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.
Let $Y(s; t)$ be a spatio-temporal process of interest, where $s \in D$, with $D$ some spatial domain in $\mathbb{R}^d$ (usually $d = 2$) and $t \in \{1, 2, \ldots\}$ a discrete set of times;

For any given time $t = t_0$, we assume that $Y(\cdot; t_0)$ is a $n_y$-dimensional (stationary) spatial stochastic process.

If available, we denote the $n_x$-dimensional vector $X(\cdot; t)$ as a predictor of $Y(\cdot; t)$

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

**It is explicitly assumed that \( X \) can be used as a predictor of \( Y \).**

The model is specified hierarchically:

- Measurements \( [Y, X] \) are described through latent factors \( [f, g] \).

- \( [f, g] \) describe the dependence in time and between coupled variables through hyperparameters.

- Hyperparameters specify the spatial dependence.
Consider the measurement equation for the variable of interest \( Y(t) \),

\[
y(t) = H_y g(t) + u_y(t)
\]

For any site \( s \), the mean level of the space-time Gaussian process is

\[
E [ Y(s, t) ] = \sum_{i=1}^{m} h_{yi}(s)g_i(t)
\]

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

\( u_y(t) \) is a zero mean Gaussian process

Same holds for \( X \) with \((H_y, g, u_y, m)\) replaced by \((H_x, f, u_x, r)\).

**Main point:** dimension reduction for \( Y \) and \( X \) \((m \ll n_y \text{ and } r \ll n_x)\).
The model for the temporal dynamic of the latent variables is

\[ g(t) = \sum_{i=1}^{p} B_i g(t - i) + \sum_{j=0}^{q} C_j f(t - j) + \xi(t) \]

\[ f(t) = \sum_{k=1}^{s} R_k f(t - k) + \eta(t) \]

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

\[ g(t) = \sum_{i=1}^{p} B_i g(t - i) + \sum_{j=0}^{q} C_j f(t - j) + \xi(t) \]  
(1)

\[ f(t) = \sum_{k=1}^{s} R_k f(t - k) + \eta(t) \]  
(2)

\[ x(t) = m_x(t) + H_x f(t) + u_x(t) \]  
(3)

\[ y(t) = m_y(t) + H_y g(t) + u_y(t) \]  
(4)

Specification is completed with priors for hyperparameter \((\Theta, H_y, H_x)\).

The model can be written in state-space form:

\[ z(t) = H\alpha(t) + u(t) \]
\[ \alpha(t) = \Phi\alpha(t - 1) + \zeta(t) \]

\[ z(t) = (y(t), x(t)), \alpha(t) = (d(t), ..., d(t - p + 1)) \] and \(d(t) = (g(t), f(t))\).
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 \( H_y \) is \( h_{yj} = (h_{yj}(s_1), \ldots, h_{yj}(s_{n_y}))' \), \( j = 1, \ldots, m \)

   \( h_{yj} \)'s are conditionally independent Gaussian Random Fields (GRF)

   \[
   h_{yj} \sim MVN \left( m_j^{(h_y)}, \Sigma_j^{(h_y)} \right)
   \]

   \( m_j^{(h_y)} \) is a \( n_y \)-dimensional mean vector and

   \( \Sigma_j^{(h_y)} \) is a parameterized spatial covariance matrix.

   Similar reasoning used for \( h_{xj} \)'s - columns of \( H_x \).
1. \( H_y \) and \( H_x \) cannot be specified ahead of time and must be considered as a parameter with prior distribution

\[
h_{yj} \sim MVN \left( m_{j}^{(h_y)}, \Sigma_{j}^{(h_y)} \right)
\]

2. Many specifications for the mean level \( m_{j}^{(h_y)} \) can be entertained:
   - constant mean level: \( m_{j}^{(h_y)} = \mathbf{1}_{n_y} \beta_{j}^{(h_y)} \)
   - regression model: \( m_{j}^{(h_y)} = D^{(h_y)} \beta_{j}^{(h_y)} \), \( D^{(h_y)} \) - design matrix

Specifications considered in our work are:
a) \( D^{(h_y)} \) is the design matrix of a surface regression model (i.e. its entries are expressed as function of site’s spatial coordinates);
b) \( D^{(h_y)} = H_x \), useful to link the spatial structures of \( Y \) and \( X \).
Space-Time Modelling of Coupled Spatio-Temporal Environmental Variables
By letting $\Sigma_j^{(h_y)} = 0$, the spatial pattern $h_{yj}$ 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 $|s| \to \infty$.

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).

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

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

In this case, according to the probabilistic corollary of Mercer’s theorem (Obled, Creutin; 1986), the measurement equations (3) and (4) are obtained from the Karhunen-Loéve expansion (KLE).
For simplicity, conditionally conjugate prior distributions are used for all parameters defining the dynamic factor model. To summarize:

- Inverse Gamma are used for model variances (e.g. the state and measurement noise, partial sill etc);

- The Normal distribution truncated to the interval \([a,b]\) is used for the autoregressive part of the state equation;

- The Normal distribution is used for the mean vector of the spatial patterns \(h_{yi}(s), i = 1, \ldots, m\);

- For the **range parameter** we define: \(IG(2,b)\), where \(b = d_0/(-2ln(0.05))\) and \(d_0\) max distance among the spatial sites;
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 → 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.
The interpolation of the spatial patterns on a new site $s_0$ 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 $s_0$

$$
\hat{y}(s_0, t) = \hat{h}_y(s_0)' \hat{g}(t)
= \sum_{i=1}^{m} \hat{h}_{yi}(s_0) \hat{g}_i(t)
$$

where $\hat{h}_{yi}(s_0)$ are the predicted spatial patterns at site $s_0$.

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

$y^o$ - observed vector at locations $\{s_1, \ldots, s_n\}$,
$y^u$ - vector of measurements to be predicted at new locations $S^u = \{s_{n+1}, \ldots, s_{n+n_u}\} \subset S$.

**Main task**: obtain predictive distribution $p(y^u|y^o)$

Task is achieved by noticing that

$$p(y^u|y^o) = \int p(y^u|H^u_y, \Theta)p(H^u_y, \Theta|y^o)dH^u_yd\Theta$$

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

First step requires the posterior distribution of \((H_y^u, \Theta)\)

\[
p(H_y^u, \Theta | y^o) = \int p(H_y^u | H_y^o, \Theta) p(H_y^o, \Theta | y^o) \, dH_y^o \, d\Theta
\]

where

\[
p(H_y^u | H_y^o, \Theta) = \prod_{j=1}^m p(h_{yj}^u | h_{yj}^o, m_j^{(h_y)}, \tau_{y,j}^2, \phi_{y,j}).
\]

Posterior calculations give \(p(H_y^o, \Theta | y^o)\) and samples \(\{(H_y^{o(l)}, \Theta^{(l)})\}\).

Standard multivariate normal results give \(p(h_{yj}^u | h_{yj}^o, m_j^{(h_y)}, \tau_{y,j}^2, \phi_{y,j})\) and samples \(\{H_y^{u(l)}\}\) - Bayesian kriging

Second step uses the measurement distribution \(p(y^u | H_y, \Theta)\) to give samples \(\{y^{u(l)}\}\)
Bayesian Prediction of $Y$

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

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

**Main task:** obtain predictive distribution $p(z(T + k)|Z)$

Task is achieved by noticing that

$$p(z(T + k)|Z) = \int p(z(T + k)|\alpha(T + k), H, \Theta) p(\alpha(T + k)|\alpha(T), H, \Theta) \times p(\alpha(T), H, \Theta|Z) d\alpha(T + k) d\alpha(T) dH d\Theta$$

$$(z(T + k)|\alpha(T + k), H, \Theta) \sim N(H\alpha(T + k), \Sigma_u),$$

$$(\alpha(T + k)|\alpha(T), H, \Theta) \sim N(\mu_k, V_k),$$

$$\mu_k = \Phi^k \alpha(T) \text{ and } V_k = \sum_{j=1}^{k} \Phi^{j-1} \Psi(\Phi^{j-1})', \text{ for } k > 0, \text{ with } \Phi^0 = I.$$
Bayesian Forecasts of $Y$

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

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

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

is a Monte Carlo approximation to $p(z(T + k)|Z)$.

A sample $\{z(T + k)^{(1)}, \ldots, z(T + k)^{(M)}\}$ from $p(z(T + k)|Z)$ obtained by sampling $z(T + k)^{(j)} \sim p(z(T + k)|\alpha(T + k)^{(j)}, H^{(j)}, \Theta^{(j)})$, for $j = 1, \ldots, M$. 
Thus, the fit of the variable of interest $Y$ can be obtained as

$$\hat{y}(t) = H_y \hat{g}(t)$$

where

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

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

$$y(t|T) = H_y \hat{g}(t|T), \quad (t > T)$$
Unconditional Temporal forecasts

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

$$f(t) = \sum_{k=1}^{s} R_k f(t-k) + \eta(t)$$

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

$$\begin{bmatrix}
I_m & -C_0 \\
0 & I_r
\end{bmatrix}
\begin{bmatrix}
g(t) \\
f(t)
\end{bmatrix}
= 
\begin{bmatrix}
B_1 & C_1 \\
0 & R_1
\end{bmatrix}
\begin{bmatrix}
g(t-1) \\
f(t-1)
\end{bmatrix}
+ \cdots + 
\begin{bmatrix}
B_p & C_p \\
0 & R_p
\end{bmatrix}
\begin{bmatrix}
g(t-p) \\
f(t-p)
\end{bmatrix}
+ 
\begin{bmatrix}
\xi(t) \\
\eta(t)
\end{bmatrix}
$$

(6)

where it is assumed without loss of generality that $p \geq \max(s, q)$, $C_i = 0$ for $i > q$ and $R_j = 0$ 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 \( g(t) \) conditional on a specific "future" (with respect to \( Y \)) path of \( f(t) \) obtained as

\[
f(t) = H_x^\dagger x(t), \quad (t > T)
\]

where \( H_x^\dagger \) is, for example, the generalized inverse of \( H_x \).

**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 \([D^{(ho)}', D^{(hu)}']' = H_x\) and try to exploit the spatial structure of the predictor to obtain conditional interpolations.
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.

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

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

$Y(s, t)_{rep}$ is a replication of $Y(s, t)$ under the model assumed.
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 ($\sim 0.2$) correlations (between the corresponding time series).

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

$D^{(h_x)}$ and $D^{(h_y)}$ 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 $D^{(h_y)} = H_x$. 
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: $D^{(hy)} = H_x$, $m = 9$, $r = 8$ and $p = q = 1$. Thus latent factor $f(t)$ follows a VAR(1) while $g(t)$ 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.
Two simpler models:
M1: common observation variance for all the stations
\[(\Sigma_{u_x} = \sigma_{u_x}^2 I; \Sigma_{u_y} = \sigma_{u_y}^2 I)\].
M2: constant mean for the Gaussian random field factor loadings
\[(m_j^{(h \cdot)} = \beta_j^{(h \cdot)} I)\].

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

**Figure:** Factor loadings interpolation: the contour values represent the range of the posterior means for $h_{y1}$ (a), $h_{y2}$ (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

Figure: Factor loadings interpolation: the contour values represent the range of the posterior means for $h_{y3}$ (c) and $h_{y4}$ (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 sites; the 95% credible interval limits are represented by dashed lines.
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 ($\times$). 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

<table>
<thead>
<tr>
<th>Interpolation</th>
<th>RMSE</th>
<th>MAE</th>
<th>CP 95% interval</th>
<th>AIW 95% int</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional</td>
<td>15.740</td>
<td>10.501</td>
<td>0.958</td>
<td>70.924</td>
</tr>
<tr>
<td>Unconditional</td>
<td>17.218</td>
<td>12.962</td>
<td>0.951</td>
<td>92.108</td>
</tr>
</tbody>
</table>

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.
**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 $f(t)$ follow a random walk. 

The remaining 5 components in $g(t)$ follow an ADL(1, 1).
The posterior means of the effective ranges for the first two factor loadings, \( h_{y1} \) and \( h_{y2} \), 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

<table>
<thead>
<tr>
<th>Interpolation</th>
<th>RMSE</th>
<th>MAE</th>
<th>CP 95% interval</th>
<th>AIW 95% interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional</td>
<td>10.191</td>
<td>6.872</td>
<td>0.952</td>
<td>27.172</td>
</tr>
<tr>
<td>Unconditional</td>
<td>12.016</td>
<td>8.542</td>
<td>0.946</td>
<td>36.722</td>
</tr>
</tbody>
</table>

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.

Figure: Conditional forecasts (continuous line) and true data (x) at six selected sites.
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 (×). 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.
we have presented a flexible Bayesian spatio-temporal model which is useful when coupled variables are available;

the state-space model formulation allows for temporal, spatial and spatio-temporal predictions;

it provides for a wealth of explanation through factors, their loadings, their evolutions, etc;

the model is particularly useful when the predictor variable is "richer" (both in space and in time) than the response;

reversible jump MCMC can be used for estimating the number of factor loadings

model extensions: 1) multivariate case, 2) exponential family.
Thank you!

dani@im.ufrj.br

dme.ufrj.br/dani