Computational Statistics and Data Analysis 55 (2011) 1319–1330

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Generalized spatial dynamic factor models

Hedibert Freitas Lopes [a,*], Dani Gamerman [b], Esther Salazar [c]

[a] The University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL, 60637, United States

[b] Instituto de Matemática, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, Rio de Janeiro, RJ, 21945-970, Brazil

[c] Duke University, CIEMAS Building, Durham, NC, 27708, United States

ARTICLE INFO

Article history:
Received 22 April 2010
Received in revised form 18 September 2010
Accepted 19 September 2010
Available online 8 October 2010

Keywords:
Exponential family
Factor model
Gaussian process
Markov chain Monte Carlo
Reversible jump
Sampling schemes

ABSTRACT

This paper introduces a new class of spatio-temporal models for measurements belonging to the exponential family of distributions. In this new class, the spatial and temporal components are conditionally independently modeled via a latent factor analysis structure for the (canonical) transformation of the measurements mean function. The factor loadings matrix is responsible for modeling spatial variation, while the common factors are responsible for modeling the temporal variation. One of the main advantages of our model with spatially structured loadings is the possibility of detecting similar regions associated to distinct dynamic factors. We also show that the new class outperforms a large class of spatial-temporal models that are commonly used in the literature. Posterior inference for fixed parameters and dynamic latent factors is performed via a custom tailored Markov chain Monte Carlo scheme for multivariate dynamic systems that combines extended Kalman filter-based Metropolis–Hastings proposal densities with block-sampling schemes. Factor model uncertainty is also fully addressed by a reversible jump Markov chain Monte Carlo algorithm designed to learn about the number of common factors. Three applications, two based on synthetic Gamma and Bernoulli data and one based on real Bernoulli data, are presented in order to illustrate the flexibility and generality of the new class of models, as well as to discuss features of the proposed MCMC algorithm.

1. Introduction

Nonlinear and non-Gaussian spatio-temporal models have been extensively used in various areas of science, from epidemiology to meteorology and environmental sciences, amongst others. For example, Hooten and Wikle (2008) use Poisson processes to model US rare bird counts and Fernandes et al. (2009) use zero-inflated Poisson processes to model Rio de Janeiro's rainfall, while Jin et al. (2005) use, for areal data, generalized multivariate conditional autoregressive (CAR) to model lung and esophagus cancer mortality rates. Their success is partially due to the cheap availability of increasingly faster computers and partially due to the development of better and more efficient MCMC schemes for general state space models.

Factor analysis has already been used to spatially model multivariate exponential family data. In such cases, factor analysis appears in its traditional static form with latent factors used to reduce the dimension of the data at each location/region. The factors are then treated as measures and modeled via standard generalized spatio-temporal models, such as the ones listed above. Knorr-Held and Best (2001), for instance, use Poisson processes to model a common component derived from bivariate disease data, while Wang and Wall (2003) use CAR structures to model common factors derived from city-level multivariate mortality rates from various types of cancer in Minnesota.

* Corresponding author. Tel.: +1 773 834 5458.
E-mail address: hlopes@chicagobooth.edu (H.F. Lopes).

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.09.020

By modeling spatial variation via spatially structured factor loadings, we entertain the possibility of identifying similarity regions that share common time series components. These components are modeled by latent and dynamic factors. Standard factor analysis summaries, such as the decomposition of the variance into common and idiosyncratic factors, will then lead to spatial clustering of the measured time series.

More specifically, we introduce a new class of spatio-temporal models for nonlinear and non-Gaussian measurements where spatial and temporal variations are considered conditionally independent and modeled via a latent and dynamic factor analysis framework. The conditional spatial variation will be modeled by the columns of the factor loadings matrix, while the common factors capture the time-series behavior of the data. In this sense, we extend the work of Lopes et al. (2008) who focused exclusively on normal data. The new class is shown to encompass a large class of spatial-temporal models that are commonly used and, more importantly, differs from them in two major aspects: (i) it avoids the curse of dimensionality commonly present in large spatio-temporal data and (ii) it facilitates the formation of spatial clusters which further avoids dimensionality issues.

Posterior inference when the number of factor is fixed will be based on a customized MCMC scheme that combines sampling and blocking ideas from Gamerman (1998) and Knorr-Held (1999), while posterior inference for the number of factors is based on the reversible jump MCMC of Lopes and West (2004).

The remainder of the paper is as follows. The generalized spatial dynamic factor model is formally introduced in Section 2. Section 3 details the sampling schemes for fully Bayesian posterior inference. Synthetic and real data illustrations are presented in Sections 4 and 5, respectively. Final remarks are listed in Section 6.

2. Modeling

Let $\{s_1, \ldots, s_N\}$ be the N spatial locations in the region of study S, where $S \subset R^d$ (usually, $d = 2$), and $y_t = (y_{t1}, \ldots, y_{tN})$ be the N-dimensional vector of measurements at time t, for $t = 1, \ldots, T$. The generalized spatial dynamic factor model (GSDFM, hereafter) is a hierarchical model with a first level measurement equation for conditionally independent univariate observations y_{ti} in the one-parameter natural exponential family, i.e.

$$p(y_{ti} \mid \eta_{ti}, \psi) = \exp\{\psi[y_{ti}\eta_{ti} - b(\eta_{ti})] + c(y_{ti}, \psi)\} \tag{1}$$

where η_{ti} is the natural parameter and ψ is a dispersion parameter (kept known). The mean and variance of y_{ti} are, respectively, $b'(\eta_{ti})$ and $b''(\eta_{ti})/\psi$. Sections 4 and 5 provide illustrative examples. In the gamma model $G(a, b)$ with mean a/b, for example, $y_{ti} \sim G(\psi, b_{ti})$, so $\eta_{ti} = -b_{ti}/\psi$, $b(\eta_{ti}) = -\log b_{ti}$ and $c(y_{ti}, \psi) = (\psi - 1)\log y_{ti} - \log \Gamma(\psi)$.

The natural parameter η_{ti} is deterministically defined by a linear combination of spatial and temporal components through the link function υ, i.e. $\eta_{ti} = \upsilon(\theta_{ti})$. The GSDFM is then completed by specifying the spatio-temporal dependence of the θ_{ti}s.

The temporal behavior of y_t is modeled by the following two levels of hierarchy

$$\theta_t = \mu_t + \beta f_t \tag{2}$$

$$f_t \mid f_{t-1} \sim N(\Gamma f_{t-1}, \Lambda) \tag{3}$$

where μ_t 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), $f_0 \sim N(m_0, C_0)$, for known hyperparameters m_0 and C_0, and Λ is the evolutional variance. The dynamic evolution of the common factors is mainly driven by Γ. Both μ_t and Γ can accommodate seasonalities, stationary and non-stationary trends as well as covariate effects (see Sections 2.3 and 2.4).

The spatial variation of y_t is modeled via the columns of the factor loadings matrix β. More specifically, the jth column of β, denoted by $\beta_j = (\beta_{1j}, \ldots, \beta_{Nj})'$, for $j = 1, \ldots, m$, is modeled as a conditionally independent, distance-based Gaussian process or a Gaussian random field (GRF), i.e.

$$\beta_j \sim N(\kappa_j, \tau_j^2 R_j) \tag{4}$$

where κ_j is a N-dimensional mean vector. The (l, k)-element of $R_j = R(\phi_j)$ 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_j is univariate when the correlation function is exponential $\rho_{\phi_j}(d) = \exp\{-d/\phi_j\}$ or spherical $\rho_{\phi_j}(d) = (1 - 1.5(d/\phi_j) + 0.5(d/\phi_j)^3)1_{\{d/\phi_j \le 1\}}$, and bivariate when it is power exponential $\rho_{\phi_j}(d) = \exp\{-(d/\phi_{j,1})^{\phi_{j,2}}\}$ or Matérn $\rho_{\phi_j}(d) = 2^{1-\phi_{j,2}}\Gamma(\phi_{j,2})^{-1}(d/\phi_{j,1})^{\phi_{j,2}}\mathcal{K}_{\phi_{j,2}}(d/\phi_{j,1})$, where $\mathcal{K}_{\phi_{j,2}}(\cdot)$ is the modified Bessel function of the second kind and of order $\phi_{j,2}$. In each one of the above families, the range parameter $\phi_{j,1} > 0$ controls how fast the correlation decays as the distance between locations increases, while the smoothness parameter $\phi_{j,2}$ controls the differentiability of the underlying process. See Lopes et al. (2008) and references therein for further details. The above setting assumes that observed locations remain unchanged throughout the study period but anisotopic observation processes can be easily contemplated. The isotopic setting is retained here simply to avoid unnecessarily cumbersome notation.

This class of models leads to a non-separable spatio-temporal covariance structure (Cressie and Huang, 1999). This characteristic is directly associated with the linear predictor θ_t where both spatial and temporal covariance structures can not be separately identified when $m > 1$. Without loss of generality, when $\mu_t = 0$, $d = \|s_i - s_j\|$ and Γ and Λ are diagonal matrices, it follows that $C(d, h) = \mathrm{cov}(\theta_{it}, \theta_{j,t+h})$ is given by

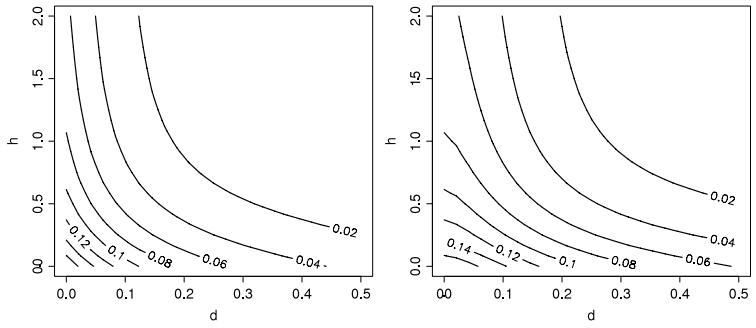

Fig. 1. Contours of $C(d, h)$ for spatial lag d, temporal lag h, $m = 3$, $\Lambda = \text{diag}(0.1, 0.15, 0.07)$, $\Gamma = \text{diag}(0.5, 0.1, 0.9)$, $\tau_1^2 = 0.1$, $\tau_2^2 = 0.6$, $\tau_3^2 = 0.2$, $\phi_{1,1} = 0.2$, $\phi_{2,1} = 0.5$, $\phi_{3,1} = 0.1$ and $\kappa_1 = \kappa_2 = \kappa_3 = 0$ considering exponential (left) and Matérn (right) correlation functions with smoothness parameter, $\phi_{j,2}$, equal to one.

$$C(d, h) = \sum_{k=1}^{m} (\lambda_k \gamma_k^h)(1 - \gamma_k^2)(\tau_k^2 \rho(d, \phi_k) + \kappa_{ik}\kappa_{jk}).$$

Fig. 1 shows a typical model setting where non-separability can be visually assessed.

2.1. Identification issues

One of the major concerns in factor analysis is to define a unique model free from identification problems. In the proposed model the product βf_t is uniquely identified but β and f_t individually are not, since for any orthogonal matrix K, $\beta f_t = (\beta K')(K f_t)$. A common approach is to impose some constraints on β. However, the spatial structure imposed into the columns can be exploited in order to avoid the problem. Then, in both the mean and variance of the Gaussian processes we adopt priors centered around pre-fixed values, as defined below.

2.2. Prior distribution

The GSDFM is represented by Eqs. (1)–(4). It is assumed in this subsection that $\mu_t = 0$ for all t, $\kappa_j = 0$ for all j and that Γ and Λ are both diagonal matrices with diagonal components in γ and λ, respectively. More general cases are left for further subsections. Also, let $y = (y_1, \ldots, y_T)'$, $f = (f_1, \ldots, f_T)'$, $\tau = (\tau_1^2, \ldots, \tau_m^2)'$ and $\phi = (\phi_1, \ldots, \phi_m)'$. Then, the posterior distribution of the model parameters $\xi = (\beta, f, f_0, \gamma, \lambda, \tau, \phi)$ is given by the product of the likelihood $p(y|\xi) = p(y|\psi, f, \beta)$ from Eq. (1) and the prior distribution

$$p(\xi) = p(f|\gamma, \lambda)p(\beta|\tau, \phi)p(\gamma)p(\lambda)p(\tau)p(\phi)p(f_0) \tag{5}$$

where $p(f|\gamma, \lambda)$ and $p(\beta|\tau, \phi)$ are given by Eqs. (3) and (4), respectively. The prior distributions for the components of λ and γ are inverse gammas and normal distributions truncated in the stationarity region $[-1, 1]$. A mixture of these truncated normal densities with point masses at one to allow for non-stationary dynamic factors were also considered. The prior distributions for the components of the Gaussian process parameters τ and ϕ are also inverse gammas with parameters $(2, n_\tau)$ and $(2, b)$, with means n_τ and b, respectively. The shape hyperparameter b is fixed at $\max(dist)/(-2\log(0.05))$, where $\max(dist)$ is the maximum distance between any two spatial locations. Also, the hyperparameter n_τ is conveniently chosen to avoid identifiability problems.

2.3. Seasonal components

A frequent feature of space-time data is the presence of seasonality, especially in environmental or climate-related data. This requires either the incorporation of seasonal components into the complete model or pre-deseasonalization of the data. The latter underestimates the uncertainty by assuming perfect knowledge of the seasonality. Therefore, we prefer to handle all sources of uncertainty simultaneously into a unified framework.

Seasonal effects may be dynamic or static and can also be incorporated into GSDFM in a number of forms. One of them is through common dynamic factors. The seasonal effect is thought to be the same throughout the region of study but possibly varying in time. In this case, the common seasonal factors loadings act as mixing location weights for the several Gaussian processes, thus implying different magnitudes for the seasonal patterns for different 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 $\Gamma_{l,11} = \Gamma_{l,22} = \cos(2\pi l/p)$ and $\Gamma_{l,12} = -\Gamma_{l,21} = \sin(2\pi l/p)$, for $l = 1, \ldots, h \leq p/2$, where h 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 vector autoregressive conditional covariance matrix Λ is no longer diagonal since the seasonal factors are correlated. In practice, few harmonics are required to adequately describe the seasonal pattern of many datasets and the dimension of this component is typically small. Similarly, dynamic

seasonal components could be easily incorporated into the mean level. The incorporation of static seasonal components into the model would be even simpler.

2.4. Covariate effects

Many specifications for the mean level of both GRFs 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, a few possibilities are a constant mean level model with $\mu_t = \mu$; a regression model with $\mu_t = x_t'\alpha$; and a dynamic regression model where $\mu_t = x_t'\alpha_t$ and $\alpha_t \sim N(\alpha_{t-1}, W)$, with W incorporated in ξ with inverse gamma or inverse Wishart prior, depending on the dimension of α_t. The simplest dynamic regression consists of a single time-varying intercept when $x_t = 1_N$ is common across locations. This leads to the model with a common component, denoted GSDFM-CC. Similarly, covariates can be included in the factor loading's prior specification, κ_j in Eq. (4), to entertain deterministic spatial variability. A few alternatives are $\kappa_j = 0$; $\kappa_j = \zeta_j 1_N$; and $\kappa_j = x_j'\xi_j$, where more flexibility can be brought up by allowing different covariates for each Gaussian random field.

2.5. Interpolation and forecasting

Two of the main interests when fitting spatio-temporal models is the ability to spatially interpolate over the whole region of study and perform a h-step ahead forecast into the future. Let $z_{t+h} = (y_{t+h}, \tilde{y}_{t+h})$ be the vector of measurements (observed and missing) for both gauged and ungauged locations $\{s_1, \ldots, s_N\}$ and $\{\tilde{s}_1, \ldots, \tilde{s}_{\tilde{N}}\}$ and time period $t+h$, for $h \geq 0$. Therefore, Bayesian spatial interpolation is posterior inference for $\tilde{y}_{t+h}$ for $h = 0$ and $\tilde{N} > 0$, while a Bayesian h-step ahead forecast is posterior inference for y_{t+h} for $h > 0$ and $\tilde{N} = 0$. When both $h > 0$ and $\tilde{N} > 0$, posterior inference comprises spatial interpolation for ungauged locations, a h-step ahead forecast for gauged locations and both spatial interpolation and a h-step ahead forecast for ungauged locations.

Bayesian kriging for $\tilde{\beta}_j$, the vector of factor loadings associated with jth common factor and $\tilde{N}$ ungauged locations, is given by

$$(\tilde{\beta}_j|\beta_j) \sim N(\tilde{\kappa}_j + \tilde{R}_j' R_j^{-1}(\beta_j - \kappa_j),\ \tau_j^2(R_{\tilde{N}} - \tilde{R}_j' R_j^{-1}\tilde{R}_j)) \tag{6}$$

where $\tilde{\kappa}_j$ is the mean vector of $\tilde{\beta}_j$, $\tilde{R}_j$ is an $N \times \tilde{N}$ matrix containing the spatial correlations between the components of β_j and $\tilde{\beta}_j$, for $j = 1, \ldots, m$, and $R_{\tilde{N}}$ is an $\tilde{N} \times \tilde{N}$ matrix containing the spatial correlations between the components of $\tilde{\beta}$. The resulting joint distribution of $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_m)$, i.e. $p(\tilde{\beta}|\tilde{\kappa}, \tilde{\phi}, \xi)$, can be used for spatial interpolation of measurement $\tilde{y}_t$ at an ungauged location and time t via

$$p(\tilde{y}_t|y) = \int p(\tilde{y}_t|\tilde{\beta}, f_t)p(\tilde{\beta}|\tilde{\kappa}, \tilde{\phi}, \xi, y)p(\tilde{\kappa}, \tilde{\phi}, \xi|y)\mathrm{d}\tilde{\beta}\mathrm{d}\tilde{\kappa}\mathrm{d}\tilde{\phi}\mathrm{d}\xi, \tag{7}$$

for $t = 1, \ldots, T$. Draws from (7) can be obtained in three steps. Firstly, $(\xi, \tilde{\kappa}, \tilde{\phi})$ is sampled from their joint posterior distribution via MCMC (Section 3). Secondly, conditionally on ξ the columns of $\tilde{\beta}$ are mutually independent and independent of y and can thus be sampled from (6). Thirdly, $\tilde{y}_t$ is sampled from (1).

Similarly, based on Eq. (3), it is easy to show that a h-step ahead forecast for the dynamic factors is given by

$$p(f_{t+h}|\xi) \sim N(\Gamma^h f_t,\ \Omega_h) \tag{8}$$

where $\Omega_h = \sum_{j=1}^{h} \Gamma^{h-j}\Lambda(\Gamma^{h-j})'$. Therefore, a h-step ahead forecast for measurements y_{t+h} from gauged locations is given by

$$p(y_{t+h}|y) = \int p(y_{t+h}|\beta, f_{t+h})p(f_{t+h}|\xi, y)p(\xi|y)\mathrm{d}f_{t+h}\mathrm{d}\xi. \tag{9}$$

Draws from (9) can be obtained in three steps. Firstly, ξ is sampled from its joint posterior distribution via MCMC (Section 3). Secondly, conditionally on ξ the common factors f_{t+h} are independent of y and can be sampled from (8). Thirdly, y_{t+h} is sampled from (1). Spatial interpolation and a h-step ahead forecast for measurements at ungauged locations can be easily obtained by combining Eqs. (6)–(9).

2.6. Unknown number of factors

Another important contribution of the paper is the fully Bayesian approach towards selecting the number of common dynamic factors. Therefore, posterior model probabilities (PMP) can be used to select the one model with highest PMP and proceed with posterior inference for the parameters of such model. See the examples of Sections 4 and 5. Alternatively, one could obtain posterior summaries based on Bayesian model averaging (Hoeting et al., 1999). For example, Bayesian kriging

and forecasting can be based on a GSDFM(m), i.e. $p(\tilde{y}_t|y, m)$ (Eq. (7)) and forecasting $p(y_{t+h}|y, m)$ (Eq. (9)) or could be based on averaged across competing factor models, i.e.

$$p(\tilde{y}_t|y) = \sum_{m\in\mathcal{M}} \pi_m p(\tilde{y}_t|y, m) \quad \text{and} \quad p(y_{t+h}|y) = \sum_{m\in\mathcal{M}} \pi_m p(y_{t+h}|y, m) \tag{10}$$

where $\pi_m = Pr(m|y)$ for $m \in \mathcal{M}$. Usually $\mathcal{M}$ equals $\{1, \ldots, m_{\max}\}$, where $m_{\max}$ is the maximum number of common factors. In the next section we describe how π_m can be estimated via a customized reversible jump MCMC algorithm for large scale nonlinear and non-normal dynamic models.

3. Computational aspects

The resulting posterior distribution is not analytically tractable and approximation techniques must be used. Bayesian kriging, forecasting and model order selection, and other posterior summaries, are made feasible via a customized Markov chain Monte Carlo algorithm with the main steps outlined in this section. Most of the parameters are either sampled from normal or inverse gamma full conditional distributions or by simple Metropolis–Hastings steps. Sampling the common dynamic factors is more challenging and, for that, we propose a block sampling scheme that combines techniques such as extended Kalman filter and block sampling. This proposal can be seen as an alternative to sampling state vectors in dynamic generalized linear models.

3.1. Sampling static components

It is straightforward to show that the full conditional distributions of the components of (γ, f_0) and (λ, τ) are normal and inverse gamma distributions, respectively. Similar comments follow for parameters α and ζ present in the regression structure for μ_t and κ.

Also, random walk Metropolis steps are implemented to sample the components of ϕ and β. More specifically, a new draw $\log \phi_j^*$ is sampled from a normal distribution with mean at the current draw and variance δ_ϕ and a new draw $\text{vec}(\beta^*)$ is sampled from a multivariate normal distribution with mean at the current draw and variance-covariance matrix $\delta_\beta I_{Nm}$. Empirical acceptance rates are controlled by tuning parameters δ_ϕ and δ_β.

3.2. Sampling dynamic components

In Gaussian SDFM, the common dynamic factors are linearly related to the observations and the whole set of common factors, $f = (f_1, \ldots, f_T)'$, can be sampled in a single block by the well established forward-filtering, backward-sampling scheme (FFBS, Carter and Kohn, 1994, and Frühwirth-Schnatter, 1994). In generalized SDFM, joint sampling of the dynamic factors in a single block typically leads to negligible acceptance rates, specially for large T. Sampling them in T individual blocks also leads to slow mixing of the Markov chains, a well know problem in dynamic systems with highly autocorrelated states.

Instead, ideas from Gamerman (1998) and Knorr-Held (1999) can be combined to construct extended Kalman filter-like proposal distributions and blocking strategies, respectively. More specifically, the $(T \times m)$ matrix of common dynamic factors f, is broken into B matrices $\tilde{f}_b = (f_{l_b}, \ldots, f_{u_b})'$ of dimensions $n_b \times m$ with $\sum_{b=1}^{B} n_b = T$. The presence of a dynamic regression for the μ_t in (2) with time-varying parameters α_t is easily handled by replacement of the factors f_t by (f_t, α_t), for all t. The full conditional distribution of $\tilde{f}_b$ is given by

$$p(\tilde{f}_b|y, \beta, \gamma, \lambda, \tilde{f}_{-b}) \propto p(\tilde{f}_b|\beta, \gamma, \lambda, \tilde{f}_{-b}) \prod_{t=l_b}^{u_b} p(y_t|f_t, \beta) \tag{11}$$

where $\tilde{f}_{-b} = (\tilde{f}_1', \ldots, \tilde{f}_{b-1}', \tilde{f}_{b+1}', \ldots, \tilde{f}_B')$. Note that the full conditional distribution of $\tilde{f}_b$ depends on $\tilde{f}_{-b}$ only through its neighboring nodes f_{l_b-1} and f_{u_b+1}, for $b \neq 1, B$, f_0 and f_{u_1+1}, for $b = 1$, and f_{l_B-1}, for $b = B$.

3.2.1. Conditional prior distribution

The conditional prior distribution $p(\tilde{f}_b|\beta, \gamma, \lambda, \tilde{f}_{-b})$ can be easily derived from the joint prior distribution $p(f|\gamma, \lambda)$. More specifically, $p(f|\gamma, \lambda) \propto \exp(-0.5 f' K f)$ where the matrix K is a so-called penalty matrix such that $K = G'Q^{-1}G$ is a $mT \times mT$ symmetric matrix, G is a $m(T-1) \times mT$ matrix given by

$$G = \begin{pmatrix} -\Gamma & I_m & 0 & \ldots & 0 \\ 0 & -\Gamma & I_m & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -\Gamma & I_m \end{pmatrix}$$

and Q is a $m(T-1) \times m(T-1)$ block diagonal matrix with elements Λ. Therefore, it can be shown that the conditional prior distribution $p(\tilde{f}_b|\beta, \gamma, \lambda, \tilde{f}_{-b})$ follows a normal distribution $N(\mu_b, \Sigma_b)$ with moments

$$\mu_b = \begin{cases} K^{-1}_{l_b,u_b} K_{u_b,T} f_{u_b+1,T}, & b = 1 \\ K^{-1}_{l_b,u_b} K_{1,l_b-1} f_{1,l_b-1}, & b = T \\ K^{-1}_{l_b,u_b} (K_{u_b,T} f_{u_b+1,T} + K_{1,l_b-1} f_{1,l_b-1}), & \text{otherwise} \end{cases}$$

and $\Sigma_b = K^{-1}_{l_b,u_b}$ where K_{l_u,u_b} denotes the submatrix out of K between the rows and columns $m(l_b-1)+1$ and mu_b associated with the elements of $\tilde{f}_b$ and f_{1,l_b-1} and $f_{u_b+1,T}$ are the matrices localized in the left and right of K_{l_u,u_b}. More details and the generalization for autoregressive priors of order k can be found in Knorr-Held (1999).

3.2.2. Block sampling scheme

A proposal draw for $\tilde{f}_b$ is obtained in two steps. Firstly, the likelihood contribution from $p(y_t|f_t, \beta)$ is approximated by $N(\beta f_t, \hat{V}_t)$ for working observation $\hat{y}_t(\hat{\theta}_t) = h'(\hat{\theta}_t)^{-1}(y_t - h(\hat{\theta}_t)) + \hat{\theta}_t$ and working variance $\hat{V}_t = h'(\hat{\theta}_t)\hat{\Sigma}_t^{-1}(\hat{\theta}_t)h(\hat{\theta}_t)$, $t = 1, \ldots, T$ where $\hat{\theta}_t = \beta f_t$, $h(\cdot)$ is the inverse of the link function and $\hat{\Sigma}_t(\hat{\theta}_t) = b''(\hat{\theta}_t)$ (see Gamerman (1998) for more details). Secondly, a sweep of the FFBS algorithm is performed for each one of the blocks.

The block $\tilde{f}_b$ is sampled using Metropolis–Hastings steps. More specifically, at the l-th iteration a candidate value $\tilde{f}_b^c$ is sampled from the proposal distribution $q(\tilde{f}_b^{(l-1)} \to \tilde{f}_b^c) = q(\tilde{f}_b^c)$ given by the approximate full conditional $p(\tilde{f}_b|\hat{y}, \xi, \tilde{f}_{-b})$ such that

- For $b = 1$

$$q(\tilde{f}_1^c) = p(f^c_{u_1}|f_{u_1+1}, \hat{D}_{u_1}, \xi) \prod_{t=1}^{u_1-1} p(f_t^c|f_{t+1}^c, \hat{D}_t, \xi).$$

- For $1 < b < B$

$$q(\tilde{f}_c^c) = p(f^c_{u_b}|f_{u_b+1}, \hat{D}_{u_b}, \xi) \prod_{t=l_b+1}^{u_b-1} p(f_t^c|f_{t+1}^c, \hat{D}_t, \xi) p(f^c_{l_b}|f^c_{l_b+1}, f_{l_b-1}\hat{D}_{l_b}, \xi).$$

- For $b = B$

$$q(\tilde{f}_B^c) = p(f^c_{u_B}|\hat{D}_{u_B}, \xi) \prod_{t=l_B+1}^{u_B-1} p(f_t^c|f_{t+1}^c, \hat{D}_t, \xi) p(f^c_{l_B}|f^c_{l_B+1}, f_{l_B-1}\hat{D}_{l_B}, \xi).$$

In this case, $\hat{D}_t = \{\hat{y}_1, \ldots, \hat{y}_t\}$ and $\hat{y}_t = \hat{y}_t(\hat{\theta}_t^{(l-1)})$ where $\hat{\theta}_t^{(l-1)} = \beta f_t^{(l-1)}$. Finally, the block $\tilde{f}_b^c$ is accepted with probability

$$\alpha_b = \min\left\{1, \frac{p(\tilde{f}_b^c|y, \xi, \tilde{f}_{-b}^{(l-1)})}{p(\tilde{f}_b^{(l-1)}|y, \xi, \tilde{f}_{-b}^{(l-1)})} \frac{q(\tilde{f}_b^c \to \tilde{f}_b^{(l-1)})}{q(\tilde{f}_b^{(l-1)} \to \tilde{f}_b^c)}\right\}.$$

A few important aspects should be pointed out: (i) fixed blocks can cause problems specifically for factors f_{l_b} and f_{u_b} in the edges of the blocks; (ii) changing the block configuration in each MCMC iteration may be a good alternative and either deterministic or random breakpoints can be considered; (iii) block sizes must be large enough in order to ensure reasonable acceptance rates, to allow for appropriate chain mixing. Thus, in order to determine the optimal number of blocks, a previous tuning MCMC run may be necessary.

The proposed draw is either accepted or rejected based on the ratios of posterior density at proposed and current values and of the approximated extended Kalman filter posterior density at proposed and current values. See Gamerman and Lopes (2006), chapter 6, and the review by Migon et al. (2005), for example, for further details about MCMC schemes in generalized dynamic linear models.

3.3. Sampling the number of factors

Uncertainty about the number of factors is approximately assessed by a reversible jump Markov chain Monte Carlo step adapted from Lopes and West (2004). The proposal distribution for the model parameters conditional on m is given by an independence proposal kernel

$$q_m(\xi_m) = q(f|m_f, \delta_1 V_f) q(\beta|m_\beta, \delta_2 V_\beta) q(\gamma|m_\gamma, \delta_3 V_\gamma) q(\lambda|\delta_4, \delta_4 m_\lambda) q(\phi|\delta_5, \phi_5 m_\phi) q(\tau|\delta_6, \tau_6 m_\tau)$$

where δs are tuning parameters. The marginal proposal densities are normal for f, β and γ and inverse gamma for λ, ϕ and τ where ms and Vs are Monte Carlo approximations for the parameters moments based on preliminary draws for each factor model indexed by m.

Simulated studies suggest that the proposed algorithm has a good performance in choosing the optimal number of common factors. Comparisons between models are based on PMPs that can also be used for Bayesian model averaging. Further details are provided by Lopes and West (2004), Lopes et al. (2008) and by the illustrations in Sections 4 and 5.

Table 1
Gamma synthetic data: model comparison. Sum of squared errors (SSE), sum of absolute errors (SAE), average squared prediction error (ASE_P), average absolute prediction error (AAE_P) and posterior model probability (PMP). The best model for each criterion appears in *italic*.

Model	SSE	SAE	ASE_P	AAE_P	PPM
GSDFM(1)	2581.00	1604.10	1.47	0.92	0.22
GSDFM(2)	*2363.38*	*1557.50*	*1.35*	*0.89*	*0.29*
GSDFM(3)	2507.27	1587.27	1.43	0.91	0.20
GSDFM(4)	2675.68	1639.14	1.53	0.94	0.10
GSDFM(5)	2522.59	1606.57	1.44	0.92	0.19
GSDM-CC	5464.65	2613.67	3.12	1.49	–

4. Simulations

An initial assessment of the methodology was performed with synthetic data. This exercise was devised to ensure appropriate identification of the generating structure and evaluation of the algorithms proposed for extraction of information from the posterior distribution. $N = 25$ (Gamma) and $N = 30$ (Bernoulli) locations were randomly selected on the unit square over 100 time periods. Model performance is assessed with a number of evaluation criteria based on a few fit and forecasting error summarizers and also posterior model probabilities. The number m of factors is explicitly incorporated into the notation of the model, hereafter referred to as GSDFM(m). The last 10 time points were left out of the analysis for evaluation of the forecasting performance of the models and thus $T = 90$. Both examples are based on GSDFM(2), whose prior distributions are $\gamma_j \sim N(0, 1)$, $\lambda_j \sim IG(0.01, 0.01)$, $\tau_j^2 \sim IG(2, 0.5)$ and $\phi_{j,1} \sim IG(2, b)$, for $j = 1, \ldots, 5$, where $b = \max(dist)/(-2\log(0.05))$ such that $\max(dist)$ is the maximum distance between 2 observed locations. The prior distribution of f_0 is $N(0_m, I_m)$.

4.1. Example 1: data with gamma distribution

Observations were generated from a Gamma distribution with $\psi = 1.5$ and θ_{ti}s following a GSDFM(2). The logarithmic link was used instead of the canonical link function. Thus, the natural parameter η_{ti} depends on the linear predictor θ_{ti} through $\eta_{ti} = -\exp(-\theta_{ti})$. Both columns of the factor loadings matrix followed GRFs with zero mean and Matérn correlations with $\phi_{1,1} = 0.15$, $\phi_{2,1} = 0.1$, smoothness parameter $\phi_{j,2} = 1$ and $\tau = (0.25, 1.0)$. The parameters driving the dynamic factor evolution are $\Gamma = \mathrm{diag}(0.8, 0.65)$ and $\Lambda = \mathrm{diag}(0.05, 0.1)$.

Common factors were sampled in blocks of size 15, 9 and 6 time points, i.e. $B = 5$, 10 or 15. Acceptance rates were lowest (<10%) with 5 blocks. Given that computation time is smaller with fewer blocks, $B = 10$ blocks were chosen. Posterior samples were generated according to the algorithm described in the previous section with 2 parallel chains of size 20,000 where the first 5000 were left out as a burn-in period and every 5th iteration was stored thereafter, leading to a total of 3000 draws. The statistic $\hat{R}$ of Gelman and Rubin (1992) was used to assess chain convergence alongside visual inspection of the traces of a few parameters. All algorithms were coded in Ox version 3.4 (Doornik, 2002). The estimated sample autocorrelation function shows a very fast decay indicating appropriate mixing of the algorithm proposed. The empirical acceptance rates for these models are around 30% for the factor loadings and for the blocks of common factors.

Another generalized dynamic spatial model with common component, denoted GSDM-CC, was also tested. It has the same observation equation but the predictor is given by the simpler structure $\theta_t = \mu_t 1_N + \delta_t$, with $\delta_t \sim N(0, \tau^2\rho_\phi(\cdot))$ and $\mu_t \sim N(\mu_{t-1}, W)$. This is supposed to serve as a benchmark for comparison purposes. Table 1 shows these values for a number of models. The largest posterior model probability among GSDFMs is obtained for $m = 2$ as expected. The other performance measures also indicate model GSDFM(2) both in terms of fit and prediction. The worse performance was obtained with GSDM-CC indicating it can not appropriately handle more complex spatio-temporal dependence.

Parameters are well estimated throughout. Figures are not shown for conciseness. Occasionally, one of the factors is overestimated but this effect is compensated by under-estimation of the corresponding loading matrix. For all cases, the spatial pattern of the factor loadings is retained. Marginal credibility intervals contain the generated value in all hyperparameters. The only exceptions are the parameters associated with the variances of the loadings and their factors where their under(over)-estimation leads to corresponding under(over)-estimation of their respective variances.

4.2. Example 2: data with Bernoulli distribution

Data was generated from a Bernoulli distribution with probability p_{ti}, denoted $y_{ti} \sim Bernoulli(p_{ti})$, logistic link $\theta_{ti} = \log\{p_{ti}/(1 - p_{ti})\}$ and θ_{ti}s following a GSDFM(2)-CC such that $\theta_t = \mu_t 1_N + \beta f_t$. Both columns of the factor loadings matrix followed GRFs with means $\kappa_j = \zeta_j 1_N$ such that $\zeta = (-1.0, 1.0)$ and Matérn correlations with $\phi_{1,1} = 0.15$, $\phi_{2,1} = 0.2$, smoothness parameter $\phi_{j,2} = 1$ and $\tau = (0.81, 1.0)$. The parameters driving the dynamic factor evolution are $\Gamma = \mathrm{diag}(0.75, 0.9)$ and $\Lambda = \mathrm{diag}(0.1, 0.05)$. The same prior distributions from Example 1 were used in addition to $\zeta \sim N((-1.0, 1.0)', 5I_2)$ and $W \sim IG(2, 0.1)$. Factors were sampled in $B = 9$ blocks. The empirical acceptance rates for these models are 64% for the factor loadings and 25% for the blocks of factors.

Table 2
Bernoulli synthetic data: model comparison. Average squared errors (ASE), average absolute errors (AAE), average squared prediction error (ASE_P), average absolute prediction error (AAE_P) and posterior model probability (PMP). The best model for each criterion appears in *italic*.

Model	ASE	AAE	ASE_P	AAE_P	PPM
GSDFM-CC(1)	0.2001	0.4373	0.2380	0.4471	0.20
GSDFM-CC(2)	*0.1949*	*0.4000*	*0.2330*	*0.4190*	*0.38*
GSDFM-CC(3)	0.2134	0.4212	0.2412	0.4360	0.27
GSDFM-CC(4)	0.2221	0.4410	0.2490	0.4522	0.15
GSDM-CC	0.3520	0.6450	0.3998	0.6993	–

In total, 5 models were entertained: GSDFM(m)-CC, for $m = 1, 2, 3, 4$ and the GSDM-CC of the previous exercise. Table 2 shows that model GSDFM(2)-CC has the largest posterior probability as expected and better performance on the other evaluation criteria. Once again, model GSDM-CC showed the worst performance. Once again, results are not presented here for conciseness but all parameters of GSDFM(2)-CC are well estimated and chains show indications of the fast mixing of the algorithm.

5. Modeling rainfall in Minas Gerais, Brazil

Modeling the occurrence of rain is important in the study of climatology because it helps the understanding of the structure of precipitation probabilities. Rainfall data has been extensively studied in the literature. One approach is to the relative frequency of rainy days as in Stern and Coe (1984). One of the earlier studies of the frequency of rainy days is Gabriel and Neumann (1962), where data from Israel is analyzed. They found a strong persistence of consecutive rainy days and obtained a good fit by using first order Markov chains. These ideas were developed further and, in the last decade, Hughes et al. (1999) proposed a non-homogeneous hidden Markov model to relate atmospheric measurements to the occurrence of rain through latent states. A more recent approach is provided by Fernandes et al. (2009). They proposed a mixture model between a point mass at 0 and a continuous distribution and used dynamic models to describe the components of the mixture. The probability of the occurrence of rain arises naturally as the mixing probability.

This section illustrates the capabilities of the GSDFM to describe the spatial-temporal dependence of analysis of the occurrence of rain. The models capture the spatial correlation between monitoring stations and the dynamic structure of the rain probabilities. This will be shown by contrasting the proposed models against some currently available models.

5.1. Data

Data contains the $T = 365$ daily occurrences of rain in 2005 measured at 17 meteorological stations in the state of Minas Gerais, Brazil. This dataset is available from SIMGE (Meteorology and Hydrological Resource System from Minas Gerais, in Portuguese). It transmits data from the Brazilian satellites SCD-1 and SCD-2. Further details, including the data used here can be found at http://www.simge.mg.gov.br. The occurrence of rain is defined as any rain measurement larger than 1 mm as in Kitagawa (1987). Fig. 2 exhibits the geographical locations of the stations. Latitude and longitude were transformed into UTM coordinates. The largest distance between the stations is 1036 km, between stations HO and NA.

The state of Minas Gerais has an area of 586,528 km^2 and is located in the southeastern region of Brazil. The predominant climates are tropical and subtropical, in the highest and southernmost regions of the state. In the northern and northeastern regions of the state, scarce rainfall and high temperatures occasionally leads to droughts. Fig. 3 shows the frequency of rain in each month for several of the stations. It shows that rain is less frequent around winter months (April to October) throughout the state and that the northern and northeasterly stations present less rainy days than other regions. Also, southern and southwesterly stations present more rainy days in January, September and October than the other stations. These indications seem to suggest a north-south divide in the state.

All stations contain missing data, with amounts as high as 91, 37, 37, 29, 26 and 23 days for stations HO, SF, MA, NA, IT and PR, respectively. These observations are routinely imputed with the MCMC algorithm by treating them as additional parameters and sampling from their Bernoulli distributions. Stations Pirapora (PI) and Caratinga (CA) were removed from the analysis for evaluation of the interpolation performance of the models and therefore only $N = 15$ were considered for the analysis.

5.2. Models

Kitagawa (1987) and Gamerman (1998) modeled the number of rainy days in city of Tokyo in the period 1983-1984. Their objective was to illustrate the use of dynamic generalized linear models with binomial response in the temporal domain. Here, these models are extended to include also the spatial domain. The advantages are: (i) identification through latent factors of geographical regions of similar behaviour; and (ii) the model ability to interpolate beyond measured stations to the entire region of interest.

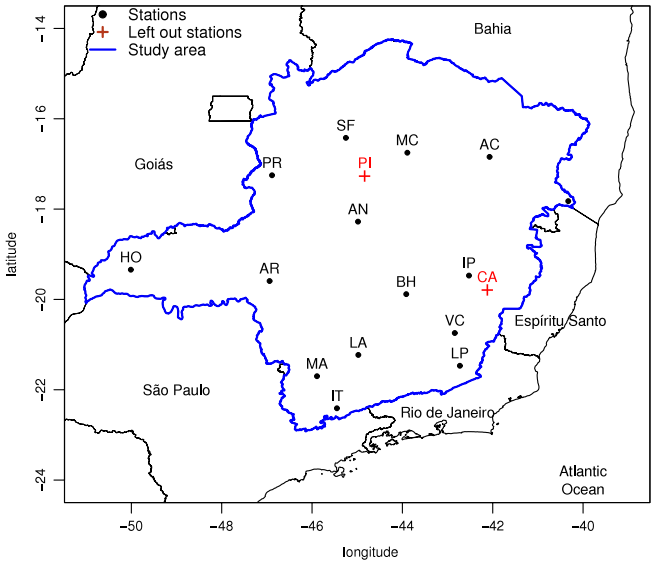

Fig. 2. Monitoring rain stations in the state of Minas Gerais, southeast Brazil. The solid line contours the state. Pirapora (PI) and Caratinga (CA) states (+) were retained for a spatial interpolation exercise.

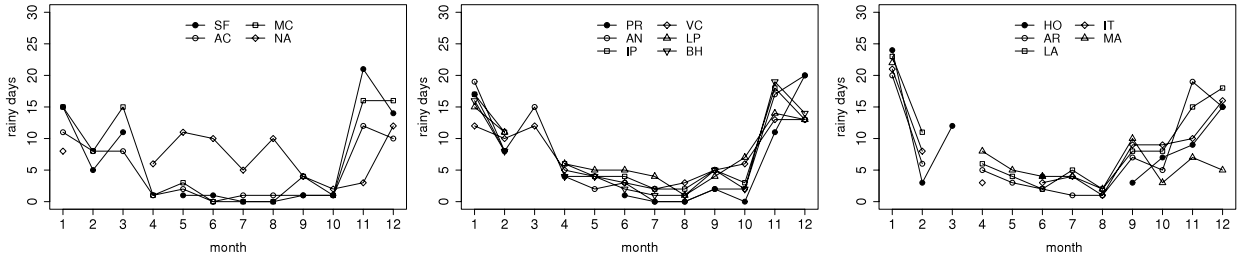

Fig. 3. Monthly counts of rainy days for north and northeast stations (left), center and southeast stations (center) and south and southwest stations (right).

The general model considered is given by Eqs. (1)–(4) with a Bernoulli likelihood in (1) with a logistic link and Matérn correlation functions with smoothness parameter equal to one for all factors. Models differ only by the number m of factors and specification of μ_t and κ_j. Two classes of models were considered:

1. Standard GSDFM(m): $\mu_t = 0\ \forall t$, and $\kappa_j = \zeta_j 1_N$, for $j = 1, \ldots, m$;
2. GSDFM(m)-CC: $\theta_t = \mu_t 1_N + \beta f_t$, $\mu_t \sim N(\mu_{t-1}, W)$, $\mu_0 \sim N(\tilde{\mu}_0, \tilde{C}_0)$ and $\kappa_j = \zeta_j 1_N$, for $j = 1, \ldots, m$.

Two models with simpler spatio-temporal specifications were considered for comparison:

1. Generalized dynamic model (GDM-CC): $\theta_t = \mu_t 1_N$, $\mu_t \sim N(\mu_{t-1}, W)$ and $\mu_0 \sim N(\tilde{\mu}_0, \tilde{C}_0)$;
2. Generalized spatial dynamic model (GSDM-CC): $\theta_t = \mu_t 1_N + \delta$, $\mu_t \sim N(\mu_{t-1}, W)$, $\mu_0 \sim N(\tilde{\mu}_0, \tilde{C}_0)$ and $\delta \sim N(0, \tau^2 R(\phi))$.

Note that GSDM-CC is a special case of GSDFM(1)-CC when the single factor equals 1 and GDM-CC is a special case of GSDM-CC, when no spatial component is present. Factor models considered up to 4 factors but significant results were never found for the 4th factor and this was dropped. Relatively vague prior distributions were assumed for all hyperparameters with $\lambda_j \sim IG(0.01, 0.01)$, $\mu_j \sim N(0, 1)$, $\tau_j^2 \sim IG(2, 0.5)$, $\phi_{j,1} \sim IG(2, b)$, for $j = 1, \ldots, m$ and $W \sim IG(0.01, 0.01)$, where $b = \max(dist)/(-2\log(0,05))$. The prior mixture was used for the autoregressive factor coefficients with prior weight 0.5 to the point mass at 1, thus allowing for non-stationary factors. Prior distributions for the initial states were $f_0 \sim N(0_m, I_m)$ and $\mu_0 \sim N(0, 1)$. Finally the prior distributions for parameters of models GDM-CC and GSDM-CC were the same as those used in the GSDFM-CC.

Acceptance rates of the Metropolis–Hastings algorithm decrease as the number of factors increase. Thus, factors were sampled in 15 and 20 blocks for the GSDFM and GSDFM-CC. Common components μ_t were jointly sampled in 10 blocks. Two chains of size 30,000 were generated with a burn-in period of 10,000 and thinning of 5 iterations leading to a posterior sample of 4000 draws after previously described checks for convergence were made. All computations were carried out in Ox version 3.4 on a Windows desktop with a dual core 1.8 GHz processor and 500 Mb memory. The processing time was approximately 40 h.

Table 3
Rain in Minas Gerais: comparison criteria. Mean square error (MSE), mean absolute error (MAE), logarithm of the model likelihood (LML) and posterior model probability (PMP). The best models for each criteria appear in *italic* font. Superscript 1 indicates in-sample MSE and MAE. Superscript 2 indicates MSE and MAE for interpolated stations PI and CA.

Model	MSE^1	MAE^1	MSE^2	MAE^2	LML	PMP
GSDFM(1)	0.1243	0.2530	0.1045	0.2285	−2118.0	0.15
GSDFM(2)	0.1057	0.2219	*0.0973*	*0.2164*	−1908.9	0.39
GSDFM(3)	*0.1039*	*0.2175*	0.1024	0.2239	*−1881.1*	*0.46*
GSDFM(1)-CC	0.1045	0.2128	0.1046	0.2178	−1849.5	0.09
GSDFM(2)-CC	*0.0923*	*0.1986*	*0.0958*	0.2103	*−1740.0*	*0.57*
GSDFM(3)-CC	0.0996	0.2074	0.0992	*0.2101*	−1766.8	0.34
GSDM-CC	0.1244	0.2545	0.1080	0.2425	−2171.0	–
GDM-CC	0.1250	0.2570	–	–	−2198.9	–

Table 4
Rain in Minas Gerais: posterior summary for static parameters of GSDFM(2)-CC.

Parameter	Mean	s.d.	Percentiles		
			2.5%	50%	97.5%
W	0.59	0.12	0.41	0.58	0.80
γ_1	0.35	0.07	0.22	0.35	0.48
γ_2	0.57	0.06	0.46	0.57	0.68
λ_1	0.75	0.11	0.56	0.74	1.00
λ_2	0.74	0.08	0.60	0.73	0.92
μ_1	−0.37	0.49	−1.23	−0.41	0.74
μ_2	0.25	0.69	−1.16	0.25	1.59
τ_1^2	0.91	0.56	0.33	0.75	2.49
τ_2^2	2.19	1.30	0.84	1.84	5.80
$\phi_{1,1}$	250.72	135.46	72.15	221.05	603.37
$\phi_{2,1}$	379.96	174.41	156.60	351.26	840.38

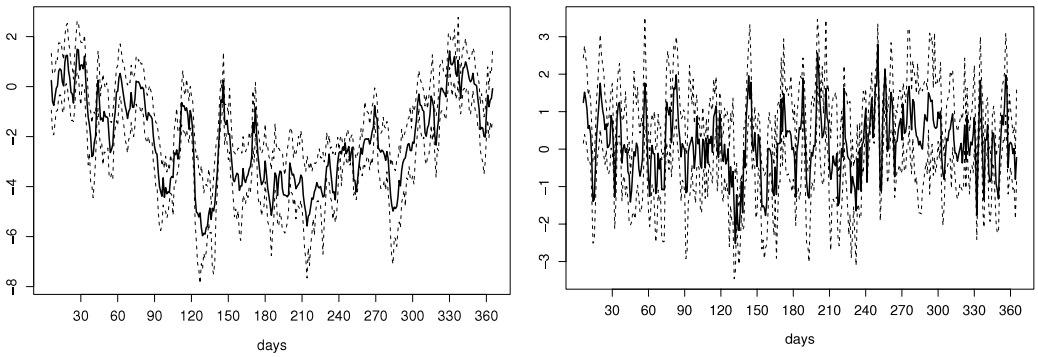

Fig. 4. Daily posterior estimates of common component (left) and second dynamic factor (right) estimated from model GSDFM(2)-CC. Solid lines are posterior means and dashed lines are 95% credibility intervals.

5.3. Results

Table 3 shows a number of performance criteria for the models considered. Models GSDFM(3) and GSDFM(2)-CC show the best results in their class with substantially high posterior probabilities. Between these two models, all other criteria suggest the latter. The models are also substantially better than the benchmark models considered. Model results hereafter refer to GSDFM(2)-CC.

Table 4 shows posterior summaries and some sampling figures. Acceptance rates are 27% for loadings, range from 8% and 49% for blocks of factors, and range from 52% and 57% for the range parameters. The common component shows smaller temporal variability than the factors. The factors are stationary with $\hat{p}(\gamma_1 = 1|y) = \hat{p}(\gamma_2 = 1|y) = 0$. The means of the GRF for the loadings seem to be null but their range parameters indicate strong spatial correlation with mean values 0.643 and 0.796 for locations 200 km apart.

Fig. 4 shows the estimated trajectories of the dynamic components of the model. The common components seems to capture the overall yearly trend with a decrease from April to September. It also seems to indicate an inter-monthly cycle within each month. The first factor exhibits an irregular variation with a low AR(1) coefficient. The second factor shows a more irregular but also more persistent pattern. It also shows a reduced level in the first days of May that may be related to the fact that this period marks the beginning of the drier season.

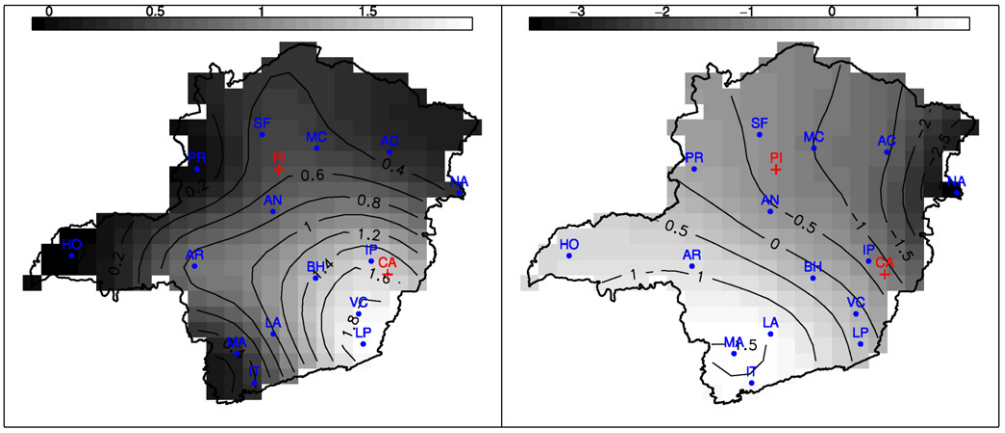

Fig. 5. Spatial interpolation of the two columns of the factor loadings matrix. Values above each contour line indicate the levels of the posterior means.

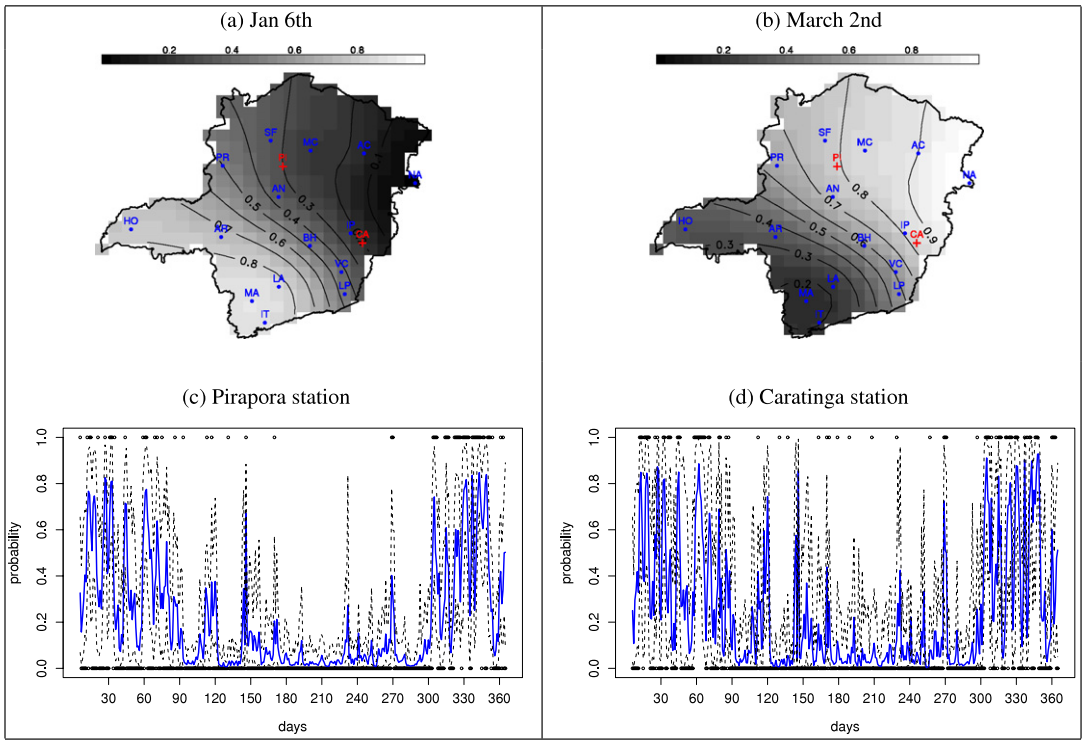

Fig. 6. (a) and (b) Mean probability maps for two typical days in 2005 for gauged stations (dot) and ungauged stations (+). (c) and (d) Daily posterior probability of rain for ungauged Pirapora and Caratinga stations in 2005. Dots are observed rain indicators, solid lines are rain mean probabilities and dashed lines are 95% credibility intervals.

The more revealing figure of factor models in general is the interpretation of the factor loadings. Fig. 5 shows the loadings for the entire state interpolated via Bayesian kriging. The plots indicate a smooth variation in the southeasterly and center-southwesterly directions for the first and second factors respectively. These findings reflect the initial impression from the data and the geographical structure of the climate in this region. The overall conclusion indicates a more stable rain pattern in the southeast than in the center and southwest, where rain is more likely.

Fig. 6(a) and (b) show rain probability maps for the whole state for two days of the year 2005. Smooth variations in space and time can be observed when sequentially analyzing the days of the year. Finally, Fig. 6(c) and (d) show the result of the spatial interpolation. The interpolated rain probabilities seem to follow the general trend of the observed binary time series at each of the two locations.

6. Conclusions

In this paper, many specifications of GSDFM were used to characterize spatio-temporal variation. These were shown to provide a meaningful interpretation in the analysis of climatological data presented both in terms of spatial and temporal variations. The factors enabled a decomposition of the geographical region of interest. Bayesian kriging played an important role for the characterization of the latent factors over the entire region, beyond the observed locations.

A distinctive feature of the models is their ability to estimate the number of factors in the non-normal setting and allow for estimation with a chosen number of factors or averaging over all possible number of factors. Sampling is cumbersome especially for larger number of factors but is possible, and the idea of forming blocks of factors seems to help MCMC convergence. Alternative approximating schemes can also be used and are currently under investigation.

The models can be extended in a number of directions. The most obvious one is the incorporation of covariates. In climatological studies, many other variables are typically available and can be used following ideas in Hughes et al. (1999). Other possible extensions are the consideration of dynamic loadings (Lopes and Carvalho, 2007) to account for changes in the composition of the factors as time passes and analysis of other forms of spatial data configuration, such as aerial data.

Acknowledgements

The first author would like to thank The University of Chicago Booth School of Business for financially supporting his research, while the other authors acknowledge financial support from CNPq and FAPERJ grants. This paper is based on the unpublished Ph.D. Thesis of the third author under the supervision of the first and second authors.

References

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.
Cressie, N., Huang, H., 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of the American Statistical Association 94, 1330–1340.
Doornik, J.A., 2002. Object-Oriented Matrix Programming using Ox, 3rd ed. Timberlake Consultants Press and Oxford, London.
Fernandes, M.V., Schmidt, A.M., Migon, H.S., 2009. Modelling zero-inflated spatio-temporal processes. Statistical Modelling 9 (1), 3–25.
Frühwirth-Schnatter, S., 1994. Data augmentation and dynamic linear models. Journal of Time Series Analysis 15, 183–202.
Gabriel, K.R., Neumann, J., 1962. A Markov chain model for daily rainfall occurrences at Tel Aviv. Quarterly Journal Royal Meteorological Society 88, 90–95.
Gamerman, D., 1998. Markov chain Monte Carlo for dynamic generalised linear models. Biometrika 85, 215–227.
Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd ed. Chapman and Hall/CRC, Boca Raton, USA.
Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–511.
Hoeting, J., Madigan, D., Raftery, A., Volinsky, C., 1999. Bayesian model averaging (with discussion). Statistical Science 14, 382–401.
Hooten, M., Wikle, C., 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. Environmental and Ecological Statistics 15, 59–70.
Hughes, J.P., Guttorp, P., Charles, S.P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. Journal of the Royal Statistical Society: Series C (Applied Statistics) 48, 15–30.
Jin, X., Carlin, B., Banerjee, S., 2005. Generalized hierarchical multivariate CAR models for areal data. Biometrics 61, 950–961.
Kitagawa, G., 1987. Non-Gaussian state-space modeling of non-stationary time series. Journal of the American Statistical Association 82, 1032–1041.
Knorr-Held, L., 1999. Conditional prior proposals in dynamic models. Scandinavian Journal of Statistics 26, 129–144.
Knorr-Held, L., Best, N.G., 2001. A shared component model for detecting joint and selective clustering of two diseases. Journal of the Royal Statistical Society, Series A 164, 73–85.
Lopes, H.F., Carvalho, C.M., 2007. Factor stochastic volatility with time varying loadings and Markov switching regimes. Journal of Statistical Planning and Inference 137, 3082–3091.
Lopes, H.F., Salazar, E., Gamerman, D., 2008. Spatial dynamic factor models. Bayesian Analysis 3, 759–792.
Lopes, H.F., West, M., 2004. Bayesian model assessment in factor analysis. Statistica Sinica 14, 41–67.
Migon, H.S., Gamerman, D., Lopes, H.F., Ferreira, M.A.R., 2005. Dynamic Models. In: Dey, D.K., Rao, C. (Eds.), Handbook of Statistics: Bayesian Thinking, Modeling and Computation, vol. 25. Noth Holland.
Stern, R.D., Coe, R., 1984. A model fitting analysis of daily rainfall data. Journal of the Royal Statistical Society, Series A 147, 1–34.
Wang, F., Wall, M.M., 2003. Generalized common spatial factor model. Biostatistics 4, 569–582.
West, M., Harrison, P., 1997. Bayesian Forecasting and Dynamic Models. Springer, New York.