Space-Time Modelling of Coupled Spatio-Temporal Environmental Variables

L. **Ippoliti*** P. **Valentini*** D. **Gamerman**†

* **DMQTE** - *University G. d'Annunzio - Chieti-Pescara*

† **DME** - UFRJ, Rio de Janeiro

Primary interests in spatio-temporal modelling

In recent years there has been a tremendous growth in the statistical models and techniques to analyze spatio-temporal data.

Nowadays, there are many monographs, review articles and papers dealing with very specific cases of interest;

1. Often, the primary interests in analyzing space-time data are to **smooth and predict** in time (and/or in space) one (or more) variable(s) of interest observed over a certain spatial domain;
2. In many important areas also **descriptive analyses** to detect meaningful patterns are of interest.

Notation

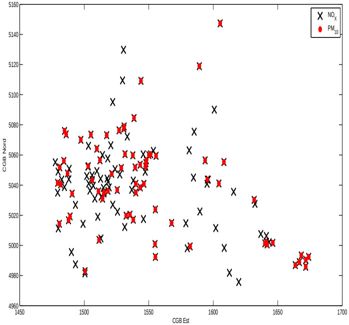
1. Let $Y(\mathbf{s};t)$ be a spatio-temporal process of interest, where $\mathbf{s} \in D$, with D some spatial domain in $\Re^d$ (usually $d = 2$) and $t \in \{1, 2, \ldots\}$ a discrete set of times;
2. For any given time $t = t_0$, we assume that $Y(\cdot; t_0)$ is a n_y-dimensional (stationary) spatial stochastic process.
3. If available, we denote the n_x-dimensional vector $X(\cdot; t)$ as a predictor of $Y(\cdot; t)$

$X(\mathbf{s};t)$ and $Y(\mathbf{s};t)$ are called **coupled variables**

Measurement dimensions n_y and n_x may change with time.

Our proposal: a state-space model for coupled variables.

Why use coupled variables?



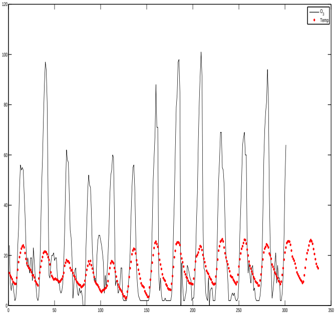


Figure: Left. Lombardy monitoring network (Italy): the sites for PM_{10} are represented by circles ($\circ$) while those for NO_x are represented by crosses ($\times$). Right. An example of hourly Ozone - O_3 - (continuous line) and air temperature (dashed lines) time series observed at Plateros station (Mexico City).

The Spatio-temporal Model for Coupled Variables

Consider two spatio-temporal processes $X(\mathbf{s}; t)$ and $Y(\mathbf{s}; t)$.

It is explicitly assumed that X can be used as a predictor of Y.

The model is specified hierarchically:

1. Measurements $[\mathbf{Y}, \mathbf{X}]$ are described through latent factors $[\mathbf{f}, \mathbf{g}]$.
2. $[\mathbf{f}, \mathbf{g}]$ describe the dependence in time and between coupled variables through hyperparameters
3. Hyperparameters specify the spatial dependence

Specifying the Measurement Equation

Consider the measurement equation for the variable of interest $Y(t)$,

$$\mathbf{y}(t) = \mathbf{H}_y \mathbf{g}(t) + \mathbf{u}_y(t)$$

For any site $\mathbf{s}$, the mean level of the space-time Gaussian process is $E\left[Y(\mathbf{s},t)\right] = \sum_{i=1}^{m} h_{yi}(\mathbf{s}) g_i(t)$

$h_{yi}(\mathbf{s}), i = 1, \ldots, m,$ - *spatial patterns* to be specified.

$\mathbf{u}_y(t)$ is a zero mean Gaussian process

Same holds for X with $(\mathbf{H}_y, \mathbf{g}, \mathbf{u}_y, m)$ replaced by $(\mathbf{H}_x, \mathbf{f}, \mathbf{u}_x, r)$.

Main point: dimension reduction for Y and X ($m \ll n_y$ and $r \ll n_x$).

Modelling the Latent Variables

The model for the temporal dynamic of the latent variables is

$$\begin{aligned} \mathbf{g}(t) &= \sum_{i=1}^{p} \mathbf{B}_i \mathbf{g}(t-i) + \sum_{j=0}^{q} \mathbf{C}_j \mathbf{f}(t-j) + \boldsymbol{\xi}(t) \\ \mathbf{f}(t) &= \sum_{k=1}^{s} \mathbf{R}_k \mathbf{f}(t-k) + \boldsymbol{\eta}(t) \end{aligned}$$

The $VAR(s)$ and $VARX(p,q)$ structures directly allow for a state-space representation of the full model.

State-space models provide a powerful tool for handling **inference** and **prediction problems**.

Complete model specification

The following dynamic (or state-space) model is thus specified

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

Specification is completed with priors for hyperparameter $(\boldsymbol{\Theta}, \mathbf{H}_y, \mathbf{H}_x)$.

The model can be written in state-space form

$$\begin{aligned}
\mathbf{z}(t) &= \mathbf{H}\boldsymbol{\alpha}(t) + \mathbf{u}(t) \\
\boldsymbol{\alpha}(t) &= \Phi\boldsymbol{\alpha}(t-1) + \boldsymbol{\zeta}(t)
\end{aligned}$$

$\mathbf{z}(t) = (\mathbf{y}(t), \mathbf{x}(t))$, $\boldsymbol{\alpha}(t) = (\mathbf{d}(t), ..., \mathbf{d}(t-p+1))$ and $\mathbf{d}(t) = (\mathbf{g}(t), \mathbf{f}(t))$.

Specifying the Spatial Patterns: Stochastic Specification

Spatial patterns aim at imposing similarity between nearby sites.

This can be achieved via stochastic or deterministic forms.

1. Stochastic Specification

$j-$th column of $\mathbf{H}_y$ is $\mathbf{h}_{yj} = \left(h_{yj}(\mathbf{s}_1), \ldots, h_{yj}(\mathbf{s}_{n_y})\right)'$, $j = 1, \ldots, m$

$\mathbf{h}_{yj}$'s are conditionally independent Gaussian Random Fields (GRF)

$$\mathbf{h}_{yj} \sim MVN\left(\mathbf{m}_j^{(h_y)}, \mathbf{\Sigma}_j^{(h_y)}\right)$$

$\mathbf{m}_j^{(h_y)}$ is a n_y-dimensional mean vector and
$\mathbf{\Sigma}_j^{(h_y)}$ is a parameterized spatial covariance matrix.

Similar reasoning used for $\mathbf{h}_{xj}$'s - columns of $\mathbf{H}_x$.

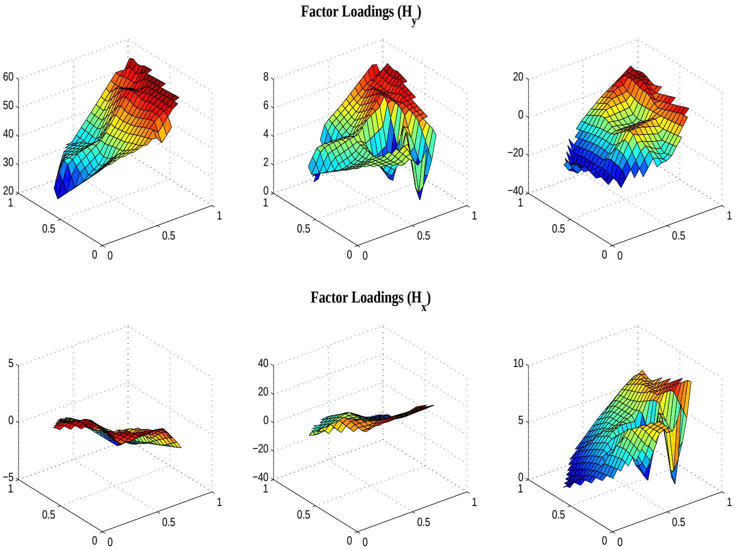
Specifying the Spatial Patterns: Stochastic Specification

1. $\mathbf{H}_y$ and $\mathbf{H}_x$ cannot be specified ahead of time and must be considered as a parameter with prior distribution

$$\mathbf{h}_{yj} \sim MVN\left(\mathbf{m}_j^{(h_y)}, \mathbf{\Sigma}_j^{(h_y)}\right)$$

2. Many specifications for the mean level $\mathbf{m}_j^{(h_y)}$ can be entertained:
   - constant mean level: $\mathbf{m}_j^{(h_y)} = \mathbf{1}_{n_y}\beta_j^{(h_y)}$
   - regression model: $\mathbf{m}_j^{(h_y)} = \mathbf{D}^{(h_y)}\boldsymbol{\beta}_j^{(h_y)}$, $\mathbf{D}^{(h_y)}$ - design matrix

Specifications considered in our work are:
a) $\mathbf{D}^{(h_y)}$ is the design matrix of a surface regression model (i.e. its entries are expressed as function of site's spatial coordinates);
b) $\mathbf{D}^{(h_y)} = \mathbf{H}_x$, useful to link the spatial structures of Y and X.


Factor Loadings (H_y)
Factor Loadings (H_x)

Specifying the Spatial Patterns: Deterministic Specification

By letting $\boldsymbol{\Sigma}_j^{(h_y)} = \mathbf{0}$, the spatial pattern $\mathbf{h}_{yj}$ is defined through a **deterministic specification**.

There are various choices for the space of drift functions and classical examples include:

i. **polynomials**: these are simple functions and do not depend on the choice of sites; however they have the disadvantage to be rather *wild* in their oscillatory behaviour and grow rapidly as $|\mathbf{s}| \to \infty$.

ii. **trigonometric functions**: useful for data which are periodic in space.

iii. **principle splines**: for principal kriging functions (Mardia et al., 1998; Sahu and Mardia, 2005).

Specifying the Spatial Patterns

iv. **Empirical Orthogonal Functions**: Wikle and Cressie(1999) model the principal fields using EOF's.

Assuming $Y(t)$ is a Gaussian process with covariance function $Q(\mathbf{s}, \mathbf{s}')$, then **the principal fields represent a realization of the eigenfunctions**, $w_k(\mathbf{s})$, of the following homogeneous integral equation

$$\int_D Q(\mathbf{s}, \mathbf{s}') w_k(\mathbf{s}) d(\mathbf{s}) = \lambda_k w_k(\mathbf{s}) \quad k \in \mathbb{N}.$$

In this case, according to the probabilistic corollary of Mercer's theorem (Obled, Creutin; 1986), the measurment equations (3) and (4) are obtained from the Karhunen-Loéve expansion (KLE).

Prior Distribution for hyperparameters

For simplicity, conditionally conjugate prior distributions are used for all parameters defining the dynamic factor model. To summarize:

1. Inverse Gamma are used for model variances (eg. the state and measurement noise, partial sill etc);
2. The Normal distribution truncated to the interval [a,b] is used for the autoregressive part of the state equation;
3. The Normal distribution is used for the mean vector of the spatial patterns $h_{y_i}(\mathbf{s}), i = 1, \ldots, m$;
4. For the **range parameter** we define: $\mathrm{IG}(2, b)$, where $b = d_0/(-2 ln(0.05))$ and d_0 max distance among the spatial sites;

Posterior Inference

The joint posterior distribution is analytically intractable.
Exact posterior inference is performed by a customized MCMC algorithm.

Common factors are jointly sampled by **forward filtering backward sampling** (FFBS) algorithm (Carter and Kohn, 1994, and Frühwirth-Schnatter, 1994).

All other full conditional distributions are multivariate normal or inverse gamma distributions $\rightarrow$ easy sampling
Exception: parameters characterizing the spatial correlations, sampled via a Metropolis-Hastings step.

Missing data: are sampled at each MCMC iteration from their full conditional distribution, obtained via the measurement equation.

Spatial Interpolation: point estimation

The interpolation of the spatial patterns on a new site $\mathbf{s}_0$ is not a difficult task and, in general, one could apply a relatively simple interpolation scheme.
The specification of the measurement equation shows how to get the point spatial prediction of Y at an unmonitored site $\mathbf{s}_0$

$$\begin{aligned}\hat{y}(\mathbf{s}_0, t) &= \hat{\mathbf{h}}_y(\mathbf{s}_0)'\hat{\mathbf{g}}(t) \\ &= \sum_{i=1}^{m} \hat{h}_{yi}(\mathbf{s}_0)\hat{g}_i(t)\end{aligned} \tag{5}$$

where $\hat{h}_{yi}(\mathbf{s}_0)$ are the predicted spatial patterns at site $\mathbf{s}_0$.

The latent variables $g_i(t)$ can be estimated (and predicted in time) within the Kalman filter (KF).

Bayesian Spatial Interpolation of Y

$\mathbf{y}^o$ - observed vector at locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_{n_y}\}$,
$\mathbf{y}^u$ - vector of measurements to be predicted at new locations
$S^u = \{\mathbf{s}_{n_y+1}, \ldots, \mathbf{s}_{n_y+n_u}\} \subset S$.

Main task: obtain predictive distribution $p(\mathbf{y}^u|\mathbf{y}^o)$

Task is achieved by noticing that

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

Two steps are required:
1) interpolation of loadings;
2) conditional prediction of measurements.

Spatial Interpolation: computations

First step requires the posterior distribution of $(\mathbf{H}_y^u, \mathbf{\Theta})$

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

where $p\left(\mathbf{H}_y^u|\mathbf{H}_y^o, \mathbf{\Theta}\right) = \prod_{j=1}^m p(\mathbf{h}_{yj}^u|\mathbf{h}_{yj}^o, \mathbf{m}_j^{(h_y)}, \tau_{y,j}^2, \phi_{y,j})$.

Posterior calculations give $p\left(\mathbf{H}_y^o, \mathbf{\Theta}|\mathbf{y}^o\right)$ and samples $\{(\mathbf{H}_y^{o(l)}, \mathbf{\Theta}^{(l)})\}$.

Standard multivariate normal results give $p(\mathbf{h}_{yj}^u|\mathbf{h}_{yj}^o, \mathbf{m}_j^{(h_y)}, \tau_{y,j}^2, \phi_{y,j})$ and samples $\{\mathbf{H}_y^{u(l)}\}$ - *Bayesian kriging*

Second step uses the measurement distribution $p(\mathbf{y}^u|\mathbf{H}_y, \mathbf{\Theta})$ to give samples $\{\mathbf{y}^{u(l)}\}$

Bayesian Prediction of Y

Forecasts for future Y's obtained through the state space formulation.

$\mathbf{z}(T+k)$ - vector of measurements to be predicted k-steps ahead of time

Main task: obtain predictive distribution $p\left(\mathbf{z}(T+k)|\mathbf{Z}\right)$

Task is achieved by noticing that

$$p\left(\mathbf{z}(T+k)|\mathbf{Z}\right) = \int p\left(\mathbf{z}(T+k)|\boldsymbol{\alpha}(T+k),\mathbf{H},\boldsymbol{\Theta}\right) p\left(\boldsymbol{\alpha}(T+k)|\boldsymbol{\alpha}(T),\mathbf{H},\boldsymbol{\Theta}\right) \times p\left(\boldsymbol{\alpha}(T),\mathbf{H},\boldsymbol{\Theta}|\mathbf{Z}\right) d\boldsymbol{\alpha}(T+k)\; d\boldsymbol{\alpha}(T)\; d\mathbf{H}\; d\boldsymbol{\Theta}$$

$(\mathbf{z}(T+k)|\boldsymbol{\alpha}(T+k),\mathbf{H},\boldsymbol{\Theta}) \sim N\left(\mathbf{H}\boldsymbol{\alpha}(T+k),\boldsymbol{\Sigma}_u\right)$,

$(\boldsymbol{\alpha}(T+k)|\boldsymbol{\alpha}(T),\mathbf{H},\boldsymbol{\Theta}) \sim N(\boldsymbol{\mu}_k,\mathbf{V}_k)$,

$\boldsymbol{\mu}_k = \boldsymbol{\Phi}^k\boldsymbol{\alpha}(T)$ and $\mathbf{V}_k = \sum_{j=1}^{k} \boldsymbol{\Phi}^{j-1}\boldsymbol{\Psi}(\boldsymbol{\Phi}^{j-1})'$, for $k>0$, with $\boldsymbol{\Phi}^0 = \mathbf{I}$.

Bayesian Forecasts of Y

Assume $\left\{(\mathbf{H}^{(1)}, \mathbf{\Theta}^{(1)}, \boldsymbol{\alpha}(T)^{(1)}), \ldots, (\mathbf{H}^{(M)}, \mathbf{\Theta}^{(M)}, \boldsymbol{\alpha}(T)^{(M)})\right\}$ is a sample from $p\,(\mathbf{H}, \mathbf{\Theta}, \boldsymbol{\alpha}(T)|\mathbf{Z})$

Easy to draw $\boldsymbol{\alpha}(T+k)^{(j)} \sim p\left(\boldsymbol{\alpha}(T+k)|\boldsymbol{\alpha}(T)^{(j)}, \mathbf{H}^{(j)}, \mathbf{\Theta}^{(j)}\right)$, $j = 1, \ldots, M$ and

$$\hat{p}\,(\mathbf{z}(T+k)|\mathbf{Z}) = M^{-1} \sum_{j=1}^{M} p\left(\mathbf{z}(T+k)|\boldsymbol{\alpha}(T+k)^{(j)}, \mathbf{H}^{(j)}, \mathbf{\Theta}^{(j)}\right)$$

is a Monte Carlo approximation to $p\,(\mathbf{z}(T+k)|\mathbf{Z})$.

A sample $\left\{\mathbf{z}(T+k)^{(1)}, \ldots, \mathbf{z}(T+k)^{(M)}\right\}$ from $p\,(\mathbf{z}(T+k)|\mathbf{Z})$ obtained by sampling $\mathbf{z}(T+k)^{(j)} \sim p\left(\mathbf{z}(T+k)|\boldsymbol{\alpha}(T+k)^{(j)}, \mathbf{H}^{(j)}, \mathbf{\Theta}^{(j)}\right)$, for $j = 1, \ldots, M$.

Temporal forecasts: "simple"

Thus, **the fit** of the variable of interest Y can be obtained as

$$\hat{\mathbf{y}}(t) = \mathbf{H}_y \hat{\mathbf{g}}(t)$$

where

$$\hat{\mathbf{g}}(t) = E\left[\mathbf{g}(t)|\cdot\right] = \sum_{i=1}^{p} \hat{\mathbf{B}}_i \mathbf{g}(t-i) + \sum_{j=0}^{q} \hat{\mathbf{C}}_j \mathbf{f}(t-j)$$

When **temporal predictions** are required, then the transition equation of the state-space specification provides the basis to predict future values of $\mathbf{g}(t)$ - and $\mathbf{f}(t)$; thus, the temporal forecasts of Y can be obtained as

$$\mathbf{y}(t|T) = \mathbf{H}_y \hat{\mathbf{g}}(t|T), \quad (t > T)$$

Unconditional Temporal forecasts

If the path of the exogenous variables X is unknown at $t > T$, then forecast of the corresponding latent variable $\mathbf{f}(t)$ can be obtained through its VAR(s) specification:

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

where $s \leq p$ and $\mathbf{u}(t)$ is white noise. In this case, the joint generation process for $\mathbf{g}(t)$ and $\mathbf{f}(t)$ as

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

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

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 $j > s$.

Conditional Temporal forecasts: point prediction

The forecaster may occasionally know the values of the exogenous variable.

In this case, temporal forecasts of $\mathbf{g}(t)$ **conditional** on a specific "future" (with respect to Y) path of $\mathbf{f}(t)$ obtained as

$$\mathbf{f}(t) = \mathbf{H}_x^{\dagger}\mathbf{x}(t), \quad (t > T)$$

where $\mathbf{H}_x^{\dagger}$ is, for example, the generalized inverse of $\mathbf{H}_x$.

NB: when $n_x > n_y$, it may be of interest to interpolate Y on the sites in which X is available. In this case, we may set $\left[\mathbf{D}^{(h_y o)\prime}, \mathbf{D}^{(h_y u)\prime}\right]' = \mathbf{H}_x$ and try to exploit the spatial structure of the predictor to obtain conditional interpolations.

Model Selection

Important issue: selection of m and r, the number of common factors, and of the orders, p, q and s, of the autoregressive components.

Several Bayesian selection methods have been proposed

Here, we consider three simple approaches: AIC, BIC and PMCC (Laud and Ibrahim, 1995; Gelfand and Ghosh, 1998; Sahu and Mardia, 2005a)

PMCC stands for minimization of the predictive model choice statistics

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

summation is taken over all the ($n_y \times T$ - $\#missing$) observations

$Y(\mathbf{s}, t)_{rep}$ is a replication of $Y(\mathbf{s}, t)$ under the model assumed.

Real Data Examples

The **first data set** consists of daily mean concentrations of PM_{10} and NO_x variables observed in the Lombardy region (Italy).

The **second data set** represents hourly measurements of O_3 and temperature variables observed over Mexico City.

In these examples, the interest is in temporal and spatial predictions of PM_{10} and O_3 variables.

Modelling PM_{10} and NO_x in Lombardy Region

Space: 20 sites where both variables are jointly available
Time: daily data from January-October 2008.
Source: Environmental Agency (ARPA) of Lombardy Region.

Latitudes and longitudes expressed according to UTM coordinates, measured in kilometres.
Data transformed to operate on a logarithmic scale.

Performance test: last week and the temporal series of two monitoring sites (Arese and Meda) excluded from estimation
Two cases considered for interpolation and forecasting comparisons:

i) *conditional predictions* : the NO_x variable is available;
ii) *unconditional predictions*: the NO_x variable is not available.

Modelling PM_{10} and NO_x: exploratory analysis

Time series corresponding to the 18 sites are highly correlated.
Some features of the data suggest it will be difficult to predict all the data satisfactorily.
Example: in summer, some close-by sites show small (~ 0.2) correlations (between the corresponding time series).

Spatial maps of X and $Y \Rightarrow$ spatial mean can be a spatial quadratic surface.

$\mathbf{D}^{(h_x)}$ and $\mathbf{D}^{(h_y)}$ are functions of the site coordinates and define a six-parameter quadratic trend surface (Cressie, 1993).

However, for the Y variable, we also considered setting $\mathbf{D}^{(h_y)} = \mathbf{H}_x$.

Modelling PM_{10} and NO_x: model comparison

Different values of p, q, s, m, r tested $(1 \leq p, q, s \leq 2$ and $1 \leq m, r \leq 12)$.

Competing models compared using PMCC, AIC and BIC.

PMCC criterion suggests: $\mathbf{D}^{(h_y)} = \mathbf{H}_x$, $m = 9$, $r = 8$ and $p = q = 1$. Thus latent factor $\mathbf{f}(t)$ follows a VAR(1) while $\mathbf{g}(t)$ follow a VARX(1, 1).

The residuals from the fit do not show any temporal correlation, also confirmed by Ljung-Box statistic.

MCMC algorithm: 75,000 iterations, posterior inference based on last 50,000 draws at every 10th iteration.

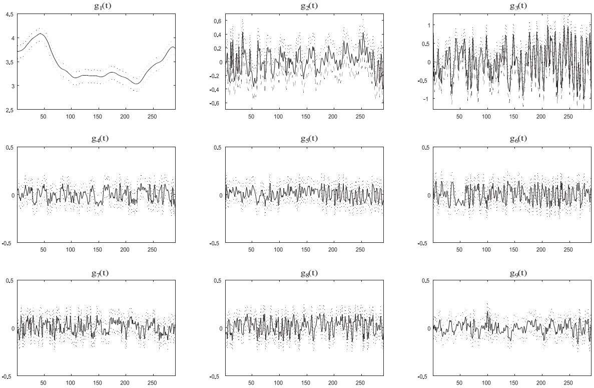
Modelling PM_{10} and NO_x: benchmark comparison

Two simpler models:
M1: common observation variance for all the stations
$(\mathbf{\Sigma}_{u_x} = \sigma^2_{u_x}\mathbf{I};\ \mathbf{\Sigma}_{u_y} = \sigma^2_{u_y}\mathbf{I})$.
M2: constant mean for the Gaussian random field factor loadings
$(\mathbf{m}_j^{(h\cdot)} = \beta_j^{(h\cdot)}\mathbf{1})$.

PMCC, AIC and BIC prefers our model:
Differences between our model and M1: AIC $= 311.34$ and BIC $= 64.72$
Differences between our model and M2: AIC $= 779.21$ and BIC $= 32.11$

Modelling PM_{10} and NO_x: factors

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



Modelling PM_{10} and NO_x: factors and their loadings

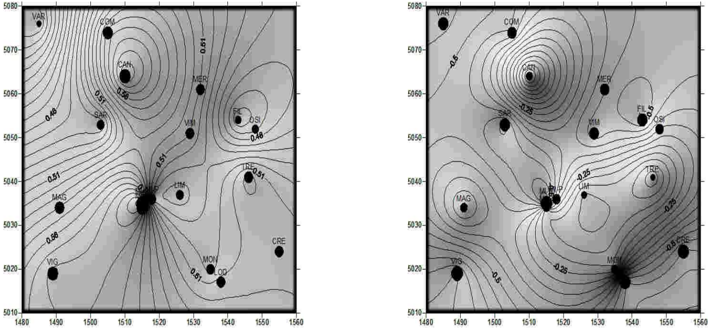
First common factor $g_1(t)$ accounts for 38% of the variability.
It shows a non-stationary process; represents large-scale variability.

The remaining factors are zero mean temporally stationary processes of more limited variation.
The third factor $g_3(t)$ accounts for 15% of the variability (2nd most important component).

Common factors explain around 71% of the total variability.

Ranges for correlation functions:
1st factor for $PM_{10} = 63$ kilometers
Smaller for other factors of PM_{10}
Vary from 31 kilometers to 58 kilometers for factors of NO_x.

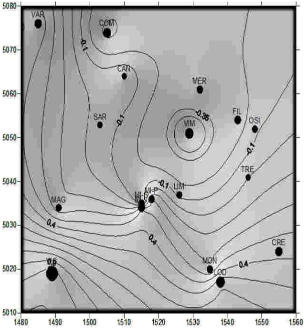
Modelling PM_{10} and NO_x: factor loadings



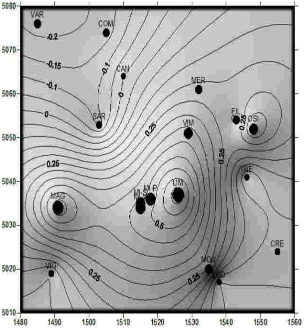
(a) (b)

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

Modelling PM_{10} and NO_x: factor loadings



(a)



(b)

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

Modelling PM_{10} and NO_x: factor loadings

The weights of the 1st factor loading are very similar to each other. It represents the grand mean and accounts for the global time-trend variability of all the series.

The loadings of the 2nd factor do not show any specific pattern.

The 3rd factor loading is a contrast between two groups (9 sites each): one mainly located in the north-east.

The 4th factor loading shows higher values in the city of Milan and represents a contrast between the group of 4 sites in the north-western part, and the remaining 14 sites.

Modelling PM_{10} and NO_x: Unconditional Forecasts

Unconditional Forecasting results - NO_x is not available at sites in S^u - are shown for six of the 18 observed series.

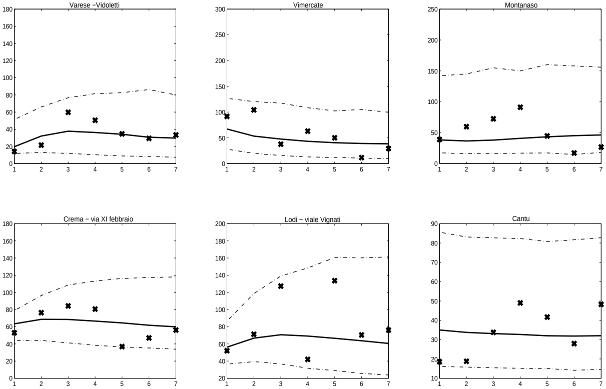


Figure: Forecasts (continuous line) and true data (x) at six selected sites; the 95% credible interval limits are represented by dashed lines.

Modelling PM_{10} and NO_x: Conditional Forecasts

Conditional Forecasting results - NO_x is available at sites in S^u - are shown for the same six series.

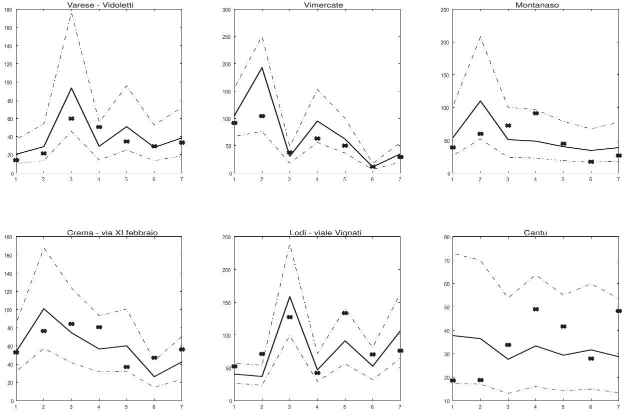


Figure: Conditional forecasts (continuous line) and true data (x) at six selected

Modelling PM_{10} and NO_x: Unconditional Interpolations

Unconditional interpolation results - NO_x is not available at sites in S^u - at Arese (top panel) and Meda (bottom panel).

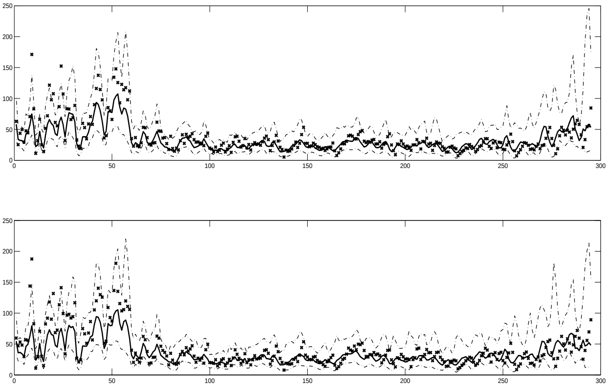


Figure: Unconditional interpolated values (continuous line) and true data (x). The 95% credible interval limits are represented by dashed lines.

Modelling PM_{10} and NO_x: Conditional Interpolations

Conditional interpolation results - NO_x is available at Arese (top panel) and Meda (bottom panel).

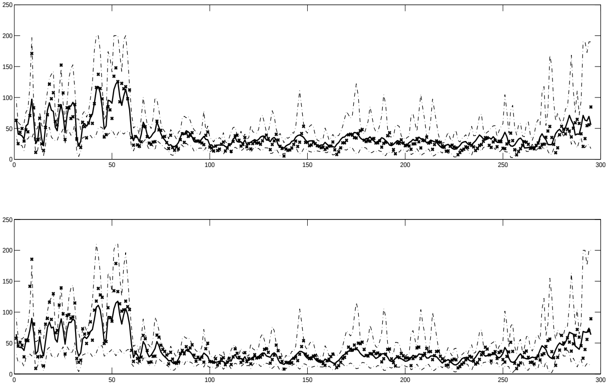


Figure: Conditional interpolated values (continuous line) and true data (x). The 95% credible interval limits are represented by dashed lines.

Modelling PM_{10} and NO_x: Interpolation measures summary

Interpolation	RMSE	MAE	CP 95% interval	AIW 95% int
Conditional	15.740	10.501	0.958	70.924
Unconditional	17.218	12.962	0.951	92.108

The summation is taken over all the $(n_u \times T)$ observations.

Coverage probabilities of the 95% intervals are close to the nominal rate for the two cases.

Conditional interpolation: better results (smaller RMSE, MAE **and** AIW)

Modelling Ozone Levels at Mexico City

Space: 22 sites for O_3 and 16 sites for air temperature (14 joint sites).
Time: hourly data from 06 February 2009 2pm to 12 February 2009 5pm.
Source: Huerta et al. (2004)

O_3 data transformed to operate on a square root scale.

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.

Exploratory analysis of the cyclical behaviour: presence of a peak with wavelengths of 24 h (daily cycle) and a peak with a wavelength of 12 h.

Modelling Ozone Levels at Mexico City

Forecasting: the last 48 hours have been excluded from the estimation procedure and used only for prediction purposes.

Interpolation: the time series of two monitoring sites for O_3 have also been excluded from the analysis.

We have $n_y = 20,\ n_x = 16,\ T = 100$; $n_u = 2$, and a forecast period $T_{48} = \{T+1, \ldots, T+48\}$.

Model specification: PMCC $\Rightarrow$ $m = 7$ and $r = 6$.

1st two components for both variables represent the harmonic cycles with a wavelength of 24 h and 12 h.
The remaining 4 components in $\mathbf{f}(t)$ follow a random walk.
The remaining 5 components in $\mathbf{g}(t)$ follow an ADL$(1,1)$.

Modelling Ozone Levels at Mexico City

The posterior means of the effective ranges for the first two factor loadings, $\mathbf{h}_{y1}$ and $\mathbf{h}_{y2}$, corresponding to the harmonic cycle of 24 h and 12 h, are 33.4 and 31.2 kilometers since the covariograms decay to 0.05 for $\hat{\phi}_{y1} = 11.15$ and $\hat{\phi}_{y2} = 10.42$, respectively.

Similar findings have also been found for the temperature; in fact 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).

Modelling O_3 at Mexico City: Interpolation measures summary

Interpolation	RMSE	MAE	CP 95% interval	AIW 95% interval
Conditional	10.191	6.872	0.952	27.172
Unconditional	12.016	8.542	0.946	36.722

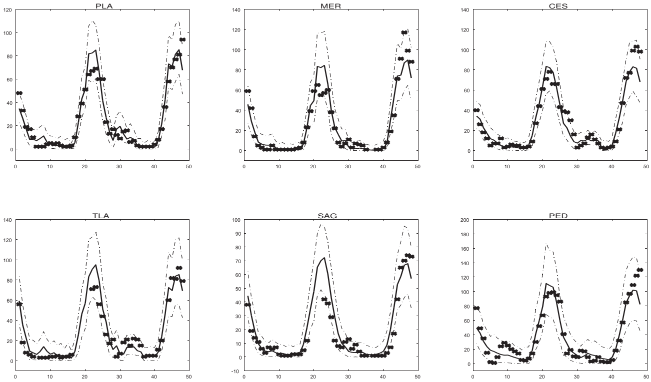
The summation is taken over all the $(n_u \times T)$ observations.

Coverage probabilities of the 95% intervals are close to the nominal rate for the two cases.

Conditional interpolation: better results (smaller RMSE, MAE **and** AIW)

Modelling O_3 at Mexico City: Conditional Forecasts

Conditional Forecasting results - Temperature is available at sites in S^u - are shown for a selection of the three best (top row) and worst (bottom row) forecasts.



Conditional interpolation results - Temperature is available at Tláhuac (TAH) (top panel) and Santa Úrsula (SUR) (bottom panel).

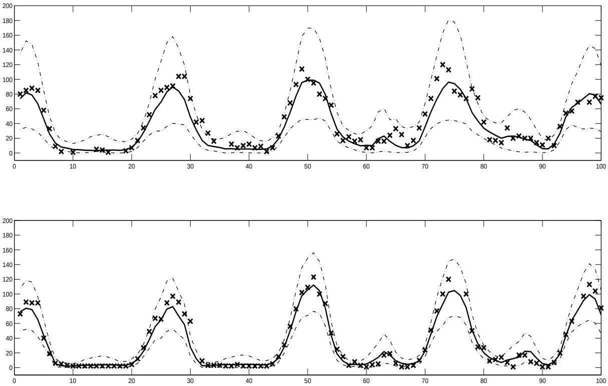
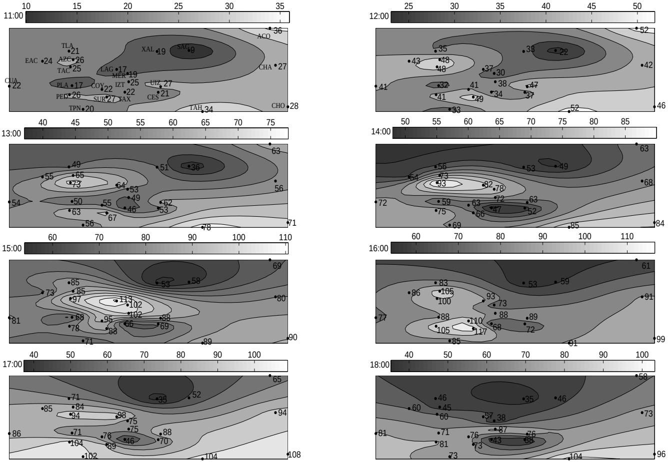


Figure: Conditional interpolated values (continuous line) and true data (x). The 95% credible interval limits are represented by dashed lines.

Modelling O_3 at Mexico City: Interpolation on a (40×30) grid

Reference period: February 7th 2009, from 11 a.m. to 6 p.m.



Discussion and ... further work

1. we have presented a flexible Bayesian spatio-temporal model which is useful when coupled variables are available;
2. the state-space model formulation allows for temporal, spatial and spatio-temporal predictions;
3. it provides for a wealth of explanation through factors, their loadings, their evolutions, etc;
4. the model is particularly useful when the predictor variable is "richer" (both in space and in time) than the response;
5. reversible jump MCMC can be used for estimating the number of factor loadings
6. model extensions: 1) multivariate case, 2) exponential family.

Thank you!

dani@im.ufrj.br

dme.ufrj.br/dani