysis to help explaining both at the level of the response and at the level of the factor loadings matrix. It also was able to capture the local behavior of SO_2 as well as its different temporal components for different locations. The comparison with benchmark models support our initial claim that the proposed spatial dynamic factor model has superior performance when based on a host of predictive measures. It reinforces the need for more complex structures, such as those proposed in our paper, to adequately address issues associated with real data analysis. Standard models are not capable of handling the spatio-temporal heterogeneity present in environmental application. We anticipate that the same is true in many other areas of applications.

The flexibility of the spatial dynamic factor model is promising and a few generalizations are currently under investigation, such as time-varying factor loadings to dynamically link the latent spatial processes (Lopes 2000, Lopes and Carvalho 2007 and Gamerman, Salazar, and Reis 2007). Another interesting direction is to allow binomial and Poisson responses by replacing the first level normal likelihood by an exponential family representation. In this case, the spatial dynamic factors would be used to model transformations of mean functions. Finally, non-diagonal idiosyncratic covariance matrix and more general dynamic factor structures can be considered to incorporate, for example, remaining spatial correlation and AR(p) structures, respectively. One can argue that the availability of well known and reliable statistical tools coupled with highly efficient, and by now well established, MCMC schemes and plenty of room for extensions will make this area of research flourish in the near future.

Appendix

The full conditional distribution of all parameters in model (9) are listed here. Namely, the idiosyncratic variances, σ, the common factor dynamics, γ, the common factors' variances, λ, the loadings means, μ, the spatial hyperparameters, τ_j^2 and ϕ_j, the factor loadings matrix, β, and the common factors, f_t, for $t = 1, \ldots, T$. Throughout this appendix $[\theta]$ and $p(\theta|\ldots)$ denote, respectively, the full conditional distribution and full conditional density of θ conditional on all other parameters. Also, for $m \times n$ and $s \times t$ matrices A and B, the Kronecker product $A \otimes B$ is the $ns \times nt$ matrix that inflates matrix A by multiplying each of its components by the whole matrix B.

Idiosyncratic variances From the likelihood presented in Section 2.4, it can be shown that $y_i|F, \sigma_i^2, \beta_i \sim N(F\beta_i, \sigma_i^2 I)$, $i = 1, \ldots, N$, where y_i is the i^{th} column of y, β_i is the i^{th} row of β. Therefore, $[\sigma_i^2] \sim IG((T + n_\sigma)/2, ((y_i - F\beta_i)'(y_i - F\beta_i) + n_\sigma s_\sigma)/2)$.

Common factors variances $[\lambda_j] \sim IG((T-1+n_\lambda)/2, (\sum_{t=2}^{T}(f_{jt} - \gamma_j f_{j,t-1})^2 + n_\lambda s_\lambda)/2)$.

Loadings means $[\mu_j] \sim N(m^*_{\mu_j}, S^*_{\mu_j})$, $m^*_{\mu_j} = S^*_{\mu_j}\left[\tau_j^{-2}\beta'_{(j)}R^{-1}_{\phi_j}1_N + m_\mu S_\mu^{-1}\right]$ and $S^{*-1}_{\mu_j} = \tau_j^{-2}1'_N R^{-1}_{\phi_j}1_N + S_\mu^{-1}$.

Factor loadings The factor loadings matrix is jointly sampled. To that end, Equation (1) is rewritten as $y_t = f_t^*\beta^* + \epsilon_t$, where $f_t^* = f_t' \otimes I_N$ and $\beta^* = (\beta'_{(1)}, \ldots, \beta'_{(m)})'$ are $N \times Nm$ and $Nm \times 1$ matrices, where $A \otimes B$ denotes the Kronecker product of matrices A and B. Similarly, the prior distribution of β^* is $\beta^* \sim N(\mu_{\beta^*}, \Sigma_{\beta^*})$, where $\mu_{\beta^*} = \mu \otimes 1_N$, $\Sigma_{\beta^*} = \Sigma_\beta \otimes R_\phi$ and $\Sigma_\beta = \text{diag}(\tau_1^2, \ldots, \tau_m^2)$. From standard Bayesian multivariate regression (Box and Tiao (1973)), it can be shown that $[\beta^*] \sim N(\tilde{\mu}_{\beta^*}, \tilde{\Sigma}_{\beta^*})$, where $\tilde{\Sigma}_{\beta^*}^{-1} = \sum_{t=1}^T f_t^{*\prime}\Sigma^{-1}f_t^* + \Sigma_{\beta^*}^{-1}$ and $\tilde{\mu}_{\beta^*} = \tilde{\Sigma}_{\beta^*}\left(\sum_{t=1}^T f_t^{*\prime}\Sigma^{-1}y_t + \Sigma_{\beta^*}^{-1}\mu_{\beta^*}\right)$.

Common factors dynamics It follows from (2) that $f_{jt} \sim N(\gamma_j f_{j,t-1}, \lambda_j)$, $j = 1,\ldots,m$ and $t = 2,\ldots,T$. Therefore, $p(\gamma_i|\ldots) \propto \prod_{t=2}^T p(f_{jt}|f_{j,t-1},\gamma_i,\lambda_i)\ p(\gamma_i|m_\gamma, S_\gamma,\alpha)$, so i) if $\alpha = 1$, $[\gamma_j] \sim Ntr_{(-1,1)}(m_{\gamma_j}^*, S_{\gamma_j}^*)$ where $S_{\gamma_j}^{*-1} = \lambda_j^{-1}\sum_{t=2}^T f_{j,t-1}^2 + S_\gamma^{-1}$ and $m_{\gamma_j}^* = S_{\gamma_j}^*\left[\lambda_j^{-1}\sum_{t=2}^T f_{jt}f_{j,t-1} + m_\gamma S_\gamma^{-1},\right]$, and ii) if $\alpha \in (0,1)$ draw γ_j with probability α^* using the normal distribution $Ntr_{(-1,1)}(m_{\gamma_j}^*, S_{\gamma_j}^*)$ or let $\gamma_i = 1$ with probability $1 - \alpha^*$ where $\alpha^* = A/(A+B)$, $A = \alpha C S_\gamma^{-1/2} S_{\gamma_j}^{*1/2} \exp\{-0.5[\lambda^{-1}\sum_{t=2}^T f_{jt}^2 + S_\gamma^{-1}m_\gamma - S_{\gamma_j}^{*-1}m_{\gamma_j}^{*2}]\}$, $C = [\Phi((1-m_{\gamma_j}^*)/S_{\gamma_j}^{*1/2}) - \Phi((-1-m_{\gamma_j}^*)/S_{\gamma_j}^{*1/2})][\Phi((1-m_\gamma)/S_\gamma^{1/2}) - \Phi((-1-m_\gamma)/S_\gamma^{1/2})]^{-1}$, $B = (1-\alpha)\exp\{-0.5\lambda_j^{-1}\sum_{t=2}^T(f_{jt} - f_{j,t-1})^2\}$ and Φ is the one-sided probability from the standard normal.

Common factors The vectors of common factors, i.e. $f_1,\ldots,f_T$, are sampled jointly by means of the well known *forward filtering backward sampling* (FFBS) scheme of Carter and Kohn (1994) and Frühwirth-Schnatter (1994), which explores, conditionally on β and Θ, the following backward decomposition $p(F|y) = \prod_{t=0}^{T-1} p(f_t|f_{t+1}, D_t)p(f_T|D_T)$, where $D_t = \{y_1,\ldots,y_t\}$, $t = 1,\ldots,T$ and D_0 represents the initial information. Starting with $f_0 \sim N(m_0, C_0)$, it can be shown that $f_t|D_t \sim N(m_t, C_t)$, where $m_t = a_t + A_t(y_t - \tilde{y}_t)$, $C_t = R_t - A_tQ_tA_t'$, $a_t = \Gamma m_{t-1}$, $R_l = \Gamma C_{t-1}\Gamma' + \Lambda$, $\tilde{y}_t = \beta a_t$, $Q_t = \beta R_t\beta' + \Sigma$ and $A_t = R_t\beta' Q_t^{-1}$, for $t = 1,\ldots,T$. f_T is sampled from $p(f_T|D_T)$. This is the forward filtering step. For $t = T-1,\ldots,2,1,0$, f_t is sampled from $p(f_t|f_{t+1},D_t) = f_N(f_t;\tilde{a}_t,\tilde{C}_t)$, where $\tilde{a}_t = m_t + B_t(f_{t+1} - a_{t+1})$, $\tilde{C}_t = C_t - B_tR_{t+1}B_t'$ and $B_t = C_t\Gamma' R_{t+1}^{-1}$. This is the backward sampling step.

Spatial hyperparameters By combining the inverse gamma prior density form (5) or the reference prior density from (6) with the likelihood function from (9), it follows that $[\tau_j^2] \sim IG(n_{\tau_j}^*/2, n_{\tau_j}^* s_{\tau_j}^*/2)$, where $n_{\tau_j}^* = N + n_\tau$ and $n_{\tau_j}^* s_{\tau_j}^* = (\beta_{(j)} - \mu_j 1_N)' R_{\phi_j}^{-1}(\beta_{(j)} - \mu_j 1_N) + n_\tau s_\tau$ when inverse gamma prior distributions are used, and $n_{\tau_j}^* = N$ and $n_{\tau_j}^* s_{\tau_j}^* = (\beta_{(j)} - \mu_j 1_N)' R_{\phi_j}^{-1}(\beta_{(j)} - \mu_j 1_N)$ when reference prior distributions are used. The full conditional density of ϕ_j has no known form and a Metropolis-Hastings step is implemented. A candidate draw $\tilde{\phi}_j$ is generated from a log-normal distribution with location $\log\phi_j$ and scale Δ_ϕ, i.e., $q_j(\phi_j,\cdot) = f_{LN}(\cdot;\log\phi_j,\Delta_\phi)$. Δ_ϕ is a *tuning parameter* and is frequently used to calibrate the proposal density. The candidate draw is accepted with probability

$$\alpha(\phi_j,\tilde{\phi}_j) = \min\left\{1, \frac{f_N(\beta_{(j)};\mu_j 1_N, \tau_j^2 R_{\tilde{\phi}_j})\pi_P(\tilde{\phi}_j)\,\tilde{\phi}_j}{f_N(\beta_{(j)};\mu_j 1_N, \tau_j^2 R_{\phi_j})\pi_P(\phi_j)\,\phi_j}\right\},$$

where π_P is either an inverse gamma prior, i.e., π_{IG} or the reference prior, i.e., π_R.

References

Akaike, H. (1974). "New look at the statistical model identification." *IEEE Transactions in Automatic Control*, AC–19: 716–723. 770

Bai, J. (2003). "Inferential theory for factor models of large dimensions." *Econometrica*, 71: 135–71. 762

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data.*. CRC Press/Chapman & Hall. 759, 766

Berger, J. O., De Oliveira, V., and Sansó, B. (2001). "Objective Bayesian analysis of spatially correlated data." *Journal of the American Statistical Association*, 96: 1361–1374. 766, 775

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Massachusetts: Addison-Wesley. 783

Brown, P. J., Vannucci, M., and Fearn, T. (1998). "Multivariate Bayesian variable selection and prediction." *Journal of the Royal Statistical Society, Series B*, 60: 627–41. 765

Calder, C. (2007). "Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment." *Environmental and Ecological Statistics*, 14: 229–47. 761, 762, 764, 773

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models." *Biometrika*, 81: 541–553. 768, 784

Carvalho, C. M., Chang, J., Lucas, J., Wang, Q., Nevins, J., and West, M. (2008). "High-dimensional sparse factor modelling - Applications in gene expression genomics." *Journal of the American Statistical Association (to appear)*. 763

Christensen, W. and Amemiya, Y. (2002). "Latent variable analysis of multivariate spatial data." *Journal of the American Statistical Association*, 97(457): 302–317. 760

— (2003). "Modeling and prediction for multivariate spatial factor analysis." *Journal of Statistical Planning and Inference*, 115: 543–564. 760

Clyde, M. (1999). "Bayesian model averaging and model search strategies." In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 6*, 157–85. London: Oxford University Press. 768

Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley. 762

Cressie, N. and Huang, H. (1999). "Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions." *Journal of the American Statistical Association*, 94(448): 1330–1340. 764

Dawid, A. P. (1981). "Some matrix-variate distribution theory: Notational considerations and a Bayesian application." *Biometrika*, 68: 265–274. 765

Dellaportas, P., Forster, J., and Ntzoufras, I. (2002). "On Bayesian model and variable selection using MCMC." *Statistics and Computing*, 12: 27–36. 769

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). "The Generalized dynamic-factor model: identification and estimation." *Review of Economics and Statistics*, 82: 540–54. 762

Frühwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models." *Journal of Time Series Analysis*, 15(2): 183–202. 768, 784

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. Chapman&Hall/CRC, 2nd edition. 759

Gamerman, D., Moreira, A. R. B., and Rue, H. (2003). "Space-varying regression models: specifications and simulation." *Computational Statistics and Data Analysis*, 42: 513–33. 761

Gamerman, D., Salazar, E., and Reis, E. (2007). "Dynamic Gaussian process priors, with applications to the analysis of space-time data (with discussion)." In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 8*. London: Oxford University Press. (in press). 783

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). "Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing." *Journal of the American Statistical Association*, 100: 1021–35. 762

George, E. I. and McCulloch, R. E. (1992). "Variable selection via Gibbs sampling." *Journal of the American Statistical Association*, 79: 677–83. 769

Gneiting, T. (2002). "Nonseparable, Stationary Covariance Functions for Space-Time Data." *Journal of the American Statistical Association*, 97(458): 590–600. 764

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). "Probabilistic forecasts, calibration and sharpness." *Journal of the Royal Statistical Society, Series B*, 69(2): 243–268. 777

Godsill, S. (2001). "On the relationship between Markov chain Monte Carlo methods for model uncertainty." *Journal of Computational and Graphical Statistics*, 10: 1–19. 769

Gschlößl, S. and Czado, C. (2005). "Spatial modelling of claim frequency and claim size in insurance." Discussion Paper 461, Ludwig-Maximilians-Universität München, Munich. 777

Hogan, J. W. and Tchernis, R. (2004). "Bayesian factor analysis for spatially correlated data, with application to summarizing area-Level material deprivation from census data." *Journal of the American Statistical Association*, 99: 314–324. 760

Lasinio, G. J., Sahu, S. K., and Mardia, K. V. (2005). "Modeling rainfall data using a Bayesian kriged-Kalman model." Technical report, School of Mathematics, University of Southampton. 762

Lopes, H. F. (2000). "Bayesian Analysis in Latent Factor and Longitudinal Models." Unpublished PhD. Thesis, Institute of Statistics and Decision Sciences, Duke University. 783

— (2003). "Expected Posterior Priors in Factor Analysis." *Brazilian Journal of Probability and Statistics*, 17: 91–105. 762

Lopes, H. F. and Carvalho, C. M. (2007). "Factor stochastic volatility with time varying loadings and Markov switching regimes." *Journal of Statistical Planning and Inference*, 137: 3082–3091. 762, 763, 766, 783

Lopes, H. F. and Migon, H. S. (2002). "Comovements and Contagion in Emergent Markets: Stock Indexes Volatilities." In Gatsonis, C., Kass, R., Carriquiry, A., Gelman, A., Verdinelli, Pauler, D., and Higdon, D. (eds.), *Case Studies in Bayesian Statistics, Volume 6*, 287–302. 762, 763

Lopes, H. F. and West, M. (2004). "Bayesian model assessment in factor analysis." *Statistica Sinica*, 14: 41–67. 759, 762, 768

Mardia, K. V., Goodall, C. R., Redfern, E., and Alonso, F. J. (1998). "The kriged Kalman filter (with discussion)." *Test*, 7: 217–85. 761, 762

Nobre, A. A., Schmidt, A. M., and Lopes, H. F. (2005). "Spatio-temporal models for mapping the incidence of malaria in Pará." *Environmetrics*, 16: 291–304. 761

Peña, D. and Poncela, P. (2004). "Forecasting with nonstationary dynamic factor models." *Journal of Econometrics*, 119: 291–321. 760, 766

— (2006). "Nonstationary dynamic factor analysis." *Journal of Statistical Planning and Inference*, 136: 1237–1257. 766

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). "Bayesian model averaging for linear regression models." *Journal of the American Statistical Association*, 92: 179–91. 768

Reich, B. J., Fuentes, M., and Burke, J. (2008). "Analysis of the effects of ultrafine particulate matter while accounting for human exposure." Technical report, Department of Statistics, North Carolina State University. 762

Rencher, A. C. (2002). *Methods of Multivariate Analysis*. New York: Wiley, 2nd edition. 779

Robert, C. R. and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York: Springer. 759

Sahu, S. K. and Mardia, K. V. (2005). "A Bayesian kriged-Kalman model for short-term forecasting of air pollution levels." *Applied Statistics*, 54: 223–44. 762

Salazar, E. (2008). “Dynamic Factor Models with Spatially Structured Loadings.” Unpublished PhD. Thesis (in portuguese), Institute of Mathematics, Federal University of Rio de Janeiro. 771

Sansó, B., Schmidt, A. M., and Nobre, A. A. (2008). “Bayesian Spatio-temporal models based on discrete convolutions.” *Canadian Journal of Statistics*, 36: 239–58. 762, 764

Schmidt, A. M. and Gelfand, A. (2003). “A Bayesian Coregionalization Model for Multivariate Pollutant Data.” *Journal of Geophysics Research*, 108 (D24): Available at `http://www.agu.org/pubs/crossref/2003/2002JD002905.shtml`. 766

Schwarz, G. (1978). “Estimating the dimension of a model.” *Annals of Statistics*, 6: 461–464. 770

Stein, M. L. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer. 762

Stroud, R., Müller, P., and Sansó, B. (2001). “Dynamic Models for Spatiotemporal Data.” *Journal of the Royal Statistical Society, Series B*, 63(4): 673–689. 764

Wang, F. and Wall, M. M. (2003). “Generalized common spatial factor model.” *Biostatistics*, 4: 569–582. 759

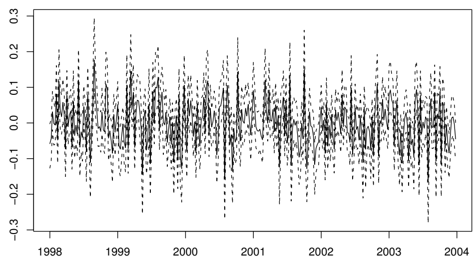
West, M. (2003). “Bayesian factor regression models in the ”large p, small n” paradigm.” In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (eds.), *Bayesian Statistics 7*, 723–732. London: Oxford University Press. 763

West, M. and Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Model*. New York: Springer-Verlag. 764

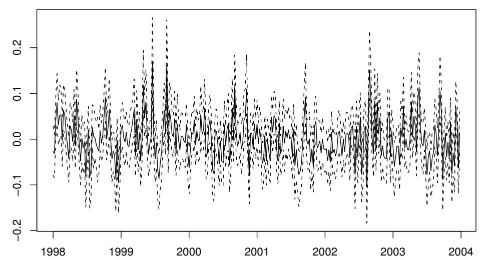
Wikle, C. K. and Cressie, N. (1999). “A dimension-reduced approach to space-time Kalman filtering.” *Biometrika*, 86: 815–829. 761, 762, 764

Acknowledgments

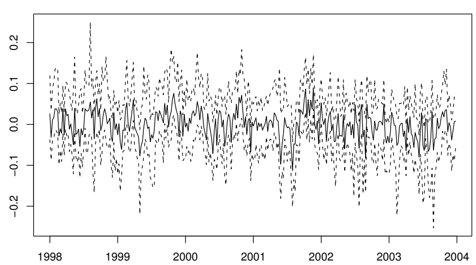
The authors would like to thank Kate Calder, Ed George, Athanasios Kottas, Daniel Peña, Bruno Sansó and Alexandra Schmidt for invaluable comments on preliminary versions of the article that significantly improved its quality. The first author acknowledges the financial support of the University of Chicago Graduate School of Business, while the second and third authors acknowledge the financial support of the research agencies FAPERJ and CNPq.



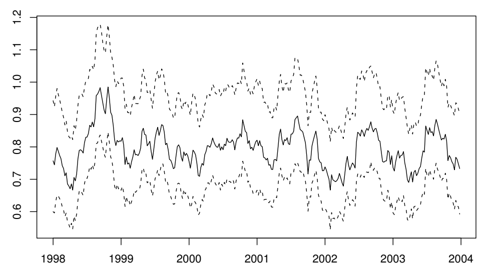
(a) 1st common factor



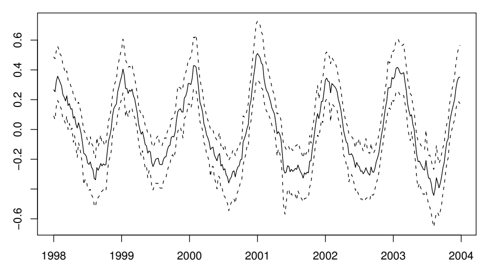
(b) 2nd common factor



(c) 3rd common factor

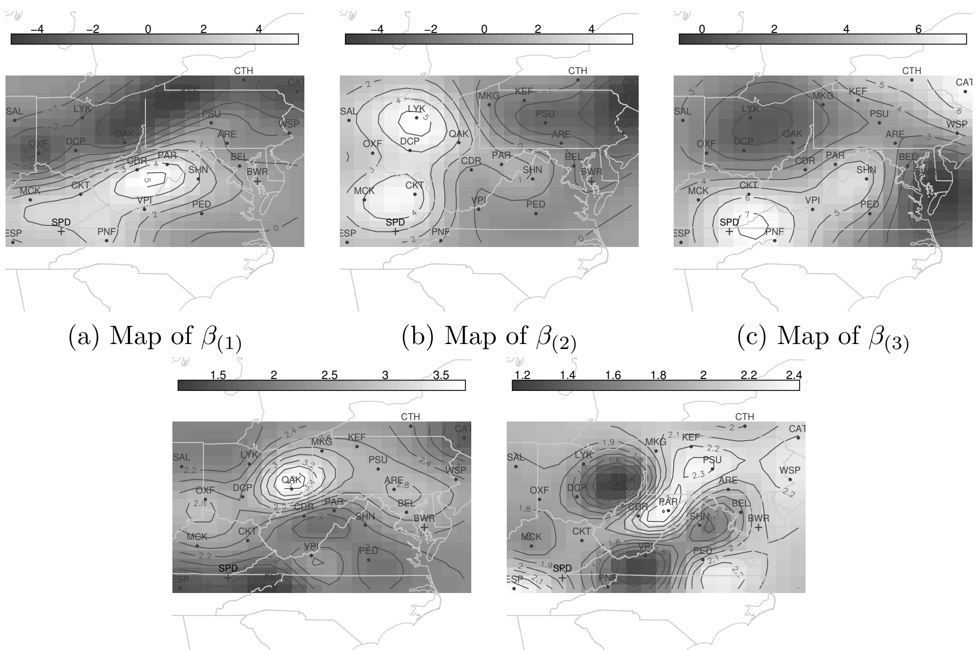


(d) 4th common factor



(e) Seasonal factor

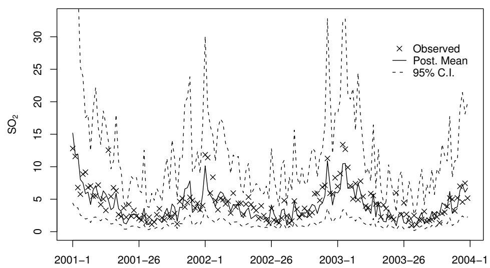
Figure 6: *CASTNet data:* Posterior means of the factors following the SSDFM(4,1)-cov-GP model. Solid lines represent the posterior means and dashed lined the 95% credible interval limits.



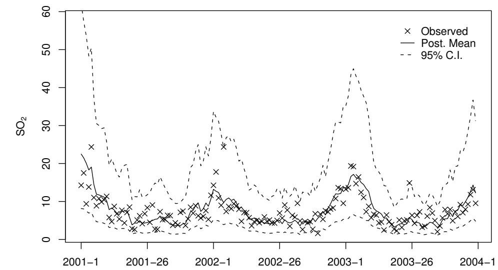
(a) Map of $\beta_{(1)}$ (b) Map of $\beta_{(2)}$ (c) Map of $\beta_{(3)}$

(d) Map of $\beta_{(4)}$ (e) Map of $\beta_{(5)}$

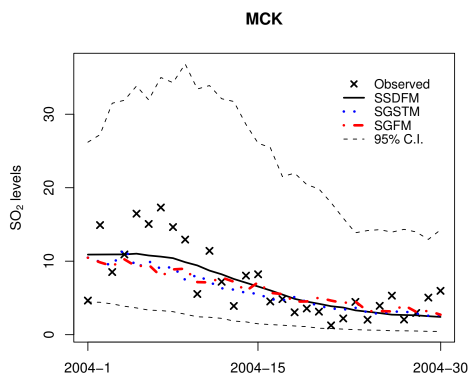
Figure 7: *CASTNet data:* Bayesian interpolation for loadings factors. Values represent the range of the posterior means.



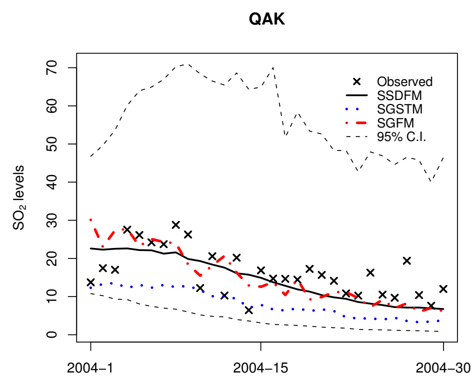
(a) Interpolated values at SPD station



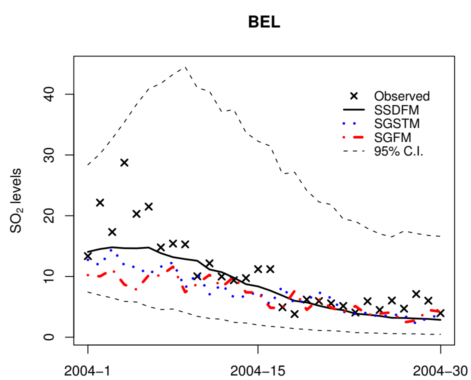
(b) Interpolated values at BWR station.



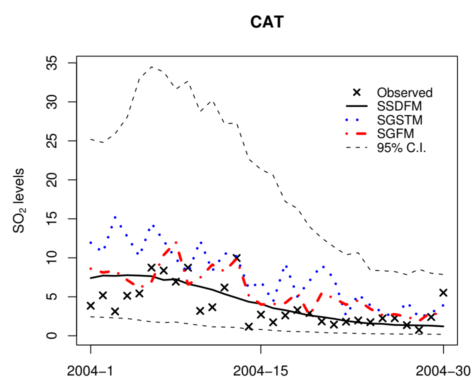
(c) Forecasted value at MCK station



(d) Forecasted value at QAK station.



(e) Forecasted value at BEL station



(f) Forecasted value at CAT station.

Figure 8: *CASTNet data:* (a)-(b)Interpolated values at stations SPD and BWR left out from the sample. (c)-(f) Forecasted values for the period 2004:1–2004:30. Solid, dotted and dotted-dashed lines represent the posterior mean of SSDFM(4,1)-COV-GP, SGSTM and SGFM(4) specifications respectively. Dashed lines represent the 95% credible interval limits of the model SSDFM(4,1)-COV-GP, $\times$ the observed values, while dotted and dotted-dashed lines represent forecasts with the benchmark models.

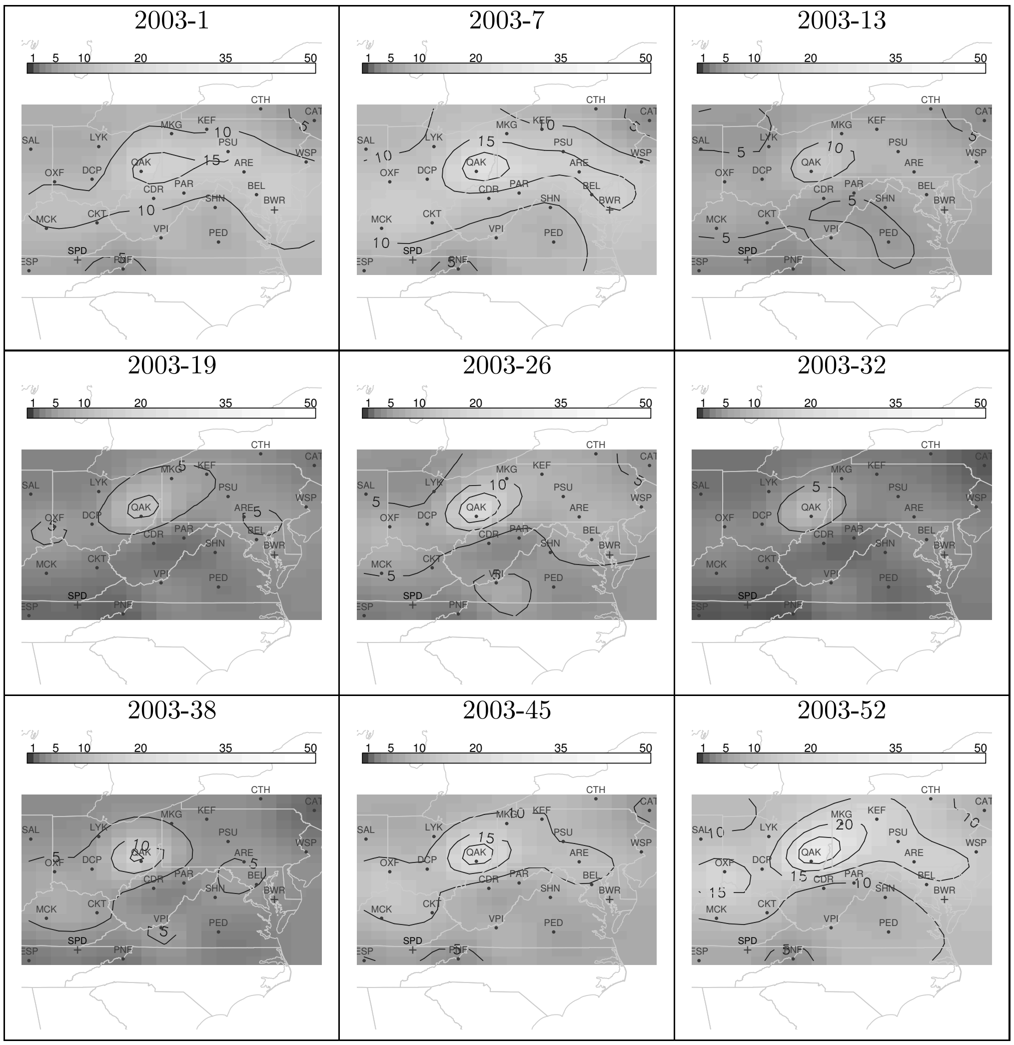


Figure 9: *CASTNet data:* Map of SO_2 concentrations using SSDFM(4,1)-cov-GP model.