Space-Time Modelling of Coupled Spatio-Temporal Environmental Variables

L. IPPOLITI *- P. VALENTINI
Department of Quantitative Methods and Economic Theory
University G. d'Annunzio, Chieti-Pescara, Italy

D. GAMERMAN
Depto. de Métodos Estatísticos - UFRJ, Rio de Janeiro

Abstract

This article proposes a dynamic factor model for spatio-temporal coupled environmental variables. The model is discussed in a state-space framework which results useful for conditional interpolation and forecast of the variables of interest. The role of the measurement matrix in spatial interpolation is considered and the proposal of a stochastic specification is discussed. Full probabilistic inference for the model parameters is facilitated by Markov Chain Monte Carlo algorithms. Standard MCMC for dynamic linear models are adapted to our model specification and predictive and interpolation results are discussed for two different data sets with variables measured at two different scales.

Key Words: MCMC, Bayesian inference, Dynamic Factor Model, Spatio-temporal models, State-space models, Spatial patterns.

1 Introduction

In recent years, spatio-temporal models have received widespread popularity and have been largely developed thorough applications in many scientific fields. Within environmental sciences, for example, modelling is one of the main activities to evaluate air quality and to prepare plans and programmes as requested by many directives (eg. the Ambient Air Quality Assessment and Management Directive 96/62/EC; see for example Beattie et al., 2002) on air quality assessment and management. To this end several efforts have been made with the aim of providing some understanding on the dynamic of pollutant variables.
Spatio-temporal models have been developed with no single approach considered uniformly as being the most appropriate for a specific problem. References within this broad framework, with a wide range of applications, include for example Sansó and Guenni (1999), Brown et al. (2000), Allcroft and Glasbey (2003) and Sahu and Mardia (2005). The text book by Banerjee, Carlin and

*Corresponding Author: DMQTE, Viale Pindaro 42, 65127 Pescara, e-mail: ippoliti@unich.it

Gelfand (2004) also provides an excellent starting point for researchers. Although the book primarily covers hierarchical modeling and analysis of spatial data with an emphasis towards making Bayesian inference, it discusses spatio-temporal modeling in some detail and discusses a range of topics including multivariate modeling, spatial epidemiology, areal data modeling and many more. The choice of the approach is generally dictated by the objective of the study, whether it be obtaining air quality forecasts, estimating trends or increasing the scientific understanding of the underlying mechanisms. In building our model, we are partly guided by the degree of problem-solving as a criterion; specifically, we are interested in the spatial, temporal and spatio-temporal predictions of a specific pollutant when a predictor for this variable is available. Thus, the aim is to model the dynamic of two fields, i.e. coupled variables, that tend to occur synchronous with one another. For example, fine Particulate Matter (PM_{10}), is an ubiquitous pollutant with adverse effects on human health. It contains particles formed in the atmosphere from gaseous emissions such as sulfates (eg. SO_2), carbon (eg. CO) and nitrates (eg. NO_x). The physical relationships existing among these variables and PM_{10} suggest that they might be used as predictors for PM_{10} and, in practice, their correlations should not be neglected in a model building strategy. In practice, this is particularly important when the number of the monitoring sites for the predictors is larger than the ones for the response variable.
Figure 1 (left), for example, shows the network of monitoring stations (Lombardy region - Italy) for the PM_{10} - sites represented by "$\circ$" - and NO_x variables - sites represented by "$\times$". As it can be seen, many of the monitoring stations recording simultaneously the two variables share the same spatial coordinates; however, it is also worth noticing that the number of the stations for the NO_x is larger than the one related to the PM_{10} sites. Thus, because of the larger information available in space for the NO_x, predictions of the PM_{10} concentrations on the "x"-labeled spatial sites can be obtained conditional on the known values of the NO_x.

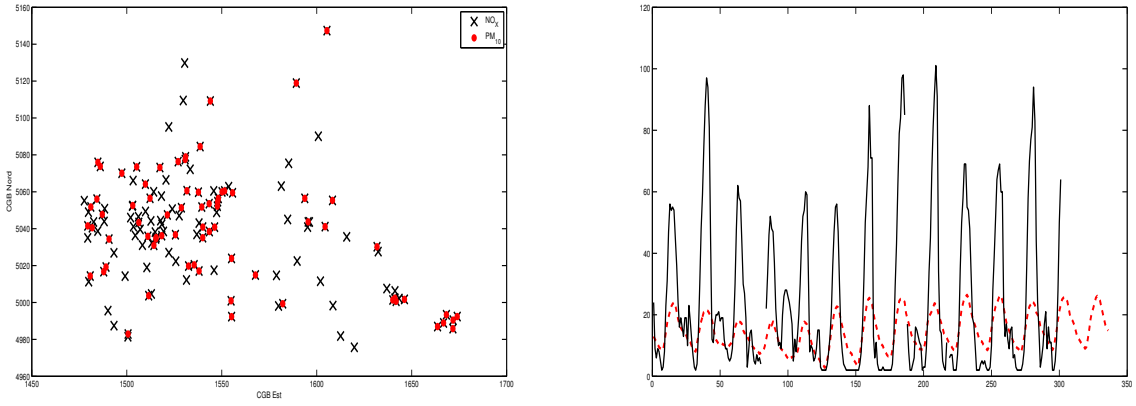

Figure 1: Left: Lombardy monitoring network; each "$\circ$" represents a site for PM_{10} while each "$\times$" represents the position of NO_x sites . Right: Ozone (O_3) and hourly temperatures time series observed over Mexico City.

The same feature, may of course also happen in time. In fact, for example, consider the two time series shown in Figure 1 (right) observed at a specific spatial site. The data, discussed in Section 7.2, refer to the series of hourly readings of measurements of Ozone (O_3) - continuous line - and temperatures - dashed line - observed over Mexico City (Huerta, Sansó and Stroud, 2004). It can be seen that since the time series of the temperature (predictor) is longer than the one of the variable of interest, temporal forecasts (or interpolation) of O_3 can be obtained conditional on the known

values of the covariate.
Motivated by these examples, in this paper we are interested in the development of latent regression models which are useful for spatial and temporal predictions of a pollutant of interest. Specifically, by exploiting the information provided by an available predictor, we discuss a modelling strategy for coupled environmental variables.
The model is developed in a state-space representation which represents a powerful way to provide full probabilistic inference for the model parameters, interpolation and forecast of the variable of interest. To account for spatial interpolation, the spatial dependence is incorporated in the measurement matrix and we describe its construction by discussing a stochastic specifications. Temporal variability instead, is incorporated by allowing the temporal model parameters to evolve in time through the state equation. This provides a natural formulation to obtain k-steps ahead temporal forecasts of the process. Full probabilistic inference for the model parameters is facilitated by Markov Chain Monte Carlo algorithms.
The remainder of the paper is organized as follows. In section 2, we describe the general dynamic latent model. The modelling of the spatial dependence is described in section 3 where we also specify in detail the components of the proposed model. In section 4 we propose a state-space formulation while in section 5 we provide prior and posterior specifications for the model parameters. In section 6 we describe the forecasting and interpolation strategies while in section 7 we present applications on two real data sets produced by monitoring networks with different features. Finally, section 8 concludes the paper with a discussion.

2 The General Model

Consider two spatio-temporal processes $X(\mathbf{s};t)$ and $Y(\mathbf{s};t)$, where $\mathbf{s} \in S$, with S some spatial domain in two dimensional Euclidean space $\mathcal{R}^2$ and $t \in \{1,2,\ldots\}$ a discrete index of times. It is explicitly assumed that X is a predictor of Y and thus, we let Y denote the specific process of interest to be predicted both in space and in time (although, in some cases, there could be interest in predicting both variables simultaneously).
The relationship existing between the two variables can be modeled in several ways. In this paper, for the modeling of the time series dynamics of the variable of interest, we consider the following dynamic factor model

$$\begin{aligned}
\mathbf{g}(t) &= \sum_{i=1}^{p} \mathbf{B}_i \mathbf{g}(t-i) + \sum_{j=0}^{q} \mathbf{C}_j \mathbf{f}(t-j) + \boldsymbol{\xi}(t) && (1)\\
\mathbf{f}(t) &= \sum_{k=1}^{s} \mathbf{R}_k \mathbf{f}(t-k) + \boldsymbol{\delta}(t) && (2)\\
\mathbf{x}(t) &= \mathbf{m}_x(t) + \mathbf{H}_x \mathbf{f}(t) + \mathbf{u}_x(t) && (3)\\
\mathbf{y}(t) &= \mathbf{m}_y(t) + \mathbf{H}_y \mathbf{g}(t) + \mathbf{u}_y(t) && (4)
\end{aligned}$$

where $\mathbf{y}(t)$ and $\mathbf{x}(t)$ are $(n_y \times 1)$ and $(n_x \times 1)$ time series vectors observed on n_x and n_y spatial sites, respectively; $\mathbf{m}_y(t)$ and $\mathbf{m}_x(t)$ are $(n_y \times 1)$ and $(n_x \times 1)$ mean components modelling the smooth large-scale temporal variability, $\mathbf{H}_y$ $(n_y \times m)$ and $\mathbf{H}_x$ $(n_x \times r)$, are measurement matrices retaining information on the spatial structure of the random fields, $\mathbf{B}_i$ $(m \times m)$, $\mathbf{C}_j$ $(m \times r)$, and $\mathbf{R}_k$ $(r \times r)$ are coefficient matrices modelling the temporal evolution of the latent vectors $\mathbf{g}(t) = (g_1(t), \ldots, g_m(t))'$ and $\mathbf{f}(t) = (f_1(t), \ldots, f_r(t))'$, respectively. Throughout the paper it is

assumed that both m and r are of several order of magnitude smaller than n_y and n_x, respectively. Also, notice that n_y and n_x need not be the same number and, even in the case in which $n_y = n_x$, it is not needed that the two variables have to be observed on the same sites. This is one of the advantages of our model formulation.

Finally, $\boldsymbol{\xi}(t)$, $\boldsymbol{\delta}(t)$, $\mathbf{u}_x(t)$ and $\mathbf{u}_y(t)$ are error terms. Specifically, we assume $\boldsymbol{\xi} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\xi)$, $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\delta)$, $\mathbf{u}_x \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{u_x})$, $\mathbf{u}_y \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{u_y})$ and restrict the variance matrices to be diagonal.

Factor analysis has previously been used to model multivariate spatial data. For example, Wang and Wall (2003), Christensen and Amemiya (2002, 2003), Hogan and Tchernis (2004) provide some evidence on how factor analysis can be used for potentially reducing the overall dimension of the response vector observed at each location. In this paper, however, the observations are univariate and factor analysis is used to identify possible clusters of locations whose temporal behavior is primarily described by a potentially small set of common dynamic latent factors. As described in Section 3, one of the key aspects of the proposed model is that flexible and spatially structured prior information regarding such clusters can be directly introduced by the columns of the factor loadings matrix.

3 The Spatial Dynamic Factor Model

A key-property of much spatio-temporal data is that observations at nearby sites and times will tend to be similar to one another. This underlying smoothness characteristic of the space-time process can be captured by estimating the state process and hence filtering out the measurement noise. It is customary in the dynamic factor literature to refer to the unobserved (state) processes as the common factors and to refer to the coefficients that link the factors with the observed series as the factor loadings. However, because of their spatial nature, the factor loadings are equivalently defined here as spatial patterns.

The specification of the spatial patterns could be done using a finite-dimensional space of regression or deterministic drift functions, or using autocorrelation to make nearby values spatially similar (Kent and Mardia, 2002). In most cases the specification is restricted to non-stochastic factor loadings matrices and some model examples are provided in Gamerman et al. (2003), Nobre et al. (2005), Stroud et al. (2001), Mardia et al. (1998), Wikle and Cressie (1999), Sahu and Mardia (2005) and Calder (2007).

In this paper, stochastic forms are considered in the specification of factor loadings which can easily incorporate external information through regression functions. The approach was first introduced by Lopes et al. (2008) but, in their specification, they only considered a simplified version referred to a single variable case. The possibility of specifying two measurement equations results in a significant advantage in terms of spatial interpolation and this makes an important difference with respect to other regression-based spatio-temporal models proposed in literature (see for example, Huerta et al., 2004). Notice that as shown in Lopes et al. (2008) the model leads to nonseparable forms (between space and time) for its covariance function.

3.1 Specifying the Spatial Patterns

By assuming that the latent factors, $\mathbf{f}(t)$ and $\mathbf{g}(t)$, are able to capture the temporal variation of the space-time field, the spatial dependencies can be modeled by the columns of the matrices $\mathbf{H}_y$ and $\mathbf{H}_x$. Specifically, following Lopes *et al.* (2008), it is assumed that the $j-$th column of $\mathbf{H}_y$ (or $\mathbf{H}_x$), $\mathbf{h}_{yj} = \big(h_{yj}(\mathbf{s}_1), \ldots, h_{yj}(\mathbf{s}_{n_y})\big)'$, $j = 1, \ldots, m$, can be modeled as a conditionally

independent Gaussian Random Field - GRF; i.e. $\mathbf{h}_{yj} \sim MVN\left(\mathbf{m}_j^{(h_y)}, \mathbf{\Sigma}_j^{(h_y)}\right)$, where $\mathbf{m}_j^{(h_y)}$ is a n_y-dimensional deterministic mean vector and $\mathbf{\Sigma}_j^{(h_y)}$ is a parameterized spatial covariance matrix. Specifically, we have $\mathbf{\Sigma}_j^{(h_y)} = \tau_{yj}^2 \mathbf{R}(\phi_{yj})$, where τ_{yj}^2 is a scale parameter and $\mathbf{R}(\phi_{yj})$ a $(n_y \times n_y)$ matrix suitably defined through a decaying spatial correlation parameter ϕ_{yj}. Assuming isotropy, in this paper we consider an exponential correlation function, $r(|\mathbf{s}_l - \mathbf{s}_k|) = exp\{-d_{lk}/\phi_{yj}\}$, where $d_{lk} = |\mathbf{s}_l - \mathbf{s}_k|$ is the Euclidean distance between sites $\mathbf{s}_l$ and $\mathbf{s}_k$.
Several alternative patterns can be represented by properly specifying the covariance structure and the mean function of each GRF; of course, compared with the deterministic approach (for a discussion see for example, Kent and Mardia, 2002), it follows that in our case both $\mathbf{H}_y$ and $\mathbf{H}_x$ cannot be specified ahead of time and must be considered as parameters. However, restrictions on $\mathbf{H}_y$ and $\mathbf{H}_x$ are needed to define a unique model free from identification problems. Several restrictions can be considered. The solution adopted here is to constrain the measurement matrices so that they are lower triangular matrices, assumed to be of full rank. This form provides both identification and useful interpretation of the factor model and, for example, is used by Geweke and Zhou (1996), Aguilar and West (2000), Lee and Shi (2000) and Reich et al. (2009). As also discussed by Lopes and West (2004), an advantage of using this specification is that the order of the sites in the measurement matrices is a modelling decision that has no effect on the resulting theoretical model nor on predictive inferences under the model. Note that as regards the identifiability problems $\mathbf{H}_y$ and $\mathbf{H}_x$ are subject to the specified 0/1 constraints on values in the the upper triangle and diagonal matrices, so the prior density applies only to the remaining, uncertain elements (Aguilar and West, 2000). Furthermore, the Gaussian process prior on the remaining loading imposes restrictions on them (loadings of locations nearby must be similar with high probability). This additional information provides additional constraints and strengthens the identifiability condition.

3.2 Covariate Effects

Many specifications for the mean level of the processes can be entertained, with the most common ones based on time-varying as well as location-dependent covariates. Here, we assume that $\mathbf{m}_y(t)$ and $\mathbf{m}_x(t)$ are zero vectors and proceed by estimating $\mathbf{m}_j^{(h_y)}$ and $\mathbf{m}_j^{(h_x)}$ only. Spatially-varying covariates are considered in explaining the mean level of the GRFs and some simple specifications might be: a) $\mathbf{m}_j^{(h\cdot)} = \mathbf{0}$, b) $\mathbf{m}_j^{(h\cdot)} = \beta_j^{(h\cdot)}\mathbf{I}$, c) $\mathbf{m}_j^{(h_x)} = \mathbf{D}^{(h_x)}\boldsymbol{\beta}_j^{(h_x)}$, and $\mathbf{m}_j^{(h_y)} = \mathbf{D}^{(h_y)}\boldsymbol{\beta}_j^{(h_y)}$, where $\mathbf{D}^{(\cdot)}$ is a matrix of covariates and $\boldsymbol{\beta}_j$ a vector of regression parameters. In the latter case c), more flexibility is brought up by allowing potentially different covariates for each Gaussian random field. Furthermore, for spatial prediction purposes, it might also be useful to link the spatial structure of Y and X and set $\mathbf{D}^{(h_y)} = \mathbf{H}_x$. This specification, results particularly useful when $n_x > n_y$ and we wish to predict Y on some of the sites in which X is available (see Figure 1 - left).

4 The Latent Processes

While the spatial structure of the two processes is modelled through the specification of the measurement equation, the temporal dynamic of the process is modelled by the specification of two dynamic stochastic processes, described in (1) and (2), for the unobserved state variables $\mathbf{g}(t)$ and $\mathbf{f}(t)$. Specifically, equation (2)

$$\mathbf{f}(t) \;=\; \sum_{k=1}^{s} \mathbf{R}_k \mathbf{f}(t-k) + \boldsymbol{\delta}(t)$$

represents a Vector Autoregressive - VAR(s) process for which we assume the following assumption:

ASSUMPTION 1. Let L be the backshift operator, then for each $t \in \mathbb{N}$, $\mathbf{f}(t) = [f_1(t), f_2(t), \ldots, f_r(t)]'$ admits the following one-sided moving average representation:

$$\mathbf{f}(t) = \sum_{j=0}^{\infty} \tilde{\mathbf{R}}_j \boldsymbol{\delta}(t-j), \qquad \tilde{\mathbf{R}}_0 = \mathbf{I}_r, \tag{5}$$

where the $\tilde{\mathbf{R}}_j$ are $(r \times r)$ matrices satisfying (Lütkepohl, 2005; sec. 2.1 and 6.1)

$$\sum_{j=0}^{\infty} \tilde{\mathbf{R}}_j L^j = (\mathbf{I}_r - L)^{-d} \left(\sum_{j=0}^{\infty} \tilde{\mathbf{G}}_j L^j \right)$$

and $Det \left[\sum_{j=0}^{\infty} \tilde{\mathbf{G}}_j z^j \right] \neq 0, \quad |z| \leq 1$, where d is the order of integration.
Equation (1) instead, provides an autoregressive distributed lag - ARDL$(p; q)$ - (Greene, 2003; Lütkepohl, 2005) specification for $\mathbf{g}(t)$ that, expressed in its *rational lag* form, also appears as a linear combination of two multivariate innovation processes:

$$\mathbf{g}(t) = \boldsymbol{\Phi}(L)\boldsymbol{\delta}(t) + \tilde{\mathbf{B}}(L)\boldsymbol{\xi}(t)$$

where, $\boldsymbol{\Phi}(L) = \tilde{\mathbf{B}}(L)\mathbf{C}(L)\tilde{\mathbf{R}}(L)$, and for a generic polynomial in the backshift operator, $\tilde{\mathbf{A}}(L) = \mathbf{A}(L)^{-1}$.

4.1 The State Space Formulation

The specification of equation (2) in the model formulation is dictated by the necessity of predicting in time the latent vector $\mathbf{f}(t)$ to produce $k-$step ahead forecasts of $\mathbf{g}(t)$. Thus, to this end, we may specify the joint generation process for $\mathbf{g}(t)$ and $\mathbf{f}(t)$ as

$$\begin{bmatrix} \mathbf{I}_m & -\mathbf{C}_0 \\ \mathbf{0} & \mathbf{I}_r \end{bmatrix} \begin{bmatrix} \mathbf{g}(t) \\ \mathbf{f}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{C}_1 \\ \mathbf{0} & \mathbf{R}_1 \end{bmatrix} \begin{bmatrix} \mathbf{g}(t-1) \\ \mathbf{f}(t-1) \end{bmatrix} + \cdots$$

$$\cdots + \begin{bmatrix} \mathbf{B}_p & \mathbf{C}_p \\ \mathbf{0} & \mathbf{R}_p \end{bmatrix} \begin{bmatrix} \mathbf{g}(t-p) \\ \mathbf{f}(t-p) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\xi}(t) \\ \boldsymbol{\delta}(t) \end{bmatrix} \tag{6}$$

where it is assumed without loss of generality that $p \geq max(s, q)$, $\mathbf{C}_i = \mathbf{0}$ for $i > q$ and $\mathbf{R}_j = \mathbf{0}$ for $j > s$. Since $\mathbf{v}(t)$ is a white noise, premultiplying the left and right hand parts by

$$\begin{bmatrix} \mathbf{I}_m & -\mathbf{C}_0 \\ \mathbf{0} & \mathbf{I}_r \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_m & \mathbf{C}_0 \\ \mathbf{0} & \mathbf{I}_r \end{bmatrix}$$

shows that the joint generation process of $\mathbf{g}(t)$ and $\mathbf{f}(t)$ is a VAR(p) process of the type

$$\mathbf{d}(t) \;=\; \mathbf{F}_1\mathbf{d}(t-1) + \ldots + \mathbf{F}_p\mathbf{d}(t-p) + \mathbf{K}\boldsymbol{\epsilon}(t)$$

where

$$\mathbf{d}(t) = \begin{bmatrix} \mathbf{g}(t) \\ \mathbf{f}(t) \end{bmatrix}, \quad \mathbf{F}_i = \begin{bmatrix} \mathbf{B}_i & \mathbf{C}_i + \mathbf{C}_0\mathbf{R}_i \\ \mathbf{0} & \mathbf{R}_i \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} \mathbf{I}_m & \mathbf{C}_0 \\ \mathbf{0} & \mathbf{I}_r \end{bmatrix}, \quad \boldsymbol{\epsilon}(t) = \begin{bmatrix} \boldsymbol{\xi}(t) \\ \boldsymbol{\delta}(t) \end{bmatrix}.$$

The presence of both measurement and latent variable equations naturally leads to the state-space representation (Hamilton, 1994) of model (1)-(4). The motivation behind casting the described dynamic model in a state-space form is primarily in the possibility of using the Kalman filter algorithm to produce a recursive estimation of the underlying unobserved variables, given the observed data. The linear Gaussian state-space model is thus described by the following *state* and *measurement* equations

$$\boldsymbol{\alpha}(t) \;=\; \boldsymbol{\Phi}\,\boldsymbol{\alpha}(t-1) + \boldsymbol{\Xi}\,\boldsymbol{\zeta}(t) \tag{7}$$
$$\mathbf{z}(t) \;=\; \mathbf{H}\,\boldsymbol{\alpha}(t) + \mathbf{u}(t) \tag{8}$$

where $\boldsymbol{\alpha}(t)$ is the state vector, $\boldsymbol{\Phi}$ is the nonsingular transition matrix, $\boldsymbol{\Xi}$ is a constant input matrix, $\mathbf{Z}(t)$ is the measurement vector and $\mathbf{H}$ is the measurement matrix. The sequences $\boldsymbol{\zeta}(t)$ and $\mathbf{u}(t)$ are assumed to be mutually independent zero mean Gaussian random variables with covariances $E\{\boldsymbol{\zeta}(t_i)\boldsymbol{\zeta}(t_j)'\} = \boldsymbol{\Psi}\delta_{ij}$ and $E\{\mathbf{u}(t_i)\mathbf{u}(t_j)'\} = \boldsymbol{\Sigma}_u\delta_{ij}$, where $E\{\cdot\}$ denotes the expectation and δ_{ij} the Kronecker delta function. In (7) and (8) we have the following specification:

$$\boldsymbol{\alpha}(t) = \begin{bmatrix} \mathbf{d}(t) \\ \mathbf{d}(t-1) \\ \vdots \\ \mathbf{d}(t-p+1) \end{bmatrix}, \; \boldsymbol{\Phi} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \mathbf{F}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix}, \boldsymbol{\zeta}(t) = \begin{bmatrix} \boldsymbol{\epsilon}(t) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{z}(t) = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{x}(t) \end{bmatrix}, \; \mathbf{H} = \begin{bmatrix} \mathbf{H}_y & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_x & \cdots & \mathbf{0} \end{bmatrix}, \; \boldsymbol{\Xi} = \begin{bmatrix} \mathbf{K} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \mathbf{u}(t) = \begin{bmatrix} \mathbf{u}_y(t) \\ \mathbf{u}_x(t) \end{bmatrix}.$$

4.2 Large Scale Dynamic Factors

Trend, periodic or cyclical behaviors are present in many applications and can be directly entertained by the dynamic model framework embedded in the proposed model. The temporal large scale variation can be incorporated into the proposed model either through the common dynamic factors or through the mean level. In the latter, the same pattern is assumed for all locations while, in the former, common factors receive different weights for different columns of the factor loading matrix, so allowing different trend/seasonal patterns for the spatial locations. For example, a

seasonal common factor of period N can be easily accommodated by specifying matrices of the type

$$\boldsymbol{\Upsilon}_j = \begin{bmatrix} \cos(2\pi j/N) & \sin(2\pi j/N) \\ -\sin(2\pi j/N) & \cos(2\pi j/N) \end{bmatrix}, \; j = 1, 2, \ldots, h = N/2$$

and $h = N/2$ is the number of harmonics needed to capture the seasonal behavior of the time series (see Lopes et al. 2008; West and Harrison, 1997, Chapter 8, for further details). In practice, fewer harmonics are required in many applications to adequately describe the seasonal pattern of many data sets and the dimension of this component is typically small.
By considering trend models to be of the form $\nabla^k \gamma(t) = \omega(t)$, where ∇^k is the k-th order difference operator and $\omega(t)$ a normally distributed zero-mean sequence with unknown variance φ^2, locally linear trend models can also be easily included in the model formulation; for example, for $k = 2$ we define trends of the type (Kitagawa and Gersch, 1996)

$$\gamma(t) = 2\gamma(t-1) - \gamma(t-2) + \omega(t).$$

Thus, by specifying the matrix

$$\boldsymbol{\Gamma} = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$$

a trend component can be easily introduced in the dynamic of the state vector.
Inference for the trend and seasonal model is done using the algorithm proposed below with (conceptually) simple additional steps. For instance, posterior samples for the involved variance matrices are obtained from inverted Wishart distributions, as opposed to the usual inverse gamma distributions. However, for the sake of notation, the following sections present the inferential procedures based on the more general equations (1-4).

5 Inference and Computations

5.1 Prior Information

Full probabilistic inference for the model parameters is carried out by elicitating the following independent prior distributions.
Let $\boldsymbol{\sigma}_y^2 = \{\sigma_{u_y,i}^2\}_{i=1}^{n_y}$, $\boldsymbol{\sigma}_x^2 = \{\sigma_{u_x,i}^2\}_{i=1}^{n_x}$, $\boldsymbol{\sigma}_\xi^2 = \{\sigma_{\xi,i}^2\}_{i=1}^{m}$, and $\boldsymbol{\sigma}_\delta^2 = \{\sigma_{\delta,i}^2\}_{i=1}^{r}$; then we assume that $p(\sigma_{y,i}^{-2}) = f_G(\sigma_{y,i}^{-2}|s_{y0}^{-2}, \nu_{y0})$, $p(\sigma_{x,i}^{-2}) = f_G(\sigma_{x,i}^{-2}|s_{x0}^{-2}, \nu_{x0})$, $p(\sigma_{\xi,i}^{-2}) = f_G(\sigma_{\xi,i}^{-2}|s_{\xi0}^{-2}, \nu_{\xi0})$ and $p(\sigma_{\delta,i}^{-2}) = f_G(\sigma_{\delta,i}^{-2}|s_{\delta0}^{-2}, \nu_{\delta0})$, with $s_{y0}^2 = s_{x0}^2 = s_{\xi0}^2 = s_{\delta0}^2 = 1$, $\nu_{y0} = \nu_{x0} = 0.002$ and $\nu_{\xi0} = \nu_{\delta0} = 0.02$.

For the spatial patterns, the parameters $\mathbf{m}_j^{(h_y)}$, τ_{yj}^2 and ϕ_{yj}, $j = 1, \ldots, m$, have the following prior specification: $\mathbf{m}_j^{(h_y)} \sim N(\mathbf{m}^\beta, \mathbf{S}_\beta)$, $\tau_{yj}^{-2} \sim G(a_\beta, b_\beta)$ and $\phi_{yj}^{-1} \sim G(2, \nu)$, where a_β, b_β are known hyperparameters, and $\nu = \max\{d_{lk}\}/(-2ln(0.05))$ (see, Lopes et al. 2008; Banerjee et al., 2004; Schmidt and Gelfand, 2003). In the applications discussed below, we consider the diffuse choice of $a_\beta = 2$, $b_\beta = 1$ and the same parametrization elicited for τ_{yj}^2 and ϕ_{yj}, was also chosen for τ_{xk}^2 and ϕ_{xk}, $k = 1, \ldots, r$.

As regards the prior specification for the regression parameters in equations (1) and (2), let $\boldsymbol{\Delta}_1 = vec[\mathbf{R}_1, \ldots, \mathbf{R}_s]$ and $\boldsymbol{\Delta}_2 = vec[\mathbf{A}_1, \ldots, \mathbf{A}_p, \mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_q]$, where $vec(\cdot)$ denotes the vectorize

operator; then we assume that $\boldsymbol{\Delta}_1|\boldsymbol{\Sigma}_\delta \sim N(\boldsymbol{\Delta}_{10}, \mathbf{V}_{10})$ and $\boldsymbol{\Delta}_2|\boldsymbol{\Sigma}_\xi \sim N(\boldsymbol{\Delta}_{20}, \mathbf{V}_{20})$, and set $\boldsymbol{\Delta}_{10} = \boldsymbol{\Delta}_{20} = 0.05\mathbf{1}$ and $\mathbf{V}_{10} = \mathbf{V}_{20} = \mathbf{I}$. However, for the regression parameters many specifications can be considered and for a discussion see, for example, Lopes et al. (2008).

Finally, the prior for the latent process $\boldsymbol{\alpha}(t)$ is provided by the transition equation and completed by $\boldsymbol{\alpha}(0) \sim N(\mathbf{a}_0, \boldsymbol{\Sigma}_{\alpha 0})$, where we set the mean of the initial state $\mathbf{a}_0$ to zero and choose the initial variance matrix $\boldsymbol{\Sigma}_{\alpha 0}$ to be a function of the system matrices.

5.2 Posterior inference

Posterior inference for the proposed class of spatial dynamic factor models is facilitated by Markov Chain Monte Carlo algorithms. Standard MCMC for dynamic linear models are adapted to our model specification such that, conditional on r and m, posterior, predictive and interpolation analysis are readily available. We provide here some information on the relevant conditional distributions.
Define $\boldsymbol{\sigma}^2 = [\boldsymbol{\sigma}_y^2, \boldsymbol{\sigma}_x^2, \boldsymbol{\sigma}_\delta^2, \boldsymbol{\sigma}_\xi^2]$, $\boldsymbol{\tau}_x^2 = [\tau_{x1}^2, \ldots, \tau_{xr}^2]$, $\boldsymbol{\tau}_y^2 = [\tau_{y1}^2, \ldots, \tau_{ym}^2]$, $\boldsymbol{\phi}_x = [\phi_{x1}, \ldots, \phi_{xr}]$, $\boldsymbol{\phi}_y = [\phi_{y1}, \ldots, \phi_{ym}]$, $\boldsymbol{\alpha} = [\boldsymbol{\alpha}(0), \boldsymbol{\alpha}(1), \ldots, \boldsymbol{\alpha}(T)]$, $\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(T)]$ and $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(T)]$. Then by denoting with "u" the unobserved data, posterior inference is based on summarizing the joint posterior distribution

$$p(\mathbf{Y}^u, \mathbf{X}^u, \boldsymbol{\tau}_y^2, \boldsymbol{\tau}_x^2, \boldsymbol{\phi}_y, \boldsymbol{\phi}_x, \boldsymbol{\alpha}, \boldsymbol{\sigma}^2|\mathbf{Y}, \mathbf{X}).$$

The common factors are jointly sampled by means of the well known forward Filtering backward sampling (FFBS) algorithm (Carter and Kohn 1994; Frühwirth-Schnatter 1994). All other full conditional distributions are "standard" multivariate normal distributions or inverse gamma distributions, except the parameters, ϕ_j, characterizing the spatial correlations which are sampled based on a Metropolis-Hastings step. Specific details for the implementation of the involved full conditional distributions can be found, for example, in Lopes et al. (2008) and Lütkepohl (2005).

6 Uses of the Model

In this sections we provide specific details on how to obtain temporal forecasts and spatial predictions of the variable of interest Y.

6.1 Forecasting

One of the main objectives of the analysis might be that of obtaining temporal forecasts of the variable Y. These can be obtained through the state space formulation which naturally provides the framework to learn about the h-steps ahead predictive density, $p\left[\mathbf{z}(T+h)|\mathbf{Z}\right]$, of the joint process $\mathbf{Z} = [Y\ X]$. Thus, let $\boldsymbol{\Theta} = [\boldsymbol{\sigma}^2, \boldsymbol{\tau}_x^2, \boldsymbol{\tau}_y^2, \boldsymbol{\phi}_x, \boldsymbol{\phi}_y]$, then

$$\begin{aligned} p\left[\mathbf{z}(T+h)|\mathbf{Z}\right] &= \int p\left[\mathbf{z}(T+h)|\boldsymbol{\alpha}(T+h), \mathbf{H}, \boldsymbol{\Theta}\right] p\left[\boldsymbol{\alpha}(T+h)|\boldsymbol{\alpha}(T), \mathbf{H}, \boldsymbol{\Theta}\right] \\ &\qquad p\left[\boldsymbol{\alpha}(T), \mathbf{H}, \boldsymbol{\Theta}|\mathbf{Z}\right] d\boldsymbol{\alpha}(T+h)\ d\boldsymbol{\alpha}(T)\ d\mathbf{H}\ d\boldsymbol{\Theta} \end{aligned}$$

where $\left[\mathbf{z}(T+h)|\boldsymbol{\alpha}(T+h), \mathbf{H}, \boldsymbol{\Theta}\right] \sim N\left[\mathbf{H}\boldsymbol{\alpha}(T+h), \boldsymbol{\Sigma}_{u_y}\right]$, $\left[\boldsymbol{\alpha}(T+h)|\boldsymbol{\alpha}(T), \mathbf{H}, \boldsymbol{\Theta}\right] \sim N(\boldsymbol{\mu}_h, \mathbf{V}_h)$, $\boldsymbol{\mu}_h = \boldsymbol{\Phi}^h\boldsymbol{\alpha}(T)$ and $\mathbf{V}_h = \sum_{j=1}^h \boldsymbol{\Phi}^{h-1}\boldsymbol{\Psi}(\boldsymbol{\Phi}^{h-1})'$, for $h \geq 0$.

Then, if $\left\{(\mathbf{H}^{(1)}, \boldsymbol{\Theta}^{(1)}, \boldsymbol{\alpha}(T)^{(1)}), \ldots, (\mathbf{H}^{(M)}, \boldsymbol{\Theta}^{(M)}, \boldsymbol{\alpha}(T)^{(M)})\right\}$ is a sample from $p\left[\boldsymbol{\alpha}(T), \mathbf{H}, \boldsymbol{\Theta}|\mathbf{Z}\right]$ it is easy to draw $\boldsymbol{\alpha}(T+h)^{(j)}$ from $p\left[\boldsymbol{\alpha}(T+h)|\boldsymbol{\alpha}(T)^{(j)}, \mathbf{H}^{(j)}, \boldsymbol{\Theta}^{(j)}\right]$, for all $j = 1, \ldots, M$, such that $\hat{p}\left[\mathbf{z}(T+h)|\mathbf{Z}\right] = M^{-1}\sum_{j=1}^{M} p\left[\mathbf{z}(T+h)|\boldsymbol{\alpha}(T+h)^{(j)}, \mathbf{H}^{(j)}, \boldsymbol{\Theta}^{(j)}\right]$ is a Monte Carlo approximation to $p\left[\mathbf{z}(T+h)|\mathbf{Z}\right]$. Analogously, a sample $\left\{\mathbf{z}(T+h)^{(1)}, \ldots, \mathbf{z}(T+h)^{(M)}\right\}$ from $p\left[\mathbf{z}(T+h)|\mathbf{Z}\right]$ is obtained by sampling $\mathbf{z}(T+h)^{(j)}$ from $p\left[\mathbf{z}(T+h)|\boldsymbol{\alpha}(T+h)^{(j)}, \mathbf{H}^{(j)}, \boldsymbol{\Theta}^{(j)}\right]$, for $j = 1, \ldots, M$.

6.1.1 Conditional Forecasting

The forecasting procedure described in the previous section is obtained under the hypothesis that the predictor X is unknown for the period of interest. However, as shown in Figure 1 (right), occasionally the forecaster may know the "future" values (with respect to Y) of the exogenous variable. When this is the case, we may produce temporal forecasts of $\mathbf{g}(t)$ conditional on a specific path of $\mathbf{f}(t)$. In the following, we thus propose a simple procedure to obtain $\mathbf{g}(T+h)|\mathbf{f}(t)$, thus avoiding the use of equation (2) to obtain h-steps ahead forecasts of $\mathbf{f}(t)$. Suppose that for the period $T+1, T+2, \ldots, T+h$, X is known such that $\mathbf{X}_h = [\mathbf{x}(T+1), \mathbf{x}(T+2), \mathbf{x}(T+h)]$. Then, h-step forecasts of $\mathbf{g}(t)$ may be obtained conditional on $\mathbf{f}_h = [\mathbf{f}(T+1), \mathbf{f}(T+2), \ldots, \mathbf{f}(T+h)]$, with $\mathbf{f}_h = \mathbf{H}_x^{\dagger}\mathbf{X}_h$ where $\mathbf{H}_x^{\dagger}$ denotes the pseudo-inverse of $\mathbf{H}_x$.

6.2 Interpolation

In this section we are now interested in spatial interpolation at the n_u locations where the response variable Y has not yet been observed. More precisely, let $\mathbf{y}^o$ denote the vector of observations from locations in S and $\mathbf{y}^u$ denote the vector of measurements to be predicted in locations defined by $S^u = \{\mathbf{s}_{n_y+1}, \ldots, \mathbf{s}_{n_y+n_u}\}$. Also, let $\mathbf{h}_{yj} = \{\mathbf{h}_{yj}^{o'}, \mathbf{h}_{yj}^{u'}\}$ be the j-column of the factor loadings matrix $\mathbf{H}_y$ with $\mathbf{h}_{yj}^o$ corresponding to $\mathbf{y}^o$ and $\mathbf{h}_{yj}^u$ corresponding to $\mathbf{y}^u$, respectively. Interpolation consists of finding the posterior distribution of $\mathbf{h}_{yj}^u$ (Bayesian kriging)

$$p\left(\mathbf{h}_y^u|\mathbf{y}^o\right) = \int p\left(\mathbf{h}_y^u|\mathbf{h}_y^o, \boldsymbol{\Theta}\right) p\left(\mathbf{h}_y^o, \boldsymbol{\Theta}|\mathbf{y}^o\right) d\mathbf{h}_y^o \, d\boldsymbol{\Theta}$$

where $p\left(\mathbf{h}_y^u|\mathbf{h}_y^o, \boldsymbol{\Theta}\right) = \prod_{j=1}^{m} p(\mathbf{h}_{yj}^u|\mathbf{h}_{yj}^o, \mathbf{m}_j^{(h_y)}, \tau_{y,j}^2, \phi_{y,j})$. Standard multivariate normal results can be used to derive, for $j = 1, \ldots, m$, the distribution of $p(\mathbf{h}_{yj}^u|\mathbf{h}_{yj}^o, \mathbf{m}_j^{(h_y)}, \tau_{y,j}^2, \phi_{y,j})$. Conditionally on $\boldsymbol{\Theta}$,

$$\begin{bmatrix} \mathbf{h}_{yj}^o \\ \mathbf{h}_{yj}^u \end{bmatrix} \sim MVN \left[\begin{pmatrix} \mathbf{D}^{(h_y o)} \\ \mathbf{D}^{(h_y u)} \end{pmatrix} \boldsymbol{\beta}_j^{(h_y)}; \tau_{y,j}^2 \begin{pmatrix} \mathbf{R}^{(o,o)}(\phi_{yj}) & \mathbf{R}^{(o,u)}(\phi_{yj}) \\ \mathbf{R}^{(u,o)}(\phi_{yj}) & \mathbf{R}^{(u,u)}(\phi_{yj}) \end{pmatrix} \right]$$

where $\mathbf{R}^{(u)}(\phi_{yj})$ is the correlation matrix of dimension n_u between ungauged locations, $\mathbf{R}^{(u,o)}(\phi_{yj})$ is a matrix of dimension $(n_u \times n_y)$ where each element represents the correlation between ungauged location $\mathbf{s}_i$ and gauged location $\mathbf{s}_j$, for $i = 1, \ldots, n_u$ and $j = 1, \ldots, n_y$. Therefore, it follows that the conditional mean and variance are

$$E\left[\mathbf{h}_y^u|\mathbf{h}_y^o, \boldsymbol{\Theta}\right] = \mathbf{D}^{(h_y u)}\boldsymbol{\beta}_j^{(h_y)} + \mathbf{R}^{(u,o)}(\phi_{yj})\mathbf{R}^{(o,o)}(\phi_{yj})^{-1}(\mathbf{h}_{yj}^o - \mathbf{D}_j^{(h_y o)}\boldsymbol{\beta}_j^{(h_y)})$$

$$Var\left[\mathbf{h}_y^u|\mathbf{h}_y^o, \boldsymbol{\Theta}\right] = \mathbf{R}^{(u,u)}(\phi_{yj}) - \mathbf{R}^{(u,o)}(\phi_{yj})\mathbf{R}^{(o,o)}(\phi_{yj})^{-1}\mathbf{R}^{(o,u)}(\phi_{yj}),$$

and the usual Monte Carlo approximation to $p(\mathbf{h}_y^u|\mathbf{y}^o)$ is $\hat{p}(\mathbf{h}_y^u|\mathbf{y}^o) = L^{-1}\sum_{l=1}^{L} p(\mathbf{h}_y^u|\mathbf{h}_y^{o(l)}, \boldsymbol{\Theta}^{(l)})$, where $\{(\mathbf{h}_y^{o(1)}, \boldsymbol{\Theta}^{(1)}), \ldots, (\mathbf{h}_y^{o(L)}, \boldsymbol{\Theta}^{(L)})\}$ is a sample from $p(\mathbf{h}_y^o, \boldsymbol{\Theta}|\mathbf{y}^o)$. If $\mathbf{h}_y^{u(l)}$ is drawn from $p(\mathbf{h}_y^u|\mathbf{h}_y^{o(l)}, \boldsymbol{\Theta}^{(l)})$, for $l = 1, \ldots, L$, then $\{\mathbf{h}_y^{u(1)}, \ldots, \mathbf{h}_y^{u(L)}\}$ is a sample from $p(\mathbf{h}_y^u|\mathbf{y}^o)$.
As a by-product, the expectation of nonobserved measures $\mathbf{y}^u$ can be approximated by $\hat{E}[\mathbf{y}^u|\mathbf{y}^o] = L^{-1}\sum_{l=1}^{L} \mathbf{h}_y^{u(l)}\boldsymbol{\alpha}^l$.
Notice that when $n_x > n_y$, as in Figure 1 (left), it might be interesting to predict Y (at least) on some of the sites in which X is available. In this case, whether the spatial correlation between X and Y is quite high it may be worth trying to set $\left[\mathbf{D}^{(h_y o)}; \mathbf{D}^{(h_y u)}\right] = \mathbf{H}_x$ so that the interpolation procedure exploits the spatial information structure of the predictor.

7 Applications

In this section, we discuss the application of our model (1-4) to the two real data sets introduced in section 1. The first data set (section 7.1) is related to the daily mean concentrations of PM_{10} and NO_x observed in the Milan district; the second one (section 7.2) , instead, represents hourly measurements of O_3 and temperature variables observed over Mexico City. In both cases the interest would be in temporal forecasts and spatial interpolation of PM_{10} and O_3.
In both applications the choice of the number of components, m and r, as well as the orders p, q and s, of the autoregressive components in equations (1) and (2), is based on the following well-known predictive model choice criterion (see for example, Laud and Ibrahim, 1995; Gelfand and Ghosh, 1998; Sahu and Mardia, 2005)

$$PMCC = \sum \left\{(Y(\mathbf{s},t) - E[Y(\mathbf{s},t)_{rep}])^2 + Var[Y(\mathbf{s},t)_{rep}]\right\} \tag{9}$$

where the summation is taken over all the $(N \times T)$ observations except for the missing observations and $Y(\mathbf{s},t)_{rep}$ is a future observation corresponding to $Y(\mathbf{s},t)$ under the model assumed.

7.1 Modeling PM_{10} and NO_x in the Milan district

A quite large number of monitoring sites located overall the Lombardy Region (Italy) for PM_{10} and NO_x was shown in Figure 1. However, due to the large amount of missing data for many of the series, we have to restrict our analysis to a smaller number for which both variables are simultaneously available. Specifically, we consider 20 sites and the period of the data covers the months January-October 2008, for a (20×297) data matrix. The raw data are provided by the Environmental Agency (ARPA) of Lombardy Region. Latitudes and longitudes are expressed according to the universal transverse Mercator (UTM) coordinates and these are measured in kilometers.
To test the model capability to perform temporal forecasting and spatial interpolation, the last week of the observed data and the temporal series of two monitoring sites have been excluded from the estimation procedure and used only for prediction purposes. Thus, we have $n_x = n_y = 18$, $T = 290$; $n_u = 2$, $S^u = \{\mathbf{s}_{n_y+1}, \mathbf{s}_{n_y+2}\}$, and a forecast period $T_h = \{T+1, \ldots, T+7\}$. For interpolation and forecasting comparison purposes, the following two different cases are also considered:

i) the NO_x variable is assumed to be available for the sites in S^u; that is, the time series of $X(\mathbf{s}_1^u, t)$ and $X(\mathbf{s}_2^u, t)$, for $t = 1, \ldots, 297$, are available at sites and time points for which both spatial interpolations and temporal forecasts for Y are required;

ii) the NO_x variable is assumed not to be available at S^u.

Because concentration data are always positive, it is convenient to operate on a logarithmic scale to remove the effect of heteroschedasticity. For this data set, missing data are in a moderate amount (less than 2%) and they will be reconstructed at each MCMC iteration by sampling from the full conditional distribution of Y.
An exploratory analysis of the data shows quite large correlations (many above 0.75) among the time series of the spatial sites; however, some features of the data suggest that it will be difficult to predict all the data satisfactorily. For example, in the period August-October, some close sites show highly disparate values with correlations close to 0.2.
Examination of graphical representations of X and Y, such as contour plots drawn at several time points, highlights the presence of a spatial quadratic surface. Thus, the entries of the design matrices, $\mathbf{D}^{(h_x)}$ and $\mathbf{D}^{(h_y)}$, for the specification of $\mathbf{m}_j^{(h_x)}$ and $\mathbf{m}_j^{(h_y)}$, respectively, are functions of the spatial coordinates and are defined to represent a six-parameter quadratic trend model (Cressie, 1993). However, for the Y variable, we also consider the effect of setting $\mathbf{D}^{(h_y)} = \mathbf{H}_x$.
Different values for p, q and s, ranging from 1 to 2, and an increasing number of factors (never larger than 12 for each variable) have been considered for model specification. The MCMC algorithm described above was also run for parameter estimation for a total of $75,000$ iterations and posterior inference was based on the last $50,000$ draws using every 10th member of the chains. The MCMC chains of the parameters were monitored to detect possible problems in convergence although, no such problems were found in the implementation.
Competing models were compared on the base of the predictive criterion (9) and results for $PMCC$ are shown in table 1. The optimal choice was found for $m = 9$ and $r = 8$ and $\mathbf{D}^{(h_y)} = \mathbf{H}_x$; henceforth we work with this model specification; the first component represents a local linear temporal trend (see section 4.2) while the others are temporally stationary components. Specifically, $\mathbf{f}(t)$ follows a VAR(1) while $\mathbf{g}(t)$ is an ADL$(1, 1)$. The dynamic evolution of the factors is characterized by considering $\mathbf{R}_1$ and $\mathbf{B}_1$ as diagonal matrices, while $\mathbf{C}_0$ and $\mathbf{C}_1$ are full matrices.

$PMCC$	r					
m	7	8	9	10	11	12
7	455.2	450.2	456.1	453.0	457.1	459.4
8	440.5	438.4	430.3	425.1	437.2	444.1
9	399.9	392.6	401.1	400.0	411.8	418.2
10	408.1	401.5	409.1	405.1	417.4	419.6
11	412.4	410.4	414.3	409.2	428.1	439.3
12	418.6	416.7	420.2	413.8	442.7	451.1

Table 1: Values of the predictive model choice criterion - $PMCC$ -for various values of m and r

The MCMC estimates of the components $g_i(t),\ i = 1, \ldots, 9$ along with their 95% credibility intervals are shown in Figures 2 for the chosen model.

Factors may be identified according to their relative weight in the explanation of the data variability. On average the larger proportion of the data variability is associated with the trend factor. It accounts for 38% while the third factor appears after that with around 15%. Over all, the common factors explain around 71% of the variability.
Table 2 presents posterior summaries for the spatial dependence of the factor loadings. For the PM_{10}, the posterior mean of the correlation parameter of the first factor loading corresponds to

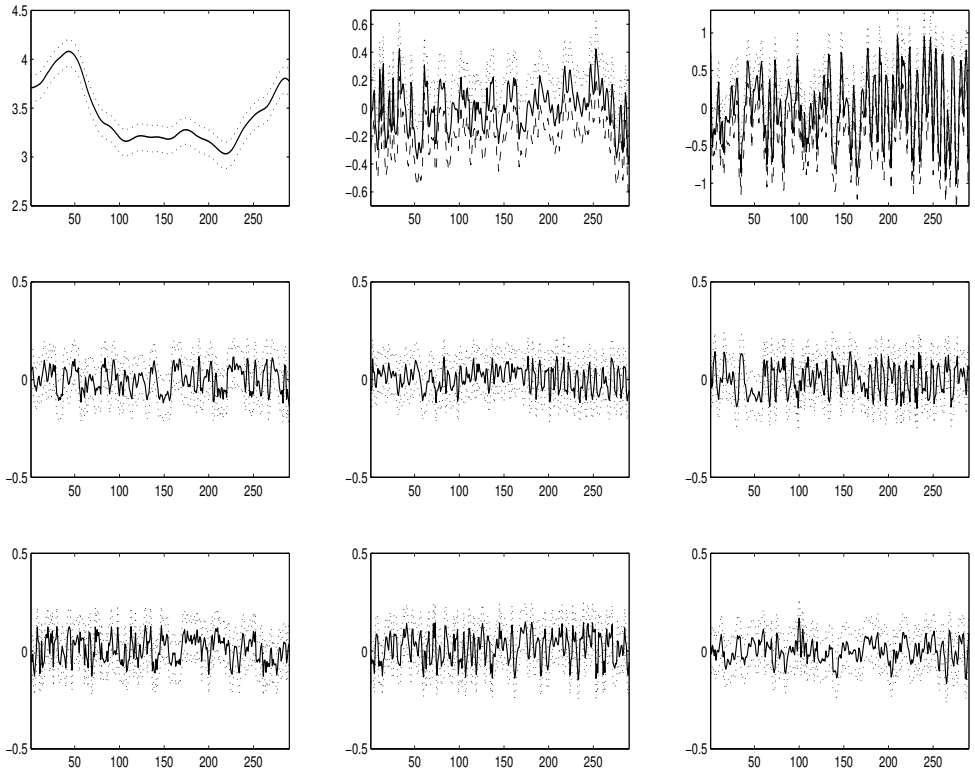

Figure 2: Marginal posterior means for $g_i(t),\ i = 1, \ldots, 9$ (in lexico-graphic order) and 95% credible intervals (dashed line).

an approximate range of 63 kilometers in spatial dependence since the covariogram decays to 0.05 for $\hat{\phi}_{y1} = 21$. For the other factor loadings the spatial correlations at 40 kilometers vary from 0.03 (for the second factor loading) to 0.06 (for the third factor loading). On the other hand, the NO_x estimated ranges vary approximately from 31 kilometers to 58 kilometers. We believe that these estimates might reflect the characteristics of the area under study which is flat (i.e it does not show specific geographical morphologies) and highly industrialized.

To provide some examples, Figures 3 and 4 present the mean surfaces for the first 4 columns of the standardized (Bollen, 1989) factor loadings matrix $\mathbf{H}_y$ obtained by interpolation, as explained in Section 6.2. In each map, we also show the monitoring network where the size of each site is proportional to the absolute value of the corresponding factor loading.

Since all the weights of the first factor loading are very similar to each other, the first factor represents the grand mean and accounts for the global time-trend variability of all the series. Also the estimated loadings of the second factor do not show any specific spatial pattern and it can thus be interpreted as a common stationary component. Instead, the third factor loading is a contrast between two groups of 9 sites each, with one group mainly located in the north-eastern part. Finally, the fourth factor loading shows the larger values for the sites centered on the city of Milan and also represents a contrast between the two groups totaling 14 sites and the remaining 4 sites in the north-eastern corner of the area of study.

	1^{st} F.L.	2^{nd} F.L.	3^{rd} F.L.	4^{th} F.L.	5^{th} F.L.
Mean	23.89	12.26	15.38	15.47	13.39
Median	21.05	11.86	14.63	14.32	12.46
95% C.I.	[14.91,29.93]	[7.29,18.84]	[8.11,22.57]	[8.60,23.70]	[7.61,23.50]
	6^{th} F.L.	7^{th} F.L.	8^{th} F.L.	9^{th} F.L.	
Mean	14.64	12.37	13.51	15.14	
Median	14.05	11.70	13.09	13.63	
95% C.I.	[7.75,21.02]	[7.73,19.58]	[8.18,22.37]	[8.06,24.10]	

Table 2: Posterior summary for the decay parameters characterizing the columns of the factor loadings matrix (PM_{10}). C.I. and F.L. stand for credibility interval and factor loading, respectively.

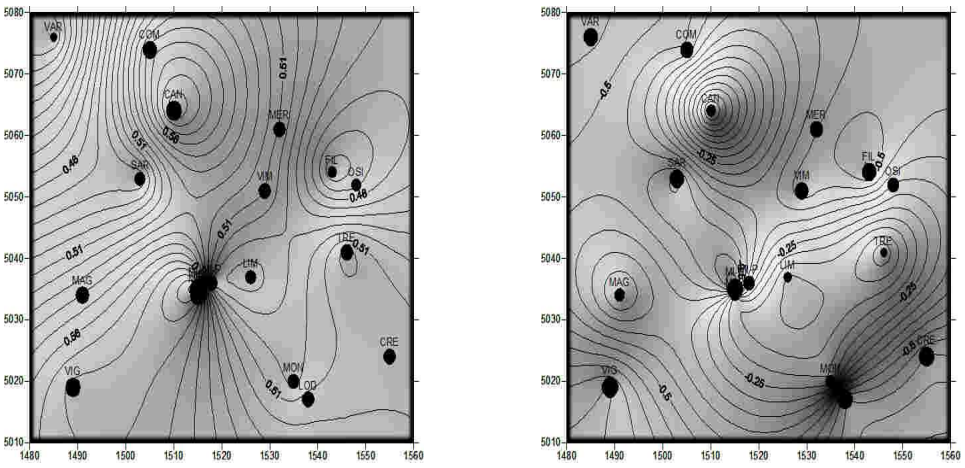

Figure 3: Factor loadings interpolation. The contour values represent the range of the posterior means for $\mathbf{h}_{y1}$ and $\mathbf{h}_{y2}$, respectively. The size of each site is proportional to the absolute value of the corresponding factor loading.

The autocorrelations of the raw residuals, the differences between the observed and the fitted, are also given in Figure 5 and they do not show any temporal pattern. This is also confirmed by the Ljung-Box statistics whose null hypothesis is rejected for all the residual series.
Forecasting and interpolation results, for the case ii) described above, where the NO_x is not available at sites in S^u, are first presented in Figures 6 and 7. Specifically, Figure 6 provides some examples of forecasts obtained for six of the 18 observed series. As it can be seen, the "unconditional" forecasts, compared with the true values, seem to provide quite good results but in some cases they tend to be very similar to the mean of the data. Figure 7 instead, shows "unconditional" interpolation results for sites in S^u. We can notice the ability of the model in tracking the behavior of the series although some difficulties in the fit are observed at the beginning of the series.
Forecasting and interpolation results, for the case i) instead, are shown in Figures 8 and 9. As can be seen, the "conditional" (on the known values of X) forecast and interpolation approach now exhibits more encouraging out-of-sample properties of the model, with data points being more accurately forecast and interpolated for several steps ahead and out-of-sample monitoring stations, respectively. It is worth noting here that none of the 95% credibility intervals, either based on forecasting or interpolation are symmetric, and that in general, they appear narrower than in the

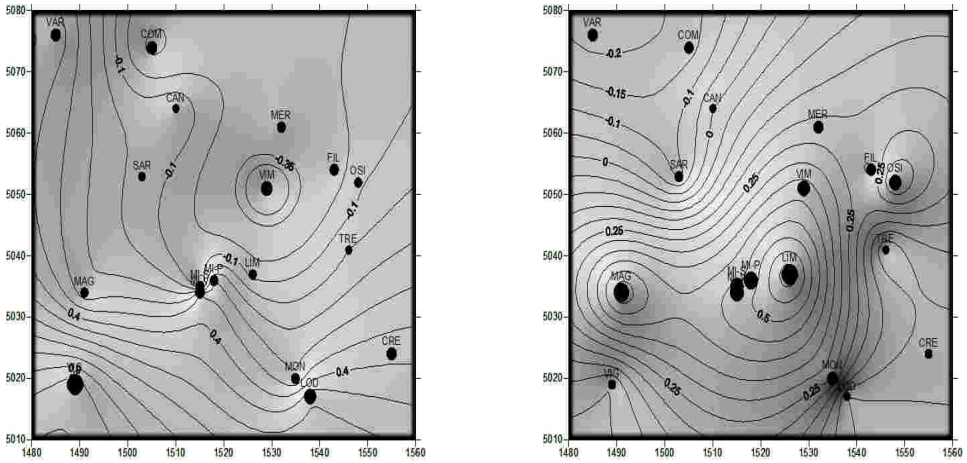

Figure 4: Factor loadings interpolation. The contour values represent the range of the posterior means for $\mathbf{h}_{y3}$ and $\mathbf{h}_{y4}$, respectively. The size of each site is proportional to the absolute value of the corresponding factor loading.

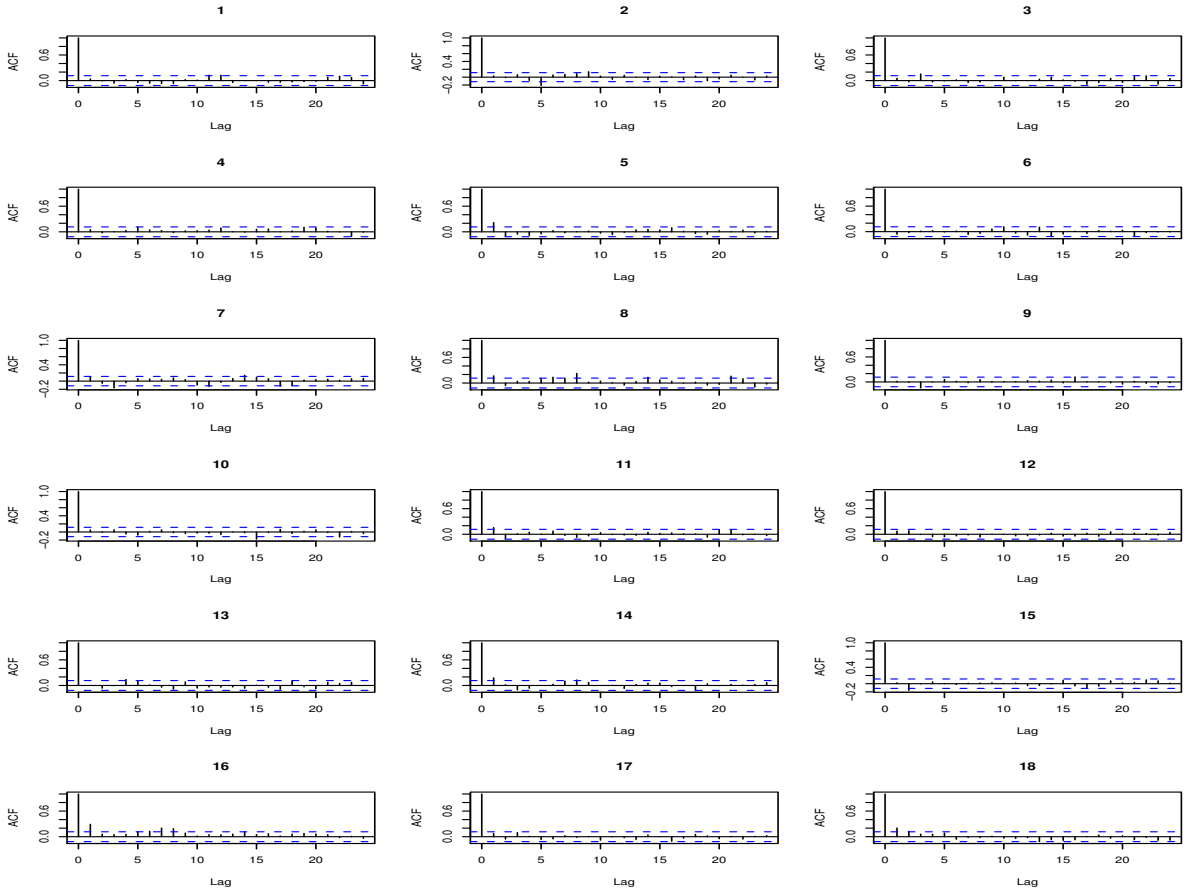

Figure 5: Plot of the autocorrelations of the raw residuals.

unconditional case.

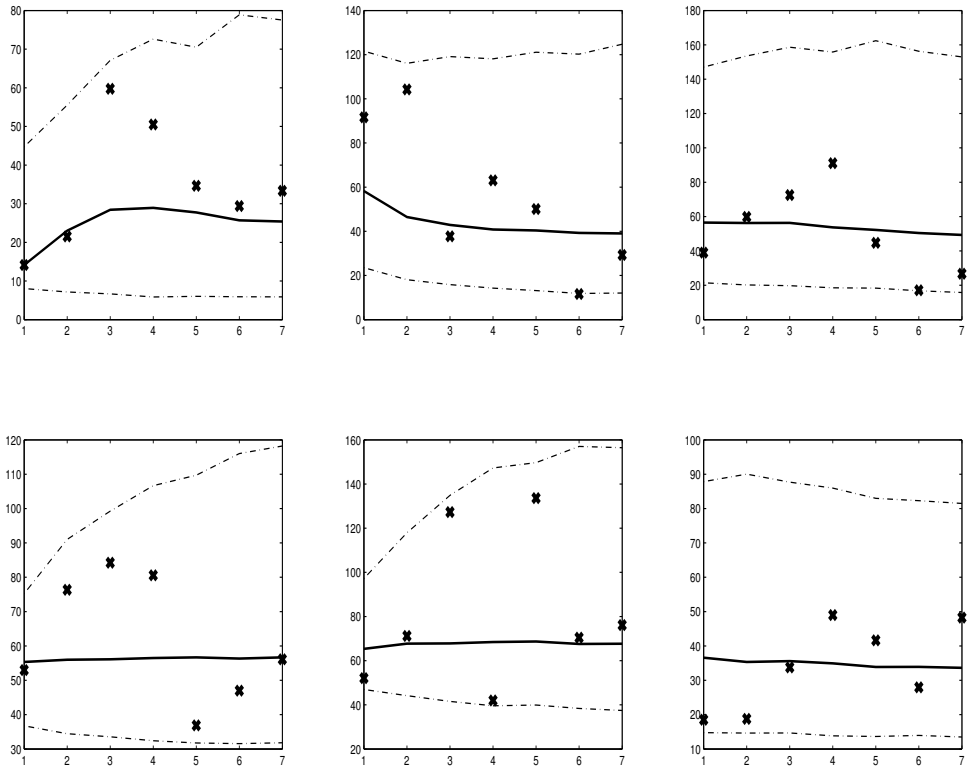

Figure 6: Comparison of "unconditional" forecasts and true values for a selection of six sites. True data (x), interpolated values (continuous line), 95% credible interval limits (dashed line).

7.2 Modeling Ozone Levels at Mexico City

In this section we describe the main results obtained in modeling Ozone levels. Specifically, we consider hourly readings of concentrations of O_3 and air temperature which, in our model formulation, is used as a covariate. The O_3 variable is observed at 20 monitoring sites while the air temperature is observed at 16 stations of which 12 share the same coordinates of the O_3 sites. The data cover the period from February 6th (14:00 p.m.) to February 12th (17:00 p.m.). The monitoring sites are scattered irregularly in Mexico City and the network is named Red Automatica de Monitoreo Ambiental (RAMA) de la Ciudad de México. The coordinates are expressed in kilometers. The exploratory analysis of the cycle behavior of the time series essentially confirms the presence of a peak corresponding to a daily cycle with wavelengths of 24 h and a peak corresponding to a harmonic cycle with a wavelength of 12 h. This feature was also noticed by Huerta et al. (2004). The exploratory analysis also confirms that the variability of the mean level across stations is important and that the cyclical behavior of the series differs in amplitude according to location. This site-specific feature can be naturally captured by the components of the factor loadings matrices.

As in section 7.1, to test the forecasting performance of the model, the last 48 hours have been excluded from the estimation procedure and used only for prediction purposes. Furthermore, to provide interpolation results conditional on known values of the temperature, the time series of two

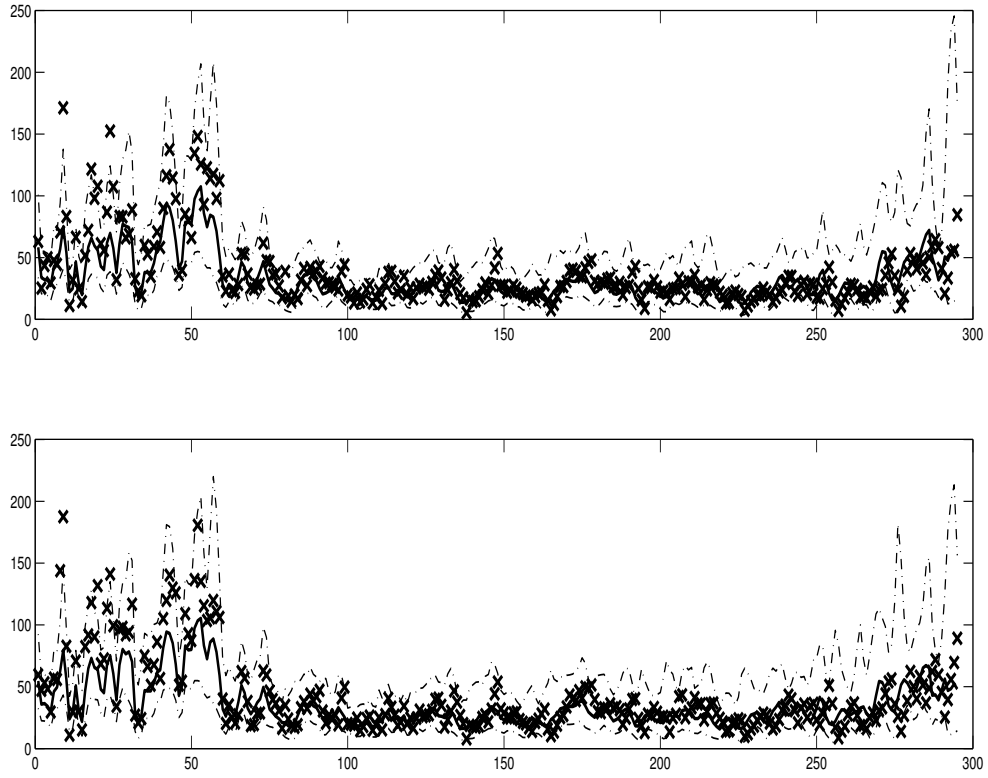

Figure 7: "Unconditional" interpolation results for sites in S^u. True data (x), interpolated values (continuous line), 95% credible interval limits (dashed line).

monitoring sites for O_3 have also been excluded from the analysis. Thus, we have $n_y = 18,\ n_x = 16,\ T = 100$; $n_u = 2,\ S^u = \{\mathbf{s}_{n_y+1}, \mathbf{s}_{n_y+2}\}$, and a forecast period $T_h = \{T+1, \ldots, T+48\}$. Different model specifications, with an increasing number of factors have been tested and, according to the predictive criterion (9), the optimal choice was found for $m = 7$ and $r = 6$. For both variables, the first two components represent the harmonic cycles with a wavelength of 24 h and 12 h, respectively. The remaining 4 components in $\mathbf{f}(t)$ follow a random walk, while the remaining 5 components in $\mathbf{g}(t)$ follow an ADL$(1, 1)$. Parameter estimation was carried out by MCMC for a total of $50,000$ iterations and posterior inference was based on the last $30,000$ draws using every 10th member of the chains. The MCMC chains of the parameters were monitored to detect possible problems in convergence although, no such problems were found in the implementation. The posterior mean for correlation of the first two factor loadings, corresponding to the harmonic cycle of 24 h and 12 h with respect to O_3, consist of an approximate range of 33.4 and 31.2 kilometers in spatial dependence since the covariogram decays to 0.05 for $\hat{\phi}_{y1} = 11.15$ and $\hat{\phi}_{y2} = 10.42$, respectively. Similar results are also found for the spatial dependence of the temperature since the estimated factor loadings corresponding to the two cycles show that the spatial correlations at 30 kilometers vary from 0.06 (for the harmonic cycle of 12 h) to 0.07 (for the harmonic cycle of 24 h).

Forecasting and interpolation results (on the original scale), conditional on the known values of the air temperature, are shown in Figures 10 and 11, respectively. Specifically, Figure 10 provides

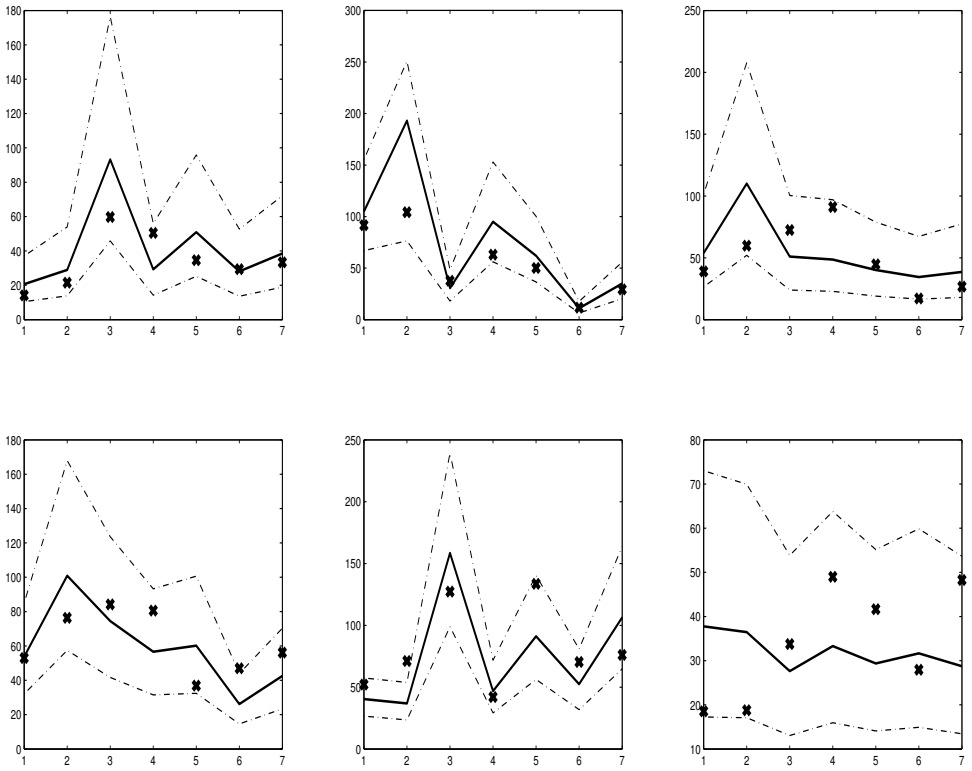

Figure 8: Comparison of "conditional" forecasts and true values for a selection of six sites. True data (x), interpolated values (continuous line), 95% credible interval limits (dashed line).

a selection of the three best (top row) and worse (bottom row) forecasts (mean with corresponding the 95% predictive probability intervals) obtained for the 18 observed series. As can be seen, in both cases, the model represents the cyclical patterns and non-stationarities of the data adequately, with the actual observed values falling within the range of the forecast intervals.

8 Discussion

In this paper we have discussed the modeling of coupled environmental variables by means of Bayesian dynamic factor models with smooth factor loadings. Specifically, we use factor analysis ideas to frame and exploit both the spatial and the temporal structure of the observed processes. The spatial variation is brought into the model through the columns of the factor loadings matrix, while the time series dynamics are captured by the common dynamic factors which follow time series decomposition processes, such as local and global trends, cycle and seasonality.
The matrix of factor loadings plays the important role of weighing the common factors in general factor analysis and is here incumbent of modeling spatial dependence. A key feature of the proposed model is its ability to encompass several existing models, which are restricted in most cases to non-stochastic factor loadings matrices. Furthermore, unlike many existing spatio-temporal models, our model formulation also works in cases in which X and Y are observed on different

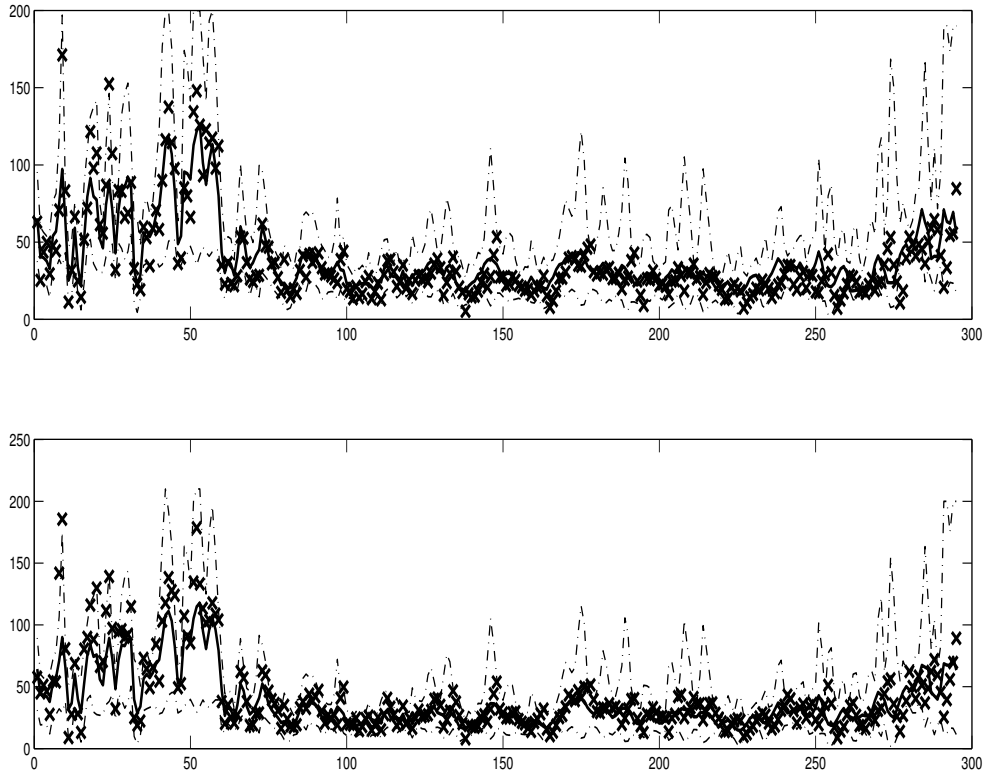

Figure 9: "Conditional" interpolation results for sites in S^u. True data (x), interpolated values (continuous line), 95% credible interval limits (dashed line).

spatial sites.
Another main advantage of the model formulation is that it enables to consider cases in which both the spatial and the temporal series of X may be longer than that of Y. As noticed, this is particularly useful to produce improved spatial and temporal predictions by "conditioning" on known values of the predictor. For example, this seems to be particularly effective for the Ozone - air temperature case where, in general, the air temperature is either available or easily predictable.
The model has been implemented in a Bayesian set-up using MCMC sampling. Despite the large number of parameters implied by the model formulation we did not find convergency problem of the algorithm. The implemented MCMC code was also validated through an extensive simulation study where, both model parameter estimation and prediction capabilities of the model were considered. In general, simulation results (not provided here) show that: i) all the parameters are well estimated and all true values fall within the marginal 95% credibility intervals, ii) accurate estimates of both factor loadings and common dynamic factors can be obtained, iii) the $PMCC$ criterion is able to correctly select the order of the simulated factor model and, iv) excellent results can be achieved if we would be able to produce predictions conditional on the known values of the predictor.
As a final consideration, we emphasize that the model can be indifferently used for both continuous and lattice spatial data and that it can be easily extended allowing both X and Y to be multivariate multidimensional spatio-temporal processes. This will constitute an extension of the model

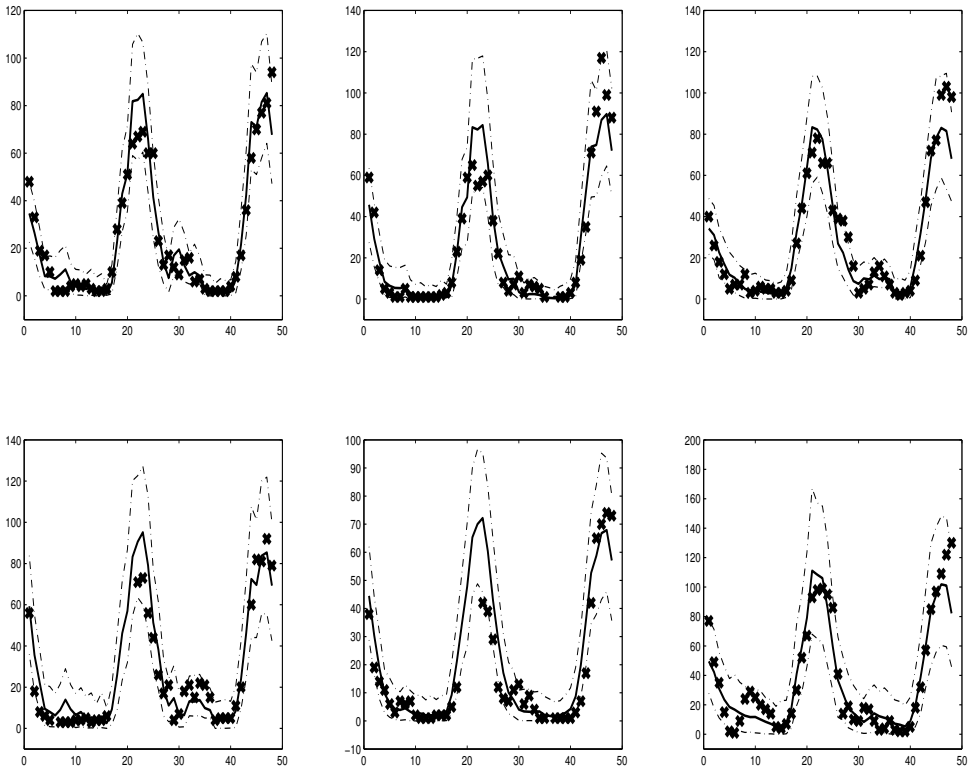

Figure 10: Comparison of "conditional" forecasts and true values: the best and worst forecasts are shown in the top and bottom rows, respectively. True data (x), interpolated values (continuous line), 95% credible interval limits (dashed line).

discussed in Fontanella et al. (2007) and will be a topic for future work. In this case however, computational problems related to the inversion of high dimensional covariance matrices might arise; in fact, the computational burden increases with the dimension of T, n_y, n_x, m and r. No such problems were found here for the two case studies.

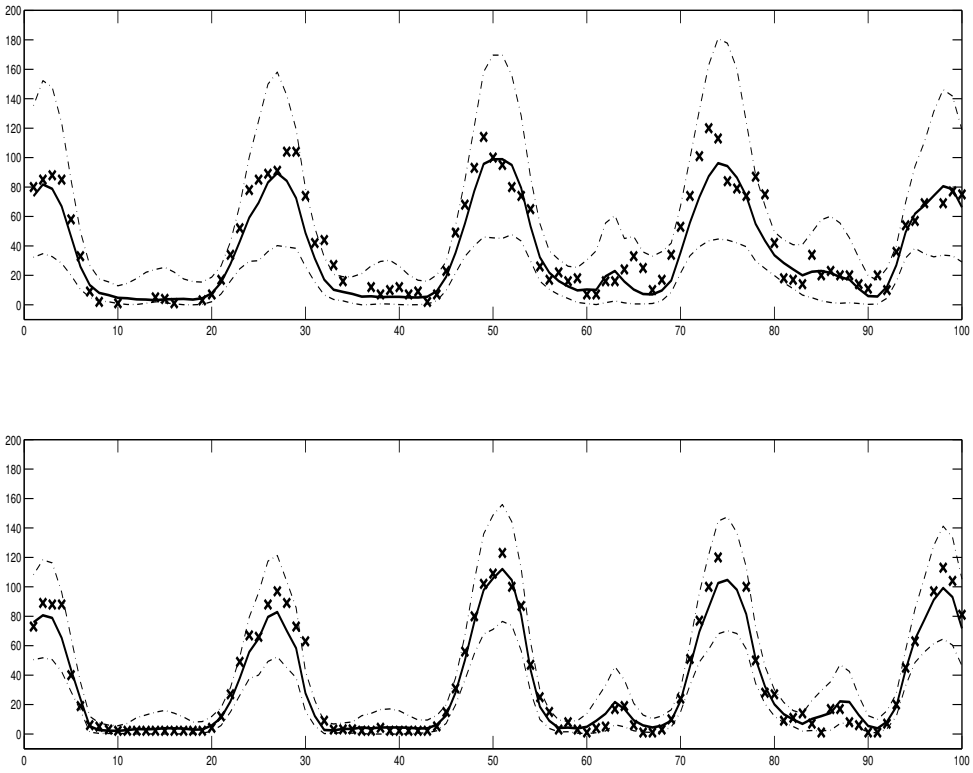

Figure 11: "Conditional" interpolation results for sites in S^u. True data (x), interpolated values (continuous line), 95% credible interval limits (dashed line).

ACKNOWLEDGMENTS

The authors would like to thank R. J. Bhansali, A. Gelfand, B. Sansó and S. Sahu for invaluable comments on preliminary versions of the article that significantly improved its quality. Particular thanks also go to E. Salazar for helpful discussion on the implementation of MCMC algorithms. The first and second authors acknowledge the financial support of the University G. d'Annunzio of Chieti-Pescara received for the cooperative agreement with the Depto. de Métodos Estatísticos - UFRJ, Rio de Janeiro. The third author thanks the financial support of research grants from CNPq and FAPERJ.

References

Aguilar O, West M. 2000. Bayesian dynamic factor model and portfolio allocation. *Journal of Business and Economic Statistics*, **18**(3), 338-357.

Allcroft, D. J. and Glasbey, C. A. (2003) A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **52**, 487498.

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

Beattie CI, Longhurst JWS, Woodfield NK. 2002. Air quality action plans: early indicators of urban local authority practice in England. *Environmental Science & Policy*, **5**, 463-470.

Bollen KA. (1989). *Structural Equations With Latent Variables*. Wiley: New York.

Brown, P. E., Diggle, P. J., Lord, M. E. and Young, P. C. (2001) Spacetime calibration of radar rainfall data. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **50**, 221241.

Calder, C. (2007). Dynamic factor process convolution models for multivariate spacetime data with application to air quality assessment. *Environmental and Ecological Statistics*, **14**, 229-247.

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

Gelfand A. E. and Ghosh S.K. (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika* **85**, 1-11.

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

Christensen, W. F. and Amemiya, Y. (2003). Modeling and prediction for multivariate spatial factor analysis. *Journal of Statistical Planning and Inference*, **115**, No. 2, 543-564.

Fontanella L., Ippoliti L. and Valentini P. (2007). Environmental pollution analysis by dynamic structural equation models. *Environmetrics*, **18**, 265-283.

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

Geweke JF, Zhou G. 1996. Measuring the pricing error of the arbitrage pricing theory. *The Review of Finantial Studies*, **9**, 557-587.

Gamerman, D., Moreira, A. R. B., and Rue, H. (2003). Space-varying regression models: specifications and simulation. *Computational Statistics and Data Analysis*, **42**, 513-533.

Greene, W. H. (2003). *Econometric Analaysis*, 5th ed. New Jersey:Prentice Hall.

Hamilton, J.D. (1994). *Time Series Analysis*, Princeton: University Press, Princeton

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

Huerta G., Sansó B., Stroud J.R. (2004). A spatiotemporal model for Mexico City ozone levels. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **53**(2), 231-248.

Kent, J.T., Mardia, K.V. (2002). Modelling Strategies for Spatial-Temporal Data. In: *Spatial Cluster Modelling. Eds A. Lawson and D. Denison*. 213-226, London: Chapman and Hall.

Kitagawa, G. and Gersch, W. (1996), *Smoothness Priors Analysis of Time Series*, Springer, New York.

Lee SY, Shi JQ. 2000. Joint Bayesian analysis of factor scores and structural parameters in the factor analysis model. *Annals of the Institue of Statistical Mathematics*, **52**(4), 722-736.

Laud, P. W. and Ibrahim, J. G. (1995) Predictive model selection. *Journal of the Royal Statistical Society, Series B*, **57**, 247-262.

Lopes, H.F., Salazar, E., and Gamerman, D.(2008). Spatial Dynamic Factor Analysis. *Bayesian Analysis*, **3**(4), 759-792.

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

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Berlin: Springer.

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

Mardia K.V., Goodall C., Redfern E.J., and Alonso F.J. (1998) The Kriged Kalman filter (with discussion). *Test*, **7**, 217-252.

Reich B. J., Fuentes M., Burke J. (2009) Analysis of the effects of ultrafine particulate matter while accounting for human exposure. *Environmetrics*, **20**(2),131-146.

Sahu, S. K. and Mardia, K. V. (2005). A Bayesian kriged-Kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **54**,223-244.

Sansó, B., and Guenni, L. (1999) Venezuelan rainfall data analysed by using a Bayesian spacetime model. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*,, **48**, 345362.

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

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

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

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

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