Spatially Varying Autoregressive Processes

Aline A. Nobre, Bruno Sansó and Alexandra M. Schmidt *

Abstract

We develop a class of models for processes indexed in time and space that are based on autoregressive (AR) processes at each location. We use a Bayesian hierarchical structure to impose spatial coherence for the coefficients of the AR processes. The priors on such coefficients consists of spatial processes that guarantee time stationarity at each point in the spatial domain. The AR structures are coupled with a dynamic model for the mean of the process, which is expressed as a linear combination of time-varying parameters. We use satellite data on sea surface temperature for the North Pacific to illustrate how the model can be used to separate trends, cycles and short term variability for high frequency environmental data .

KEY WORDS: **spatio-temporal model, AR processes, process convolutions**

1 Introduction

The time series in Figure 1 could very well correspond to the differences between monthly temperature means over a period of ten years and monthly temperature readings at a meteorological station. In such case, some questions relevant to the climate dynamics of the area would be: Are there periodic cycles in the data? Can we detect a trend in the level or the amplitude of the signal? The abundance of information from remote sensors, like satellites, and automatic environmental data networks has produced an explosion of,data with high time frequency and high space resolution. Answering the former questions for relevant spatial domains is key for the understanding of long term environmental changes. Thus effective statistical tools are needed to identify spatially coherent signals from noisy collections of geo-referenced time series and detect significant cycles and trends.

*Aline A. Nobre is a Researcher of the Scientific Computing Program, at Oswaldo Cruz Foundation, Av. Brasil, 4365 Manguinhos, Rio de Janeiro-RJ CEP:21.045-900 Brazil. aline@fiocruz.br. Bruno Sansó is Professor of Statistics and Chair, Department of Applied Mathematics and Statistics, University of California at Santa Cruz, 1156 High St. MS: SOE2, Santa Cruz, CA-95064, U.S.A. bruno@ams.ucsc.edu, www.ams.ucsc.edu/∼bruno. Alexandra M. Schmidt is Associate Professor at the Instituto de Matemática, Universidade Federal do Rio de Janeiro, Caixa Postal 68530 Rio de Janeiro - RJ CEP:21.945-970 Brazil. alex@im.ufrj.br, http://www.dme.ufrj.br/∼alex/ .

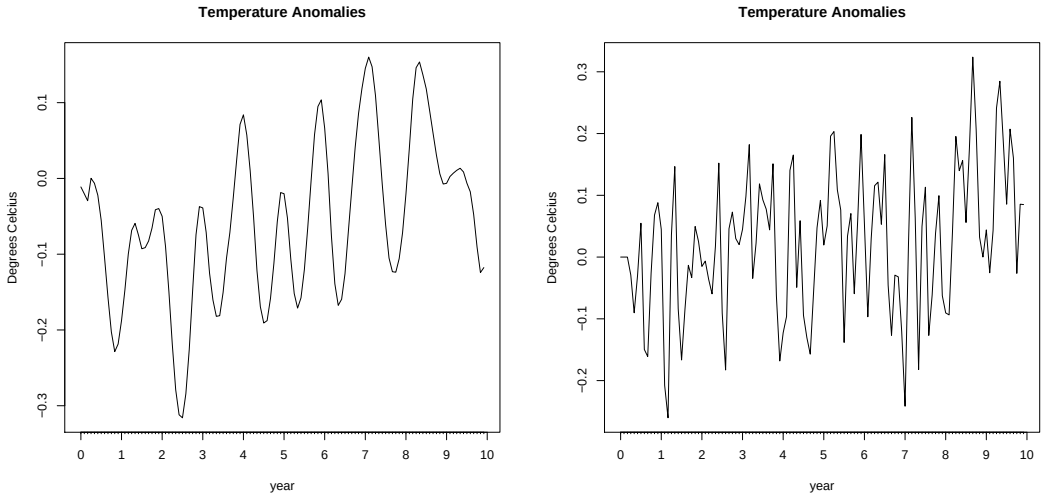


Figure 1: Simulated temperature data from two AR(3) processes. The roots of the characteristic polynomial of the process in the left panel are $G_1 = 0.9, G_2 = 0.9e^{i\pi/6}, G_3 = 0.9e^{-i\pi/6}, \sqrt{-1} = i$ and the variance is $\sigma^2 = 0.1^2$. For the left panel $G_1 = 0.7, G_2 = 0.7e^{i\pi/2}, G_3 = 0.7e^{-i\pi/2}$ and $\sigma^2 = 0.1^2$.

The data presented in Figure 1 do not correspond to any actual observations. They have been simulated from two autoregressive processes of order three (AR(3)). They correspond to stationary processes, so they have no trends or cycles. It is clear that a naive least squares fit would produce seemingly significant positive slopes in both cases. Thus, careful inference on the stochastic structure of the process generating the data is needed. Autoregressive processes (see, for example, Priestley, 1981) are well known times series models that provide great flexibility in spite of their simple formulation. A univariate process x_t that is an AR(p) can be written as

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2).$$

Using the backwards operator B, $Bx_t = x_{t-1}$, we factorize this expression as

$$\prod_{i=1}^{p} (1 - G_i B) x_t = \varepsilon_t.$$

G_i are the roots of the characteristic polynomial. The process is stationary when $|G_i| < 1$, $\forall i$ (see, for example Box et al., 1994). The fact that the simulations in Figure 1 show possible cycles and trends should not come as a surprise, as positive real roots provide some persistence and complex roots provide quasi periodicities. This features are likely to be present in long environmental time series with high frequency data. On the other hand, when roots are outside the region of stationarity, the process shows an explosive behavior. When an AR(p) is fitted to time series corresponding to nearby locations, we need to make

sure that the coefficients are in the stationarity region. Furthermore, the condition should hold for the coefficients of spatial interpolators, to prevent them from being explosive. The goal of this paper is to build spatio-temporal models that satisfy those conditions.

The building block of our proposed methods is given by the following model: Let $s \in S \subset \mathbb{R}^d$ denote location, then, for a space-time process, we have

$$x_t(s) = \sum_{i=1}^{p} \phi_i(s) x_{t-i}(s) + \varepsilon_t(s) \quad \text{or} \quad \prod_{i=1}^{p} (1 - G_i(s)B) x_t(s) = \varepsilon_t(s),$$

with $\text{cov}(\varepsilon_t(s), \varepsilon_t(s')) = \tau^2 \Sigma_{ss'}$, where Σ is a covariance matrix. We obtain time stationarity and spatial coherence by assuming appropriate priors on $G_i(s)$. Imposing stationarity on the original coefficients $\phi_i(s)$ would be hopeless due to the difficulty of describing the stationarity region (Huerta and West, 1999). Notice that, since the coefficients of the AR(p) vary with s, the process is not spatially stationary. Also, there is no separability between the space and the time components. That is, the joint space-time covariance function can not be factorized into a space covariance multiplied by a time covariance.

The literature on spatio-temporal modeling has been experiencing a significant growth in the last 10 years. Here, we concentrate on Gaussian spatio-temporal processes. Usually the main concern lies on proposing flexible covariance structures which are non-separable and non-stationary in both dimensions. Cressie and Huang (1999); Gneiting (2002); Stein (2005) present examples of non-separable stationary covariance functions for space-time processes. Cressie and Huang (1999) propose classes of covariances that are obtained by inverting spectral densities of the form $h(\omega, v) = \rho(\omega, v)k(\omega)$, where ω denotes the spatial frequency and v the temporal one. Gneiting (2002) defines a class of covariances that provides a direct construction of valid covariance functions, without the need of inverting the spectral density. Stein (2005) considers classes of covariance functions that are smooth everywhere, except possibly at the origin.

Another way of obtaining flexible covariance structures is through the use of conditional dynamic linear models (West and Harrison, 1997). There are several examples of spatio-temporal applications based on state space models in the literature. Some recent examples are Wikle et al. (1998); Sansó and Guenni (2000); Shaddick and Wakefield (2002); Higdon (2002); Huerta et al. (2004); Calder (2005); Sahu and Mardia (2005); Lemos and Sansó (2009). Sansó et al. (2008) consider a linear representation of a spatio-temporal process inspired by corregionalization models for multivariate spatial data (Wackernagel, 2001). They mention the idea of using spatial processes for the roots of the characteristic polynomial of an AR(p). This paper carefully develops that approach by extending to the spatial domain the priors for autoregressive processes proposed in Huerta and West (1999).

The paper is organized as follows. Section 2 introduces in detail our proposed model. As the inference procedure follows the Bayesian paradigm, the prior distributions of the parameters are also discussed therein. Section 3 presents the inference procedure used to obtain samples from the resultant posterior distribution. The next Section illustrates the proposed model by analyzing satellite data on sea surface temperature. Finally, Section 5 contains a discussion of the methods proposed in the paper.

2 A spatio-temporal model

Denote the observed process as $y_t(s)$, with $s = s_1, \ldots, s_n \in S \subset \mathbb{R}^d$ and denote $\boldsymbol{y}_t = (y_t(s_1), \ldots, y_t(s_n))'$ the vector of observations at time t. analysis and assume that all models are conditional on observing most applications it is very natural to consider spatio-temporal processes where the mean evolves in time. Let the mean of $\boldsymbol{y}_t$ be $\boldsymbol{\mu}_t$. We assume that $\boldsymbol{\mu}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t$ where $\boldsymbol{F}_t \in \mathbb{R}^{k\times n}$ corresponds to a matrix of observed covariates and $\boldsymbol{\theta}_t \in \mathbb{R}^k$ is an unobserved latent variable. Consider a covariance function $C(s, s')$ and the corresponding covariance matrix $\boldsymbol{\Sigma}_{i,j} = C(s_i, s_j)$. C can be treated with full generality, but we will focus on the isotropic case where $C(s, s') = \sigma^2\rho(||s - s'||; \lambda, \kappa)$, where ρ is a correlation function with a parametric form that depends on two univariate parameters, λ and κ. Typical examples are the power exponential and the Matérn classes of correlations (Banerjee et al., 2004). We factorize $\boldsymbol{\Sigma}$ as $\boldsymbol{K}\boldsymbol{K}'$, where $\boldsymbol{K}$ is a square root matrix of $\boldsymbol{\Sigma}$. We assume that $\boldsymbol{y_t}$ can be decomposed into a time-varying mean plus a spatially correlated autoregressive vector $\boldsymbol{x}_t$. Thus

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{K}\boldsymbol{x}_t && (1)\\
\boldsymbol{x}_t &= \sum_{j=1}^{p} \boldsymbol{\Phi}_j\boldsymbol{x}_{t-j} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N_n(0, \tau^2\boldsymbol{I}_n) && (2)\\
\boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim N_k(0, \boldsymbol{W}_t) && (3)
\end{aligned}$$

where $t = p+1, \ldots, T$ and $\boldsymbol{\Phi}_j = diag(\phi_j(s_1), \ldots, \phi_j(s_n)) \in \mathbb{R}^{n\times n}$. Equation (1) corresponds to an observation equation where the measurement error is "colored" both spatially and temporally. Equation (2) defines the autoregressive structure of the measurement error. Finally Equation (3) defines the evolution of the underlying process. A spatially and temporally dependent measurement error is a realistic assumption in many environmental processes, in particular when the observations are obtained using remote sensing. Alternatively, the interest could be focused on assuming that the process of interest is a spatially-varying autoregression and that the actual observations, say $\boldsymbol{Y}_t$ is subject to a measurement errors that is simple white noise. We would then add the equation $\boldsymbol{Y}_t = \boldsymbol{y}_t + \boldsymbol{\eta}_t,\ \boldsymbol{\eta_t} \sim N(0, \kappa^2\boldsymbol{I}_n)$ as a top layer in the hierarchy defined by Equations (1)–(3). This additional layer requires a straightforward extension of our proposed sample based inference. We will not focus on this as it does not provide any conceptual gains.

Collapsing Equations (1) and (2) we obtain that

$$\boldsymbol{y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{K}\left(\sum_{j=1}^{p} \boldsymbol{\Phi}_j\boldsymbol{x}_{t-j} + \boldsymbol{\epsilon}_t\right) = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \sum_{j=1}^{p} \boldsymbol{\Phi}_j(\boldsymbol{y}_{t-j} - \boldsymbol{F}_{t-j}'\boldsymbol{\theta}_{t-j}) + \boldsymbol{v}_t$$

and thus

$$\boldsymbol{y}_t^* = \sum_{j=1}^{p} \boldsymbol{\Phi}_j\boldsymbol{y}_{t-j}^* + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \sim N_n(0, \tau^2\boldsymbol{\Sigma}),$$

where $\boldsymbol{y}_t^* = \boldsymbol{y}_t - \boldsymbol{F}_t'\boldsymbol{\theta}_t$. We observe that only the product $\sigma^2\tau^2$ is identifiable in the likelihood. Our choice is to fix one of the two variance parameters, in particular, we set $\tau^2 = 1$. Thus,

$\boldsymbol{y}_t^*$ is a multivariate AR(p) with spatially varying coefficients, $\boldsymbol{\Phi}_j$, $j = 1, \ldots, p$ and spatially correlated variance, $\boldsymbol{\Sigma}$. In essence, at each location s we have a time series modeled through an autoregressive process with a common order for all locations but with different coefficients. We refer to this model as a Spatially-Varying Autoregression of order p (SVAR(p)). Suppose that the order of the process is $p = R + 2C$, where R and C are, respectively, the number of real and complex roots of the characteristic polynomial. Then the AR(p) model at location s can written as

$$\prod_{j=1}^{R}(1 - a_j(s)B) \prod_{j=R+1}^{R+C} (1 - r_j(s)e^{i\omega_j(s)}B)(1 - r_j(s)e^{-i\omega_j(s)}B)y_t^*(s) = v_t(s) \tag{4}$$

Here $i = \sqrt{-1}$, $a_j(s)$ is the j-th real root, $r_j(s)$ and $\omega_j(s)$ are, respectively, the modulus and the argument of the j-th complex root.

To obtain a stationary process at each location s we need to impose the conditions $|a_j(s)| < 1,\ 0 < r_j(s) < 1$ and $-\pi < \omega_j(s) < \pi\ , \forall j$ and $\forall s$. To achieve this we need to consider bounded spatial processes indexed in s. A natural approach would be to consider Gaussian processes transformed to a bounded interval by a function like the logistic, as proposed in Sansó et al. (2008). While this will provide the flexibility of Gaussian processes, it has the problem of being highly sensitive to the estimation of the parameters that define the processes' covariance structure. To see why this is the case, consider the logistic transformation of a univariate normal. For some choices of the normal variance we can have bimodality in the transformed distribution, due the accumulation of mass in the tails of the normal that is mapped to the extremes of the unit interval. This phenomenon is exacerbated in a multivariate setting. In the next two sections we consider alternative prior spatial specifications.

2.1 Prior processes for the real roots

The idea of representing a spatial process using a kernel convolution has been applied successfully in the recent literature. A very clear explanation of process convolutions is presented in Higdon (2007). In a similar fashion to Lee et al. (2007), where a process convolution is used to obtain a positive process, we consider the convolution of bounded variables to obtain a bounded process. Thus we consider a kernel k, a collection of locations $u_1, \ldots, u_M \in S$ and let

$$a_j(s) = 2\sum_{m=1}^{M} k(s - u_m)z_j^a(u_m) - 1\ \ , j = 1, \ldots, R\ ,$$

where $z_1^a(u_m) \sim Beta(\alpha_1, \beta_1), m = 1, \ldots, M$, and $z_j^a(u_m)|z_{j-1}^a(u_m) \sim Beta(\alpha_j, \beta_j)I_{[0, z_{j-1}^a(u_m)]}$ $(z_j^a(u_m))$ for $m = 1, \ldots, M$ and $j = 2, \ldots, R$. In other words, for the first real root, we consider a collection of M beta-distributed random variables. We use a kernel and an affine transformation to create a process bounded in $(-1, 1)$. For the second real root, we condition on the beta-distributed variables corresponding to the first root and generate from a

truncated beta. We then transform to the interval $(-1, 1)$. We repeat the process conditioning on the previous root until the R-th one. This guarantees that the roots are ordered in descending order. The ordering is important for the identifiability of the parameters in the model.

2.2 Prior processes for the complex roots

To obtain a prior spatial process for the modulus and the argument of the complex root we first consider a transformation of the moduli and arguments in the second term of the product in Equation (4). Let $\phi_{1j}(s) = 2r_j(s)\cos\omega_j(s)$ and $\phi_{2j}(s) = -r_j(s)^2$. The stationarity conditions translate into the restrictions $\phi_{1j}(s) \in (l_{1j}(s), l_{2j}(s))$ and $\phi_{2j}(s) \in (-1, 0)$ where $l_{1j}(s) = -2\sqrt{-\phi_{2j}(s)}$ and $l_{2j} = 2\cos(4\pi/t)\sqrt{-\phi_{2j}(s)}$. We proceed in a similar fashion to the approach taken for the real roots. That is we convolve bounded random variables to obtain bounded random fields. We assume that

$$\phi_{2j}(s) = \sum_{m=1}^{M} k(s - u_m) z_j^{\phi}(u_m) - 1$$

and

$$\phi_{1j}(s) = (l_{2j}(s) - l_{1j}(s)) \sum_{m=1}^{M} k(s - u_m) z_j^{\phi}(u_m) + l_{1j}(s)$$

where $j = 1, \ldots, C$. Then, for each of the collections of variables $z^{\phi}(\cdot)$, we have, $z_1^{\phi}(u_m) \sim Beta(\alpha_1, \beta_1)$ and $z_j^{\phi}(u_m)|z_{j-1}^{\phi}(u_m) \sim Beta(\alpha_j, \beta_j) I_{[0, z_{j-1}^{\phi}(u_m)]}(z_j^{\phi}(u_m))$ for $m = 1, \ldots, M$ and $j = 2, \ldots, R$. As in the case of the real roots, we impose ordering for the latent variables in order to achieve identifiability.

3 Fitting the model

Inference for the parameters of the SVAR(p) is based on sampling the posterior distribution using Markov chain Monte Carlo (Gamerman and Lopes, 2006). For this we need to explore the full conditional distributions of all model parameters. We start by noting that, conditional on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p$, the likelihood is proportional to

$$|\boldsymbol{\Sigma}|^{-(T-p)/2} \exp\left\{ -\frac{1}{2} \sum_{t=p+1}^{T} \left(\boldsymbol{y}_t^* - \sum_{j=1}^{p} \boldsymbol{\Phi}_j \boldsymbol{y}_{t-j}^* \right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}_t^* - \sum_{j=1}^{p} \boldsymbol{\Phi}_j \boldsymbol{y}_{t-j}^* \right) \right\}. \tag{5}$$

The full conditional of the parameters of the covariance function, denoted as $\boldsymbol{\Lambda} = (\sigma^2, \lambda, \kappa)$, given $\boldsymbol{\theta}_t$ and $\boldsymbol{\Phi}_j$ is the product of the equation (5) and the prior $p(\boldsymbol{\Lambda})$. We sample these parameters using Metropolis-Hastings steps.

Let $\boldsymbol{d}_t = \boldsymbol{y}_t - \sum_{j=1}^{p} \boldsymbol{\Phi}_j \boldsymbol{y}_{t-j}$ then, conditioning on $\boldsymbol{\Lambda}$ and $\boldsymbol{\Phi}_j$, we have that

$$\boldsymbol{d}_t = \sum_{j=0}^{p} -\boldsymbol{\Phi}_j \boldsymbol{F}_{t-j}' \boldsymbol{\theta}_{t-j} + \boldsymbol{v}_t = BD(\boldsymbol{\Phi}_j \boldsymbol{F}_{t-j}') \boldsymbol{\Theta}_t + \boldsymbol{v}_t \tag{6}$$

where $\boldsymbol{\Phi}_0$ is equal to $-\boldsymbol{I}$, $BD(\cdot)$ denotes a block diagonal matrix and $\boldsymbol{\Theta}_t = (\boldsymbol{\theta}_t, \ldots, \boldsymbol{\theta}_{t-p})'$. Equation (6) corresponds to the observation equation of a dynamic linear model (West and Harrison, 1997). The evolution equation is given by $\boldsymbol{\Theta}_t = BD(\boldsymbol{G})\boldsymbol{\Theta}_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, BD(\boldsymbol{W}))$. Then $\boldsymbol{\Theta}_t$ can be sampled using a Forward Filtering Backward Sampler (FFBS) as proposed by Carter and Kohn (1994) and Frühwirth-Schnater (1994).

Using $G_j(s)$ to denote a generic root of the polynomial in Equation (4), we have that, if $j = 1, \ldots, R$, $G_j(s) = a_j(s)$. For $j = R+1, \ldots, R+C$ we have $G_j(s) = (\phi_{1j}(s), \phi_{2j}(s))$, the pair of complex conjugated roots parametrized by $(r_j(s), \omega_j(s))$. We use bold faces to denote the vectors of processes evaluated at $s_1, \ldots, s_n$. Denote G_{-j} the set of root processes excluding $G_j(s)$, then given $\boldsymbol{\Lambda}$, $\boldsymbol{\theta}_t$ and G_{-j}, we define the vector-valued time series $\boldsymbol{u}_{jt} = \prod_{\substack{i=1 \\ i \neq j}}^{R+C} (1 - \boldsymbol{G}_j B)\boldsymbol{y}_t^*$. Thus, for $j = 1, \ldots, R$, we have that $\boldsymbol{u}_{jt}$ is a multivariate AR(1) process with coefficients $\boldsymbol{a}_j$ and variance $\boldsymbol{\Sigma}$.

Let A_j, $j = R+1, \ldots, C$ be the index set of all other roots. Given $\boldsymbol{\Lambda}$, $\boldsymbol{\theta}_t$ and G_{-j}, we can compute the filtered time series $\boldsymbol{u}_{jt} = \prod_{i \in A_j} (1 - \boldsymbol{G}_i B)\boldsymbol{y}_t^*$. It follows that, in this case, $\boldsymbol{u}_{jt}$ is a multivariate AR(2) process with coefficients $\boldsymbol{\phi}_{1j} = 2\boldsymbol{r}_j \cos \boldsymbol{\omega}_j$ and $\boldsymbol{\phi}_{2j} = -\boldsymbol{r}_j^2$ and variance $\boldsymbol{\Sigma}$. Here vector notation is used to denote component-wise operations. The resulting full conditional of the roots does not have a closed form and we use Metropolis-Hastings algorithm to obtain samples from this distribution. Given the roots $G_j(s)$, we compute the implied AR coefficients, $\Phi_j(s)$, by solving the equation (4). For example, for $p = 2$ real roots we have $\Phi_1(s) = a_1(s) + a_2(s)$ and $\Phi_2(s) = -a_1(s) * a_2(s)$.

To complete the model fitting procedure we need to consider inference for missing values. Within a Bayesian framework, missing values are treated as additional model parameters and sampled within the MCMC. We denote $\boldsymbol{y}_t = (\boldsymbol{y}_t^u, \boldsymbol{y}_t^g)'$ a partition of the vector $\boldsymbol{y}$ for each time t where $\boldsymbol{y}_t^u$ is a vector with n_{ut} rows containing the missing values in time t and $\boldsymbol{y}_t^g$ is a vector with n_{gt} rows containing the observations in t such as $n_{ut} + n_{gt} = n$, $\forall t$. We can compute the conditional distribution $\boldsymbol{y}_t^u | \boldsymbol{y}_t^g$ using the property of the multivariate normal as described in Anderson (1994).

3.1 Spatial interpolation and temporal prediction

Let $\boldsymbol{y}_t^u$ be the vector of the observations for ungaged locations each time t. Spatial predictions can be obtained by considering the distribution of $(\boldsymbol{y}_t, \boldsymbol{y}_t^u)$ conditional on the parameters and the first p observations. This distribution is given as

$$\begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_t^u \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{F}_t^y \boldsymbol{\theta}_t + \sum_{j=1}^p \boldsymbol{\Phi}_j^y (\boldsymbol{y}_{t-j} - \boldsymbol{F}_{t-j}^y \boldsymbol{\theta}_{t-j}) \\ \boldsymbol{F}_t^u \boldsymbol{\theta}_t + \sum_{j=1}^p \boldsymbol{\Phi}_j^u (\boldsymbol{y}_{t-j}^u - \boldsymbol{F}_{t-j}^u \boldsymbol{\theta}_{t-j}) \end{pmatrix}; \begin{pmatrix} \boldsymbol{\Sigma}^y & \boldsymbol{\Sigma}^{yu} \\ \boldsymbol{\Sigma}^{uy} & \boldsymbol{\Sigma}^u \end{pmatrix} \right),$$

where $\boldsymbol{F}_t^u$ and $\boldsymbol{\Phi}_j^u$ correspond, respectively, to the regression matrix and the AR(p) coefficients vector for ungaged locations. Similar notation is used to split the covariance n four blocks. We then have that

$$\boldsymbol{y}_t^u | \boldsymbol{y}_t, \Theta \sim N(\boldsymbol{\mu}_t^u + \boldsymbol{\Sigma}^{uy}(\boldsymbol{\Sigma}^y)^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_t^y), (\boldsymbol{\Sigma}^u - \boldsymbol{\Sigma}^{uy}(\boldsymbol{\Sigma}^y)^{-1}\boldsymbol{\Sigma}^{yu})),$$

where $\boldsymbol{\mu}_t^y = \boldsymbol{F}_t^y\boldsymbol{\theta}_t + \sum_{j=1}^p \boldsymbol{\Phi}_j^y(\boldsymbol{y}_{t-j} - \boldsymbol{F}_{t-j}^y\boldsymbol{\theta}_{t-j})$ and $\boldsymbol{\mu}_t^u = \boldsymbol{F}_t^u\boldsymbol{\theta}_t + \sum_{j=1}^p \boldsymbol{\Phi}_j^u(\boldsymbol{y}_{t-j}^u - \boldsymbol{F}_{t-j}^u\boldsymbol{\theta}_{t-j})$ for all t. Note that $\boldsymbol{\Phi}_j^u$ is unknown and needs to be sampled following the approach described in Section 3. To obtain samples of $\boldsymbol{y}_t^u$ we start by drawing samples from $\boldsymbol{y}_{p+1}^u$ conditioning on samples from the posterior of the parameters, and then recursively sample forward in time.

Temporal k steps ahead prediction, assuming $p = 1$ for simplicity, is given by

$$
\begin{aligned}
&p(\boldsymbol{y}_{T+k}|\boldsymbol{y}_{1:T}) = \int p(\boldsymbol{y}_{T+k}|\boldsymbol{F}_{T+k}, \boldsymbol{y}_{T+k-1}, \boldsymbol{\theta}_{T+k}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}) p(\boldsymbol{\theta}_{T+k} \mid \boldsymbol{\theta}_{T+k-1}, \boldsymbol{W}) \\
&p(\boldsymbol{y}_{T+k-1}|\boldsymbol{F}_{T+k-1}, \boldsymbol{y}_{T+k-2}, \boldsymbol{\theta}_{T+k-1}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}) p(\boldsymbol{\theta}_{T+k-1} \mid \boldsymbol{\theta}_{T+k-2}, W) \dots \\
&\dots p(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{\boldsymbol{T}}|\boldsymbol{y}_1, \dots, \boldsymbol{y}_T) d(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\theta}_{\boldsymbol{1}}, \dots, \boldsymbol{\theta}_{\boldsymbol{T}}, \boldsymbol{W}).
\end{aligned}
$$

The integral above can be approximated by

$$
p(\boldsymbol{y}_{T+k}|\boldsymbol{y}_{1:T}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\boldsymbol{y}_{T+k}|\mathbf{F}_{T+k}\boldsymbol{\theta}_{T+k}^{(m)} + \boldsymbol{\Phi}^{(m)} \left(\boldsymbol{y}_{T+k-1} - \boldsymbol{F}_{T+k-1}\boldsymbol{\theta}_{T+k-1}^{(m)} \right), \boldsymbol{\Sigma}^{(m)}).
$$

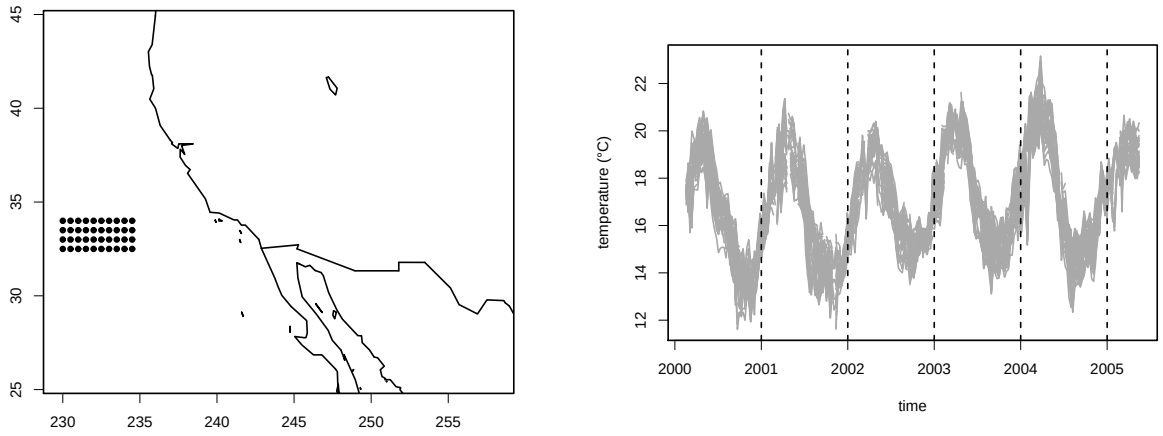
Here the superscript (m) denotes samples from the posterior of $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T$, $\boldsymbol{\Phi}$, $\boldsymbol{\Sigma}$. $\boldsymbol{y}_{T+i}^{(m)}$, $i > 0$ is obtained by propagating the samples from the posterior through the evolution equations for $\boldsymbol{y}_t$.

4 An illustrative example

We consider an illustration of our method to high frequency environmental data. We focus on satellite sea surface temperature data for a region in the Pacific Ocean off the coast of California. The measurements correspond to a 10×4 (longitude $\times$ latitude) grid with a spatial resolution of $0.5°$. This is illustrated in Panel (a) of Figure 2. The data were recorded every 8 days, from July 2000 to May 2005. Thus, for each grid cell we a time series of T=240 periods of time. The time series are illustrated in Panel (b) of Figure 2. As is common for satellite data, there are a number of missing values. Clearly, all series show a yearly seasonal pattern. Higher temperatures are observed in April and lower ones around October of each year. A possible increasing long-term trend is present in the data, as well as possible changes in the amplitude of the cycles. These two issues are of relevance for climatic studies. Thus, the separation of such long-term trends from high frequency variabilities is very relevant. Another important goal of the analysis is to fill in the gaps in the satellite measurements and produce fields at a resolution of $0.25°$, which is the current standard of the World Ocean Atlas 2001 version 2 (Boyer et al., 2005).

4.1 Preliminary analysis

Notice that the multivariate dynamic linear model (DLM), as described in (West and Harrison, 1997) is a particular case of the model in Equations (1)-(3) when $p = 0$. Therefore, we start by fitting a multivariate DLM to the temperature data. We denote this model as



(a) Gauged sites (b) Observed time series

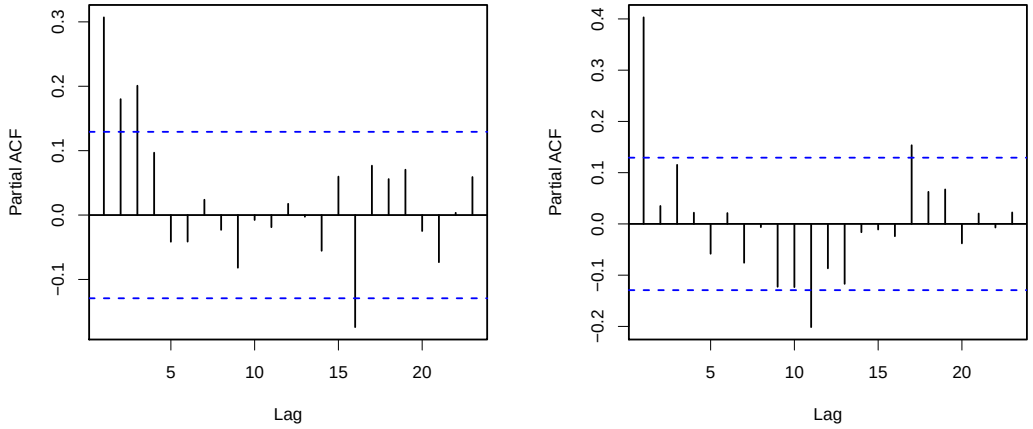
Figure 2: Panel (a) Map with the coast of California together with 40 sites. Panel (b) observed temperature measurements taken every 8 days, starting from July 2000 until May 2005.

M0. More specifically, we assume a time-varying mean structure comprising a baseline and a seasonal component with an annual cycle, such that $\boldsymbol{\theta}_t = (\beta_t, \gamma_{1t}, \gamma_{2t})$. As the multiple time series are observed at different locations, we assume the covariance matrix, $\boldsymbol{\Sigma}$ based on an exponential correlation function, that is $\Sigma(s, s') = \sigma^2 \exp\{-||s - s'||/\lambda\}$, where $||s - s'||$ is the Euclidean distance between locations s and s', λ, and σ^2 are parameters of the model. More specifically, this model is described by

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{F}'\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \\ \boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{W}) \ . \end{aligned}$$

We used the same prior distributions considered for other models, as described in Section 4.3. We run a MCMC with 50,000 iterations, considered 20,000 as burn in and kept every 30th iteration. Panels of Figure 3 present the partial autocorrelation function of the residuals from this model for two different sites. Both panels suggest that there is some structure left even after removing the trend of each time series. Therefore next Subsection analyzes the fitting considering different versions of SVAR(p).

We propose five different models for capturing this extra structure left in the data, based on three different values for $p = 1, 2, 3$. We denote M1 as the model with one real root ($R = 1$), M2 the model with 2 real roots ($R = 2$), M3 the model with one pair of complex roots ($C = 1$), M4 that with 3 real roots ($R = 3$), and M5 the model with one real root and 1 pair of complex roots ($R = 1$ and $C = 1$). Table 1 summarizes the specification of the fitted models.



(a) Site 1 (b) Site 5

Figure 3: Partial autocorrelation function of the residuals of the time series at two locations after fitting the multivariate DLM model.

4.2 Model comparison criteria

In order to compare the different models we fit to the data, we use two different criteria. One is the posterior predictive loss criterion (EPD) introduced by Gelfand and Ghosh (1998); the other is the predictive likelihood. Below we give more detail about them.

EPD The posterior predictive loss criterion (EPD) is based on replicates of the observed data obtained from each fitted model. It is given by the sum of a goodness-of-fit term (G) and a penalty term (P). Denoting $Y_{l,rep}$ as the vector of replicates of the observed data, the criterion is given by the expression $D = P + \frac{1}{2}G$, where

$$P = \sum_{l=1}^{L} Var(Y_{l,rep}|\mathbf{y}) \; G = \sum_{l=1}^{L} \left(E(Y_{l,rep}|\mathbf{y}) - Y_{l,obs}\right)^2, \tag{7}$$

and L represents the sample size from the predictive distribution. For the computation we need to solve integrals which do not have an analytical solution but they can be approximated using the samples obtained via MCMC methods.

Predictive likelihood This criterion measures the ability of the model in predicting a set of future values, $\boldsymbol{y}_f = (\boldsymbol{y}_{T+1}, \cdots, \boldsymbol{y}_{T+k})$. The predictive likelihood under model M is given by

$$p(\boldsymbol{y}_f|M, D_T) = \int p(\boldsymbol{y}_f|\boldsymbol{\theta}_f, M, D_T)p(\boldsymbol{\theta}_f|M, D_T)d_{\boldsymbol{\theta}_f} = E_{\boldsymbol{\theta}_f|M,D_T}[p(\boldsymbol{y}_f|\boldsymbol{\theta}_f, M, D_T)].$$

where $p(\boldsymbol{\theta}_f|M, D_T)$ is obtained by propagating the samples from the posterior $p(\boldsymbol{\theta}_T|D_T, M)$, through the system equation. The expression above can be approximated by using Monte Carlo estimation

$$\hat{E}_{\boldsymbol{\theta}_f|M,D_T}[p(\boldsymbol{y}_f|\boldsymbol{\theta}_f, M, D_T] = \sum_{m=1}^{M} \frac{\prod_{i=1}^{k} p(\boldsymbol{y}_{T+i}|\boldsymbol{\theta}^m_{T+i}, M, D_T)}{M}.$$

The model with the highest predictive likelihood is the best among those fitted.

4.3 Fitting SVAR(p)

As described on Subsection (4.1) we consider a time-varying baseline, β_t, and a seasonal component with an annual cycle, $(\gamma_{1t}, \gamma_{2t})$. It is reasonable to assume that the temperature varies around 18°C in the beginning of the analysis, therefore we assume $\beta_0 \sim N(18, 1)$. The priors for γ_{10} and γ_{20} are assumed to follow a normal distribution, more specifically, $N(3, 1)$. We assign a inverse gamma distribution for σ^2 with shape parameter equals 3 and scale equals 2. The prior mean of λ is based on the idea of practical range, that is, we assume the correlation is close to zero (say 0.05) at half of the maximum observed inter-location distance. The prior variance is fixed at a reasonable large value. Thus we have $\lambda \sim Ga(300, 4)$. The prior distribution for the real and complex roots is assigned by assuming a truncated beta distribution for z_j^a and z_j^ϕ. We consider $\alpha_j = \beta_j = 1$ for all the roots, which is equivalent to a uniform prior distribution in the interval $(0, 1)$. We use a Gaussian kernel to construct the beta priors assuming that we have a smooth process. The maximum distance between the grid of $M = 21$ points (7×3) and coordinates of the gauged sites was about 430 Kilometers. The parameter of the Gaussian kernel was fixed at 130 Kilometers and chosen using the same idea of pratical range.

We run the MCMC for 200,000 iterations, consider a burn in of 50,000 iterations and keep every 150-th iteration. The convergence was checked by using Geweke's convergence diagnostic proposed by Geweke (1992). Table 1 shows the values of the model comparison criteria for each of the six fitted models. It is clear that M0 results in the highest value of EPD and the lowest value of the predictive likelihood; also these values are quite different when compared to the other models, providing strong evidence that models without an autoregressive structure have a poor fit. On the other hand, M4 has the highest predictive likelihood and the lowest EPD. Based on these criteria we present the results for M4, that is, SVAR(3) with three real roots.

The posterior mean of the scale parameter σ^2 is 0.258 with 95% credible intervals (0.244-0.272), which is expected from this type of data. On the other hand, the posterior mean of the parameter of the exponential correlation function λ is equal to 190 kilometers, its provides correlations that range from 0.10 to 0.89 among all sites.

In Figure 4 we show the posterior mean and respective 95% credible intervals for the baseline and amplitude $(\sqrt{\gamma_{1t}^2 + \gamma_{2t}^2})$. There is evidence of an increasing temperature trend during the first four years of the series, until roughly 2004. The posterior mean of the baseline in early 2001 is 16.6°C and becomes 17.6°C in early 2004. This is followed by a decreasing

Model	Autoregressive structure	Predictive likelihood	EPD		
			P	G	D
M0	$p = 0$	8.70e-34	5446.4	5813.5	8353.2
M1	$p = 1$ $(R = 1)$	3.06e-22	3962.3	6352.3	7138.5
M2	$p = 2$ $(R = 2)$	1.36e-21	3814.2	6360.6	6994.5
M3	$p = 2$ $(C = 1)$	2.96e-23	4123.2	5977.1	7111.7
M4	$p = 3$ $(R = 3)$	3.01e-21	3765.4	6184.1	6857.4
M5	$p = 3$ $(R = 1$ and $C = 1)$	4.53e-22	3973.8	6106.5	7027.1

Table 1: Predictive likelihood and EPD criteria for each fitted model.

trend during the period 2004 – 2005 when the baseline mean becomes 17.1°C. As for the amplitude we see a similar, but substantially less pronounced pattern. The amplitude mean is 2.8°C in 2001 decreasing to 2.5°C in 2002. This is followed by an increase to 2.7°C in 2004.

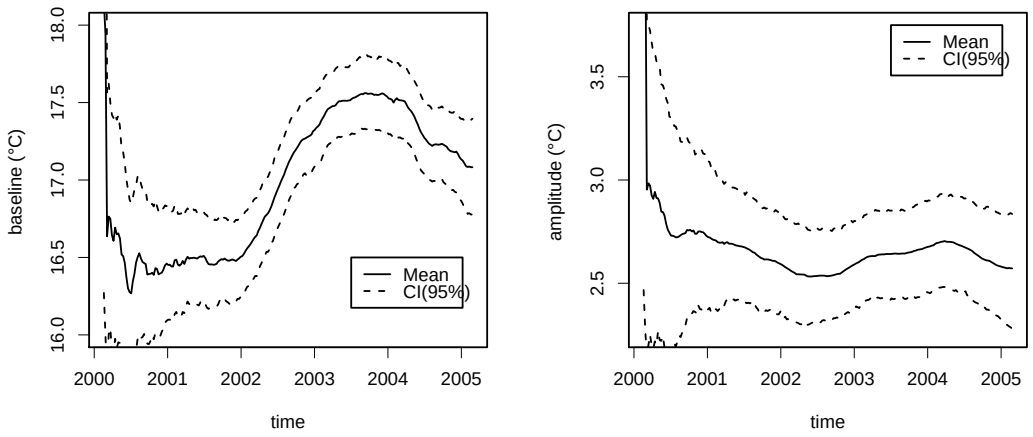


Figure 4: Posterior mean and limits of the 95% credible intervals of β_t (baseline) and α_t (amplitude).

The 95% posterior credible interval of the autoregressive parameters, $\Phi_1(s)$, $\Phi_2(s)$ and $\Phi_3(s)$ for all sites are showed in Figure 5. We notice that even after the inclusion of the mean structure, which accounts for a baseline and a seasonal component, there is some structure left in the error term. Moreover, the coefficients of the autoregression vary smoothly with location. This is illustrated in Figure 5. We observe significant difference between the coefficients of different locations. In particular we observe that locations in the east boundary of the domain have ϕ_1 coefficients that are substantially larger than those for other locations. As for ϕ_3, we observe that in most cases the the 95% posterior credible intervals include zero.

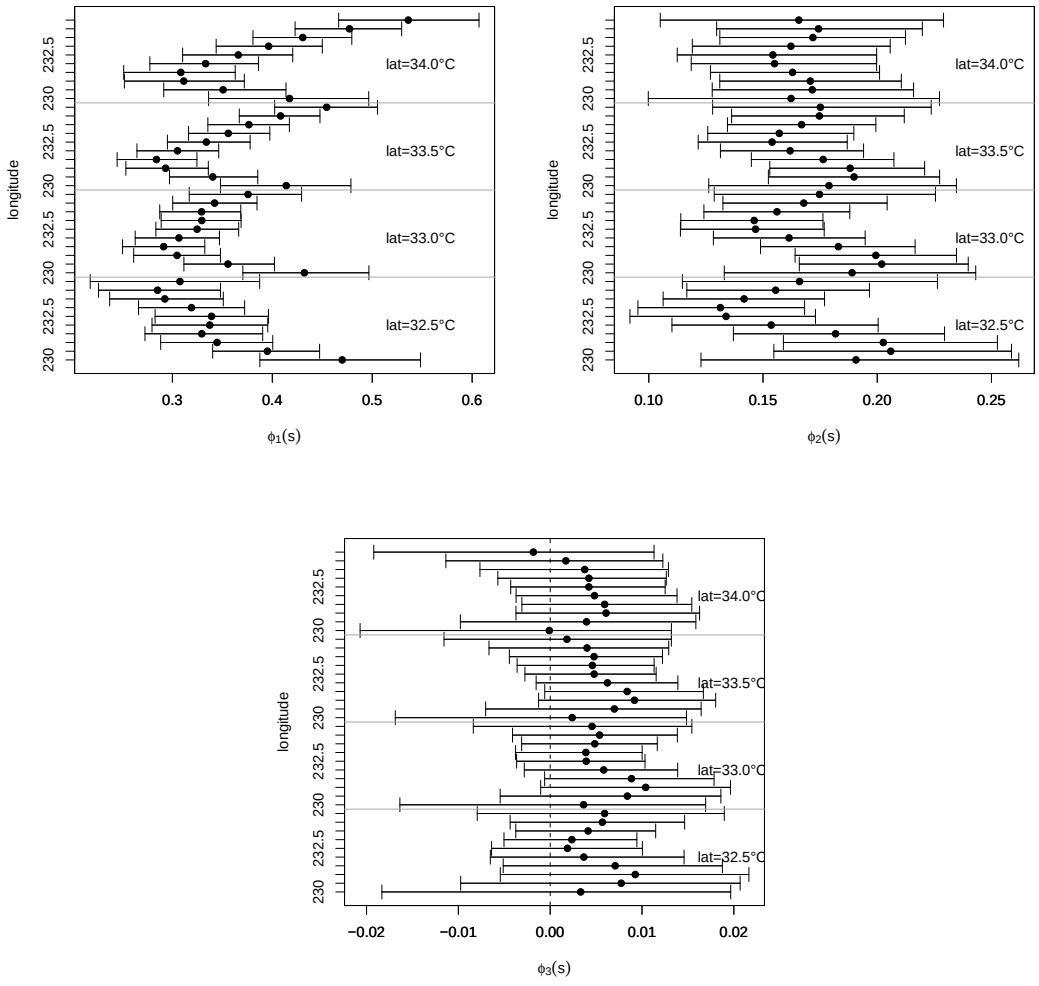
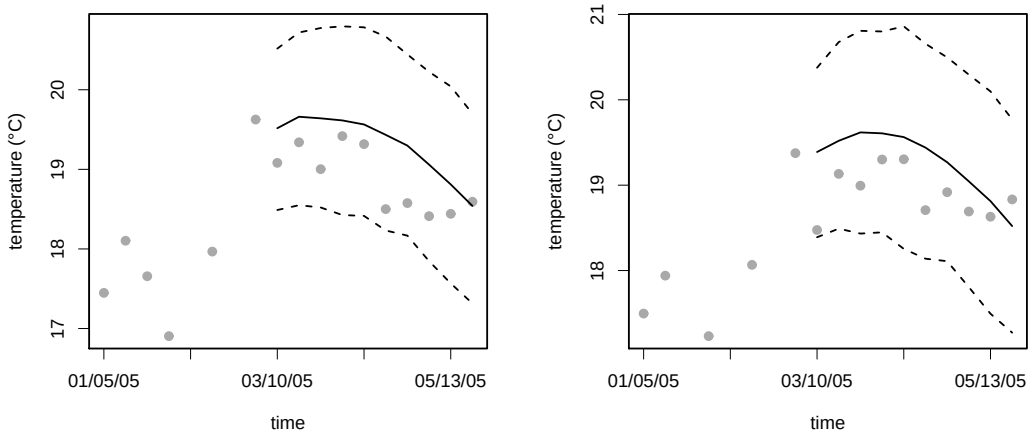


Figure 5: Posterior mean and limits of the 95% credible intervals of $\Phi_1(s)$, $\Phi_2(s)$ and $\Phi_3(s)$ based on model $M4$.

The two panels of Figure 6 show the temporal prediction for $K = 10$ periods of time. These observations were left out from the inference procedure. The solid circles represent the actual observations, the solid line represents the posterior mean and the dashed lines the posterior 95% credible interval. Apparently the model is performing well in terms of short term temporal predictions.



(a) lon=232.5°C lat=33.5°C (b) lon=232.5°C lat=33.5°C

Figure 6: Temporal prediction for $K = 10$ periods of time at two different locations. The solid line represents the posterior mean, and the dashed lines represent the limits of the 95% posterior credible intervals.

Figure 7 shows the original data with a spatial resolution of 0.5°for 8 points in time from 12/10/02 to 02/04/03. We observe missing data for the last three periods of time in some grid cells. Figure 8 presents the evolution of estimated temperature surface for the same period in a grid of 240 points (20×12) with a spatial resolution of 0.25°. We observe, as expected, that the estimated temperatures are compatible with the observed ocean temperatures and that the resulting contours are smoother than the ones based on the corser resolution data.

In addtion to the six models considered we run M4 with and independent covariance matrix ($\boldsymbol{\Sigma} = \boldsymbol{I}$). The idea here is to check if the spatial process induced by the discrete process convolutions for the AR coefficients captures all the spatial variability. We obtained a predictive likelihood of 7.03e-130. This is a very small value, especially when compared to 3.01e-21, the value of the predictive likelihood for M4. We conclude that we can not drop the assumption of spatial correlation in the errors.

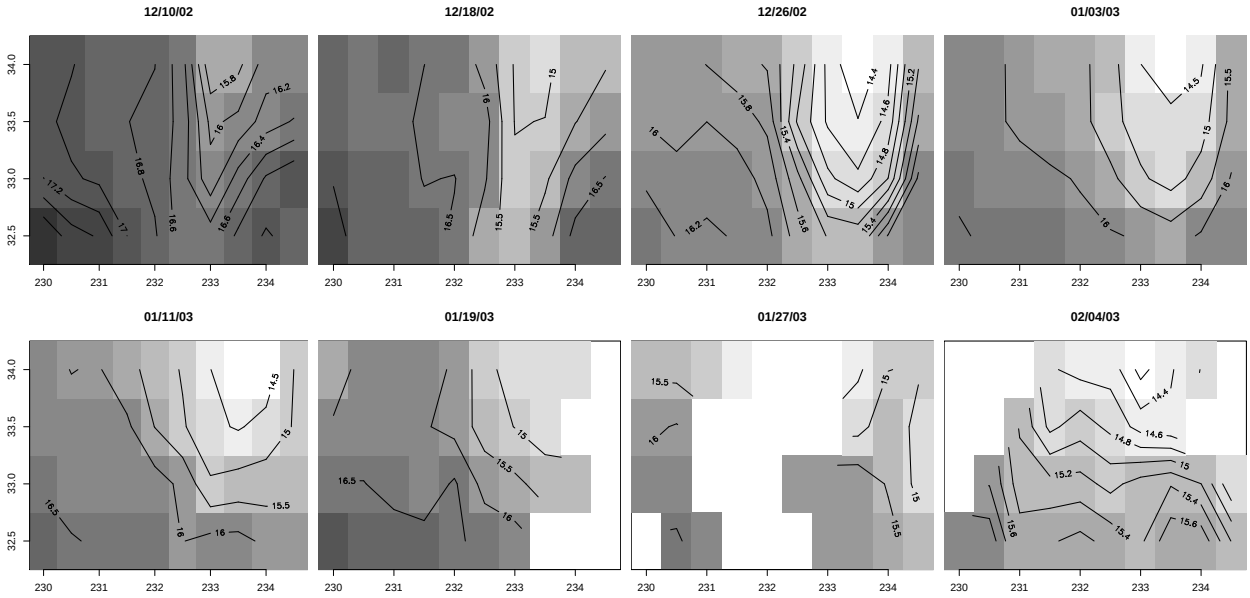


Figure 7: Original data at 0.5°spatial resolution: Temperatures for 8 points in time.

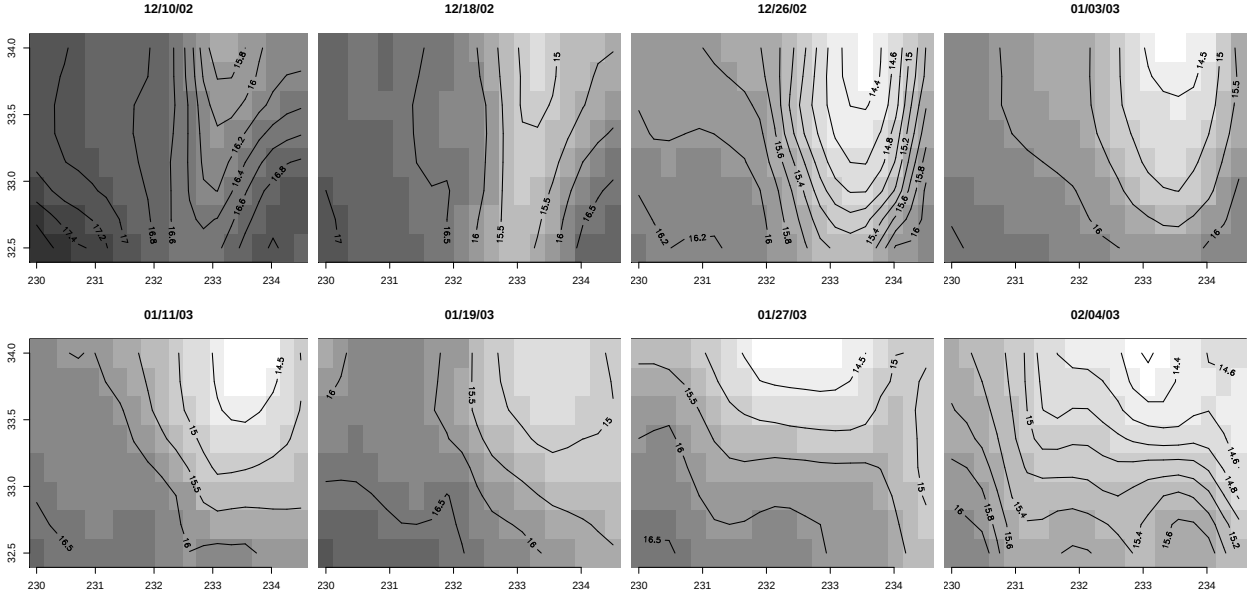


Figure 8: Spatial interpolation at 0.25°spatial resolution: Posterior mean of the estimated temperatures for 8 points in time.

5 Discussion

In this paper we have presented a spatio-temporal model that provides a spatial structure to autoregressive processes. We use bounded spatial processes as priors for the roots of the

characteristic polynomial of the AR process. This imposes smooth spatial variability on the coefficients of the autoregression as well as stationarity of the time series corresponding to any point in the spatial domain. Autoregressive processes are popular for their simple formulation together with their ability to capture persistence and stochastic cycles. So they are a valuable tool for the detection of long term trends in high frequency data.

We illustrate our method with an analysis of satellite data of ocean temperatures. We believe that this is a pertinent example, as data from remote sensors are likely to be contaminated by noise that is not white. Clearly we could include additional features in the example, like spatially-varying seasonalities and some description of the ocean dynamics. We prefer to keep the model simple to focus on the inferential methods we propose in this paper.

An obvious extension to the proposed SVAR(p) would be to consider that the order p is unknown and possibly space-varying. The simplest way to achieve this is to let p be a large number and impose priors on the coefficients that include a point at zero. This is the approach taken, in the time series context, in Huerta and West (1999). The extension to the spatial case requires that whole spatial processes be exactly zero. Unfortunately our attempts to fit models where the priors are given by mixtures of processes strongly concentrated around zero with processes like the ones described in Sections 2.1 and 2.2 have not been successful. A model comparison approach, like the one taken in our ocean data example, is our preferred way of exploring the most suitable values of p, R and C.

Acknowledgments

This work is part of Nobre's PhD thesis obtained at the Graduate Program in Statistics of UFRJ, Brazil, under the supervision of Schmidt and Sansó. During Nobre's studies she benefited from a scholarship from CNPq. Schmidt is grateful to CNPq and FAPERJ for the financial support during this work. This research was partially supported by National Science Foundation grants CMG 0417753, DMS 0504851 and DMS 0906765.

References

Anderson, T. (1994). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Inc.

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis of Spatial Data*. Chapman and Hall, New York.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice-Hall Inc.

Boyer, T., Levitus, S., Garcia, H., Locarnini, R. A., Stephens, C., and Antonov, J. (2005). Objective analyses of annual, seasonal, and monthly temperature and salinity for the World Ocean on a 0.25 grid. *Int. J. Climatol.*, 25:931–945.

Calder, C. A. (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.

Cressie, N. and Huang, H. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**:1330–1340.

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

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. Chapman and Hall, London, UK, second edition.

Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85:1–11.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In JM Bernado, JO Berger, A. D. and Smith, A., editors, *Bayesian Statistics 4*. Clarendon Press, Oxford, UK.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, **97**:590–600.

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

Higdon, D. (2007). A primer on space-time modeling from a Bayesian perspective. In Finkenstädt, B., Held, L., and Isham, V., editors, *Statistical Methods for Spatio-Temporal Systems*, chapter 6, pages 217–279. Chapman and Hall.

Huerta, G., Sansó, B., and Stroud, J. R. (2004). A spatio-temporal model for Mexico city ozone levels. *Applied Statistics*, **53**:1–18.

Huerta, G. and West, M. (1999). Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society, B*, **61**:881–899.

Lee, H., Sansó, B., Zhou, W., and Higdon, D. (2007). Inference for a proton accelerator using convolution models. *Journal of the American Statistical Association*, 103:604–613. DOI 10.1198/016214507000000833.

Lemos, R. and Sansó, B. (2009). A spatio-temporal model for mean, anomaly and trend fields of north Atlantic sea surface temperature (with discussion). *Journal of American Statistical Association*, 104:5–25. DOI 10.1198/jasa.2009.0018.

Priestley, M. (1981). *Spectral Analysis and Time Series*. Academic Press, San Diego, USA.

Sahu, S. K. and Mardia, K. V. (2005). A Bayesian kriged Kalman model for short-term forecasting of air polution levels. *Applied Statistics*, **54**:223–244.

Sansó, B. and Guenni, L. (2000). A nonstationary multisite model for rainfall. *Journal of the American Statistical Association*, **95**:1064–1089.

Sansó, B., Schmidt, A. M., and Nobre, A. A. (2008). Bayesian spatio-temporal models based on discrete convolutions. *The Canadian Journal of Statistics*, (469):310–321.

Shaddick, G. and Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Applied Statistics*, **51**:351–372.

Stein, M. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, **100**(469):310–321.

Wackernagel, H. (2001). *Multivariate Geostatistics - An Introduction with Applications*. Springer-Verlag, New York, USA, second edition.

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

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