Spatial dynamic factor analysis

Hedibert Freitas Lopes

Graduate School of Business, University of Chicago, 5807 S. Woodlawn Ave., Chicago, IL, 60637, USA.

Esther Salazar and Dani Gamerman

Instituto de Matemática, Universidade Federal do Rio de Janeiro, CP 68530, 21945-970, RJ, Brazil.

Summary. 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 identifying 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 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.
Conditionally conjugate and reference-type prior distributions are entertained for the parameters of the spatial components. Different covariance structures are also entertained. Conditionally on the number of common factors, inference is performed by standard Gibbs and Metropolis-Hastings steps. The number of factors is treated as another unknown parameter and fully Bayesian inference is performed by a reversible jump Markov Chain Monte Carlo algorithm.
The new class of models is tested against two synthetic and one real data sets. The real data was obtained from Clean Air Status and Trends Network (CASTNET) and refers to atmospheric concentration of sulfur dioxide weekly observed at 24 monitoring stations from 1998 to 2004. The factor model decomposition is shown to capture 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 just two successful examples of statistical areas that have been experiencing major attention both from the research community as well as practitioners, mainly due to increased availability of efficient computational schemes coupled with faster and easy to use (desktop) computers. In particular, Markov chain Monte Carlo (MCMC) simulation methods (Gamerman and Lopes, 2006) 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.

This paper proposes a a new class of space-time model that resembles a standard dynamic factor model (Peña and Poncela, 2004, for instance). The novelty of the proposal lies on the fact that at any given time the univariate observations from all observed locations, either areal or point-referenced, are group together in what otherwise would be the vector of attributes in standard factor analysis. Consequently, the common factors look for similarities amongst regions/sites. The columns of the factor loadings matrix are modeled using standard spatial Gaussian processes (equation (1) below).

More specifically, let N be the number of locations in a given region and let $y_t = (y_{1t}, \ldots, y_{Nt})'$ be the N-dimensional vector containing all observations at time t, for $t = 1, \ldots, T$. The basic set up of the proposed model is

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

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

where f_t is an m-dimensional vector containing the common factors, for $m < N$ (potentially m is several orders of magnitude smaller than N) and $\beta = (\beta_{(1)}, \ldots, \beta_{(m)})$ is the $N \times m$ matrix of factor loadings.

The modeling of multivariate data through spatially correlated factors has previously been done. Wang and Wall (2003), for instance, fitted a spatial factor model to the mortality rates for three major diseases at nearly one hundred counties of Minnesota. Christensen and Amemiya (2002, 2003) proposed what they called the shift-factor analysis method to model multivariate spatial data with temporal dependence modeled by autoregressive components. Also, Hogan and Tchernis (2004) fitted a one-factor spatial model and entertain 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.

Here, the observations are univariate and factor analysis is used to reducing (identifying) cluster/groups of locations/regions whose temporal dependence can be primarily described by a few common dynamic factors. Prior information about these clusters/groups is reflected in the columns of the factor loadings matrix, which are spatially structured. Where the common factor completely known, this class of structured hierarchical priors fall into the class of spatial priors for regression coefficients (see Nobre, Schmidt and Lopes, 2005, and Gamerman, Moreira and Rue, 2003, for instance). Adopting similar modeling structures, Mardia, Goodall, Redfern and Alonso (1998), Wikle and Cressie (1999) and Calder (2005) applied deterministic spatial characterization for the factor loadings.

The rest of this paper is organized as follows. Section 2 specifies in details the components of Equations (1) and (2) along with prior specification for the model parameters, as well as forecating and interpolation strategies. Posterior inference for fixed number of factors is outlined in Section 3, while uncertainty about m appears in Section 3. Simulated and real data illustrations appear in Section 4 with Section 5 listing conclusions and directions of current and future research.

2. Proposed space-time model

Recalling, equations (1) and (2) define the first level of the proposed dynamic factor model. Throughout this paper it is assumed that $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. In words, it is assumed, like in standard factor analysis, that all covariance structure present in y_t is captured by m independent common factors f_t, i.e., $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$. The dynamic evolution of the factors is characterized by $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_m)$, which could be easily extended to the non-diagonal case. It is also assumed that $f_0 \sim N(m_0, C_0)$, for known hyperparameters m_0 and C_0.

Conditional spatial dependence is introduced by the columns of the factor loadings matrix, β. More specifically, the m columns of β are conditionally independent and follow distance-based Gaussian processes or Gaussian random fields (GRF), i.e., $\beta_{(j)}(\cdot) \sim GRF\left(\mu_j, \tau_j^2 \rho_{\phi_j}(\cdot)\right)$, for $j = 1, \ldots, m$. This notation is equivalent to

$$\beta_{(j)} \sim N(\mu_j 1_N, \tau_j^2 R_{\phi_j}), \tag{3}$$

where 1_N is an N-dimensional vector with ones, R_{ϕ_j} a $N \times N$ matrix whose (l,k) entry given by $r_{lk} = \rho(|s_l - s_k|; \phi_j)$, $l, k = 1, \ldots, N$, for some suitably defined correlation functions $\rho(\cdot; \phi_j)$ possibly depending on parameters ϕ_j's, typically scalars or low dimensional quantities, $j = 1, \ldots, m$. There are many options for the correlation function. The most used ones are exponential: $\rho_\phi(d) = \exp\{-d/\phi\}$, power exponential: $\rho_\phi(d) = \exp\{-(d/\phi_1)^{\phi_2}\}$, spherical: $\rho_\phi(d) = (1 - 1.5(d/\phi_1) + 0.5(d/\phi_1)^3)1_{\{d/\phi_1 \leq 1\}}$ and Mátern: $\rho_\phi(d) = 2^{1-\phi_2}\Gamma(\phi_2)^{-1}(d/\phi_1)^{\phi_2}\kappa_{\phi_2}(d/\phi_1)$ where $\kappa_{\phi_2}(\cdot)$ is the modified Bessel function of the second kind and of order ϕ_2 In each of above families, the range parameter $\phi_1 > 0$. controls how fast the correlation decays with distance, and the smoothness parameter ϕ_2 controls the differentiability of the underlying process (for details, see Cressie, 1993 and Stein, 1999). An alternative nonparametric formulation for spatial dependence is given by Gelfand, Kottas and MacEachern (2005).

The spatial dynamic factor model is defined by equations (1) - (3). In a related article, Mardia, Goodall, Redfern and Alonso (1998) introduced the kriged Kalman filter (KKF) and named the columns of β the common fields. They split the columns of β in two sets: trend fields and principal fields, which are fixed at the beginning of the analysis and are functions of empirical orthogonal functions. Similar, and almost simultaneous, work appears in Wikle and Cressie (1999) where the columns of β are deterministic, complete and orthonormal basis functions. More recently, Calder (2005) uses smoothed kernels to deterministically derive β (see Sansó and Schmidt, 2004, for related ideas).

All these papers assume that both the number of common factors m and the entries of the factor loading matrix β are known and commonly derived by a pre-gridding principal component decomposition (Wikle and Cressie, 1999) or by a principal kriging procedure (Sahu and Mardia (2005ba,b) and Lasinio, Sahu and Mardia (2005)). In this paper both m and β are fully treated as unknown parameters and posterior inference is performed by MCMC algorithms, to be described in the next few sections.

Likelihood function

Conditional on f_t, for $t = 1, \ldots, T$, model (1) can be rewritten in matrix notation as

$$y = F\beta' + \epsilon, \tag{4}$$

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 is

$$p(y|\Theta, F, \beta, m) = (2\pi)^{-TN/2}|\Sigma|^{-T/2} etr\left\{-\frac{1}{2}\Sigma^{-1}(y - F\beta')'(y - F\beta')\right\}, \tag{5}$$

where $\Theta = (\sigma, \lambda, \gamma, \mu, \tau, \phi)$, $\sigma = (\sigma_1^2, \ldots, \sigma_N^2)'$, $\lambda = (\lambda_1, \ldots, \lambda_m)'$, $\gamma = (\gamma_1, \ldots, \gamma_m)'$, $\mu = (\mu_1, \ldots, \mu_m)'$, $\tau = (\tau_1^2, \ldots, \tau_m^2)'$, $\phi = (\phi_1, \ldots, \phi_m)'$, $etr(X) = \exp(trace(X))$. The dependence on the number of factors m is made explicit and considered as another parameter in Section 3.

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. Independent prior distributions for the hyperparameters σ, γ and λ are as follows: i) $\sigma_i^2 \sim IG(n_\sigma/2, n_\sigma s_\sigma/2)$, $i = 1, \ldots, N$; ii) $\gamma_j \sim N(m_\gamma, S_\gamma)$, $j = 1, \ldots, m$; and iii) $\lambda_j \sim IG(n_\lambda/2, n_\lambda s_\lambda/2)$, $j = 1, \ldots, m$, where $n_\sigma, s_\sigma, m_\gamma, S_\gamma, n_\lambda$ and s_λ are know hyperparameters. The parameters μ_j, ϕ_j and τ_j^2, for $j = 1, \ldots, m$, follow one of the two priors: (i) vague but proper priors and (ii) reference-type priors.

In the first case, $\mu_j \sim N(m_\mu, S_\mu)$, $\phi_j \sim IG(\epsilon, \epsilon)$ and $\tau_j^2 \sim IG(n_\tau/2, n_\tau s_\tau/2)$, $j = 1, \ldots, m$, where m_μ, S_μ, ϵ, n_τ and s_τ are known hyperparameters, i.e., $\pi(\mu_j, \tau_j^2, \phi_j) = \pi_N(\mu_j)\pi_{IG}(\tau_j^2, \phi_j)$ where

$$\pi_{IG}(\tau_j^2, \phi_j) \quad = \quad \pi_{IG}(\tau_j^2)\pi_{IG}(\phi_j) \propto \tau_j^{-(n_\tau+2)} e^{-n_\tau s_\tau/2\tau_j^2} \phi_j^{-(\epsilon+1)} e^{-\epsilon/\phi_j}, \tag{6}$$

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

$$\pi_R(\tau_j^2, \phi_j) \quad = \quad \pi_R(\tau_j^2)\pi_R(\phi_j) \propto \tau_j^{-2} \left\{ \mathrm{tr}(W_{\phi_j}^2) - \frac{1}{N-1}[\mathrm{tr}(W_{\phi_j})]^2 \right\}^{1/2}, \tag{7}$$

where $W_{\phi_j} = ((\partial/\partial\phi_j)R_{\phi_j})R_{\phi_j}^{-1}(I_N - 1(1'R_{\phi_j}^{-1}1)^{-1}1'R_{\phi_j}^{-1})$. Notice that $\pi_{IG}(\tau_j^2) = \pi_R(\tau_j^2)$ when $n_\tau = 0$.

Berger, Oliveira and Sansó 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 yield 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. Nonetheless, it seems to be a reasonable approach to follow, which is corroborated by the good empirical findings from Section 4.

Seasonal dynamic factors

Equations (1) and (2) encompasses 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.

A seasonal common factor of period p ($p = 52$ for weekly data and annual cycle) can be easily accommodated by simply letting $\beta = (\beta_{(1)}, 0, \beta_{(2)}, 0, \ldots, \beta_{(h)}, 0)$ and $\Gamma = \mathrm{diag}(\Gamma_1, \ldots, \Gamma_h)$, where

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

where $h = p/2$ is the number of harmonics need to capture the seasonal behavior of the time series.

In this context the covariance matrix Λ is no longer diagonal since the seasonal factors are correlated, i.e., $\Lambda = \text{diag}(\Lambda_1, \ldots, \Lambda_h)$. Note that the seasonal factors are weighted 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. See West and Harrison (1997, Chapter 8) for further details. For the sake of notation, the following sections present the inferential procedures based on the more general equations (1) and (2).

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.

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

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 $h > 0$. Therefore, if $\{(\beta^{(1)}, \Theta^{(1)}, f_T^{(1)}), \ldots, (\beta^{(M)}, \Theta^{(M)}, f_T^{(M)})\}$ is a sample from $p(f_T, \beta, \Theta|y)$ (Section 3), then it is easy to draw $f_{T+h}^{(j)}$ from $(f_{T+h}|f_T^{(j)}, \beta^{(j)}, \Theta^{(j)})$, for all $j = 1, \ldots, M$, such that

$$\hat{p}(y_{T+h}|y) = \frac{1}{M}\sum_{j=1}^{M} p(y_{T+h}|f_{T+h}^{(j)}, \beta^{(j)}, \Theta^{(j)})$$

is a Monte Carlo approximation to $p(y_{T+h}|y)$. If $y_{T+h}^{(j)}$ is a draw from $p(y_{T+h}|f_{T+h}^{(j)}, \beta^{(j)}, \Theta^{(j)})$, for $j = 1, \ldots, M$, then $\{y_{T+h}^{(1)}, \ldots, y_{T+h}^{(M)}\}$ is a sample from $p(y_{T+h}|y)$.

Interpolation

The interest now is in interpolating the response for N^n locations where the response variable y has not been observed. More precisely, let y^o denote the vector of observations from locations in $S = \{s_1, \ldots, s_N\}$ and y^n denote the (latent) vector of measurements from locations in $S_n = \{s_{N+1}, \ldots, s_{N+N_n}\}$. Analogously, let $\beta^o_{(j)}$ and $\beta^n_{(j)}$ be the j-column of the factor loadings matrix corresponding to the observed and non-observed values, respectively.

Interpolation consists of finding the posterior predictive distribution of β^n,

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

where $p(\beta^n|\beta^o, \Theta) = \prod_{j=1}^m p(\beta^n_{(j)}|\beta^o_{(j)}, \mu_j, \tau_j^2, \phi_j)$. The distributions $p(\beta^n_{(j)}|\beta^o_{(j)}, \mu_j, \tau_j^2, \phi_j)$, for $j = 1, \ldots, m$, can be easily obtained by using multivariate normal results. Conditionally on Θ,

$$\begin{pmatrix} \beta^o_{(j)} \\ \beta^n_{(j)} \end{pmatrix} = N\left(\mu_j 1_{N+N^n}, \tau_j^2 \begin{pmatrix} R^o_{\phi_j} & R^{o,n}_{\phi_j} \\ R^{n,o}_{\phi_j} & R^n_{\phi_j} \end{pmatrix}\right)$$

where $R^n_{\phi_j}$ is the correlation matrix of dimension N^n between ungauged locations, $R^{o,n}_{\phi_j}$ 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 $j = N+1, N+N^n$. Consequently, $\beta^n_{(j)}|\beta^o_{(j)}, \Theta \sim N\left(\mu_j 1_{N^n} + R^{n,o}_{\phi_j} {R^o_{\phi_j}}^{-1}(\beta^o_{(j)} - \mu_j 1_N); \tau_j^2 (R^n_{\phi_j} - R^{n,o}_{\phi_j} {R^o_{\phi_j}}^{-1} R^{o,n}_{\phi_j})\right)$ and a Monte Carlo approximation to $p(\beta^n|y^o)$ is given by

$$\hat{p}(\beta^n|y^o) \quad \approx \quad \frac{1}{L}\sum_{l=1}^{L} p(\beta^n|\beta^{o(l)}, \Theta^{(l)})$$

where $\{(\beta^{o(1)}, \Theta^{(1)}), \ldots, (\beta^{o(L)}, \Theta^{(L)})\}$ is a sample from $p(\beta^o, \Theta|y)$ (Section 3). If $\beta^{n(l)}$ is drawn from $p(\beta^n|\beta^{o(l)}, \Theta^{(l)})$, for $l = 1, \ldots, M$, then $\{\beta^{n(l)}, \ldots, \beta^{n(L)}\}$ is a sample from $p(\beta^n|y^o)$. As a by-product, the posterior distribution of non-observed measures y^n can be approximated by

$$\hat{p}(y^n|y^o) \approx \frac{1}{L}\sum_{l=1}^{L} f_N(y^n; \beta^{n(l)} f^{(l)}, \Sigma^{(l)}).$$

3. Posterior inference

Posterior inference for the proposed class of spatial dynamic factor models is facilitated by novel Markov Chain Monte Carlo algorithms designed for two cases: (1) known number of common factors m 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.

Fixed number of common factors

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

$$\begin{aligned} p(F, \beta, \Theta|y) \quad &\propto \quad p(y|F, \beta, \sigma)p(F|\gamma, \lambda)p(\beta|\mu, \tau, \phi)p(\sigma)p(\gamma)p(\lambda)p(\mu)p(\tau)p(\phi) \\ &= \quad \prod_{t=1}^{T} p(y_t|f_t, \beta, \sigma)p(f_0|m_0, C_0)\prod_{t=1}^{T} p(f_t|f_{t-1}, \lambda, \gamma)\prod_{j=1}^{m} p(\beta_{(j)}|\mu_j, \tau_j^2, \phi_j) \\ &\times \quad \prod_{j=1}^{m} p(\gamma_j)p(\lambda_j)p(\mu_j)p(\tau_j^2, \phi_j)\prod_{i=1}^{N} p(\sigma_i^2), \end{aligned} \tag{8}$$

which is analytically intractable. Exact posterior inference is performed by means of a customized MCMC algorithm that cycles through the following full conditionals. Throughout this section $p(\theta|\ldots)$ denotes the full conditional of θ given all other parameters. Sampling schemes to draws from the full conditionals appear in the Appendix.

Unknown number of common factors

This section extends the MCMC algorithm detailed in the previous section to account for uncertainty regarding m the number of common factors in the spatial dynamic factor model. The reversible jump MCMC (RJMCMC) scheme introduced by Lopes and West (2004) for standard static factor models is adapted.

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. For the spatial dynamic factor model, the overall proposal distribution is

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

where a, b, c, d, e, f, g e h are tuning parameters and M_x and V_x are sample means and sample variances based on the preliminary MCMC runs. By letting $p(y, m, \Psi_m) = p(y|m, \Psi_m)p(\Psi_m|m)Pr(m)$ and initial values 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 $(m', \Psi_{m'})$ with probability

$$\alpha = \min\left\{1, \frac{p(y, m', \Psi_{m'})}{p(y, m, \Psi_m)} \frac{q_m(\Psi_m)q(m', m)}{q_{m'}(\Psi_{m'})q(m, m')}\right\}. \tag{9}$$

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 $p(\Psi_m|m, y)$. The closer $q_m(\Psi_m)$ and $p(\Psi_m|m, y)$ are, the closer acceptance probability is to

$$\alpha = \min\left\{1, \frac{p(y|m')}{p(y|m)} \frac{q(m', m)}{q(m, m')}\right\},$$

which can be thought of as a stochastic model search algorithm (George and McCulloch, 1992). The adopted algorithm is a particular case of what Godsill (2001) and Dellaportas, Forster and Ntzoufras (2002) called the *Metropolised Carlin and Chib* method, 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 proposals based on the initial, auxiliary MCMC analysis 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).

4. 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 the customized

MCMC and RJMCMC algorithms implemented. The second case considers a real data set. The data was obtained from the Clean Air Status and Trends Network (CASTNET) and refers to atmospheric concentration of sulfur dioxide weekly observed at 24 monitoring stations from 1998 to 2004.

4.1. Simulated study

In this simulated study a total of $T = 100$ observations are simulated for $N = 50$ locations in the square $[0,1] \times [0,1]$, so y_t is a 50-dimensional vector. Figure 1 exhibits the 50 selected locations. The simulated data is a function of $m = 4$ dynamic common factors. The matrices Γ and Λ defining the dynamic factors' evolutions (2) are diag$(0.6, 0.5, 0.2, 0.3)$ and diag$(0.15, 0.1, 0.2, 0.07)$, respectively. The columns of the factor loadings matrix follow Gaussian processes with exponential correlations given by $\rho(d; \phi_j) = \exp(-d/\phi_j)$ for $\phi = (0.2, 0.5, 0.3, 0.1)$ and $\tau = (1, 0.6, 0.8, 0.5)$. The vector of means μ was considered fixed and equal to zero. Finally, σ_i^2 were uniformly simulated in the interval $(0.01, 0.05)$, for $i = 1, \ldots, 50$.

Figure 1 about here.

The prior distributions for σ_i^2 and λ_j are the same inverse gamma, i.e., $GI(a, b)$ with mean $|\bar{y}|$ and unit variance. The prior distribution of γ is multivariate normal, i.e., $\gamma \sim N((0.6, 0.5, 0.2, 0.3)', I_4)$. For the parameters defining the Gaussian process, two prior distributions were considered: (i) vague inverse gamma prior distributions for both τ_j^2 and ϕ_j, i.e., $IG(0.1, 0.1)$, or (ii) reference priors (see Section 2).

Models with up to six common factors were entertained and compared based on their posterior model probabilities. They were also compared based on routine information criteria, such as the AIC (Akaike, 1974) and the BIC (Schwarz, 1978), as well as mean square errors (MSE). The results are summarized in Table 1, with the $m = 4$ factor model presenting the lowest values, regardless of the prior used. The proposed RJMCMC algorithm also selects the right model by assigning the highest posterior model probability to the dynamic factor model with 4 common factors.

Table 1 about here.

Conditioning on $m = 4$, i.e., the true number of common factors, the MCMC algorihtm outlined in Section 3 was run for a total of 5000 iterations and posterior inference was based on the last 4000 draws. Table 2 presents posterior means and standard deviations for some of the model parameters. 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 (not shown here). The correctness of the fitted model is also evident by examining Figures 2 and 3, which show how accurately estimated both factor loadings matrix and common factor scores are.

Table 2 and Figures 2 and 3 about here.

4.2. SO_2 concentrations in eastern US

Here, the proposed spatial dynamic factor model is used to study the spatial and temporal variations of SO_2 concentrations across 24 monitoring stations (each coded by three letters). Figure 4 exhibits the monitoring stations and the interpolation grid. Weekly measurements are collected by the Clean Air Status and Trends Network (CASTNET), which is part of the Environmental Protection Agency (EPA). Measurements used in this section span from the first week of 1998 to the 30th week of 2004, a total of 342 observations. Stations BWR and SPD were left out of the analysis and will be used to check the performance of the model's spatial interpolation. Similarly, the last 30 weeks (2004:1-2004:30) were left out of the analysis and will be used to check the performance of the model's time series forecast. In other words, a total of 22 stations and 312 observations are used in the following analysis.

Figure 4 about here.

An important issue is the correction of the curvature of the earth in spatial data set. A procedure that eliminates the curvature effect by converting latitudes and longitudes to universal transverse mercator (UTM) coordinates was used. The converted coordinates are measured in kilometers from western-most longitude and the southern-most latitude observed in the data set.

Figure 5 shows the time series for logarithm of SO_2 for some stations. The time series exhibit seasonal components with a marked annual cycle and higher values at the beginning of each year. Besides, the trend and seasonality of the series are, at least visually, related. It will be assumed that the logarithm of the SO_2 is normally distributed and seasonal and nonseasonal dynamic factor model (Section 2) considered. Let SDFM(m) and SSDFM(m, h) stand for spatial dynamic factor model and seasonal spatial dynamic factor model, respectively, in order to distinguish the two subclasses of models that were entertained, with $m = 2, 3$ and 4 factors and $h = 1$ harmonic with cycles of 52 weeks. Gaussian processes with exponential correlation functions were used in all models.

Figure 5 about here.

Relatively vague prior distributions were used, that is $p(\sigma_i^2) = p(\lambda_j) = IG(0.01, 0.01)$, $p(\gamma_j) = N(0, 0.5^2)$, $p(\Lambda_{2,l}) = IW(Q, 10)$ where $\text{diag}(Q) = (0.1, 0.1)$ and $q_{12} = -0.05$. Reference priors were used for the parameters of the Gaussian processes. A 50000 run of the MCMC scheme was generated, discarding the first half as burn-in and retaining only every 10th step thereafter. Therefore posterior summaries are based on 2500 draws. Models with more than 4 factors were analyzed but results not included because most of the additional parameters seemed to be insignificant.

Competing models were compared based on their posterior model probabilities (through RJMCMC schemes), as well as their mean square errors (MSE) and mean absolute errors (MAE). Two separate groups of models were considered: models with seasonality and models without seasonality. From Table 3, it can be seen that MSE and MAE do not convincingly differentiate SDFM from SSDFM. Nonetheless, SDFM(3) and SSDFM(2,1) exhibit somewhat smaller indicators. On the other hand, MSE based on forecasted and interpolated values are smaller for SSDFM with two and three common factors. Finally, the posterior model probabilities for SDFM(3) and SSDFM(2,1) are the highest. Models SSDFM(2,1) and SSDFM(3,1) share similar posterior results (not shown here), with the additional common factor in SSDFM(3,1) centered at zero and with high variability. This helps explaining the higher posterior model probability for the more parsimonious model.

For illustrative purposes, the rest of the analysis if performed under the SSDFM(2,1) specification. Nevertheless, it should be mentioned that forecasting observations at gauged locations, forecasting observations at ungauged locations and many other functionals that appear in all competing models could be analyzed by combining the forecasts/interpolations of different and competing models through Bayesian model averaging, as detailed, for instance, in Raftery, Madigan and Hoeting (1997) and Clyde (1999).

Table 3 about here.

Table 4 shows posterior summaries for some of the SSDFM(2,1) model parameters. Notice that the parameters that control how fast spatial correlation decays, ϕ_js, are very high. This fact is expected since after the correction of the curvature the matrix of distances has high values in kilometers. Figure 6 exhibits the evolution of the common factors over time. As it is fairly common in factor analysis applications, the first common factor account for the global variability of the series and plays the role of the grand mean. One can argue that there is a slight decrease in SO_2 over the years. This characteristic will be further discussed later on. The second trend factor is very noisy but of limited variation. The seasonal common factor captures the cyclic behavior of the time series. It appears that the amplitude of the cycles are slightly increasing with time.

Table 4 and Figure 6 about here.

Figure 7 presents the mean surfaces for the columns of the factor loadings matrix β obtained by Bayesian kriging. Loadings for the first factor are shown to be higher in the center of the interpolated area, around station QAK. Simple exploratory data analysis indicates that the highest values of SO_2 concentrations were measured at QAK, confirming the role of the first factor as a grand mean. Smaller loadings for the second factor are obtained for the Appalachians Mountains, where fewer industrial activities take place. Loadings for the seasonal factor are smaller in state of Ohio, an industrial region. This indicates that seasonality has lower amplitude in this area.

Figure 7 about here.

Combining these findings regarding the factor loadings matrix with those related to the three common factors, one can argue that (i) the first factor is basically an averaged time series with higher loadings associated with the region of higher pollution levels; (ii) the second factor is differentiating regions of different levels of economic occupation and; (iii) the seasonal factor exhibits a cyclical pattern with amplitudes that seem to be increasing over time and to be less relevant over more industrialized regions. In summary, the proposed dynamic spatial factor model is able to meaningfully and parsimoniously separate the data spatial variation from its temporal variation. Figure 8 exhibits encouraging out-of-sample properties of the model, with data points being accurately forecasted and interpolated, for several steps ahead and out-of-sample monitoring stations, respectively.

Figure 8 about here.

5. Conclusions

This paper introduces a new class of spatial dynamic factor models that nonlinearly separates space and time variations of space-time data. The spatial variation is brought into the modeling through the columns of the factor loadings matrix, while time series dynamics are captures by the common factors. The dynamic spatial factor model is capable of separating spatial variation from temporal variation in parsimonious way. The model takes advantages 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.

Fully Bayesian inference is made feasible by novel reversible jump MCMC algorithms with within m-factor model inference facilitated by a MCMC algorithm that combines well established schemes, such as the forward filtering backward sampling algorithm.

The simulated application exploits the potentials of the proposed model and indicates that the designed MCMC and RJMCMC algorithms are practically sound. 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 real data application reiterates the potentials of the model both as an interpolation tool and a forecasting tool.

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 dynamic link the latent spatial processes (Lopes and Carvalho, 2006). 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. One can argue that the affluence 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 (8) 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 $t = 1, \ldots, T$.

Idiosyncratic variances From Equation (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. The conditional distribution of σ_i^2 is given by $p(\sigma_i^2|\ldots) \propto p(y_i|F, \beta_i, \sigma_i^2)p(\sigma_i^2)$, so $(\sigma_i^2|\ldots) \sim IG(n_{\sigma_i}^*/2, n_{\sigma_i}^* s_{\sigma_i}^*/2)$, with $n_{\sigma_i}^* = T + n_\sigma$ and $n_{\sigma_i}^* s_{\sigma_i}^* = (y_i - F\beta_i)'(y_i - F\beta_i) + n_\sigma s_\sigma$.

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, $p(\gamma_i|\ldots) \propto \prod_{t=2}^{T} p(f_{jt}|f_{j,t-1}, \gamma_i, \lambda_i) p(\gamma_i|m_\gamma, S_\gamma)$, so $(\gamma_j|\ldots) \sim N(m^*_{\gamma_j}, S^*_{\gamma_j})$, $m^*_{\gamma_j} = S^*_{\gamma_j}\left[\lambda_j^{-1}\sum_{t=2}^{T} f_{jt}f_{j,t-1} + m_\gamma S_\gamma^{-1}\right]$ and $S^{*-1}_{\gamma_j} = \lambda_j^{-1}\sum_{t=2}^{T} f^2_{j,t-1} + S_\gamma^{-1}$.

Common factors variances Similarly, $p(\lambda_i|\ldots) \propto \prod_{t=2}^{T} p(f_{jt}|f_{j,t-1}, \gamma_i, \lambda_i) p(\lambda_i|n_\lambda, n_\lambda s_\lambda)$, so $(\lambda_j|\ldots) \sim IG(n^*_{\lambda_i}/2, n^*_{\lambda_j}s^*_{\lambda_j}/2)$, $n^*_{\lambda_j} = T-1+n_\lambda$ and $n^*_{\lambda_j}s^*_{\lambda_j} = \sum_{t=2}^{T}(f_{jt}-\gamma_j f_{j,t-1})^2 + n_\lambda s_\lambda$.

Loadings means It follows from (8) that $p(\mu_j|\ldots) \propto p(\beta_{(j)}|\mu_j, \tau_j^2, \phi_j)p(\mu_j)$, so $(\mu_j|\ldots) \sim N(m^*_{\mu_j}, S^*_{\mu_j})$, $m^*_{\mu_j} = S^*_{\mu_j}\left[\tau_j^{-2}\beta'_{(j)}R^{-1}_{\phi_j}1_N + m_\mu S_\mu^{-1}\right]$ 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 †. 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 $\Sigma_\beta = \mathrm{diag}(\tau_1^2, \ldots, \tau_m^2)$. From standard Bayesian multivariate regression (Box and Tiao, 1973), it can be shown that $(\beta^*|\ldots) \sim N(\tilde{\mu}_{\beta^*}, \tilde{\Sigma}_{\beta^*})$, where $\tilde{\Sigma}^{-1}_{\beta^*} = \sum_{t=1}^{T} f_t^{*\prime}\Sigma^{-1}f_t^* + \Sigma^{-1}_{\beta^*}$ and $\tilde{\mu}_{\beta^*} = \tilde{\Sigma}_{\beta^*}\left(\sum_{t=1}^{T} f_t^{*\prime}\Sigma^{-1}y_t + \Sigma^{-1}_{\beta^*}\mu_{\beta^*}\right)$.

Common factors The vectors of common factors $f_1, \ldots, f_T$ are sampled jointly by means of the well known *forward filtering backward sampling* (FFBS) scheme of Carter and Kohn (1994) and Früwirth-Schnatter (1994), which explores, conditionally on β and Θ, the following backward decomposition $p(F|y) = p(f_T|D_T)\prod_{t=0}^{T-1} p(f_t|f_{t+1}, \ldots, f_T, D_t) = p(f_T|D_T) \prod_{t=0}^{T-1} p(f_t|f_{t+1}, D_t)$, where $D_t = \{y_1, \ldots, y_t\}$, $t = 1, \ldots, T$ and D_0 represents the initial information. Roughly speaking, the FFBS works as follows. The density $p(f_T|D_T)$ is obtained by recursively updating a standard multivariate dynamic linear model (West and Harrison, 1997). 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 $m_t = a_t + A_t(y_t - \tilde{y}_t)$, $C_t = R_t - A_tA'_tQ_t$, $a_t = \Gamma m_{t-1}$, $R_l = \Gamma C_{t-1}\Gamma' + \Lambda$, $\tilde{y}_t = \beta a_t$, $Q_t = \beta R_t \beta' + \Sigma$ 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 $\tilde{a}_t = m_t + B_t(\tilde{a}_{t+1} - a_{t+1})$, $H_t = C_t - B_t(R_{t+1} - H_{t+1})B'_t$ and $B_t = C_t\Gamma R^{-1}_{t+1}$. This is the backward sampling step.

Spatial hyperparameters By combining the inverse gamma prior density form (6) or the reference prior density from (7) with the likelihood function from (8), it follows that $(\tau_j^2|\ldots) \sim IG(n^*_{\tau_j}/2, n^*_{\tau_j}s^*_{\tau_j}/2)$, where $n^*_{\tau_j} = N + n_\tau$ and $n^*_{\tau_j}s^*_{\tau_j} = (\beta_{(j)} - \mu_j 1_N)'R^{-1}_{\phi_j}(\beta_{(j)} - \mu_j 1_N) + n_\tau s_\tau$ when inverse gamma prior distributions are used, and $n^*_{\tau_j} = N$ and $n^*_{\tau_j}s^*_{\tau_j} = (\beta_{(j)} - \mu_j 1_N)'R^{-1}_{\phi_j}(\beta_{(j)} - \mu_j 1_N)$ 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.

†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.

Acknowledgement

The authors would like to thank the participants of the talks ministered at the Department of Statistics, University of Chicago, Department of Applied Mathematics and Statistics, University of California at Santa Cruz and the participants of the eight Brazilian meeting for Bayesian inference for invaluable comments that improved the quality of this article. In particular, we would like to thank Kate Calder, Alexandra Schmidt, Ed George, Athanasios Kottas and Bruno Sansó. 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, respectively.

References

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

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

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

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

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.

Calder, C. (2005). "Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment." *Environmental and Ecological Statistics (to appear)*.

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

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

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

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

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

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

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

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

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

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.

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.

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

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

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.

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.

Lopes, H. F. and Carvalho, C. M. (2006). "Factor stochastic volatility with time varying loadings and Markov switching regimes." Technical report, Graduate School of Business, University of Chicago.

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

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

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.

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

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.

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

— (2005b). "Recent trends in modeling spatio-temporal data." In *Proceedings of the Special Meeting on Statistics and Environment*.

Sansó, B. and Schmidt, A. M. (2004). "Spatio-temporal models based on discrete convolutions." Technical report, Statistical Laboratory, UFRJ.

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

Stein, M. L. (1999). *Statistical Interpolation of Spatial Data*. New York: Springer.

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

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

	Number of common factors			
	$m = 3$	$m = 4$	$m = 5$	$m = 6$
PMP	0.000	*0.840*	0.158	0.002
AIC1	7503.1	*6296.0*	6596.5	7510.9
AIC2	8957.3	*6263.2*	6829.4	6891.91
BIC1	10839.8	*10636.5*	11940.6	13858.7
BIC2	12294.1	*10603.7*	12173.5	13239.7
MSE1	305.4	*144.47*	144.49	154.36
MSE2	295.4	*140.60*	140.70	146.20

Table 1. Posterior model probabilities (PMP) and standard information criteria. AIC1, BIC1 and MSE1 are relative to vague but proper priors, while AIC2, BIC2 and MSE2 are relative to the reference priors. PMP are computed based sole on the vague but proper priors. Best models appear in *italic*.

j	True	λ_j	True	γ_j	True	τ_j^2	True	ϕ_j
1	0.15	0.16 (0.07)	0.6	0.54 (0.09)	1.0	0.83 (0.43)	0.2	0.19 (0.08)
2	0.10	0.24 (0.14)	0.5	0.50 (0.09)	0.6	0.31 (0.23)	0.5	0.44 (0.48)
3	0.20	0.15 (0.05)	0.2	0.18 (0.09)	0.8	0.64 (0.34)	0.3	0.22 (0.16)
4	0.07	0.17 (0.06)	0.3	0.29 (0.10)	0.5	0.27 (0.11)	0.1	0.17 (0.08)

Table 2. Posterior means (posterior standard deviations in parenthesis) for the parameters in the simulated study.

Model	MSE	MAE	MSE1	MSE2	PMP
SDFM(2)	0.12	0.26	0.52	0.20	0.37
SDFM(3)	*0.08*	*0.21*	0.46	*0.16*	*0.48*
SDFM(4)	0.11	0.26	*0.45*	0.18	0.15
SSDFM(1,1)	0.11	0.24	0.76	0.15	0.43
SSDFM(2,1)	*0.08*	*0.21*	*0.22*	0.16	*0.56*
SSDFM(3,1)	0.12	0.25	0.23	*0.13*	0.01

Table 3. Comparison criteria for SDFM(m) and SSDFM(m, h) models: mean squared error (MSE), mean absolute error (MAE), and posterior model probabilities (EPD). MSE1 are based on predicted values and MSE2 are based on interpolated values at stations BWR and SPD. Best models appear in *italic*.

	Mean	S.D.	2.5%	50%	97.5%
γ_1	0.997	0.002	0.991	0.997	1.000
γ_2	0.029	0.064	-0.092	0.029	0.149
λ_1	0.011	0.004	0.005	0.011	0.020
λ_2	0.018	0.013	0.002	0.015	0.054
λ_3	0.006	0.002	0.003	0.006	0.009
λ_4	0.005	0.001	0.003	0.005	0.008
λ_{34}	-0.002	0.001	-0.004	-0.002	0.000
μ_1	0.85	0.63	-0.51	0.93	1.54
μ_2	0.33	3.20	-4.07	0.14	7.20
μ_3	2.68	0.50	1.83	2.61	3.79
τ_1^2	0.44	1.60	0.033	0.094	4.37
τ_2^2	12.60	28.00	0.69	3.84	103.00
τ_3^2	0.46	0.37	0.14	0.36	1.46
ϕ_1	1488.00	5287.00	62.17	268.23	16452.00
ϕ_2	983.08	1595.40	136.84	463.54	5826.60
ϕ_3	180.11	155.98	46.71	138.92	600.96

Table 4. Posterior summaries for SSDFM(2,1) model.

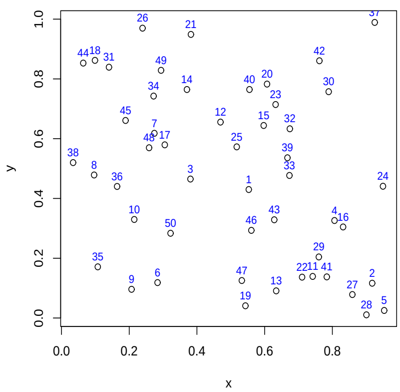
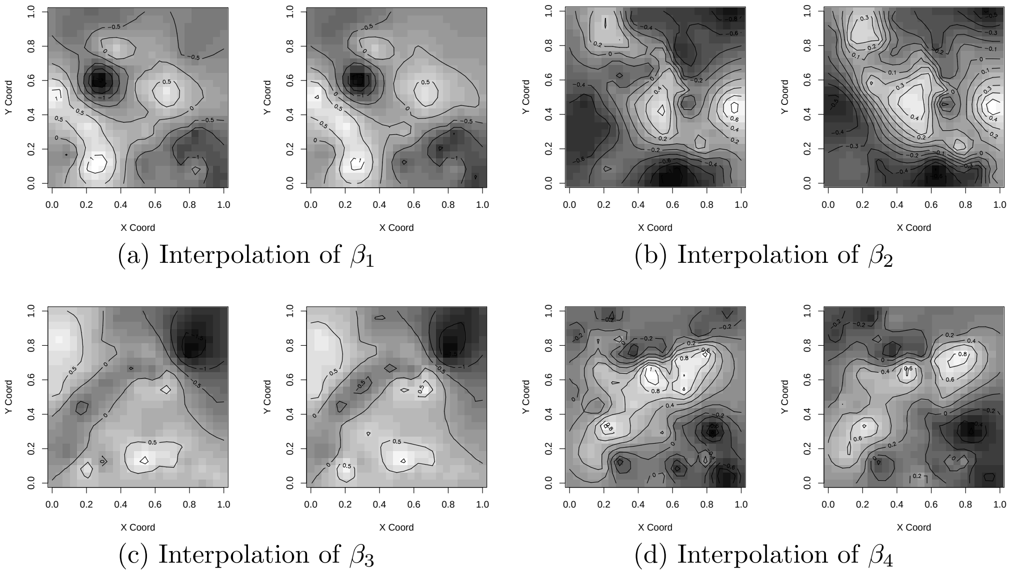


Fig. 1. Coordinates of the $N = 50$ locations on the square $[0, 1] \times [0, 1]$.



(a) Interpolation of β_1

(b) Interpolation of β_2

(c) Interpolation of β_3

(d) Interpolation of β_4

Fig. 2. Interpolation of the four 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.

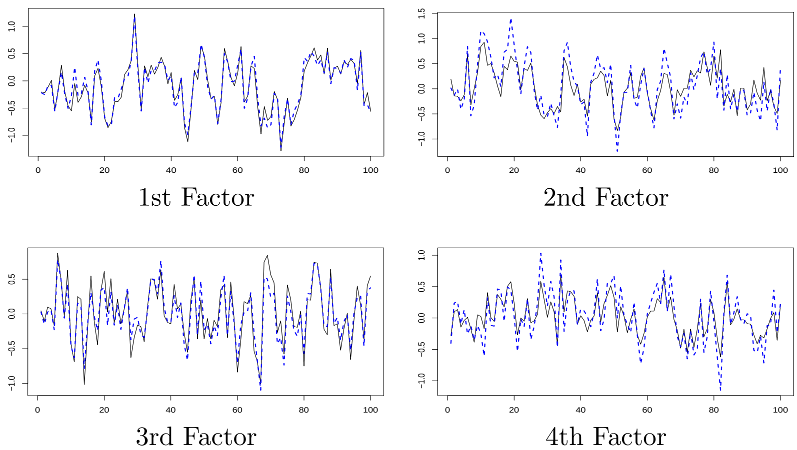


Fig. 3. True common factors (solid lines) and posterior means (dashed lines).

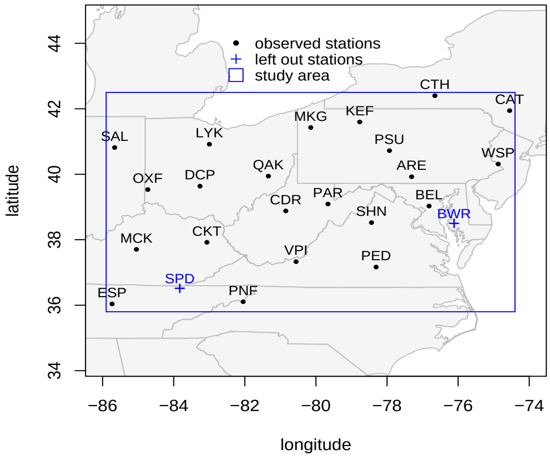
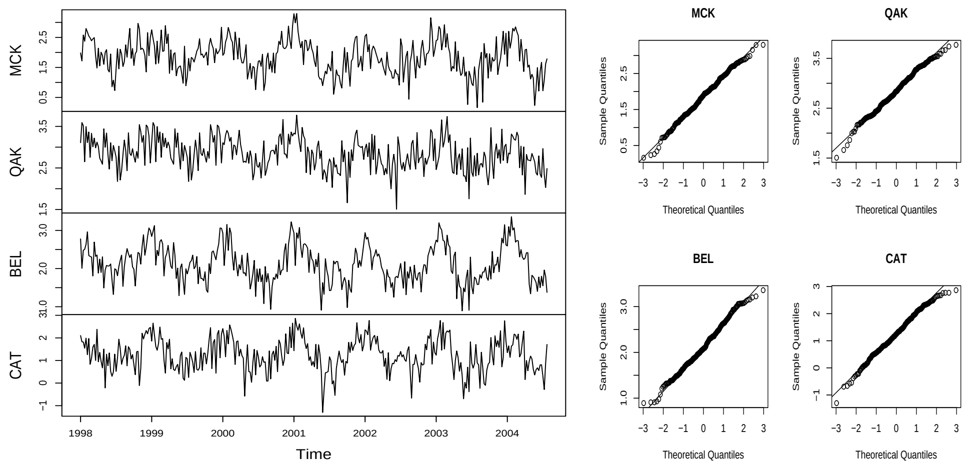
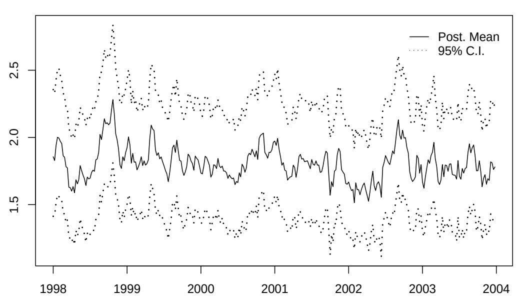


Fig. 4. CASTNET data: Location of the monitoring stations. The stations SPD and BWR were left out from the sample for interpolation purposes.

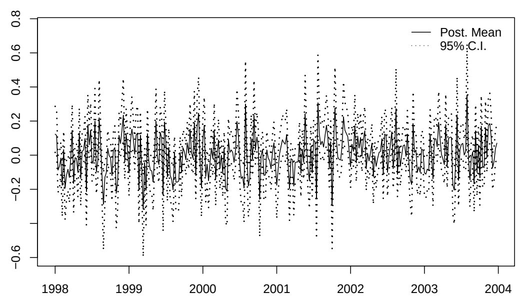


(a) $\text{Log}(\text{SO}_2)$ (b) Normal Q-Q Plot

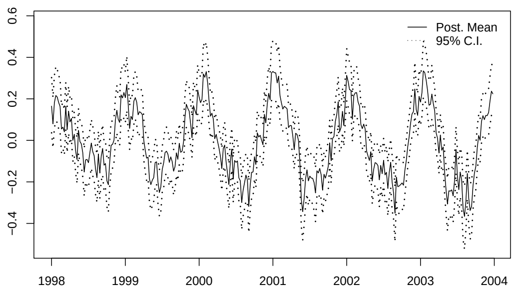
Fig. 5. CASTNET data (a) Time series of weekly $\log(\text{SO}_2)$ concentrations at MCK, QAK, BEL and CAT stations. (b) Normal Q-Q plot for time series plotted in (a).



(a) First Factor

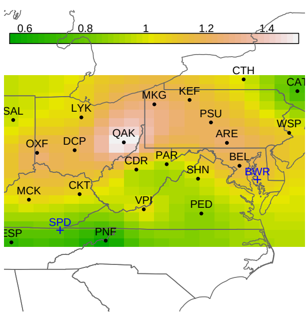


(b) Second Factor

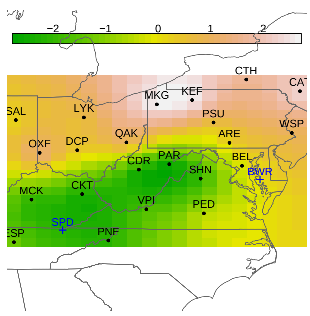


(c) Seasonal Factor

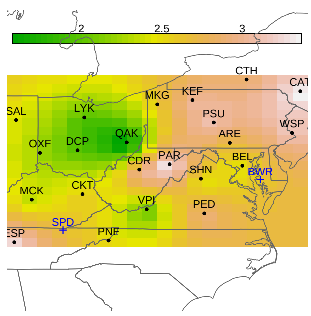
Fig. 6. Posterior means of the factors following the SSDFM(2,1) model. Solid lines represent the posterior means and dashed lined the 95% credible interval limits.



(a) Interpolation of $\beta_{(1)}$

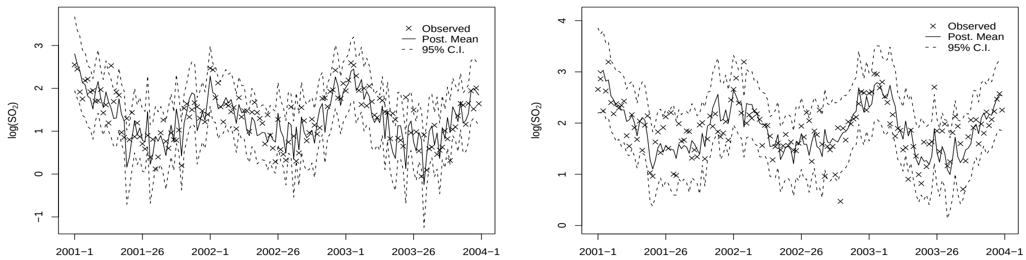


(b) Interpolation of $\beta_{(2)}$

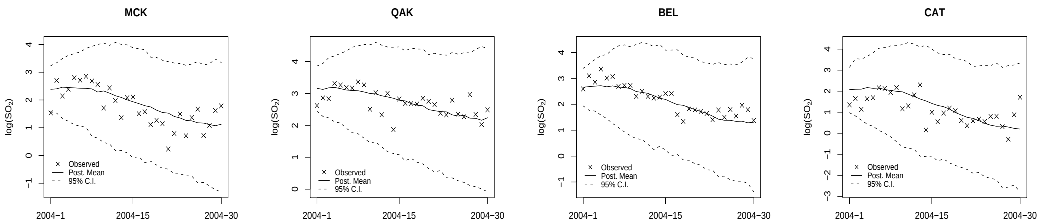


(c) Interpolation of $\beta_{(3)}$

Fig. 7. Bayesian interpolation for loadings of the first, second and seasonal factors respectively. Values represent the range of the posterior means.



(a) Interpolated values at SPD station (b) Interpolated values at BWR station.



(c) Forecasted values at some stations.

Fig. 8. (a)-(b)Interpolated values at stations SPD and BWR left out from the sample. (c) Forecasted values for the period 2004:1–2004:30. Solid lines represent the posterior mean, dashed lines to the 95% credible interval limits and $\times$ the observed values.