
**Università degli studi di Firenze**

Dipartimento di Statistica G. Parenti

**Dottorato di Ricerca in Statistica Applicata**
**XXI ciclo**

# Spatio-temporal modeling of environmental processes: nitrogen oxides concentrations in the Tuscany region

**Valentina Lapolla**

Tutor: **Prof. Alessandra Petrucci**

Co-tutor: **Prof. Chris Wikle**

Co-tutor: **Dr. Franco Giovannini**

Coordinatore: **Prof. Guido Ferrari**

Settore disciplinare SECS-S/01

# Contents

# Introduction

This thesis is about spatiotemporal modeling suited for environmental applications and in particular for air pollution problems.
Environmental variables have an intrinsic spatial and/or temporal nature, being realizations of an underlying spatiotemporal process. Indeed, if we consider air pollutant concentrations, these can be seen as a spatial field evolving in time. From measurement systems we only know the value of this field in a discrete set of points in space and on time average: for instance the network of air pollution monitoring stations measures the concentrations of several pollutants in a set of locations, with hourly or daily time step. However, what is crucial to know for environmental and health policies are the spatial features in the whole domain of interest, checking space and time trends and correlation structure. Inferring what happens in space and time over a region of interest starting from the knowledge on a limited set of observations is thus a significant task for public organizations and authorities dealing with environmental problems.
This work is done in collaboration with the Environmental Protection Agency-Tuscany region (ARPAT), that has the role of monitoring and assessing the quality of the environment, to check and verify the main sources of pollutants, and to estimate the risks for environment and public health. This collaboration has focused on many issues, from data collection to cooperation with air pollution experts to build a modeling framework suited for the particular characteristic of the process of interest, that could be useful in real applications for exposure and health assessment, regulatory activity in the management and planning of air emissions, policies on preservation and recovery of the environment.
According to ARPAT we focused on the case of nitrogen oxides. Nitrogen dioxide is toxic by inhalation and there is evidence that long-term exposure to $NO_2$ at high concentrations has adverse health effects. $NO_2$ and other nitrogen oxides are also precursor of ozone and particulate matter, whose effects on human health are well documented.
We deal with both the spatial bivariate relationship between nitrogen monoxide and nitrogen dioxide and the spatiotemporal evolution of nitrogen oxides as a whole.

Environmental processes are generally very complex: they include very intricate spatiotemporal processes, occurring on a wide variety of scales and arising from the interactions of many subprocesses involving physical, biological and chemical mechanisms. In these applications very large and heterogeneous datasets are often involved, from automatic monitoring networks to satellite data, with a series of a priori informations and knowledge about the process dynamics. Thus it is important to adopt a modeling strategy that could account for all these sources of information in a coherent manner while at the same time keeping computational tractability.
Many efforts have been made in the study of air pollution, both in a deterministic and in statistical framework, but the problem of determining the concentrations of a pollutant in a given point in space and time is someway still open.
From a deterministic point of view, the governing equations are theoretically known: starting from these relationships a lot of different algorithms and models have been developed with different modeling hypothesis and empirical relationships. Besides the limits given by modeling assumptions, the needed input informations, emissions, meteorological and boundary conditions etc, are often hard, even if not impossible, to be known with the required accuracy. Moreover it is also hard to establish how the uncertainty on input data is reflected on the estimated concentrations. In this kind of modeling approach measured concentrations are only used as a calibration tool. Otherwise the statistical approach uses measured concentrations to make inference about the concentrations in an arbitrary unobserved point in space and time, by using other informations, emissions, meteorological conditions etc, as potential covariates.
In statistical modeling of air pollution problems and similar environmental systems we have to face many different challenges.
First of all the complexity of the problem is such that it is very difficult to think the model in terms of joint distribution, specifying the appropriate multivariate spatiotemporal covariance structures. A hierarchical approach could simplify the specifications of the model by factorizing the joint distributions into a series of conditional models, linked together in a probabilistically valid framework. Thus complicated dependency could be mitigated by conditioning. Although either a classical or Bayesian perspective can be adopted for hierarchical models, the Bayesian approach is usually needed for more complex models. Here we refer to the hierarchical Bayesian context.
Dimensionality is another very important issue when dealing with spatiotemporal models. In fact it is often the case that the vectors of data representing the process are of very high dimension, thus the numerical implementation is hard and the computational aspects often challenging, and a reduced dimensional set up is often needed.
Another very important question is how to account in the model for the knowledge we have about the process, for example arising from empirical

relationships and physical laws. In the case of dispersion of pollutants in air we know the dynamics driving the process and connecting concentration values to emissions and meteorological conditions. The challenge here is to find a modeling approach that includes this dynamics in its structure and that allows for statistical inference and tractability.
Modeling techniques dealing with spatiotemporal processes where the spatial and temporal structures can be modeled separately (separable models) and/or the spatial temporal structure does not change with location and time (stationary) are well developed and studied in literature. Environmental processes are often neither separable nor stationary, and these assumptions could turn to be very weak in this field.
Moreover multivariate interactions could add more complexity to this framework and appropriate modeling strategies are needed in this case as well.
Finally monitoring locations are chosen for regulatory and health protection reasons, thus the most of these are located in the main cities, near heavy traffic roads, with a very non homogeneous distribution. Measured points chosen following these criteria are far from an optimal choice for inferential goals.
Although in relative recent literature many efforts have been made to deal with these issues we are far from having a clear and complete reference framework.

In this thesis we proceed as follows. The first chapter is about the theory of spatial modeling, with a focus on the review of recent developed methods to deal with nonstationarity and high dimensionality. In the second chapter the multivariate issue is introduced and an application of the described methods to the bivariate case of nitrogen oxides and nitrogen dioxides in Tuscany is developed. The third chapter is about spatiotemporal modeling: here we introduce dynamical systems and reduced dimension spatiotemporal models. In this chapter the question about the underlying physical process is discussed and a modeling framework able to include the governing differential equations and to account for the role of emissions and meteorological variables is presented. Here we also review a number of statistical models suited to atmospheric pollution modeling. Finally the last chapter is about the application of the described spatiotemporal modeling techniques to the case of nitrogen oxides concentrations in the Tuscany region. We propose a hierarchical Bayesian dynamical model to estimate concentrations at unobserved space and time points. The model introduced can account for a nonstationary and nonseparable structure: it is proposed in a complete and temporal reduced version, while dimension reduction in space is achieved by using a kernel convolution approach. Different kernels are evaluated, comparison of competitive models is performed within a posterior predictive approach and predictive ability of the selected model is then checked.

# Chapter 1

# Spatial modeling for point referenced data

We consider a stochastic process $Y(s) : s \in D$, where $s$ varies continuously over $D$, a fixed subset of the d-dimensional Euclidean space. This case is referred as *geostatistical*, a field that historically has been developed in relation to geological sciences, different from other spatial processes like areal data (where the spatial index is finite, with the domain being partitioned in a finite subset) or point pattern data (where the location itself is the random variable of interest). A bidimensional ($d = 2$) domain is commonly assumed for spatial processes, while the one dimensional setting is used in time series literature[1].

We are dealing with a stochastic process indexed by $s$ and we have to specify the joint distribution of an uncountable number of random variables: in practice we define a *valid* joint distribution between an arbitrary set of finite variables.
Assuming the process to have a finite mean and its variance to exist for all $s \in D$, then the process is said to be *strictly stationary* if, for any set of $n$ locations (with $n > 1$) and any vector $h \in R^d$, the distribution of $[Y(\mathbf{s_1}), \ldots, Y(\mathbf{s_n})]$ and that of $[Y(\mathbf{s_1} + \mathbf{h}), \ldots, Y(\mathbf{s_n} + \mathbf{h})]$ is the same.
If $Y(\mathbf{s})$ has finite second moments, constant mean, and the covariance function of $Y$ only depends on the distance between the two points, that is $\mu(s) = \mu$ and $Cov(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$ for all $h \in R^d$, then it is said to be weakly stationary. If this property holds the covariance function $C(\mathbf{h})$ describes the covariance properties of the process evaluated at any different

[1] It is worth noting that the main difference between the two approaches is given by the fact that, while for time we have a natural ordering, for spatial information this no longer holds and thus the concept of asymptoticity is different. In fact in time series context asymptoticity is reached allowing the time to go to infinity, while in spatial context the domain is kept fixed while the increasing number of observations fills the domain.

couple of locations.
Another type of stationarity (called *intrinsic*) assumes that the variance of the variable defined as the difference between the process at $\mathbf{s}+\mathbf{h}$ and the process at $\mathbf{s}$, $[Var(Y(\mathbf{s}+\mathbf{h})-Y(\mathbf{s}))]$ depends only on the separation vector $\mathbf{h}$.
Then we can define the so called *variogram* as $Var(Y(\mathbf{s}+\mathbf{h})-Y(\mathbf{s})) = 2\gamma(\mathbf{h})$. A valid variogram has to be negative definite. This kind of stationarity does not assume anything about the distribution of $[Y(\mathbf{s_1},\ldots,\mathbf{s_n})]$, focusing only on the first and second moment of the different variables. It can be proven the following relationship holds between the (semi)variogram and the covariance function:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \tag{1.1}$$

Some conditions are needed in order to recover $C$ from $\gamma$. If the process is ergodic (that is in this case $C(\mathbf{h})$ goes to zero when $|h| \to \infty$) and it is weakly stationary we have

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h}) = \lim_{|u|\to\infty}\gamma(u) - \gamma(\mathbf{h}) \tag{1.2}$$

Strong stationarity implies weak stationarity but not the converse and in the same way weak stationarity implies intrinsic stationarity, but the vice-versa is not true.
If the covariance function depends only on the length of the separation vector, and not on its direction, then the process is said *isotropic*.
There is a number of classes of parametric forms for isotropic and intrinsically stationarity processes, including exponential, Gaussian and Matern forms that will be discussed later in detail.
The empirical variogram

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|}\sum_{(s_i,s_j)\in N(t)}[Y(s_i)-Y(s_j)]^2 \tag{1.3}$$

where $N(t)$ is the set of pairs of points $||s_i - s_j|| = t$, with cardinality $|N(t)|$, can be used as an explorative tool for appropriately choosing the best parametric form.

## 1.1 Gaussian processes

An important class of processes are the Gaussian, where $\mathbf{Y} = (Y(s_1),\ldots,Y(s_n))^T$ for any set of $n > 1$ sites, has a multivariate normal distribution. This choice is a convenient one for more reasons than simplicity of derivation of the joint, marginal and conditional distributions once the mean and covariance function have been specified. In fact in this case to ensure validity of the joint

distribution between an arbitrary set of finite variables it is sufficient to ensure validity of the correlation function. Moreover for a Gaussian process weak stationarity assures strong stationarity.

### 1.1.1 Valid covariance functions

Defining a *valid* class of covariance functions play a key role for Gaussian processes, as it assures consistency of the model specification. To be valid a correlation function has to be positive definite in order to ensure that the variance of any linear combination of values of the process at various locations is positive.
A necessary and sufficient condition for a stationary covariance function to be positive definite is given by the Bochner's theorem, that states that a function is continuous and positive definite if and only if it is the characteristic function of a d-dimensional symmetric random variable.
Every stationary process $Y(s)$ on $R^d$ with mean and finite variance can be represented in spectral form as

$$Y(s) = \int_{R^d} e^{i\omega' s} V(d\omega) \tag{1.4}$$

where $V$ is a random measure with independent increments.The associated covariance function $C_Y(s_1, s_2) = C_Y(h)$ (where $h = s_1 - s_2$) of this stationary process $Y$ can be expressed as

$$C(h) = \int_{R^d} e^{i\omega' h} F(d\omega) \tag{1.5}$$

where $F$ is a non-negative symmetric measure on $R^d$ and $E[|V(d\omega)|^2] = F(d\omega)$. If $F$ has density with respect to Lebesgue measure then $F(d\omega) = g(\omega)d\omega$ and this density is called the spectral density $g$, which is the Fourier transform of the covariance function

$$g(\omega) = \frac{1}{2\pi} \int_{R^2} \exp(-i\omega' h) C(h) dh \tag{1.6}$$

This property can be used to generate valid correlation functions and makes possible the estimation of the spatial process in the spectral domain.
Larger class of correlation functions could also be provided by mixing, product and convolutions of valid correlation functions.
Stationary and isotropic correlation functions are the most common choice and there exist several classes of valid isotropic correlation functions. A very popular covariance function is the exponential one

$$C(h) = \sigma^2 \exp(-\phi h) \qquad h > 0 \tag{1.7}$$

where the parameter $\phi$ is associated to the effective *range*, that is the distance at which the spatial correlation reaches approximatively zero. The corresponding variogram is of the form

$$\gamma(h) = \tau^2 + \sigma^2(1 - \exp(-\phi h)) \qquad h > 0 \tag{1.8}$$

where $\tau^2$ represents the *nugget* effect, that is the limit value of the variogram when the distance tends to zero, while the asymptotic value of the variogram (called *sill*) is the nugget plus the $\sigma^2$ parameter. Another important class is represented by the Gaussian covariance function

$$C(h) = \sigma^2 \exp(-\phi^2 h^2) \qquad h > 0 \tag{1.9}$$

Both exponential and Gaussian forms are particular cases of the Matern form. This class of correlation function is one of the most common and powerful in spatial statistics, due to its flexibility and interpretability of the parameters. This covariance has the form

$$C(h) = \sigma^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} K_\nu \left(2\sqrt{\nu}\frac{h}{\rho}\right) \qquad \rho > 0, \nu > 0; \tag{1.10}$$

where $h$ is the distance, $\rho$ is the spatial range parameter, and $K_\nu(\cdot)$ is modified Bessel function of the second kind of order $\nu$, $\nu$ being the differentiability (or smoothness) parameter.
This kind of function first arose in the study of turbulence from the analysis by Von Karman and Tatarskii. Whittle [1] introduced spatial covariance structures based on stochastic partial differential equations, generalizing in two dimension the exponential family, corresponding to a Markov process in one dimension.
Starting from a diffusion process for a stationary variable he found a correlation function corresponding to the $\nu = 1$ case of that we now call the Matern function. This kind of result has been extended to the general classes of stochastic differential equations (parabolic, elliptic and hyperbolic form) by Heine [2], who found corresponding form of covariance function of the process. Matern generalized Whittle's class to derive families in any dimension $d$ (for a complete history of this correlation function see [3]).
Handcock and Stein [4] introduced this family as the Matern family, pointing out the interpretation of the two parameters, one controlling the range of correlation and the other one related to the smoothness of the process. In fact a random field with this covariance have $\nu - 1$ times mean square differentiable paths. When $\nu$ goes to infinite this function becomes the Gaussian form, with realizations infinitely differentiable. When $\nu = 0.5$ this form becomes the exponential one, that is continuous but not differentiable. While the smoothness parameter allows for more flexibility, it is often not easily estimated by the data, but it is possible to define this parameter relating to

theoretical considerations about the process that is being considered.
A $d$ dimensional spectral distribution function of the Matern class is

$$f(\omega) \propto (\alpha^2 + |\omega|_d^2)^{-\nu - d/2} \tag{1.11}$$

where $\alpha^{-1}$ can be thought as the autocorrelation range and $\nu$ is related to the degree of smoothness of the process. The greater is $\nu$, the less the weights of the higher frequencies of the spectrum.

## 1.2 The basic model

The basic geostatistical Gaussian model is of the form

$$Y(s) = \mu(s) + w(s) + \epsilon(s) \tag{1.12}$$

where the process is decomposed in a mean part and two error processes. The mean can be modeled using covariates $\mu(s) = x^T(s)\beta$, while $\epsilon$ represents an uncorrelated error term. The new term $w(s)$ is introduced to capture the residual spatial association and it is assumed to be a realization from a mean zero Gaussian spatial process. In a parametric approach the main issue in this case is to choose a valid and appropriate correlation function, typically considering families of stationary processes.
The (universal) kriging method is a particular case of this basic model. Let $\mathbf{Y}$ be an $n$x1 data vector $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))^T$ and assume for a moment $\mu = 0$, we can write

$$\mathbf{Y} = \mathbf{W} + \epsilon \tag{1.13}$$

with $\mathbf{W} \sim N(0, \Sigma_w)$, and $\epsilon \sim N(0, \Sigma_\epsilon)$. The minimum mean square error spatial prediction can be proven to be the conditional expectation of $Y(s_0)$ given the data, where $s_0$ is the new point. If the parameters are known, from a multivariate normal theory, the conditional distribution of new the point is normal with mean and variance:

$$\mu = \Sigma_{y_0,y}(\Sigma_w + \Sigma_\epsilon)^{-1}\mathbf{Y} \tag{1.14}$$

$$\Sigma = \Sigma_{y_0} - \Sigma_{y_0,y}(\Sigma_w + \Sigma_\epsilon)^{-1}\Sigma'_{y_0,y} \tag{1.15}$$

where $\Sigma_{y_0,y} = cov(\mathbf{Y_0}, \mathbf{Y})$ and $\Sigma_{y0} = \mathbf{Y}$. For example for an isotropic model with $\Sigma_w = \sigma^2 H(\phi)$, where $(H(\phi))_{ij} = \rho(\phi; d_{ij})$ with $\rho$ a valid correlation function and $\Sigma_\epsilon = \tau^2 I$, we have $\Sigma_{y_0} = \sigma^2 + \tau^2$ and $\Sigma_{y_0,y} = (\sigma^2\rho(\phi; d_{01}), \ldots, \sigma^2\rho(\phi; d_{0n}))$.
In a Bayesian setting we have

$$p(\theta|y) \propto f(y|\theta)p(\theta) \tag{1.16}$$

$$\mathbf{Y}|\theta \sim N(0, \sigma^2 H(\phi) + \tau^2 I) \tag{1.17}$$

The required priors on the parameters are often assumed independent

$$p(\theta) = p(\beta)p(\sigma^2)p(\tau^2)p(\phi) \tag{1.18}$$

and the conjugate distributions are multivariate normal for $\beta$ and inverse gamma for $\sigma^2$ and $\tau^2$, while for $\phi$ the choice depends on the class of the chosen correlation function.
The spatial surface $w|y$ is recovered via composition sampling. The predictive distribution can be obtained as

$$\begin{aligned} p(y_0|\mathbf{Y}, X, \mathbf{x}_0) &= \int p(y_0, \theta|\mathbf{y}, X, \mathbf{x}_0)d\theta & (1.19) \\ &= \int p(y_0|\mathbf{y}, \theta, X, \mathbf{x}_0)p(\theta|\mathbf{Y}, X)d\theta & (1.20) \end{aligned}$$

with $p(y_0|\mathbf{y}, \theta, X, \mathbf{x}_0)$ normal and a Gibbs sampling can be easily performed. If we are interested in a joint prediction for $m$ points, we can proceed in the same way as above.

### 1.2.1 Hierarchical formulation

The basic geostatistical model can be arranged in a hierarchical setting. The hierarchical approach is essentially based on the fact that a joint distribution of a set of random variables can be decomposed in some conditional models, that is $[X, Y, Z] = [Z|Y, X][Y|X][X]$. It is often easier to specify this conditional models than the joint ones. A general framework can be formulated as ([5])

[data | process, parameters]

[process | parameters]

[parameters]

and we can obtain the distribution of process and parameters given the data, as

[process,parameters|data] $\propto$ [data|process,par][process|par][par]

In the geostatistical setup we can write:

$$\begin{aligned} Y|W, \theta &\propto N(X\beta + W, \tau^2 I) & (1.21) \\ W|\theta_w &\propto N(0, \Sigma(\theta_w)) & (1.22) \\ \beta &\propto N(\beta_0, \Sigma_\beta) & (1.23) \\ & & [\sigma^2, \theta_w, \beta_0, \Sigma_\beta] & (1.24) \end{aligned}$$

where $\tau^2$ accounts for small scale variability (nugget) and $W$ is the correlated spatial process. Here the correlation function is assumed to be parameterized by some parameters $\theta_w$. This model is the analogous of the linear mixed model in classical modeling, and in the same way it is possible to define a generalized version for non Gaussian data.
In this setting we can perform Montecarlo sampling using the non marginalized form $[\theta|y] \propto [y|\theta, W][W|\theta][\theta]$.

## 1.3 Kernel convolution to generate processes

An alternative method to generate Gaussian processes as those we described in the previous section is by using process convolution. This method has been used to develop a general class of models for spatial data ([6], [7], for a review see [8]).
In fact it is possible to build a Gaussian process over a general spatial region $S$ by convolving a continuous white noise process with a smoothing kernel $k(s)$.
More formally let

$$Y(s) = \int_{R^d} k(s,u)V(du), \quad s \in D \tag{1.25}$$

where $k(\cdot,\cdot)$ is a square integrable kernel function $\int k^2(s,u)du < M < \infty$, and $V$ is a process with independent increments, zero mean and finite variance proportional to the volume of the increment. In particular by choosing $V$ as a $d$ dimensional Brownian motion ([9]), the process $Y(s)$ defined above is Gaussian and an arbitrary collection of $Y(s_i), i = 1, \ldots, n$ has a finite dimensional multivariate normal distribution (for a proof see for example [10]).
Due to the independent increments of $V(.)$ the covariance function of the process $Y$ can be written in terms of kernel functions alone, that is

$$C_Y(s_1, s_2) = \int_{R^d} k(s_1,u)k(s_2,u)du \tag{1.26}$$

where $E(V(du)^2) = du$.
The covariance defined above is valid, in fact

$$\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j C_Y(s_i, s_j) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \int_{R^d} k(s_i,u)k(s_j,u)du \tag{1.27}$$

$$= \int_{R^d} (\sum_{i=1}^{k} a_i k(s_i,u))^2 du \geq 0 \tag{1.28}$$

In the special case of a kernel function in $R^d$ only depending on the distance between locations

$$Y(s) = \int_{R^d} k(s-u)V(du), \quad s \in D \tag{1.29}$$

we have a stationary process and we can write the variogram in terms of the kernel

$$2\gamma(h) = \int_{R^d} (k(u) - k(u-h))^2 du \tag{1.30}$$

where $h = s_1 - s_2$. The covariance function now is

$$C_Y(h) = \int_{R^d} k(u)k(u-h)du \tag{1.31}$$

Moreover in this case, taking the square root of the Fourier transform of the covariance function and obtaining the inverse Fourier transform gives a function proportional to the kernel and this relationship is one to one for isotropic processes. In fact, under regularity conditions that assure the existence of the Fourier transform, we have

$$\hat{C}(\omega) = |\hat{K}(\omega)|^2 \tag{1.32}$$

where $\hat{C}(\omega)$ and $\hat{K}(\omega)$ are the Fourier transform (with respect to Lebesgue measure) of the covariance and kernel function respectively. Thus a kernel corresponding to a particular covariance function can be obtained as

$$k(u) = (2\pi)^{-d} \int_{R^d} \exp^{i\omega' u} \sqrt{\hat{C}(\omega)} d\omega \tag{1.33}$$

A stationary random process can be defined by a convolution process if and only if has spectral density. Although in many situations there is not a closed expression for the covariance kernel, for the (isotropic) Matern class of covariance (and so also for the Gaussian) we have a direct correspondence. For example, for Gaussian univariate covariance we have:

$$C(h) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-h^2/(2\sigma^2)} \tag{1.34}$$

$$k(u) = \frac{1}{\alpha\sqrt{2\pi}} \exp^{-u^2/(2\alpha^2)} \tag{1.35}$$

where $\sigma^2 = 2\alpha^2$.
Bivariate Gaussian kernel is proportional to that of the univariate case (as the bivariate Gaussian kernel can be decomposed as $k(|s|) \propto k(|s_x|)k(|s_y|)$ (see [11]).
A closed form for a kernel corresponding to a Matern covariance function

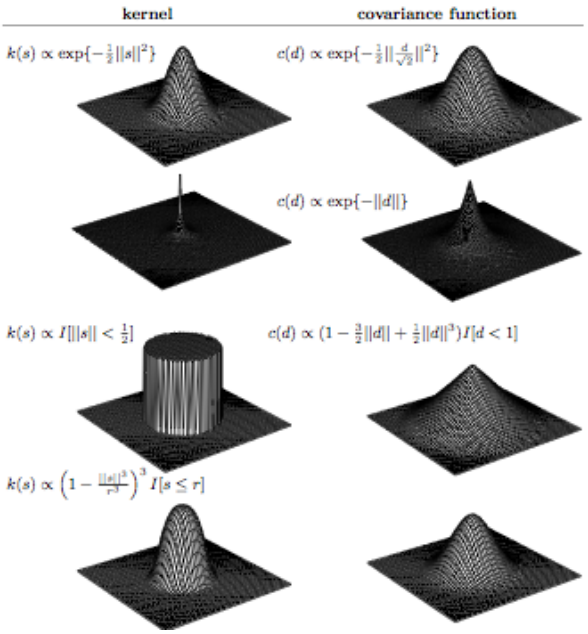


Figure 1.1: Relationship between kernel and induced covariance functions, from [11].

can be derived. Using a different parameterization of the Matern form we have:

$$C_{(\phi,\alpha,\nu)}(h) = \frac{\pi^{d/2}\phi}{2^{\nu-1}\Gamma(\nu+d/2)\alpha^{2\nu}}(\alpha|h|)^{\nu}K_{\nu}(\alpha|h|) \tag{1.36}$$

$$\alpha = 2\nu^{1/2}/\rho \tag{1.37}$$

$$\phi = \frac{\Gamma(\nu+d/2)2^{2\nu}\nu^{\nu}}{\pi^{d/2}\Gamma(\nu)\rho^{2\nu}}\sigma^2 \tag{1.38}$$

and the spectral density can be written as

$$\hat{C}_{\phi,\alpha,\nu}(\omega) = (2\pi)^d\phi(\alpha^2 + |\omega|^2)^{-\nu-d/2} \tag{1.39}$$

thus the corresponding kernel

$$k(u) = (2\pi)^{-d/2}C_{(\phi^{1/2},\alpha,\nu/2-d/4)}(u). \tag{1.40}$$

Figure 1.1 shows correspondence between kernel and covariance functions in different cases (Gaussian, exponential, spherical covariogram) in the two dimensional plane ([11]).

### 1.3.1 Discretization of convolution processes

The theoretically continuous underlying process in practice can be approximated by a discretized version without much loss if the discretization is not too coarse relative to the smoothing kernel. The approximation consists in restricting the integration over a domain and replacing the integral with a finite sum ([10]).
Consider a uniform partition of a square domain $D_r$, with $D_r \to D$, when $r \to \infty$, and the consider the centroids grid locations $u_i, i = 1, \ldots, p$. Assume $|A_{rj}|, j = 1, \ldots, p$ the area of each subsquare, then

$$\begin{aligned}
\int_{R^2} k(s-u)V(du) &= \lim_{r\to\infty} \int_{D_r} k(s-u)V(du) && (1.41)\\
&\approx \lim_{r\to\infty} \sum_{j=1}^{p} k(s-u_j) \int_{A_{rj}} V(du) && (1.42)\\
&= \lim_{r\to\infty} \sum_{j=1}^{p} k(s-u_j)V(A_{rj}) && (1.43)\\
&\approx \sum_{j=1}^{p} k(s-u_j)V_j\sqrt{(}|A_{rj}|) && (1.44)
\end{aligned}$$

where $V_j = x(u_j)$ are independent random variables with zero mean and variance $\sigma^2$. Thus

$$Y(s) \simeq \sum_{j=1}^{p} k(s, u_j)x(u_j) \tag{1.45}$$

where $x(u_j), j = 1, \ldots, p$ is a collection of $p$ independent normal random variables with zero mean and $\sigma^2$ distribution defined on the lattice.
Considering a set of locations $(s_1, \ldots, s_n)$ and being $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))$ the corresponding vector, the definition above yields $\mathbf{Y} \sim N(0, \sigma^2\Sigma)$, where

$$\Sigma_{ii'}(\theta) = \sum_{j=1}^{p} k(s_i - u_j)k(s_{i'} - u_j)|A_j| \tag{1.46}$$

This framework can also be applied with a non uniform partition of the domain, that can be useful when the sample locations are not uniform in space.
The question about how close this approximation is to the real process has to be addresses. Kullback Leibler divergence has for example been used in [10], concluding that this approximation seems to be robust, with this robustness increasing as the spatial correlation becomes stronger.
For Gaussian kernels simulation studies showed that the lattice distance has to be no more than the standard deviation of the chosen kernel ([8]).

The basic model that can be built with this decomposition is of the type [12]

$$\mathbf{Y} = \mu\mathbf{1}_n + \mathbf{K}\mathbf{x} + \epsilon \tag{1.47}$$

where $\mathbf{x}$ and $\epsilon$ are zero mean normal with variance matrix $\Sigma_x = \sigma^2 I_p$, and $\Sigma_\epsilon = \sigma_\epsilon^2 I_n$ and $K_{ij} = k(s_i - u_j)$. This model is equivalent to a linear mixed effects model and can be estimated with standard statistical software.
This modeling framework has several advantages. For example it has been used to develop nonstationary models, and can also be used in order to achieve dimension reduction.
Moreover the framework can be extended to allow $x(u_i)$ to be dependent processes ([13], [14]) and to be used in a spatiotemporal context. All these extensions will be explained in the following sections. For example allowing $x$ to have a correlation function $\rho$ yields to

$$Cov(Y(s_i), Y(s_j)) = \sigma^2 \int_{R^2}\int_{R^2} k(s_i - u)k(s_j - u')\rho(u - u')dudu' \tag{1.48}$$

and by using a change of variable it can be proven this covariance is still stationary, depending only on $s_i - s_j$.

## 1.4 Anisotropy

A stationary anisotropic process is a process in which spatial association depends upon the separation vector between locations, not only for its absolute value but also for its direction. When an anisotropic process can be reduced to isotropy by a linear transformation of the coordinates we have geometrical anisotropy, that is

$$c(s_1 - s_2) = \sigma^2\rho((s_1 - s_2)'B(s_1 - s_2)) \tag{1.49}$$

where $B$ is a positive definite matrix and $\rho$ is a valid correlation function in $R^d$. A constant value of correlation yields to an ellipse (rather than a circle). More precisely we can have different kinds of anisotropy, that is nugget anisotropy, range anisotropy and sill anisotropy ([15]). The most common case is range anisotropy where the range depends upon direction. The directional range, that is the range of the directional variogram for any separation angle, corresponding to the $\rho = 0.05$ contour, is assumed to lie on an ellipse, with major axis corresponding to the maximum range [16]. The parameters of this ellipse can be estimated by the data.
It can be proven that geometric anisotropy can be addressed by convolution kernel approach by using a kernel $K^* = K(B^{1/2}h)$.
Empirical semivariogram contour plots can be an useful tool to assess anisotropy.

## 1.5 Nonstationarity

Although the stationarity assumption can be reasonable in many applications (this assumption is assumed often after conditioning on the mean part), there are situations for which this hypothesis is not appropriate, for example if we have large and very heterogeneous domains.

While recently many methods have been developed to deal with nonstationarity, modeling non stationarity is still difficult, because a covariance matrix globally positive definite is needed and at the same time this matrix has to be specified in terms of local features to achieve non stationarity.

In a geostatistical framework there are many methods recently developed, most of which are based on the assumption of local stationarity.
For example a test for detect nonstationarity has been developed by Fuentes [17] for regularly spaced data starting from spectral representation and the estimation of the periodogram.

The convolution process provides an attractive way of introducing non stationarity. For example, Higdon ([7]) defined a nonstationary process convolving white noise processes with spatially varying kernels. Letting the kernel vary with spatial location the correlation function

$$\rho(s_1, s_2) \propto \int_{R^2} k_{s1}(u) k_{s2}(u) du \tag{1.50}$$

is for construction definite positive and the resulting process can be expressed as

$$Y(s) = \int_{R^2} k_s(u) x(u) du \tag{1.51}$$

The spatially varying kernel $k_s(u)$ (that has to be square integrable for all points) is function of some parameters that could be evaluated in a hierarchical Bayesian framework, using for example a spatial stationary process in order to let the kernels evolve smoothly over space.
In the case of a bivariate Gaussian kernel, that is

$$k_s(u) = (2\pi)^{-1} |\Sigma(s)|^{-1/2} \exp(-u' \Sigma(s)^{-1} u/2) \qquad h \in R^2 \tag{1.52}$$

the analytical expression for the covariance function can be obtained, and it is possible to account for anisotropy. In fact, due to the correspondence between the bivariate normal distribution and the standard deviation ellipse, it is possible to parameterize $\Sigma(s)$ in terms of the parameters of such ellipse. Higdon ([7]) modeled the two foci locations as spatial random fields to assure smooth variation. Alternatively one can use the major and minor axes and rotation angles, that could account for (range) anisotropy ([18]).
Starting from the analytical function for nonstationary covariance of the Higdon model when the kernel is multivariate Gaussian, Paciorek and Shervish

([19], [20]) defined a general formulation of nonstationary covariance, including a nonstationary form of the Matern covariance function.
In fact in the multivariate normal case, the covariance can be expressed as

$$C^{NS}(x_i, x_j) = \sigma^2 |\Sigma_i|^{\frac{1}{4}} |\Sigma_j|^{\frac{1}{4}} |\frac{\Sigma_i + \Sigma_j}{2}|^{-\frac{1}{2}} \exp(-Q_{ij}) \tag{1.53}$$

$$Q_{ij} = (x_i - x_j)^T \left( \frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (x_i - x_j) \tag{1.54}$$

where $\Sigma_i = \Sigma(x_i)$, the covariance matrix of the Gaussian kernel, which is centered at $x_i$ (kernel matrix). The authors demonstrated that if we have any isotropic correlation function $R^s$, positive definite on $R^p$ for every $p$, the non stationary correlation function

$$R^{NS}(x_i, x_j) = \Sigma_i|^{\frac{1}{4}} |\Sigma_j|^{\frac{1}{4}} |\frac{\Sigma_i + \Sigma_j}{2}|^{-\frac{1}{2}} R^s(\sqrt{Q_{ij}}) \tag{1.55}$$

is positive definite on $R^p$. For example the resulting nonstationary form of the Matern covariance function is

$$C^{NS}(x_i, x_j) = \sigma^2 |\Sigma_i|^{\frac{1}{4}} |\Sigma_j|^{\frac{1}{4}} |\frac{\Sigma_i + \Sigma_j}{2}|^{-1/2} \left( 2\sqrt{\nu Q_{ij}} \right)^{\nu} K_\nu \left( 2\sqrt{\nu Q_{ij}} \right) \tag{1.56}$$

This approach provides a way to have a closed form of nonstationary covariance function, based on a stationary form and local parameters. We can define the $\Sigma$ matrix over the whole domain in many ways, for example partitioning the domain in a set of non-overlapping areas $A_i$, or with a moving window, or defining a spatial process for the element of $\Sigma$, or defining a spatial process for the eigenvalues/vector processes associated with $\Sigma$. It is also possible to model the covariance matrix as a function of covariates.

Another kernel based method is that from Fuentes [21]. In this method the process considered is the convolution of a fixed kernel over independent stationary processes with different covariance parameters, that is

$$Y(s) = \int_D k(s - x) Z_{\theta(x)}(s) dx, \qquad s \in D \tag{1.57}$$

where $k$ is a stationary kernel function convolving the mean zero stationary spatial process $Z_{\theta(x)}(s)$, that has a covariance function with spatially varying parameters $\theta(x)$. Thus $Y(s)$ is such that

$$var(Y(s)) = \int_{R^2} k^2(s - x) C(0; \theta(x)) dx \tag{1.58}$$

$$cov(Y(s_i), Y(s_j)) = \int_{R^2} k(s_i - x) k(s_j - x) C(s_i - s_j; \theta(t)) dt \tag{1.59}$$

In this case the whole covariance function is hard to be evaluated analytically.
Multiresolution approach to nonstationarity hase been developed in[22] by means of wavelet basis functions.
Pintore and Holmes ([23]) handle nonstationarity in the spectral domain. The term *spectral* is intended in a broad sense, that is in a decomposition of the covariance function in orthogonal basis. They used in fact both the Fourier basis from a parametric point of view and the Karhunen Loeve expansion for a non-parametric equivalent. Their method is based on the concept of local stationarity, and on the fact that a process can be locally represented, in the spectral domain, as a superposition of Fourier frequencies with suitable weight functions.
Given a parametric model for the spectral density $g(\omega;\theta)$ Pintore and Holmes define a nonstationary spectral density $g^s_{NS}(\omega)$ proportional (equal) to $g(\omega, s; \theta(s))$, i.e. of the same form of the spectral density whose parameters vary locally, demonstrating that, if $C(s,t)$ is a stationary covariance function whose spectrum is $g(\omega)$, then the function on $D$x$D$ given by

$$C_{NS}(s,t) = \int_\Omega \exp(i\omega(s-t)) g(\omega, s; \theta(s))^{1/2} g(\omega, t; \theta(t))^{1/2} d\omega \tag{1.60}$$

is a valid covariance function if and only if

$$\int_\Omega |g(\omega, s, \theta(s))| d\omega < \infty \tag{1.61}$$

for all $s \in D$.
For example for the Matern correlation function we have:

$$g^s_{NS}(\omega) = h(s)^2 (\alpha^2 + ||\omega||^2_d)^{-\nu(s)-d/2} \tag{1.62}$$

with corresponding covariance function

$$\begin{aligned} C_{NS}(s,t) &= h_{s,t} (\alpha \|s-t\|_d)^{\nu_{s,t}} K_{\nu_{s,t}}(\alpha \|s-t\|_d) && (1.63)\\ \nu_{s,t} &= 0.5(\nu(s)+\nu(t)) && (1.64)\\ h_{s,t} &= \frac{h(t)h(s)\pi^{d/2}}{2^{\nu_{s,t}-1}\Gamma(\nu_{s,t}+d/2)\alpha^{2\nu_{s,t}}} && (1.65)\\ h(t) &= \text{normalizing constant} && (1.66) \end{aligned}$$

Again the spatially varying parameter can be modeled with different strategies (in [23] the authors proposed regression splines). The parameter $\nu(s)$ allows modeling different power of the higher/lower frequencies (i.e. smoothness) at different points in space.
The two complementary approaches of [19] and [23] were generalized by Stein ([24]). He demonstrated the important result that if $\Sigma$ is a mapping from $R^p$ to positive definite *pxp* matrices, $\mu$ is a nonnegative measure on $[0, \infty)$,

and for each $x \in R^p$, $g(\cdot, x) \in L^2(\mu)$ and defining $\Sigma(x,y) = 0.5(\Sigma(x)+\Sigma(y))$ and $Q(x,y) = (x-y)'\Sigma(x,y)^{-1}(x-y)$, then

$$R(x,y) = \frac{|\Sigma(x)|^{1/4}|\Sigma(y)|^{1/4}}{|\Sigma(x,y)|^{1/2}} \int_0^\infty e^{-wQ(x,y)} g(w;x)g(w;y)\mu(dw) \tag{1.67}$$

is a valid covariance function. Thus a nonstationary form for the Matern function with both $\nu$ and $\Sigma$ spatially varying is given by

$$\begin{aligned} R(x,y) &= \frac{c(x)c(y)}{|\Sigma(x,y)|^{1/2}} M_{0.5(\nu(x)+\nu(y))}(Q(x,y)^{1/2}) && (1.68)\\ M_\nu(x) &= (x)^\nu K_\nu(x) && (1.69)\\ Q(x,y) &= (x-y)^T\Sigma(x,y)^{-1}(x-y) && (1.70)\\ \Sigma(x,y) &= 0.5(\Sigma(x)+\Sigma(y)) && (1.71) \end{aligned}$$

where $\Sigma$ is a mapping from $R^p$ to positive definite *pxp* matrices.

Finally, before concluding this section, it is useful to cite the non-parametric approach to nonstationarity in [23] using the Karhunen-Loeve expansion. In the discrete case this expansion is equivalent to a principal component analysis (PCA) and we have

$$C = VDV' \tag{1.72}$$

where $C$ is the covariance matrix, $D$ is the diagonal matrix of eigenvalues and $V$'s columns are the corresponding eigenvectors. So the $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))$ process can be represented in terms of these basis function as

$$\mathbf{Y} = V\alpha + \epsilon \tag{1.73}$$

where $\alpha \sim MVN(0, D)$, and $\epsilon \sim MVN(0, \sigma^2 I_n)$. Nonstationarity can be introduced in this context by making the random coefficients $\alpha$ vary with locations, so that $\alpha_i = \alpha(s_i) \sim MVN(0, C_i)$.

## 1.6 Reduced dimension setup for spatial processes

In many applications very large datasets are needed to be handled. Due to the increasing number of automatic measurement systems the dimensionality of the geostatistical model can be computationally infeasible, both in a frequentist and Bayesian approach. A low dimensional setting and more efficiency in computation can be reached by using a lower dimensional latent process or by choosing a different representation, as spectral forms. As demonstrated by Wikle it is possible to set a general *reduced rank* representation able to include many common methodologies like discrete kernel

convolutions of the previous section, orthogonal polynomials, empirical orthogonal functions (EOF), splines and wavelets.
Recasting the hierarchical form of the basic geostatistical model by using a lower dimensional latent process and assuming for simplicity $\mu = 0$ we have:

$$\begin{aligned} Z &= Y + \epsilon, \epsilon \sim N(0, \Sigma_\epsilon) && (1.74) \\ Y &= K\alpha + \eta, \eta \sim N(0, \Sigma_\eta) && (1.75) \\ \alpha &\sim N(0, \Sigma_\alpha) && (1.76) \end{aligned}$$

where $\alpha$ is a $p$-dimensional random effects vector, such that $p << n$ (where $n$ is the dimension of the data), so $K$ is an expansion matrix that maps the low dimensional latent process $\alpha$ to the true process of interest $Y$. This parameterization allows computational advantages in both Bayesian and frequentist estimation. The power of this structure is increased if some simplified structure can be used for the correlation matrix: for example a diagonal structure for $\Sigma_\epsilon$ is appropriate in many cases. If we know the expansion matrix $K$ we can obtain the latent spatial process at unobserved locations. Many choices are possible for this matrix, corresponding to apparently different approaches in literature.
A distinction can be made between orthogonal and non orthogonal basis functions. Fourier basis, orthogonal polynomials, eigenvectors from covariance matrix and Karhunen-Loeve expansion are examples of the first group, while splines, wavelets and kernel functions are from the second one.
The $n$x$p$ $K$ matrix can be written as

$$K = \begin{pmatrix} k_1' \\ k_2' \\ \vdots \\ k_n' \end{pmatrix} \quad (1.77)$$

where $k_i' = (k_i(1), \ldots, k_i(p))$ corresponds to the $i$th spatial location. The orthogonal basis function satisfy the orthogonality constraint, that is $k_i \perp k_j$ for each $i, j$. In this case we can define $k_i$ for any spatial location $s_i$.
In this class the Fourier basis functions have many advantages. For this class of basis fast computational algorithms exist and it can be proven that, if the spatial process is weakly stationary, the $\alpha$ coefficients are nearly uncorrelated and $\Sigma_\alpha$ can be reasonably assumed as diagonal, and by using a class of stationary covariance model, these diagonal elements are function only of the parameters of this class with the variances at a given frequency nearly an half of the power spectral density at a given frequency (examples are in [25] and [26]). Another type of orthogonal expansion is the Karhunen-Loeve, that is an optimal choice in term of minimizing the variance of truncation error. The need of solving an integral equation and the fact that the basis can

be evaluated only for locations for which there is an observation (although one can use some interpolation method), limit the use of this expansion in practice.
If there are repeated observations (for example in time), a PCA can be performed on the empirical covariance matrix. In spatial statistics the eigenvectors from this PCA decomposition are called empirical orthogonal functions, and correspond to a discretization of the KL integral for equally spaced data. By using this basis function there is no need of stationarity assumption, and the covariance matrix $\Sigma_\alpha$ in this case is still diagonal ([27]). The drawback is the same as for the KL expansion, that is the need of interpolating the eigenfunctions to obtain these in different locations, and to obtain estimates of the associated covariance matrix.
An important example of non orthogonal basis function are the kernel basis. As we saw in section 1.3 a correlated stochastic process can be written in terms of convolution of a Brownian motion process. The discrete equivalent of this approach leads to a definition of the K basis functions on some support points for the $\alpha$ process.
Consider the kernel expansion

$$Z(s) = \sum_{j=1}^{p} k(s, r_j; \theta_s)\alpha_j + \eta(s) \tag{1.78}$$

where $k(s, r_j; \theta_s)$ corresponds to some kernel function and its value for a given spatial location s depends on the location $j = 1, \ldots, p$ support points over which the process $\alpha$ is defined, and where the kernel parameters $\theta_s$ may vary in space. Assuming for simplicity these parameters constant in space, we can use the kernel to define the mapping matrix K, that is $k_i(\theta) = (k(s_i - r_1; \theta), \ldots, k(s_i - r_p; \theta))$, then

$$K(\theta) = \begin{pmatrix} k_1'(\theta) \\ k_2'(\theta) \\ \vdots \\ k_n'(\theta) \end{pmatrix} \tag{1.79}$$

Assuming $\alpha$ as $N(0, \sigma_\alpha^2 I)$, the spatial structure is obtained by smoothing the white noise process on the $p$ support points. Otherwise, thinking at the process on these points as a spatial process, a spatial covariance function can be used for $\Sigma_\alpha$ (examples are in [21], [13], [14]).

# Chapter 2

# Multivariate modeling

Point referenced spatial data are often multivariate. It is typical in fact to collect more than one variable in the same site. For example air pollution monitoring networks measure at the same time a set of pollutants and/or a set of meteorological variables. The modeling problem is due to the double dependence between measurements at a specific site and the correlation between the values of a variable at different locations.
Let $\mathbf{Y}(s)$ a $m$x1 vector of $m$ random variables collected at a specific site $s$, and denote with $\mathbf{Y} = (\mathbf{Y}(s_1), \ldots, \mathbf{Y}(s_n))$ . The $m$x$m$ cross covariance matrix

$$C(s_1, s_2) = Cov(\mathbf{Y}(s_1), \mathbf{Y}(s_2)) \tag{2.1}$$

has to model the correlation between a variable measured in two different sites, the correlation between different variables and the combination of the two. This cross covariance matrix needs not to be symmetric or definite positive, but in a limiting sense, as $s_1 \to s_2$ this cross covariance becomes the symmetric and positive definite variance covariance matrix of the variables within a specific site $\mathbf{Y}(s)$.
Moreover this matrix has to be chosen in such a way that the $nm$x$nm$ covariance matrix for $\mathbf{W}$ is symmetric and definite positive for an arbitrary number of locations.
In the context of the general model introduced earlier for univariate data, that is

$$\mathbf{Y}(s) = \mu(s) + \mathbf{W}(s) + \epsilon(s) \tag{2.2}$$

the vector $\epsilon(s)$ models the measurement error and it is assumed to be multivariate normal with $m$x$m$ covariance matrix $\Psi$, and again the zero mean $\mathbf{W}(s)$ term accounts for spatial association and it is assumed to be a realization from a gaussian spatial process.
Considering the collection of sites $(s_1, \ldots, s_n)$, $[W = W(s_i)]_{i=1}^n$ is a $mn$x1 vector and its distribution is a multivariate normal with (parametric) variance-covariance $mn$x$mn$ matrix $\Sigma_W(\theta)$.

This matrix can be decomposed in $m$x$m$ block, i.e.$\Sigma_W(\theta) = [\mathcal{K}(s_i, s_j; \theta)]_{i,j=1}^n$, where $\mathcal{K}(s_i, s_j, \theta) = [Cov(W_k(s_i), W_l(s_j))]_{k,l=1}^m$. Thus $\mathcal{K}(s_i, s_j)$ is the cross covariance matrix function of the locations $(s_i, s_j)$. The dispersion matrix of $Y$ is $\Sigma_W(\theta) + I_n \otimes \Psi$. $\Sigma_w(\theta)$ and $\mathcal{K}(s, s; \theta)$ have to be symmetric and positive definite, and $\mathcal{K}(s_i, s_j, \theta)$ has to be chosen in an appropriate way.
The easiest way to achieve this property is by using a *separable* model, that is

$$\mathcal{K}(s_i, s_j) = \rho(s_i, s_j)T \tag{2.3}$$

where $\rho$ is a valid correlation function governing the spatial association and $T$ is a $m$x$m$ positive definite matrix accounting for dependence between variables in a specific site.
It can be proven that in this case the covariance matrix for $\mathbf{Y}$ is $H \otimes T$, where $(H)_{ij} = \rho(s_i, s_j)$, is symmetric and positive definitive. This assumption assures tractability, many computational advantages and a good interpretability.
Limitations in this hypothesis are related to a symmetric cross covariance matrix $\mathcal{K}(s_i, s_j)$ and in imposing only one spatial correlation function, so that every variable has the same spatial behavior.
The hierarchical approach avoids the difficulty of specifying a valid joint covariance function. As outlined in [28] and [5] is often more natural (and always valid) to specify a conditional relationship between variables than a joint covariance. For example in a bivariate spatial process $[y_1, y_2]$ we have

$$[y_1, y_2, \theta] = [y_1|y_2, \theta_1][y_2|\theta_2][\theta_1, \theta_2] \tag{2.4}$$

The conditional relationship between $y_1$ and $y_2$ can be based on a causal relationship or just inferred. Royle and Berliner ([28]) applied this model for the conditional relationship between ozone and temperature, defining

$$\begin{aligned} \text{Ozone}|\text{temp}, B, \beta_{oz}, \Sigma_{oz|\text{temp}} &\sim N(X_{oz}\beta_{oz} + B\text{temp}, \Sigma_{oz|\text{temp}}) && (2.5) \\ \text{temp}|\beta_{\text{temp}}, \Sigma_{\text{temp}} &\sim N(X_{\text{temp}}\beta_{\text{temp}}, \Sigma_{\text{temp}}) && (2.6) \end{aligned}$$

Moreover, letting the elements of $B$ vary over space, the cross covariance and marginal covariance matrices of ozone are nonstationary. Another advantage of this approach is given by the possibility of conditioning on latent processes, that can be useful when the number of variables is high (for example see [29]).

## 2.1 The coregionalization approach

The coregionalization model is a constructive way to build valid covariance functions in the joint model, and an equivalence can be established between

this approach and the hierarchical one.
This method consists in building a rich class of valid covariance functions by linearly transforming a simple covariance structure ([30]).
In the following we describe the coregionalization model in gaussian spatial models as [31],[32] and [33].
Considering a diagonal cross correlation function $\tilde{\mathcal{K}}(s_i, s_j; \theta) = diag[\rho_i(s_i, s_j, \theta_i))]_{i=1}^m$, with $\rho_l(\cdot, \cdot; \theta_l)$ a correlation function of $W_l(s)$, we have independent spatial processes ($Cov(\tilde{W}_l(s_i), \tilde{W}_k(s_j)) = 0$ if $l \neq k$). For each process $\rho$ the correlation function controls the spatial association and can be parameterized by some parameter $\theta$, for example we can use the Matern family.
Linearly transforming this independent spatial term allows to build a richer covariance function that is still valid, i.e. $W(s) = A(s)\tilde{W}(s)$. If we consider a spatially varying matrix $A$ the covariance associated with the spatial effect is nonstationary. In this setting the cross covariance matrix is $\mathcal{K}(s_i, s_j; \theta) = A(s)\tilde{\mathcal{K}}(s_i, s_j)A^T(s')$ and, because $\tilde{\mathcal{K}}(s, s; \theta) = I_m$, we have $A(s) = \mathcal{K}^{1/2}(s, s; \theta)$. Finally the covariance matrix $\Sigma_W = [\mathcal{K}(s_i, s_j; \theta)]_{i,j=1}^n$ is

$$[A(s_i)\tilde{\mathcal{K}}(s_i, s_j; \theta)A^T(s_j)]_{i,j=1}^n = [\oplus_{i=1}^k A(s_i)][\oplus_{k=1}^m \rho_k(s_i, s_j; \theta_k)]_{i,j=1}^n[\oplus_{i=1}^k A^T(s_i)] \tag{2.7}$$

where $\oplus$ is the direct sum operator. This cross covariance is positive definite because $\tilde{\mathcal{K}}(s_i, s_j; \theta)$ is a valid cross covariance function.
If $A(s) = A$, the stationary case, we have

$$\Sigma_w = [I_n \otimes A][\oplus_{k=1}^m \rho_k(s_i, s_j; \theta_k)]_{i,j=1}^n[I_n \otimes A_T] \tag{2.8}$$

The separable case is obtained using a single correlation function for each component of $\tilde{W}(s)$, that is $\tilde{\mathcal{K}}(s_i, s_j) = \rho(s_i - s_j; \theta)I_m$. In this case the covariance matrix for the process $w$ becomes

$$\Sigma_w = R(\theta) \otimes \mathcal{K}(0, \theta) \tag{2.9}$$

where $R(\theta) = [\rho(s_i, s_j; \theta)]_{i,j=1}^n$ is the spatial term, while $\mathcal{K}$ accounts for the within site correlation.

An equivalence can be established between the joint and the conditional formulation. For example in a bivariate case $v = Aw(s)$, with A lower triangular, then $v_1(s) = a_{11}w_1(s)$, $a_{11} > 0$ and $v_2(s)|v_1(s) \sim N(\frac{a_{21}}{a_{11}}v_1(s), a_{22}^2)$. Viceversa, let $v_1(s) = \sigma_1 w_1(s)$ with $\sigma_1 > 0$ and $w_1(s)$ a mean zero spatial process with variance 1 and correlation function $\rho_1$, and $v_2(s) = \alpha v_1(s) + \sigma_2 w_2(s)$ with $\sigma_2 > 0$ and $w_2(s)$ a mean zero spatial process with variance 1 and correlation function $\rho_2$. This is equivalent to the joint formulation $v = Aw(s)$ if $a_{11} = \sigma_1$, $a_{21} = \alpha\sigma_1$, $a_{22} = \sigma_2$. A link between the two methods can be established for the specification of the priors as well.

## 2.2 Kernel convolution and convolution of covariance

Kernel convolution approach and convolution of covariance functions are other approaches useful to build multivariate models ([34], [35], [18]). The former is the multivariate version of the model of section 1.3. Let $w(s)$ a mean zero variance 1 gaussian process, with correlation function $\rho$, and let $k_h(\cdot), h = 1, \ldots, m$, a set of square integrable kernel functions on $R^2$.
Defining

$$Y_h(s) = \sigma_h \int_{R^2} k_h(s-u)x(u)du \qquad l = 1, \ldots, p \tag{2.10}$$

for the $h$th component of $\mathbf{Y}(s)$ a valid cross covariance function is obtained with component

$$C_{h,h'}(s_i, s_j) = \sigma_h \sigma'_h \int_{R^2} \int_{R^2} k_h(s_i - u) k_{h'}(s_j - u')\rho(u - u')dudu' \tag{2.11}$$

that is valid by construction. By transformation of variables we can see how this covariance depends only on $s_i - s_j$, that it is a stationary covariance function. Moreover this covariance is isotropic if $\rho$ is isotropic and the kernel functions depend only on the absolute value of the difference between locations.
However considering the discretized version of this model, with $p$ the number of reference points, we have

$$C_{h,h'}(s_i, s_j) = \sigma_h \sigma'_h \sum_{l=1}^{p} \sum_{l'=1}^{p} k_h(s_i - u_l) k_{h'}(s_j - u_{l'})\rho(u_l - u_{l'}) \tag{2.12}$$

that is no longer stationary.
The latter approach is stationary and consists on directly convolving the covariance function (and not the process). Instead of introducing a set (m) of kernel functions a set of $m$ covariance functions can be used. In fact suppose $C_1, \ldots, C_m$ are real valid covariance functions defined on $R^d$. It can be proven that, under weak condition ($C_i$ squared integrable stationary covariance function), the collection of

$$C_{ij}(s) = \int_{R^d} C_i(s-u)C_j(u)du \qquad i \neq j \tag{2.13}$$

$$C_{ii}(s) = \int_{R^d} C_i(s-u)C_i(u)du \tag{2.14}$$

are valid covariance functions for the $m$ multivariate process, with $Cov(Y_i(s), Y_j(s')) = C_{ij}(s - s')$. This model is more parsimonious than the coregionalization approach. An application to daily average of three correlated pollutants with fully Bayesian estimation can be found in [36].

## 2.3 Bivariate modeling of NO – $NO_2$ in the Tuscany region

Spatial prediction of pollutant concentrations is central in the most of environmental policies. We applied the coregionalization model to the bivariate spatial modeling of nitrogen oxide and nitrogen dioxide. Environmental Protection Agency of Tuscany (ARPAT) is interested in this kind of analysis due to the fact that while Italian law sets limit values on nitrogen dioxide concentrations, due to its toxicity, authorization procedures are based on nitrogen oxide emissions.

### 2.3.1 NO and $NO_2$

Nitrogen dioxide is toxic by inhalation and there is evidence from toxicological studies that long-term exposure to $NO_2$ at high concentrations has adverse effects. $NO_2$ and other nitrogen oxides are also precursor of ozone and particulate matter, whose effects on human health are well documented. Existing Italian law sets limit values for $NO_2$, that is for the annual mean 40 $\mu g/m^3$ (D.M. 2.04.2002 N.60, 1999/30/CE).
There is a strong relationship between NO and $NO_2$ due to the chemical mechanism of formation on one side and to the measurement technique on the other. Major sources of NO in air are combustion emissions, mainly from vehicles, while ambient air $NO_2$ is in large part derived from the oxidation of NO. During daylight NO and $NO_2$ are in equilibrium with the ratio $NO/NO_2$ determined by the intensity of sunshine (which converts $NO_2$ to NO) and ozone (which reacts with NO to give back $NO_2$) in a complex photochemical reaction. Moreover this reaction takes some time and air can travel some distance before secondary pollutants are generated, so the spatial distribution of the ratio of $NO/NO_2$ is difficult to establish. Thus correlation between NO and $NO_2$ is not negligible a priori in the modeling strategy.

The network of air pollution monitoring stations of the Tuscany region (see Fig. 2.1) measures the concentrations of several pollutants in a set of locations spread across the region. The most of monitoring stations are located in the north of the region and near the major cities, mainly for regulatory reasons. We have annual means of NO and $NO_2$ ($\mu g/m^3$) in 61 monitoring sites for the year 2003, 2004 and 2005, collected by ARPAT. Following the European regulation (2001/752/CE) we have different kind of stations with respect to the monitored zone and to monitoring type. The total concentration of NO plus $NO_2$ ($NO_x$) are measured on hourly time step simultaneously by the chemiluminescence technique. It is worth noting that, not only the processes of formation, but the measurement errors of NO and $NO_2$ are correlated as well. In fact first NO is measured, inducing a reaction with

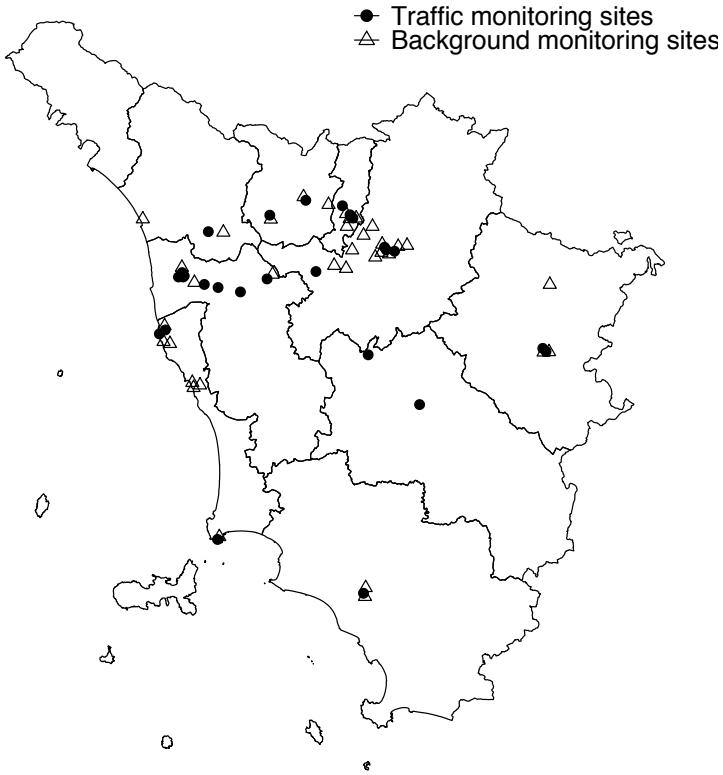


Figure 2.1: The regional monitoring network of Tuscany.

ozone to form light whose intensity is proportional to the concentration of the analyte. After that $NO_2$ in the sample is reduced to NO by a converter. The new NO concentration can again be measured by chemiluminescence: the difference of the concentrations gives the concentration of $NO_2$.
First we have done an explorative analysis of annual means of NO and $NO_2$ over the 61 stations to understand the structure of the data and to adopt proper modeling strategies. We considered in the analysis only such monitoring stations that have at least 70% of valid data over the year to perform the analysis. For the data we have, the log scale seems to be more attractive in term of normality with respect to both the original and square root scale.

The variables *station type* and *location type* supplied in the dataset are used as regression variables. The first variable accounts for different mean levels of the near source (that is *traffic*) stations with respect to the background sites, while the second variable is from administrative classification and distinguishes rural from urban areas. The traffic and background stations resulted to be strongly significant in a regression linear model both for NO and NO2, while the indication for urban characterization seems to be

weaker in particular for $NO_2$. We then examinate the spatial variation using an empirical variogram of the residuals of these models. The estimated values of these empirical variogram will be used as priors for the model we are going to estimate.

### 2.3.2 Modeling

We concentrate on a model template suitable to estimate spatial association and correlation between the two pollutants and to evaluate different hypothesis on the correlation structure. We used the coregionalization approach (as described in section 2.1) and the general template we refer to is from [37], making use of the R package spBayes ([38]).
This package allows the Bayesian estimation for a stationary multivariate gaussian model of coregionalization, with different correlation hypothesis, both separable and nonseparable.
The model is of the form

$$Y(s) = X^T(s)\beta + W(s) + \epsilon(s) \tag{2.15}$$

where $X(s)$ is a $m$x$p$ matrix of regressors and the vector $\epsilon(s)$ models the measurement error and it is assumed to be multivariate normal with $m$x$m$ covariance matrix $\Psi$. Considering the collection of sites $[s_1, \ldots, s_n]$, we have a $mn$x1 vector distributed as multivariate normal with variance-covariance matrix $\Sigma_W$, composed by $m$x$m$ cross covariance matrix functions of the locations $(s_i, s_j)$, $\mathcal{K}(s_i, s_j)$. The resulting dispersion matrix of $Y$ is $\Sigma_W(\theta) + I_n \otimes \Psi$. A process with diagonal cross covariance function $\tilde{W}(s)$, that is with an independent spatial process for each response variable, is linearly transformed by a matrix $A$, $W(s) = A(s)\tilde{W}(s)$. In the stationary case $A(s) = A$.
The Matern family with parameter $\theta = (\phi, \nu)$ is used to control the spatial association and the smoothness of each spatial process in $\tilde{W}$.

A Bayesian approach is taken and a Gibbs sampler, with Metropolis-Hastings step whenever required, is used to get estimates. MCMC model fitting is performed on the marginalized scale to reduce the number of parameters, that is after integrating out the $\tilde{W}$ process

$$Y \sim MVN(X\beta, (I_n \otimes A)\Sigma_{\tilde{W}}(I_n \otimes A)^T + I_n \otimes \Psi) \tag{2.16}$$

The model is completed by setting the priors distributions on the collection of the parameters $\Theta = (\beta, \mathrm{A}, \theta, \Psi)$.
The posterior distribution of the spatial effect $\tilde{W}$ is then recovered using

$$p(\tilde{W}|\text{data}) \propto \int p(\tilde{W}|\Theta, \text{data})p(\Theta|\text{data})d\Theta \tag{2.17}$$

Once posterior samples for $\Theta$ from $p(\Theta|data)$ are drawn by the sampling algorithm, we can sample from the gaussian distribution $p(\tilde{W}|\Theta, data)$ to obtain posterior samples from $p(\tilde{W}|\text{data})$. The posterior predictive distribution in a collection of new points $s_{0i}$, $p(\tilde{W}^*|\text{data})$ is again obtained through composition sampling using

$$p(\tilde{W}|\text{data}) \propto \int p(\tilde{W}^*|\tilde{W}, \Theta, \text{data}) p(\tilde{W}|\Theta, \text{data}) p(\Theta|\text{data}) d\Theta d\tilde{W} \quad (2.18)$$

where the first term is again multivariate normal. Predictions on $Y$ can be done for those locations in which the $X$ matrix of covariates is known sampling from the conditional expectation $E(Y^*|data) = X^*\beta^l + (I_n \otimes A^l)\tilde{W}^{*l}$, for $l = 1, \ldots, L$ or alternatively by drawing from the marginal distribution

$$p(Y^*|\text{data}) = \int p(Y^*|\Theta, \text{data}) p(\Theta|\text{data}) d\Theta \quad (2.19)$$

where again the first term is multivariate normal (see [38] for details).

To complete the model we have to assign prior distributions on the parameters. We choose a flat prior with Gibbs updating for each $\beta$ parameter. Since we used an exponential correlation function we have to estimate only the range parameter for which we assign an informative uniform prior with support greater than zero and related to the maximum distance of the points in the domain. For $\Psi$ matrix we choose either an inverse Wishart for the full matrix, or, in the case with independent non spatial error an inverse gamma for each of the diagonal elements. The same strategy is used for the cross covariance matrix $\mathcal{K}$. To define the hyperparameters of the priors we used the values suggested by the fit of the empirical variograms.

### 2.3.3 Model selection, results and validation

We use station type and zone type as a regressors for both NO and $NO_2$ in all the evaluated models. We compare six stationary models with different complexity, with hypothesis specified as follows.

The first model is the richer one. We assume a full $\mathcal{K} = AA^T$ matrix that allows to estimate the spatial covariances among the response variables within a location, a full $\Psi$ measurement error accounting for measurement correlation, and specific spatial decay $\phi$ for each response variable. The second model is the separable version of the first one, setting a common spatial decay for each response variable. The third model is again nonseparable but with a diagonal $\mathcal{K} = \text{diag}(\sigma_i^2)$ matrix, i.e. assumes independently varying spatial processes. The fourth model is the separable equivalent of the third. Finally in model 5 and 6 we check the independence of the measurement error setting $\Psi = \text{diag}(\tau_i^2)$, both in a separable or nonseparable version.

| 2003 | parameters | DIC03 | DIC04 | DIC05 |
|---|---|---|---|---|
| M1 | $\phi_m, A, \Psi$ | **-356.27** | -368.64 | -373.73 |
| M2 | $\phi, A, \Psi$ | -343.19 | -391.89 | -398.92 |
| M3 | $\phi_m, \sigma^2_m, \Psi$ | -354.99 | -378 | -384.92 |
| M4 | $\phi, \sigma^2_m, \Psi$ | -354.16 | **-430.35** | **-415.61** |
| M5 | $\phi_m, A, \tau_m$ | -350.78 | -348.58 | -343.01 |
| M6 | $\phi, A, \tau_m$ | -336.18 | -344.24 | -326.62 |

Table 2.1: DIC for the six models (calculated on unmarginalized likelihood). For each model the associated parameters are reported.

| | Spatial Variance | Non-Spatial Variance. | Spatial Behavior |
|---|---|---|---|
| M1 | not indep. | not indep. | non separable |
| M2 | not indep. | not indep. | separable |
| M3 | indep. | not indep. | non separable |
| M4 | indep. | not indep. | separable |
| M5 | not indep. | indep. | non separable |
| M6 | not indep. | indep. | separable |

For each of these models two MCMC chains were run until convergence and we saved the last 25000 iterations.
For models comparison we used DIC criterion[1] ([39]), that has nice properties for gaussian likelihood. DIC is defined as the sum of a measure of modeling fit (the deviance, minus twice of the log-likelihood of the model) and a penalty for model complexity, pD.

$$DIC = \overline{D(\Theta)} + pD \tag{2.20}$$
$$pD = \overline{D(\Theta)} - D(\overline{\Theta}) \tag{2.21}$$

So the model with lowest DIC score has to be preferred. DIC scores for different modeling hypothesis are reported in table 2.1.

Following the DIC criterion model 5 and 6, that is ignoring the correlation related to non spatial term, result to have the lowest fit, and this is reasonable due to the correlation between the measurement errors of NO and $NO_2$. Other model assumptions give more similar DIC scores, although for 2004 and 2005 model 4 has the lowest DIC and the lowest number of effective parameters, followed by model 2 scores. For 2003 model 1 is the best for DIC and pD criterion, and interpolated spatial effect estimates of this model are reported in Fig. 2.2.
For validation purposes we used the leave-one-out principle for each station and evaluated the quantile of the values predicted by the different models. In terms of numbers of correct interval prediction model 1 and 2 are the

[1]DIC criterion will be discussed in more detail in section 4.4

| | Psi correlation M1 | $\mathcal{K}$ correlation M1 | Psi correlation M4 |
|---|---|---|---|
| 2003 | 0,51 (-0,52;0,7) | 0,64 (0,05;0,74) | 0,73 (0,68;0,76) |
| 2004 | 0,76 (0,13;0,83) | 0,69 (-0,51;0,85) | 0,83 (0,82;0,87) |
| 2005 | 0,78 (-0,36;0,83) | 0,57 (-1,47;0,78) | 0,82 (0,81;0,84) |

Table 2.2: Estimated correlation parameters.

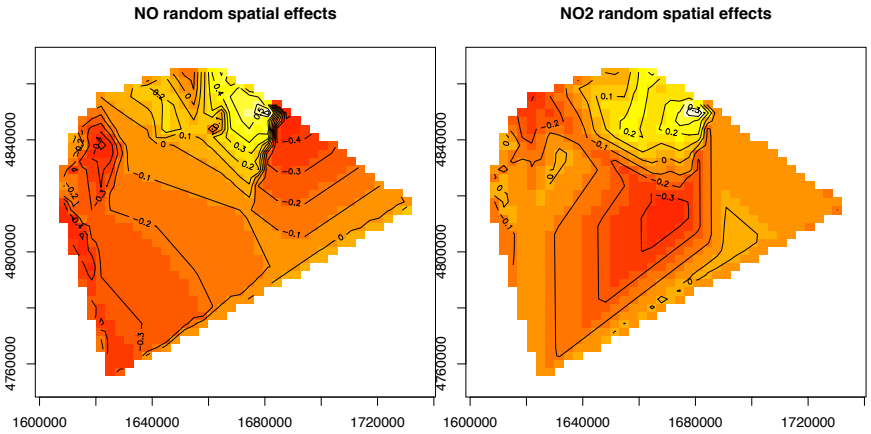


Figure 2.2: Interpolated surfaces of the recovered spatial effects on measurement sites from Model 1 for 2003 year.

best (correlation into the spatial and non spatial term), while model 4 is the preferred model in terms of RMSE for 2004 and 2005. The analysis of correlation estimated by the models supports again the non evidence of correlation in spatial term for 2004 and 2005. Instead for 2003 the correlation seems to be related to the spatial component (see Table 2.3.3 ). While the different behavior of 2003 can be explained by the particular meteorological conditions that we observed in 2003 in Italy and over the entire Europe, for 2004 and 2005 years correlation between NO and $NO_2$ associated to the spatial term seems to be not clearly supported by the data. This behavior can be related to nonstationarity in the spatial process.

The posterior distribution of each parameter of the regressors is consistent with the expected values from the exploratory analysis.

The modeling framework presented here shows good prediction performances on average, with the observed high amplitude of confidence interval probably due to the small number of measurement sites, especially in the south of the region. The model has also the ability of capturing the correlation structure between the two pollutants, either in the spatial and measurement error term, resulting in an overall good fitting and prediction ability.

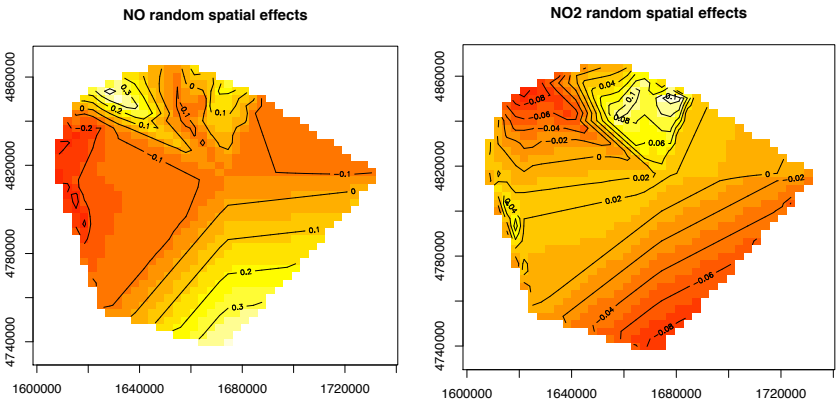


Figure 2.3: Interpolated surfaces of the recovered spatial effects, M4-2005.

# Chapter 3

# Spatiotemporal modeling

Air pollution data, like many other environmental data, have spatial and temporal nature. While the dispersion mechanism of a pollutant is deterministic in principle, in practice we rarely know all we need in order to make predictions and also in the case of ideal conditions a lot of simplified assumptions have to be done to solve the governing equations. Moreover uncertainty in measuring and manipulating the data and errors due to discrete sampling of a continuous system have to be considered. For spatiotemporal statistical models to be able to account for this uncertainty, the challenge is to capture variability in space and time and the interaction of the two, often arising from complicated dynamical physical processes.
Formally a spatiotemporal process could be viewed as a spatial process in the three dimensional space, but the different nature of time with respect to the spatial component makes this approach incorrect.
As in the spatial case both the joint and the conditional approach can be used: the two approaches will be described in the following.

## 3.1 Joint formulation

Consider the spatiotemporal observations $Y(s,t)$ defined on $s \in D$ at time t. The basic model can be formulated as

$$Y(s,t) = \mu(s,t) + w(s,t) + \epsilon(s,t) \tag{3.1}$$

where $\mu(s,t)$ denotes the mean structure, $\epsilon$ represents the residual and $w(s,t)$ is a zero mean spatiotemporal process. The mean term can be expressed as $\mu(s,t) = x(s,t)^T\beta(s,t)$, where $\beta$ can be spatially, temporally or both varying.

The spatiotemporal component can be simplified in an additive form composed by a pure spatial effect and a pure temporal term, that can be modeled with a valid covariance function in two and one dimension if the time is continuous, or with an autoregressive form if the time is discrete. Independent time series at each location or a collection of independent (in time) spatial processes are other two possible approaches. All of this methods avoid interaction between space and time, and are unsatisfactory in many practical cases.
If time is considered as continuous (that is $t \in R^+$), as in the spatial case, assuming a gaussian process what is needed is a valid correlation function $Cov(Y(s,t), Y(s',t'))$, that is for any set of locations and any set of points in time the covariance matrix of the associated random variable has to be positive definite. Notion of stationarity (and isotropy) is the same as in the spatial case and implies $Cov(Y(s,t), Y(s',t')) = c(s-s', t-t')$. In a similar way, a continuous both space time stationary spatiotemporal covariance function, can be written in terms of the spectral density $g$ of the spatiotemporal process. In fact following Bochner's theorem a continuous and symmetric function $C$ on $R^d$x$R$ is a covariance function if and only if it can be written as

$$C(h,u) = \int\int \exp^{i(h'\omega + u\tau)}\, dF(\omega,\tau) \tag{3.2}$$

where $F$ is a finite nonnegative symmetric measure on $R^d$x$R$. The class of stationary space time covariance functions on $R^d$x$R$ is identical to the class of Fourier transforms of finite nonnegative and symmetric measures on this domain. If $C$ is integrable, we can write the covariance in terms of the spectral density $g(\omega,\tau)$

$$C(h,u) = \int\int \exp^{ih'\omega + iu\tau}\, g(\omega,\tau) d\omega d\tau \tag{3.3}$$

where $h = s - s'$ and $u = t - t'$, with

$$\begin{aligned} g(\omega,\tau) &= (2\pi)^{-d-1} \int\int \exp^{-ih'\omega}\, C(h;u) dh du & (3.4)\\ &= (2\pi)^{-1} \int \exp^{-iu\tau}\, h(\omega;u) du & (3.5) \end{aligned}$$

where

$$h(\omega,u) = (2\pi)^{-d} \int \exp^{-ih'\omega}\, C(h,u) dh = \int \exp^{iu\tau}\, g(\omega,\tau) d\tau \tag{3.6}$$

A separable form, as in the multivariate case, takes the form

$$\mathrm{Cov}(Y(s,t), Y(s',t')) = C_s(s,s') C_T(t,t') \tag{3.7}$$

where the covariance is splitted into a (valid) spatial two-dimensional covariance function multiplied for a (valid) time one-dimensional covariance function. This structure is valid and assures many computational advantages.
In fact if we have a collection of $n$ locations and $T$ time points, the covariance matrix of the $nT$ vector $(Y(s_1,t_1),\ldots,Y(s_n,t_1),\ldots,Y(s_1,t_T),\ldots,Y(s_n,t_T))$ can be expressed as the kronecker product of a $T$x$T$ temporal covariance matrix times a $n$x$n$ spatial covariance matrix, and thus the calculation of the inverse is greatly simplified.
The spectral representation in this case can be written as Eqn. 3.3-3.6 with

$$h(\omega,u) = h_1(u)h_2(\omega) \tag{3.8}$$

with functions $h_1$ and $h_2$ such that $g$ is a spectral density, that is $h_1(\omega,.)$ is a continuous autocorrelation function for each $\omega \in R^d$, and $h_2(\omega) > 0$ and $\int h_2(\omega)d\omega < \infty$.
A test of separability based on this spectral representation has been proposed by Fuentes [40].
It is not always the case this separation between space and time behavior is correct and as demonstrated in [41] nonseparability is often a consequence of the governing physical laws. A notion related to separability is full symmetry, that is

$$cov(Y(s_1,t_1),Y(s_2,t_2)) = cov(Y(s_1,t_2),Y(s_2,t_1)) \tag{3.9}$$

for all $(s_1,t_1)$ and $(s_2,t_2) \in R^d$x$R$. However, in transport processes, whether subject to air or water flows, a lack of full symmetry is usually observed.
Separability is a special case of full symmetry.

Nonseparable spatiotemporal covariance functions have been studied as well. In literature a method based on spectral representations is proposed by Cressie and Huang ([42]). Considering the spectral density as in Eqn. 3.3-3.6, thus

$$g(\omega,\tau) = (2\pi)^{-1} \int \exp^{-iu\tau} h(\omega,u)du \tag{3.10}$$

$$h(\omega,u) = \int \exp^{iu\tau} g(\omega,\tau)d\tau \tag{3.11}$$

Specifying appropriate models for $h(\omega,u)$ yields to a class of valid nonseparable models. Let

$$h(\omega,u) = h_1(\omega,u)h_2(u) \tag{3.12}$$

where as above $h_1(\omega,\cdot)$ is a continuous autocorrelation function for each $\omega \in R^d$, and $h_2(\omega) > 0$ and $\int h_2(\omega)d\omega < \infty$. It can be proven that a Fourier

inversion of the spectral density built as above gives a valid continuous spatiotemporal stationary covariance function

$$C(h,u) = \int \exp^{ih'\omega} \rho(\omega;u)k(\omega)d\omega \tag{3.13}$$

This method requires that the Fourier inversion for the covariance function can be obtained in a closed form, that is not often the case.
Gneiting ([43]) overcomes this limitation through a covariance function defined as:

$$C(h,k) = \frac{\sigma^2}{\phi(|u|^2)^{d/2}} \psi\left(\frac{\parallel h \parallel^2}{\phi(|u|^2)}\right) \tag{3.14}$$

where $\psi(t)$,$t \geq 0$ is a completely monotone function and $\phi(t)$, $t \geq 0$ is a positive function with a completely monotone derivative, $h \in R^d$ represents the spatial vector and u $\in$ R is a time component. Different specifications of the functions $\psi$ and $\phi$ yield different valid covariance functions. This covariance function is fully symmetric.
Stein ([44]) pointed out a lack of differentiability in most of space time nonseparable covariance functions, that is models that away from the origin are not smoother than they are at the origin have a kind of discontinuity along certain axes that it is preferable to avoid.
Thus the author introduced a class of spectral densities, corresponding to the nonseparable covariance function

$$g(\omega,\tau) \propto [c_1(\alpha_1^2 + ||\omega||^2)^{\alpha_1} + c_2(\alpha_2 + \tau^2)^{\alpha_2}]^{-\tau} \tag{3.15}$$

for $c_1$ and $c_2$ positive, $\alpha_1^2+\alpha_2^2 > 0$, $\alpha_1$ and $\alpha_2$ positive integers and $d_1/(\alpha_1\nu)+ d_2/(\alpha_2\nu) < 2$. A particular case of this covariance is that of Jones and Zhang ([45]) ($d_1$=2, $\alpha_2 = \nu = d_2 = 1$). The covariance functions associated with this class cannot again be expressed explicitly, and fast Fourier transform can be used to obtain a numerical computation.
Another spectral density is proposed in [46],

$$g(\omega,\tau) = \gamma(\alpha^2\beta^2 + \beta^2|\omega|^2 + \alpha^2\tau^2 + \epsilon|\omega|^2\tau^2)^{-\nu} \tag{3.16}$$

where $\gamma, \alpha, \beta$ are positive, $\nu \geq (d+1)/2$ and $\epsilon \in [0,1]$; $\alpha^{-1}$ and $\beta^{-1}$ are respectively the spatial and temporal range, explaining the rate of decay of spatial/temporal correlation; $\nu$ is the parameter governing the smoothness of the process while $\epsilon$ governs the interaction between the spatial and temporal component, with $\epsilon = 1$ for the separable case. But it is again necessary to carry out a Fourier transformation, although if $\epsilon \in (0,1)$ a one dimensional transformation is enough.

Stationary space time covariance functions that are not fully symmetric can be constructed using a Lagrangian reference frame starting from a stationary

covariance spatial function $C_s$ as

$$C(h,u) = E(C_s(h - Vu)) \qquad (h,u) \in R^d xR \tag{3.17}$$

where $V$ is the random velocity for the entire field and the expectation is taken with respect to $V$. This velocity can be updated dynamically leading to a nonstationary covariance structure (see [47]). This kind of covariance function can be also constructed on the basis of diffusion equations or stochastical partial differential equations as first argued by Heine ([2]). Examples are in [45], [44], [41]. Finally a powerful way to achieve nonseparability is by blurring, or IDE models, as discussed later (section 3.5.1).

## 3.2 Hierarchical formulation

Hierarchical spatiotemporal models represent a powerful tool for modeling spatiotemporal variability. A very general framework has been proposed in [48]. The three stages presented in section 1.2.1, corresponding to data, process and parameters are valid in this context as well, although a richer five stage structure is proposed for spatiotemporal models.
Let $Z(s,t)$ be the observations at location $s$ and time $t$, where $(s,t) \in$ M, M being a grid or a lattice and $Y(s,t)$ the value of the process of interest. At a first stage a measurement equation is specified, defining the conditional distribution of $Y$ given $Z$. Typically $Z$ are assumed independent given $Y$, with a pure gaussian measurement error, but it is possible in this stage to account for misalignments between $Z$ and $Y$ or spatial and temporal averaging.
The second stage models the process $Y$ conditional on three other processes, accounting for site specific mean, large scale temporal variations, and a short time dynamical process, that is

$$Y(s,t) = \mu(s;\theta_{\mu,t}) + \gamma(t;\theta_{\gamma,s}) + \alpha(s,t;\theta_{\alpha,s,t}) + \epsilon(s,t) \tag{3.18}$$

where $\mu(s;\theta_{\mu,t})$ is a spatial trend surface with parameters that could be time varying, $\gamma(t;\theta_{\gamma,s})$ is a temporal mean with parameters that could be space varying. The key spatiotemporal process is represented by $\alpha$, accounting for nonseparable space time interactions often with dynamical behavior. This is a general framework and the terms needed for a specific application can be a subset of those above and a prior knowledge of the structure of the problem is needed for identifiability reasons. The role of the third stage in fact is that of building the structure of those single terms, while the two last stages are dedicated to the specification of the priors and hyperpriors.
The dynamic term $\alpha$ plays a crucial role and will be described in more details in the next sections.

## 3.3 Dynamic spatiotemporal models

In a dynamical process the current value of a variable depends on the values assumed at previous times. Physical processes often show this behavior and many spatiotemporal models are able to include a dynamical term. In the following we refer to discrete time processes.
Dynamical evolution is often included in a state space representation, with the dynamics modeled at a latent stage. A basic setting is to use two equations: the first one is a *measurement* equation that links the observations with the latent process, while the second one, the *transition* equation governs the dynamical evolution. A general dynamic space time statistical model can be written as in the following.
First consider the data model for the observations $Z(s,t)$:

$$Z(\cdot,t) = f_d[Y(\cdot,t);\theta_d(t)] \tag{3.19}$$

where $f_d$ is a stochastic functional that depends on the true process $Y(\cdot,t)$ and the parameters $\theta_d(t)$. The process model is:

$$Y(\cdot,t) = f_p[Y(\cdot,t-1),\ldots,Y(\cdot,1);\theta_p(t)] \tag{3.20}$$

where $f_p$ is a stochastic functional of past values of the true process and of the parameters $\theta_p(t)$. This equation describes the space-time dynamics of the process. This general framework makes no distributional or linearity assumption.
A first order dependency, that is a Markovian assumption, is often sufficient to describe the evolution of the process, and linearity and gaussian assumptions complete the hypothesis.
With a pure temporal process we can write

$$Z(t) = F_t Y(t) + \epsilon_t, \qquad \epsilon_t \sim N(0,\sigma_t^{\epsilon}) \tag{3.21}$$

$$Y(t) = H_t Y(t-1) + \eta_t, \qquad \eta_t \sim N(0,\Sigma_t^{\eta}) \tag{3.22}$$

where $Z(t)$ is an $m$x1 vector of observables and $Y(t)$ is a $p$x1 *state* vector, $F_t$ and $G_t$ are $mxp$ and $pxp$ system matrices, often assumed to be known ($F_t$ is a sort of design matrix). The resulting covariance structure can be computed explicitly $Cov(\theta_t,\theta_{t-1}) = H_t Var(\theta_{t-1})$ and $Cov(Y_t,Y_{t-1}) = F_t H_t Var(\theta_{t-1}) F_t^T$. Given the parameters, the unobserved state process can be estimated with a Kalman filter or smoother. Different approaches to extend this framework at the spatiotemporal context has been proposed in literature, and we review some of these approaches in the following sections.

### 3.3.1 Space time Kalman filter

The so called *space time Kalman filter* was developed throughout the 90s by several authors. Considering the vector of $Z$ at the $m$ spatial locations

results in a multivariate model like the one in previous section, that could be very high dimensional.
Mardia ([49]) first suggested the idea of a reduced dimension space time Kalman filter, with the state process expressed in some set of basis function, that had been applied by the same author as *kriged Kalman filter*. This framework allows for a spatially nonstationary and spatiotemporal nonseparable structure.
Wikle and Cressie ([50], [51]) developed the *space time Kalman filter* proposing the following model

$$z(s,t) = y(s,t) + \epsilon(s,t) \tag{3.23}$$
$$y(s,t) = \mu(s,t) + \nu(s,t) \tag{3.24}$$
$$\mu(s,t) = \int \omega_s(u)\mu(u,t-1)du + \eta(s,t) \tag{3.25}$$

where $\epsilon$ is a white noise representing measurement errors, $\nu$ represents a spatial structure independent in time (this component has been assumed equal to zero in kriged Kalman filter, resulting in an over-smoothing of the process) and $\eta$ is gaussian temporally white but spatially colored.
Assume that $\mu$ can be decomposed as

$$\mu(s,t) = \sum_{j=1}^{K} a_j(t)\phi_j(s) \tag{3.26}$$

where $\phi_j$ are deterministic basis functions (complete and orthonormal) and $a_j(t)$ are zero mean time series. We can expand the weight functions

$$\omega_s(u) = \sum_{l=1}^{\infty} b_l(s)\phi_l(u) \tag{3.27}$$

Truncating the infinite series above and using the orthonormality of the basis functions we can write

$$\phi(\mathbf{s})'\mathbf{a}(t) = \mathbf{b}(\mathbf{s})'\mathbf{a}(t-1) + \eta(s,t) \tag{3.28}$$

where $\phi(\mathbf{s}) = [\phi_1(\mathbf{s}), \ldots, \phi_K(\mathbf{s})]$, $\mathbf{a}(t) = [a_1(t), \ldots, a_K(t)]$ and $\mathbf{b}(\mathbf{s}) = [b_1(\mathbf{s}), \ldots, b_K(\mathbf{s})]$ For the locations $(s_1, \ldots, s_n)$ we have:

$$z(\mathbf{s},t) = y(\mathbf{s},t) + \epsilon(\mathbf{s},t) \tag{3.29}$$
$$y(\mathbf{s},t) = \sum_{j=1}^{K} a_j(t)\phi_j(\mathbf{s}) + \nu(\mathbf{s},t) \tag{3.30}$$
$$\mathbf{a}(t) = H\mathbf{a}(t-1) + J\eta(t) \tag{3.31}$$

where $\eta(t) = [\eta(\mathbf{s_1},t), \ldots, \eta(\mathbf{s_n},t)]'$, $H = JB$,$J = (\Phi'\Phi)^{-1}\Phi'$ with $\Phi = [\phi(\mathbf{s}_1), \ldots, \phi(\mathbf{s}_n)]'$ and $B = [\mathbf{b}(\mathbf{s}_1), \ldots, \mathbf{b}(\mathbf{s}_n)]'$. A physical dynamics can be

introduced in this setting by using the structure of matrices $H$ and $J$.
The optimal predictor under squared error loss for the value in a new point $Y(\mathbf{s}_0, t_0)$ is the mean of the posterior distribution $E[Y(\mathbf{s}_0, t_0)|Z, \theta]$. Under gaussian assumptions this expectation is linear in the data and can be obtained through recursive equation of Kalman filter and smoother. Estimation can be made in a fully Bayesian setting or with an empirical Bayes approach, that is by using an estimate of $\theta$. This estimate can be obtained using the EM algorithm (with restrictions on the parameter matrices, [52]).

### 3.3.2 Modeling spatiotemporal processes by means of space time dynamic coefficients

When the dynamic behavior is transferred on the coefficients of the mean part we have another class of dynamical spatial models.
This kind of models are used for example in [53], where a spatial process is expressed as a locally weighted mixture of linear regressions, and a random walk evolution is assumed for each coefficient of the regressions. Another example is in [54], modeling the spatiotemporal variability of the ozone in Mexico city. In this model random walk dynamics is assumed for the coefficients of sinusoidal components and of the coupled spatiotemporal process of temperature. A spatial correlation is superimposed both on the residual part of the process and on the coefficients of the sinusoidal component.

Gelfand et al. ([55]) developed a general framework for univariate and multivariate data, allowing general mean structures and also non stationary association structures, by adopting a spatiotemporally varying form of the coefficients and coregionalization.
The data are viewed as arising from a time series of a spatial process, the space is viewed as continuous but time is taken to be discrete. The response, $Y(s,t)$ is modeled through a measurement equation, while the transition equation involves the regression parameters of the covariates $x(s,t)$. The slope vector is decomposed into a purely temporal component $\beta_t$ and a spatiotemporal component $\beta(s,t)$. The univariate case is given by:
Measurement equation:

$$\begin{aligned} Y(s,t) &= \mu(s,t) + \epsilon(s,t); \qquad \epsilon \sim N(0, \sigma^2_\epsilon) && (3.32)\\ \mu(s,t) &= x^T(s,t)\tilde{\beta}(s,t) && (3.33)\\ \tilde{\beta}(s,t) &= \beta_t + \beta(s,t) && (3.34) \end{aligned}$$

Transition equation:

$$\begin{aligned} \beta_t &= \beta_{t-1} + \eta_t, \eta_t \sim N_p(0, \Sigma_\eta) && (3.35)\\ \beta(s,t) &= \beta(s, t-1) + w(s,t) && (3.36) \end{aligned}$$

where $w(s,t) = Av(s,t)$, $A$ is $pxp$ matrix and $v(s,t) = (v_1(s,t), \ldots, v_p(s,t))^T$ with $v_l(s,t)$ serially independent replications of a gaussian process with unit variance and correlation function $\rho(\cdot, \phi_l)$ (if $\phi_l = \phi$ the model is separable). The inference is taken in a Bayesian hierarchical framework, that is completed by prior specifications.

## 3.4 Process convolution in the spatiotemporal context

Process convolution can be used to build spatiotemporal models. If time is continuous this can be done easily by convolving a three dimensional white noise process ([56]). Otherwise a more flexible structure is introduced by convolving dynamic processes. This approach is used in [57], [18] and [14]. The key is to convolve dependent (in time) processes. Generally we can write

$$Y(s,t) = \sum_{i=0}^{n} k(s-u_i) x_{u_i}(t) \tag{3.37}$$

$$x_{u_i}(t) = f(x_{u_1}(t-1), \ldots, x_{u_m}(t-1), \beta) + \nu_{u_i}(t) \tag{3.38}$$

where $f(\cdot;\beta)$ is a parametric function of the values of the latent processes. Sansò et al. [14], explores different model specifications of spatiotemporal processes, combining process convolution and autoregressive processes. One can either consider a time series evolving in space, letting

$$y(s,t) = \int_R k(u-t;\Phi) x(u,s) du \tag{3.39}$$

where $x(\cdot,\cdot)$ is a gaussian spatial process, or, to be more physically based, consider a spatial field evolving in time, that is

$$y(t,s) = \int_{R^2} k(v-s;\theta) x(t,v) dv \qquad \forall t \in T \tag{3.40}$$

where $x(\cdot,\cdot)$ in this case is a time series process. The spectral density of $y$ in this case is

$$g_y(s,t) = |H(\omega)|^2 g_x(\tau,\omega) \tag{3.41}$$

$$H(\omega) = \int_R \exp^{-i\omega' s} k(s) ds \tag{3.42}$$

where $g_x$ is the spectral density of $x$. This model is similar to that introduced in [58] by setting

$$g_x(\tau,\omega) = (2\pi)^{-1} \int h_1(\omega,u) du \tag{3.43}$$

$$h_2(\omega) = |H(\omega)|^2 \tag{3.44}$$

This framework can include both separable and nonseparable structures. Discretization yields

$$Y_t = Kx_t \qquad t = 1, \ldots, T \tag{3.45}$$

Assuming $x$ autoregressive with correlation $\rho(t-t', \theta_\rho)$ the covariance of the process is:

$$k_{ij} k_{ij'} \rho(t - t', \theta_\rho), \tag{3.46}$$

that is a separable correlation function. If the kernels are time varying we then have a nonstationary and nonseparable structure.
The complete model structure includes also dynamic coefficients (see section 3.3.2), and takes $K$ lower triangular

$$\begin{aligned} Z_t &= F_t' \beta_t + K_t x_t & (3.47)\\ \beta_t &= G\beta_{t-1} + \eta_t \qquad \eta_t \sim N(0, \Sigma_\eta) & (3.48)\\ x_t &= \sum_{i=1}^{p} \phi_i x_{t-1} + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2 I_n) & (3.49) \end{aligned}$$

where $F_t$ is the matrix of covariates, and the latent process $x$ is AR(p).

A bivariate case is considered in [18] to model $PM_{2.5}$ and $PM_{10}$ concentrations, through the specification of two independent latent processes. Spatial anisotropy is introduced by allowing the kernel covariance of the processes to depend on wind direction and speed. The model is of the form

$$\begin{pmatrix} y_t^{2.5} \\ y_t^{10} \end{pmatrix} = \begin{pmatrix} \mu_t^{2.5} \\ \mu_t^{10} \end{pmatrix} = \begin{pmatrix} K_t^{fine2.5} 0 \\ K_t^{fine10} K_t^{coarse10} \end{pmatrix} \begin{pmatrix} x_t^{fine} \\ x_t^{coarse} \end{pmatrix} + \epsilon_t \tag{3.50}$$

$$\begin{pmatrix} x_t^{fine} \\ x_t^{coarse} \end{pmatrix} = \begin{pmatrix} x_{t-1}^{fine} \\ x_{t-1}^{coarse} \end{pmatrix} + \nu_t \tag{3.51}$$

where $y_t^{2.5}$ and $y_t^{10}$ are the measured concentrations of the two kind of particulate matter and the underlying processes $x_t^{\text{fine}}$ and $x_t^{\text{coarse}}$ are defined over two different lattices.

## 3.5 Physical modeling

In all dynamical models a central issue is represented by the structure of the transition equation. Here we focus on this problem by considering first order spatiotemporal dynamic models.
Let $\mathbf{Y}_t$ be a vector of length $n$ of the variables at the $n$ spatial locations at time $t$, we have

$$\mathbf{Y}_t = \mathbf{H}\mathbf{Y}_{t-1} + \eta_t \tag{3.52}$$

where $\mathbf{H}$ is a $nxn$ transition matrix and $\eta$ is assumed to be a spatially colored noise process with variance $\Sigma_\eta$. The matrix $\mathbf{H}$ describes how the process at $t-1$ affects the process at next time (in particular the $i-$th row of this matrix links the $i-$th location at time $t$ with every points at time $t-1$), that is

$$Y(s_i,t) = \sum_{k=1}^{n} h(i,k)Y(s_k;t-1) + \eta(s_i,t) \tag{3.53}$$

Typically this matrix can be parameterized in many ways.
The simplest one is a random walk behavior, that is, $\mathbf{H} = \mathbf{I}$ ([53], [54]), but such assumption is not often realistic. A diagonal matrix allows nonseparability if the diagonal terms are assumed different in space, but it is not able to describe propagating phenomena. A tridiagonal matrix represents a nearest neighbor structure when the value of the variable in the nearest neighbor at previous time affects the value at location $s$.
For physical and biological processes it is often the case that the underlying partial differential equations drive the process. It is possible to use this information to develop a coherent prior on $\mathbf{H}$ and $\mathbf{\Sigma}_\eta$. Examples are in [59], [60], [61] [62] and [63].
Wikle ([62]) used a diffusion PDE to motivate the spread of an ecological process. Let $\alpha$ be a spatiotemporal process on a grid driven by a diffusion partial differential equation with linear (Malthusian) growth, that is

$$\frac{\partial \alpha}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x,y)\frac{\partial \alpha}{\partial x}\right) + \left(\delta(x,y)\frac{\partial \alpha}{\partial y}\right) + \beta\alpha \tag{3.54}$$

where $\delta(x,y)$ are spatially varying diffusion coefficients and $\beta$ the growth coefficient, which distribution would be specified at the next level of the hierarchy. Discretizing this equation (with forward difference in time and centered in space) and rearranging the terms we have

$$\alpha_t = \mathbf{H}(\delta,\Delta_t,\Delta_x,\Delta_y)\alpha_{t-\Delta_t} + \mathbf{H}_B(\delta,\Delta_t,\Delta_x,\Delta_y)\alpha^B_{t-\Delta_t} + \eta_t \tag{3.55}$$

where $\alpha$ is a vectorization of the gridded process and $\delta$ is the corresponding vector of diffusion coefficients. The spatially correlated $\eta$ error accounts for uncertainties due to the discretization and other modeling errors. The transition matrix $\mathbf{H}$ is tridiagonal in this case with parameters depending on diffusion coefficients and the discretization interval in time and space. The notation above outlined the need of accounting for boundary conditions with a different sparse transition matrix $\mathbf{H}_B$ representing the transition matrix for the boundary elements of the process, that could be again modeled as a random process at a different stage ([64]). Allowing the diffusion coefficients to vary in space nonstationarity is introduced in a sensible way. Then a spatial random field can be used to model the diffusion parameters, other than the covariates that affect the rate of the diffusion mechanism.

### 3.5.1 Integrodifferential models

Integro differential equations offer another dynamical mechanism that could be more realistic in some applications and that can be also included in hierarchical Bayesian modeling. A general integrodifferential equation for a discrete time and space continuous process $y$ is given by

$$y_{t+1}(s) = \int_{-\infty}^{\infty} k_s(r) g(y_t(r)) dr \tag{3.56}$$

where $y_t(s)$ is a spatiotemporal process at spatial location $s$ and time $t$, $g(\cdot)$ is a function of the y-process (the growth at $r$ between $t$ and $t+1$), and $k_s(r)$ is the redistribution kernel, that links the process at previous time at location $r$ with the process at location $s$ at the next time. This framework allows modeling complicated dynamical processes, including long range dependence and wave fronts with shape and speed depending on the kernel behavior. Wikle ([65], [66]) demonstrated that it can be used to model extra diffusive propagation using a translation of the kernel, and has been used in order to model the complicated dynamical behavior of a precipitation front.
Considering a one dimensional gaussian kernel, we have

$$k_s(r, \theta_1, \theta_2) = \frac{1}{\theta_2 \sqrt{2\pi}} \exp\left(-0.5(r - \theta_1 - s)^2 / \theta_2\right) \tag{3.57}$$

where $\theta_1$ is a translation parameter (the kernel is centered at $\theta_1 - s$), and $\theta_2$ is a dilation parameter. Allowing these parameters to vary in space, a complicated dynamics can be modeled. Assuming $g = \gamma y_t(r)$ we have

$$y_{t+1}(s) = \gamma \int k_s(r; \theta_s) y_t(r) dr + \tilde{\eta}_{t+1}(s) \tag{3.58}$$

where $\eta$ is a spatially colored noise process and $\gamma$ is a parameter that controls explosive growth. This model is nonseparable and, by using spatially varying kernel is also nonstationary.
A stationary version of this model has been previously developed in [67], [68]. The authors explicitated the covariance of the model and analitycally solved the one dimensional case. Moreover the authors showed the time continuous equivalent of their model in terms of stochastic differential equations, and interpreted the parameters of the model in terms of physical variables in a diffusion process. The model is of the same type as above, with

$$Y(s,t) = \theta_\Delta h_\Delta \star Y(s, t - \Delta) + Z_\Delta(s,t) \tag{3.59}$$

where

$$h_\Delta \star Y(s, t - \Delta) = \int_{R^d} h_\Delta(u) Y(s - u, t - \Delta) du + Z_\Delta(s,t) \tag{3.60}$$

where $h_\Delta$ is the kernel, while $\theta$ is a scaling constant, which, to be independent from the choice of $\Delta$, takes the form $\theta_\delta = \exp(-\lambda\Delta)$. $Z$ is a noise term, spatially colored but white in time.
The blurring function $h$ must be non negative and integrate to 1: the simplest choice is a gaussian kernel $g(\mu, \Sigma)$. Brown et al. demonstrates that for a gaussian kernel function and a small time step $\delta$ the model can be rewritten as a stochastic differential equation

$$dY(s,t) = -0.5(AY(\cdot,t))(s)dt + dB(s,t) \tag{3.61}$$

where the function $A$

$$(AX(\cdot))(s) = 2\frac{\partial}{\partial s^T}X(s)\mu - tr(\frac{\partial^2}{\partial s \partial s^T}X(s))\Sigma + 2\lambda X(s) \tag{3.62}$$

is of Ornstein Uhlenbeck form, with the noise term $dB(s,t)$ being a spatially correlated Brownian motion with covariance $c_B(.)dt$. If $c_B(h) = g(h; 0, \Phi)$ with $g$ the gaussian density, the covariance of $Y$ is

$$c_Y(h,k) = \int_{-\infty}^{\infty} \exp\{-\lambda(2v + |k|)\}g(h, k\mu, (2v + |k|)\Sigma + \Phi)dv \tag{3.63}$$

This equation is analytically solved when $d = 1$, resulting in

$$c_Y(h,k) = \frac{1}{2\gamma\sigma}\exp(\frac{\gamma^2\phi^2}{2\sigma^2})\exp(a)G(\frac{\sigma\mu k - \gamma b^2 - \sigma h}{\sigma b}) + \exp(-a)G(\frac{\sigma h - \gamma b^2 - \sigma\mu k}{\sigma b}) \tag{3.64}$$

where G denotes the standard gaussian cumulative probability.

Finally we may notice that this model, in the simplified case with $\mu = 0$ and $\Sigma = I$, is related to the framework of Whittle ([69]).

We can derive an equivalence between this equation and the deterministic advection diffusion equation governing the transport and spread of a pollutant:

$$\frac{\partial c}{\partial t} + u\frac{\partial c}{\partial x} + v\frac{\partial c}{\partial y} + w\frac{\partial c}{\partial z} = \frac{\partial}{\partial x}\left(K_{xx}\frac{\partial c}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_{yy}\frac{\partial c}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_{zz}\frac{\partial c}{\partial z}\right) + E \tag{3.65}$$

where $c$ is the concentration of a pollutant, $(u, v, w)$ are respectively the $(x, y, z)$ wind components and the $K$s are the diffusion coefficients in three dimensions. Equations 3.61-3.62 are equivalent to equation 3.65, if $\mu$ is formed by wind components $(u, v, w)$ and $\Sigma$ is a diagonal matrix with $(K_{xx}, K_{yy})$ as diagonal terms, that is the diffusivity coefficients. The term with $\lambda$ could be the rate of decay (or production) of the pollutant.
It is possible to deal with this model in a reduced dimension form, recognizing the similarity between the IDE model and the reduced Kalman filter

([50]).
Considering the spectral expansion of the kernel and of the process we obtain

$$k_s(r;\theta_s) = \sum_i b_i(s,\theta_s)\phi_i(r) \tag{3.66}$$

$$y_t(s) = \sum_j \alpha_j(t)\phi_j(s) \tag{3.67}$$

where the basis functions $\phi$ are complete and orthonormal and $b_i(s,\theta_s)$ and $\alpha_j(t)$ are the random spectral coefficients for the kernel and the process. Truncating the sum at $I$ we have

$$y_{t+1}(s) = \gamma \mathbf{b}'(s,\theta_s)\alpha_{\mathbf{t}}^{(\mathbf{1})} + \tilde{\eta}_{t+1}(s) \tag{3.68}$$

where $\mathbf{b}(s;\theta_s) = [b_1(s;\theta_s), ..., b_I(s;\theta_s)]'$ and $\alpha_t^{(1)} = [\alpha_1(t), ..., \alpha_I(t)]'$. For locations $(s_1, \ldots, s_n)$ we then have:

$$\mathbf{y}_{t+1}(s) = \gamma \mathbf{B}'_\theta \alpha_t^{(1)} + \tilde{\eta}_{t+1}(s) \tag{3.69}$$

where $\mathbf{y}_{t+1} = [y_{t+1}(s_1), \ldots, y_{t+1}(s_n)]$, and $\mathbf{B}_\theta = [\mathbf{b}(s_1;\theta_{s_1}), \ldots, \mathbf{b}(s_n;\theta_{s_n})]$. Thus:

$$\alpha_{t+1}^{(1)} = \mathbf{\Phi'}_{(1)}\mathbf{B'}_\theta \alpha_t^{(1)} + \eta_{t+1}^{(1)} \tag{3.70}$$

$$\alpha_{t+1}^{(2)} = \mathbf{\Phi'}_{(2)}\mathbf{B'}_\theta \alpha_t^{(1)} + \eta_{t+1}^{(2)} \tag{3.71}$$

where $\alpha_{\mathbf{t}}^{(\mathbf{2})} = [\alpha_{I+1}(t), \ldots, \alpha_n(t)]$, $\mathbf{\Phi}_{(1)} = [\phi_1, \ldots, \phi_I]$, $\mathbf{\Phi}_{(1)} = [\phi_{I+1}, \ldots, \phi_n]$. $\eta$ is normally distributed with covariance function $\mathbf{C}_\eta^{(j)} = \mathbf{\Phi'}_{(j)}\mathbf{C}_{\tilde{\eta}}\mathbf{\Phi}$.
Assuming a gaussian kernel with parameters $\theta_1(s)$ and $\theta_2(s)$ and Fourier basis functions, then the Fourier transform of the gaussian kernel is its characteristic function:

$$b_j(s;\theta_1(s),\theta_2(s)) = \exp[i\omega_j(\theta_1(s)+s) - 0.5\omega_j^2\theta_2(s)] \tag{3.72}$$

where $\omega_j$ is the spatial frequency. In [70] the kernel parameters are modeled as a stationary spatial field.

## 3.6 Reduced dimension spatiotemporal processes

As we saw above the dimensionality of many spatiotemporal problems is often problematic or prohibitive. The spectral representation can be useful to address this problem by using fast computation algorithms or to produce a simpler structure in the dynamical evolution. In fact let $\mathbf{Y}_t = \mathbf{K}\alpha_t$, where $\mathbf{Y}$ is a vectorization of the process at $n$ spatial locations on a regular lattice,

$\mathbf{K}$ is a $nxn$ matrix of spectral basis functions (Fourier, wavelet, empirical). It is often the case that the transformed spectral process evolves dynamically, i.e. $\alpha_t = \tilde{\mathbf{H}}\alpha_{t-1} + \tilde{\eta}$, with a simpler structure in the transition matrix $\tilde{H}$ or in the error term $\tilde{\eta}$. For example for Fourier basis functions when weakly stationarity is assumed, the $\alpha$ terms are nearly uncorrelated, resulting in a diagonal covariance matrix. Moreover the transformation allows dimension reduction, due to the fact that often only a few coefficients could describe adequately the dynamics of the process.
The low rank general representation (see section 1.6) can be introduced also in the context of dynamical models, for example the reduced Kalman filter falls into this general framework.
Consider the decomposition $\mathbf{Y}_t = \mathbf{K}\alpha_t + \gamma_t$, where $\gamma_t$ could be a spatially correlated noise, and $\alpha_t$ a $p$ reduced dimension latent process. The state space representation becomes,

$$\mathbf{Z}_t = \mathbf{K}\alpha_t + \gamma_t + \epsilon_t \tag{3.73}$$

$$\alpha_t = \tilde{\mathbf{H}}\alpha_{t-1} + \tilde{\eta}_t \tag{3.74}$$

where $\epsilon$ is a pure gaussian noise, $\gamma_t \sim N(0, \Sigma_\gamma)$ accounts for spatial dependence, and the $\tilde{\eta}_t \sim N(0, \Sigma_\eta)$. As in spatial modeling many basis functions can be chosen to form the $\mathbf{K}$ matrix. Examples in literature are [18], [49], [53], [50], [61], [59], [52].

Empirical orthogonal functions (EOF) and kernel basis functions are two important classes of basis functions, one orthogonal and the other not, that we will use in the application in Chapter 4.
In the following we describe in detail empirical orthogonal basis functions that will be used in the model introduced in section 4.6. For what concerns kernel basis functions, see section 1.3.1.

### 3.6.1 Empirical orthogonal functions

As we saw in section 1.6 empirical orthogonal functions derive from a discretization of the KL expansion. This technique, popular in meteorological applications since its introduction by Lorenz in the 50s, is designed to find orthogonal basis function driven by the data. This is the analogous of PCA analysis but in the context of spatiotemporal data, working with time series and spatial patterns. Let $Y$ be a spatiotemporal process observed at locations $(s_1, \ldots, s_n)$ for T time steps. Arranging the values of that process in a $Txn$ matrix, say F, with the i-th row corresponding to a map of the process at time $i$ and the $j$th column corresponding to a time serie at location $j$. The rationale is to find an orthonormal basis $\Psi$ such that the variance of the transformation $F\Psi$ is maximized, that is $\psi_k = (\psi_k(s_1), \ldots, \psi_k(s_n))'$ is such that $var(a_k(t)) = var(\psi_k' f_t)$, where $f_t$ is the row $t$ of $F$, is maximized with

$\psi_k$ orthonormal. This can be viewed as an eigenvalue problem. After removing the mean from each time series (that is, columns) consider the covariance matrix $C = F^T F/(T-1)$. Solving the eigenvalue problem $C\Psi = \Psi\Lambda$, we find the eigenvalues $\lambda_i$ of the covariance matrix (on the diagonal of the diagonal matrix $\Lambda$) and the corresponding (orthonormal) eigenvectors on the respective column of the matrix $\Psi$. These eigenvectors are the basis we are looking for, and called in this context *empirical orthogonal functions*. Normally the eigenvectors are ordered according to the magnitude of the corresponding eigenvalue, so the first EOF corresponds to the biggest eigenvalue. Each eigenvector can be viewed as a map, and time evolution of each eigenvector can be obtained as

$$\mathbf{a}_i = F\psi_i \tag{3.75}$$

where $\psi_i$ is the $i$th EOF and $\mathbf{a}_i$ the corresponding time series, called expansion coefficients.
The original data is obtained as $F = \sum_{i=1}^{n} \mathbf{a}_i\psi_i$. Typically EOF are used to reduce dimension by truncating the reconstruction of the original data at some $j << n$, assuming that the first $j$ eigenvectors are able to capture the most relevant features of the system. The symmetric matrix $C$ can be written in terms of EOFs as

$$C = \lambda_1\psi_1\psi_1^T + \ldots + \lambda_n\psi_n\psi_n^T \tag{3.76}$$

with the $j$th eigenvalue explaining $\lambda_j/\sum_i \lambda_i$ percent of the total variance. This decomposition corresponds to a KL discretization if equal areas of influence are assumed for each observation.
The EOFs technique can be viewed as a particular case of Singular Value Decomposition (SVD), that is a general decomposition of the $T$x$n$ matrix $F$ in the product of two orthonormal matrices and one diagonal

$$F = U\Gamma V^t \tag{3.77}$$

where $U$ is a $T$x$T$ orthonormal matrix, $V$ is a $n$x$n$ orthonormal matrix and $\Gamma$ a diagonal $T$x$n$ diagonal matrix with $r = rank(F) \leq min(T, n)$ diagonal elements, called singular values of the matrix, while the column vectors of $U$ and $V$ are called singular vectors. By using this decomposition we can view the spatiotemporal process as the product of a spatial and a temporal process, by thinking $\mathbf{u}_l = (u_l(s_1), \ldots, u_t(s_n))'$ and $\mathbf{v}_l = (v_l(1), \ldots, v_l(T))'$, $F = \sum_{l=1}^{T} \gamma_l \mathbf{u}_l \mathbf{v}_l'$.
When the number of singular values is less than $T$ and $n$, some of the singular vectors are redundant and one can write $F$ with a reduced matrix $F = U_r\Gamma_r V_r^t$. It can be proven the equivalence with EOFs method, that is $C = \Psi\Lambda\Psi' = V\Gamma'\Gamma V'$, so $\Psi = V$ and the eigenvalues are equal to the square of the corresponding diagonal term of $\Gamma$.

## 3.7 Spatiotemporal models for atmospheric pollution

There is an extensive literature on the application of the statistical methods we outlined so far to the analysis and prediction of atmospheric pollution. We have already described some of these models in the previous sections. Here we review a partial selection of other examples of applications to the problem of atmospheric pollution.

A relative simple, non dynamical model but with a nonseparable structure is introduced in [71], modeling $PM_{2.5}$ daily concentrations over 124 irregularly spaced monitoring stations in US Midwest, where two separate spatiotemporal processes are introduced for urban and rural area. Let $Z(s_i, t)$ be the square root of PM concentrations, the authors defined this hierarchical structure

$$\begin{aligned} Z(s_i,t) &= Y(s_i,t) + \epsilon(s_i,t) \qquad i = 1,\dots,n, \qquad t = 1,\dots,T && (3.78)\\ Y(s_i,t) &= \mu(s_i,t) + w(s_i,t) + p(s_i)v(s_i,t) && (3.79)\\ & && (3.80) \end{aligned}$$

where the mean $\mu$ is modeled in terms of population density of an urbanity indicator and of a seasonality indicator. $w(s_i, t)$ and $v(s_i, t)$ are independent zero mean spatiotemporal processes: weighting $v(s_i, t)$ by population density a *urban* spatiotemporal behavior is added to a background spatiotemporal process. While a separable structure is used for both the processes, the combination of the two yields a nonseparable structure. A Bayesian approach is adopted for inference.

Shaddick and Wakefield ([72]) modeled a multivariate series of four pollutants on eight monitoring stations for a 4 year period using a hierarchical Bayesian dynamic linear model. The residual spatiotemporal behavior of this multivariate process is splitted into a pure stationary spatial process and a multivariate autoregressive temporal process.

A hierarchical space time modeling for $PM_{10}$ pollution has been developed in [73], using site specific and meteorological variables from a meteorological model with an additive form for the spatial and temporal residual terms. Isotropic stationary covariance for the spatial effect and a random walk dynamics for the temporal term has been used in this case: identification of the sources of variability demonstrated that the latter is the most relevant process. Sahu et al. ([74]) explained space time behavior of ozone in Ohio using increments in meteorological variables. They account for spatial variability using a stationary exponential family and a autoregressive term for the residual temporal dynamics.

A model with dynamic coefficients is used also in [54] to model the ozone levels in Mexico city with a coupled spatiotemporal process of temperature.

Let $Y_{it}$ be the square root of ozone measures on the $i$th station at time $t$, the model is of the type

$$Y_{it} = \beta_t^y + S_t'(\mathbf{a})\alpha_{it} + Z_{it}\gamma_{it} + \epsilon_{it}^y \tag{3.81}$$

where $\beta_t^y$ is a spatial trend, $S$ is a vector representing two periodical components corresponding to 12 and 24 hours period and $Z$ is the temperature which is modeled at the next step of the hierarchy. Both the $\beta$s and $\gamma$s coefficients are modeled as a random walk; an autoregressive dynamics with unit coefficient is used also for the $\alpha$s but a stationary spatially correlated error term is added. The use of conditional distribution of the ozone given the temperature is analogous to that developed in [28], where only a pure spatial context is considered.
Space time dynamic coefficients are used in [75] to model ozone dynamics, using a nonstationary and nonseparable covariance function developed by Fuentes ([46]). Nonstationarity is achieved by using a spectral representation, yielding a weighted sum of locally stationary (but eventually nonseparable) gaussian processes, with a fixed kernel used as weighting function. Space time dynamics coefficients are used also in [76] to model the different chemical components (sulfate, nitrate ecc) of fine particulate matter as a multivariate process for which a linear model of coregionalization is introduced. The authors also deal with the problem of linking two sources of the same information, that is the total mass of PM measured by a monitoring network and the sum of the partial components measured by another network. In the context of Kalman filter we cite, among others, [47], [77], and [78]. Huang and Hsu introduced transport effects in ozone modeling, using a nonseparable spatiotemporal covariance function that depends on wind speed and wind direction and thus is nonstationary in time and space. The proposed state space model is

$$Z(s,t) = S(s,t) + \epsilon(s,t) \tag{3.82}$$

$$S(s,t) = \mu(s,t) + Y(s,t) + \nu(s,t) \tag{3.83}$$

where $\mu$ is a deterministic mean process, $\nu(s,t)$ is a spatially stationary process and $Y(s,t)$ is the zero mean spatiotemporal process. This last term is then modeled as an IDE model

$$Y(s,t) = \int_D w(s,u,x(s,t-1))Y(u,t-1)du + \eta(s,t) \qquad s \in D, t \in N \tag{3.84}$$

where $\eta$ is a spatial error and $w$ is a weight function depending on the wind field $x$. The model is then rearranged in a reduced form by using empirical orthogonal functions and Kalman filtering has been used for parameter estimation.
A state space representation (a kriged Kalman model) has been used also by Sahu and Mardia ([77]) to model the $PM_{2.5}$ concentrations. At the process

level a spatial isotropic correlation is added to a time component, assumed to evolve as a stochastic time varying linear combination of some optimal functions. A fully Bayesian estimation is then performed. On the contrary Fassó and Cameletti used EM algorithm to estimate a state space model for $PM_{10}$.
A reduced rank state space model, using kernel basis, with fully Bayesian estimation, is that in [57], describing the space time evolution of a set of five correlated pollutants, in [18] applied to the correlated evolution of $PM_{10}$ and $PM_{2.5}$. As we saw in section 3.4 a discrete process convolution approach with gaussian stationary kernel is embedded in a state space dynamic model, and the bivariate behavior is modeled following the basic idea of building processes that share part of a common latent process.

Another kind of question is that of the statistical models built to compare the results of deterministic models with measurements. Fuentes et al. developed a statistical framework to evaluate the performance of air quality numerical models, also addressing the problem of misalignments and change of support between measures and predicted values ([79], [80], [81]). In a hierarchical setting both the observations and the predicted values are modeled as (stochastic) function of a hidden true process, that is then modeled using a nonstationary covariance function. The same approach is used in [82], with a hierarchical Bayesian model applied to evaluate the performance of CAMx model (see EPA) in ozone prediction.

# Chapter 4

# Application

In section 2.3.1 we described the main characteristics of nitrogen dioxide and nitrogen oxide, their toxicity, the relationship between these two pollutants and how we can model them in a bivariate spatial setting. Now we consider the spatiotemporal behavior but in a univariate setting, that is considering the total concentration of NO plus $NO_2$, together referred to as nitrogen oxides ($NO_x$).
The aim here is to build a spatiotemporal model suitable to describe the spatiotemporal dynamics of this pollutant over the Tuscany region, by using the methods described in the previous sections to allow the model to include nonseparability and nonstationarity in time and space. Basically we use the kernel convolution approach in a dynamical and hierarchical framework.
This section is organized as follows. First we briefly describe in more detail the characteristics of $NO_x$ making clear the importance of modeling this pollutant. Then, after an illustration of the available data set, we proceed to an explorative data analysis, useful for a proper choice of the modeling strategy. After that we concentrate on the modeling approach, comparing the different models proposed and evaluating the results. Finally we introduce a completely different approach as a future research field.

## 4.1 Why nitrogen oxides?

Adverse health effects, in particular for respiratory apparatus, are related to both long-term exposure to high concentrations and short term exposure to very high concentrations of nitrogen dioxide. Thus limit values are needed to prevent population from high exposure and health risks.
World Health Organization sets guideline values for $NO_2$, that are an 1-hour level of 200 $\mu g/m^3$ and an annual average of 40 $\mu g/m^3$. Existing italian law follows this guideline (D.M. 2.04.2002 N.60, 1999/30/CE).

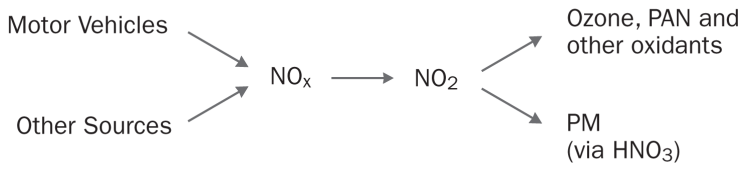


Figure 4.1: Sketch of relationship between nitrogen oxides and oxidants and particulate matter.

$NO_2$ and other nitrogen oxides are precursors for a number of harmful secondary air pollutants such as ozone and particulate matter, and play a role in the formation of acid rain. In fact $NO_2$ is subject to extensive further atmospheric transformations that lead to the formation of ozone and other strong oxidants that participate in the conversion of $NO_2$ to nitric acid and sulfur dioxide to sulphuric acid and to subsequent conversions to their ammonium neutralization salts. Thus, through the photochemical reaction sequence initiated by solar-radiation-induced activation of $NO_2$, the newly generated pollutants are an important source of nitrate, sulphate and organic aerosols that can contribute significantly to total PM10 or PM2.5 mass (see figure 4.1). $NO_x$ contributes to acid deposition and eutrophication which in turn can lead to potential changes occurring in soil and water quality, while the subsequent impacts of acid deposition can be significant.

Combustion of fossil fuels is by far the dominant source of $NO_x$ emissions. High temperatures and oxidation-rich conditions generally favor $NO_x$ formation in combustion, with formation rate being primarily a function of temperature and of the residence time of nitrogen at that temperature.
Nitrogen oxides emissions have decreased by 31% between 1990 and 2005 in European union. Despite the decreasing trend of the emissions of $NO_x$ in Europe, this is lower than that of other pollutants (see figure 4.2).

Over the past 50 years vehicular traffic has largely replaced other sources (e.g., domestic heating, local industry) as the major outdoor source of $NO_x$ from fossil fuel combustion, and hence of $NO_2$, over the entire Europe, in particular in urban environment (see figure 4.3).

In the period 1996-2005, 21-47% of urban population was potentially exposed to ambient air nitrogen dioxide concentrations above the EU limit value of 40 $\mu g/m^3$ as annual mean, even if there was a slight downwards trend over the period (see figure 4.4).

The exposure of individuals to $NO_2$ from outdoor sources depends largely on their proximity to vehicular traffic in space and time, given that mobile sources are the chief contributors to ambient $NO_2$ in European cities.
Ambient $NO_2$ concentrations measured at fixed urban sites may not ac-

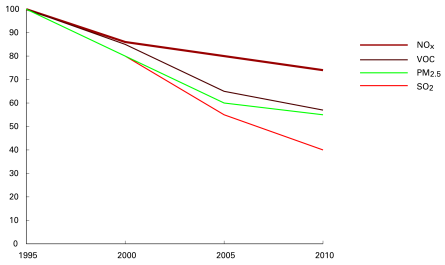


Figure 4.2: Trend in the reduction of emissions in Europe (EEA database).

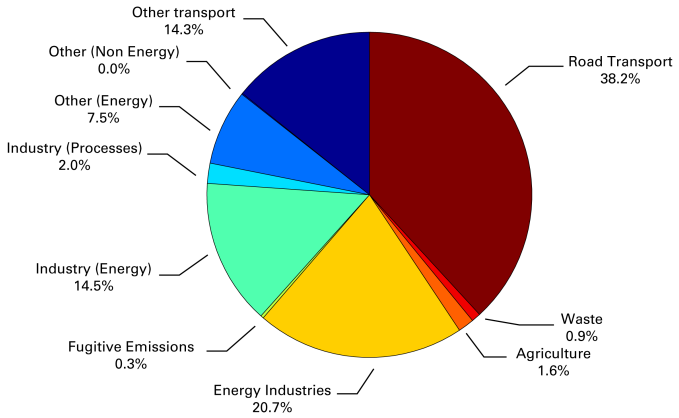


Figure 4.3: Contribute of total nitrogen oxides emissions in Europe by sources (2007) (EEA database).

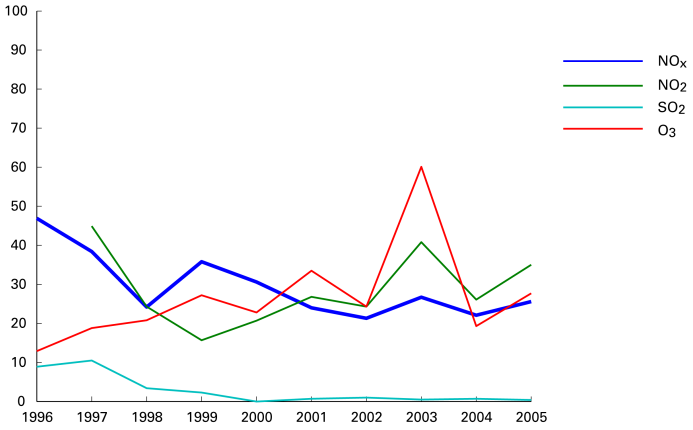


Figure 4.4: Estimated percentage of population exposed to a concentration of $NO_2$ above the limit value of 40 $\mu g/m^3$ in Europe cities (EEA).

curately reflect personal exposure to $NO_2$ from outdoor sources, because ambient $NO_2$ concentrations vary widely due to traffic patterns, the characteristics of built environment and meteorological conditions. Fixed monitoring stations are not necessarily sited with the intent of reflecting the population average exposure, therefore the accuracy with which their measurements reflect population exposure may vary.
In our case different kind of stations, that is traffic of background, are useful to capture both maximum values near major roads and medium and large scale dynamics of transport and diffusion.

## 4.2 Data set

### 4.2.1 The monitoring network of $NO_x$

As we saw in section 2.3.1 the network of air pollution monitoring stations of the Tuscany region measures the concentrations of several pollutants in a set of locations spread across the region.
There are 97 monitoring stations spread over the Tuscany region and the most are set in the context of main cities. Measured pollutants are $PM_{10}$, $SO_2$, CO, $NO_x$, $O_3$ and $H_2S$. Moreover, benzene and benzoapyrene are measured in a few special stations. Each station has been working since different date, starting from 1992 until now. Following the italian regulation we have different kind of stations with respect to monitored zone (rural, urban and suburban) and to monitoring type (background, traffic or industrial), see

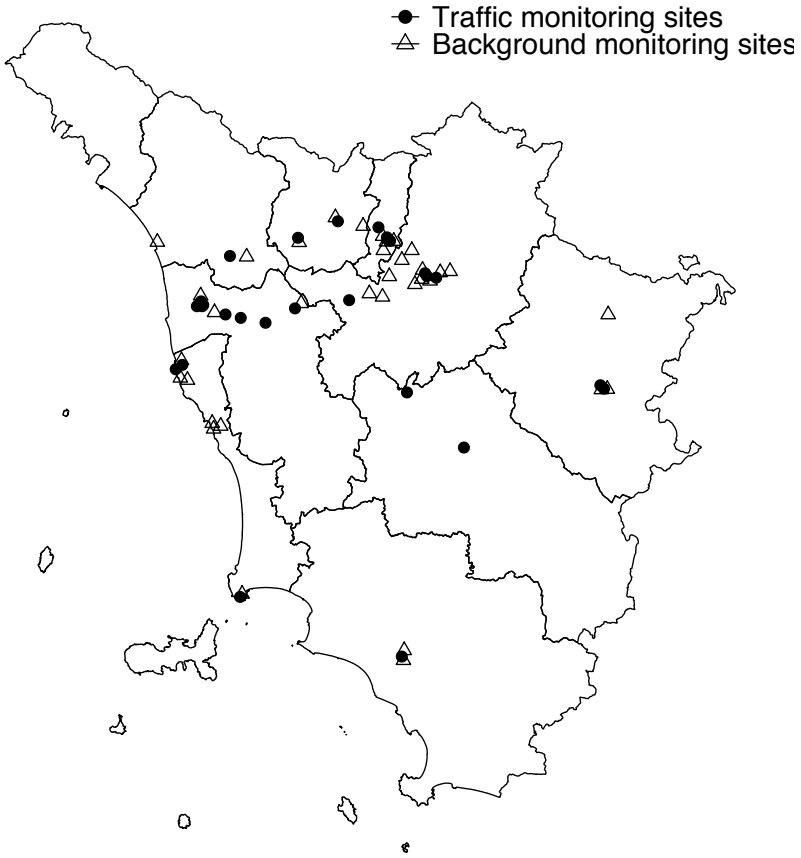


Figure 4.5: ARPAT monitoring network map, traffic stations (circles) and background stations (triangles).

figure 4.5.
We have meteorogical data for the 2005 year and we are interested in the number of $NO_x$ monitoring stations that were working in that period, that is 55 monitoring stations.
Other available data are pollutant emissions and meteorological conditions, and informations on population density and other geographical issues as morphological features and land use (see figure 4.6).

### 4.2.2 Emissions

Regional archives of emissions are available for the Tuscany region (IRSE). The estimated emissions of $SO_x$, $NO_x$, VOC, CO, $PM_{10}$, $NH_3$ and others are present in this database. The spatial resolution is that of the administrative level of municipality, but is possible to have emission values on a 1x1 km grid as well. Emissions are disaggregated by source type, i.e. point emis-

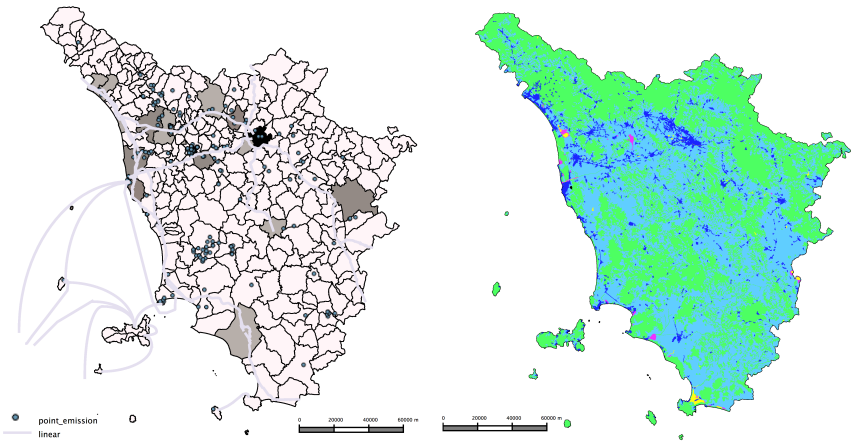


Figure 4.6: (left) Point, linear and diffused emissions at municipality and grid level. (right) Land cover (according to CORINE classification) for the region.

sions (industrial stacks), linear emissions (highways, boats etc) and diffused emissions (all the other small sources). In Fig. 4.6 the available data are shown.

### 4.2.3 Meteorological variables

The Regional Atmospheric Modeling System - RAMS - was run by the University of Florence (Dept. of Civil Eng.) for the 2005 year, from march to November. This is a limited area atmospheric model with complete physics and non-hydrostatic equations, developed at the Colorado State University. RAMS solves a set of equations that describe dynamics and thermodynamics of the atmosphere, mass and energy conservation and hydrometeors microphysics. The initial and boundary conditions used are that of the ECMWF (European Center for Medium Range Weather Forecasts) model, with the model run in diagnostic mode and data assimilation performed from satellite data. A validation activity for these data is now in progress.
The data are characterized by having a horizontal resolution of 4 X 4 km (47 X 52 grid points for the entire Tuscany region) and 14 vertical levels, from 48 to 3732 m above ground level (AGL).
Simulated variables are, among others: wind speed, temperature, pressure, relative humidity, precipitation rate, long and short radiations, for all the vertical layers (some of these are also evaluated at 2m AGL). The model also provides a set of variables to directly or indirectly obtain micrometeorological variables according to the Monin Obukhov similarity theory.

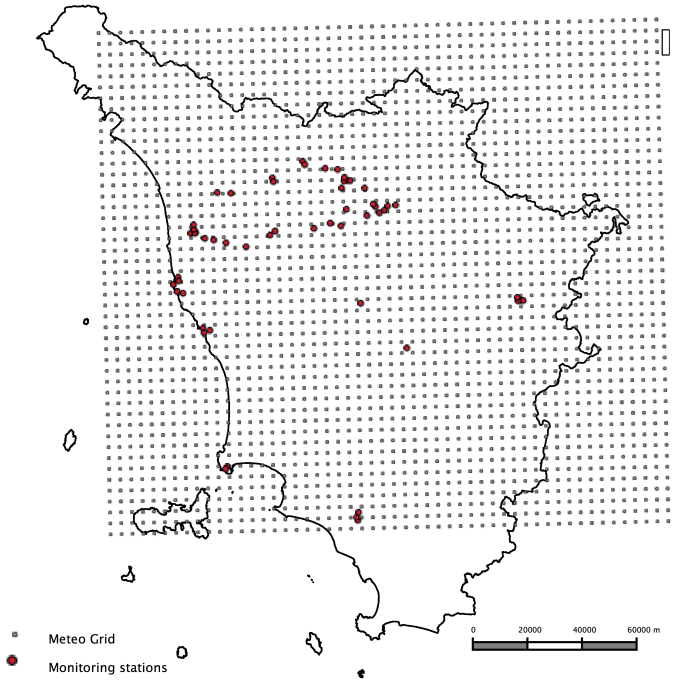


Figure 4.7: Simulation grid of the meteorological model, with a grid resolution of 4x4 km.

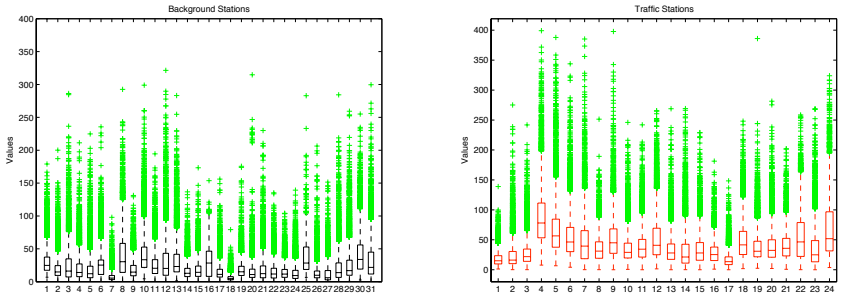


Figure 4.8: Box-plot of measured concentrations for the 55 stations (original scale), both for background stations (left) and traffic (right).

A similarity theory is an empirical method of finding universal relationships between variables that are made dimensionless using appropriate scaling factors. Similarity methods have proved very useful in the atmospheric boundary layer, where the complexity of turbulent processes precludes direct solution of the exact governing equations.
The Monin Obhukov theory is formed by a set of relationships describing the vertical behavior of nondimensionalized mean flow and turbulence properties within the atmospheric surface layer (the lowest 10% or so of the atmospheric boundary layer) as a function of some key parameters. These key parameters are the height $z$ above the surface, the buoyancy parameter ratio between inertia and buoyancy forces, the kinematic surface stress, and the surface flux of virtual temperature.
The key parameters can be used to define a set of four dimensional scales for the surface layer: a velocity scale (friction velocity), a surface-layer temperature scale, a length scale (Obukhov length) and the height above ground scale. These key scales can then be used in dimensional analysis to express all surface-layer flow properties as dimensionless universal functions of them. As all these variables play a key role in describing the turbulence status of the atmosphere, they also play an important role in the dispersion mechanism for a pollutant and thus in the measured concentration level.

## 4.3 Data Analysis

We have hourly time series of $NO_x$ ($\mu g/m^3$) concentrations over 55 monitoring stations, for the period ranging from the first of march to the 30 of November 2005, corresponding to 6600 hours. All the stations show some missing data, due to failure of the monitoring instruments. Considering the whole data set there is a total of 48870 missing data with respect to a full series of 363000 data, corresponding to about 15%. This percentage is vari-

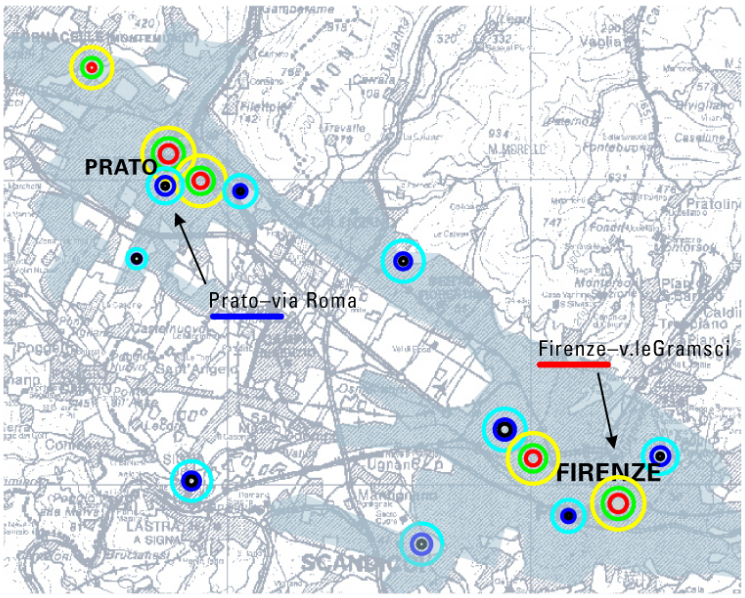


Figure 4.9: Locations of a set of monitoring stations: circles are proportional to 0.025,0.5 and 0.975 quantiles of nitrogen oxides concentrations in each stations. Different colors indicate traffic (inner circle in red) versus background stations (inner circle in black).

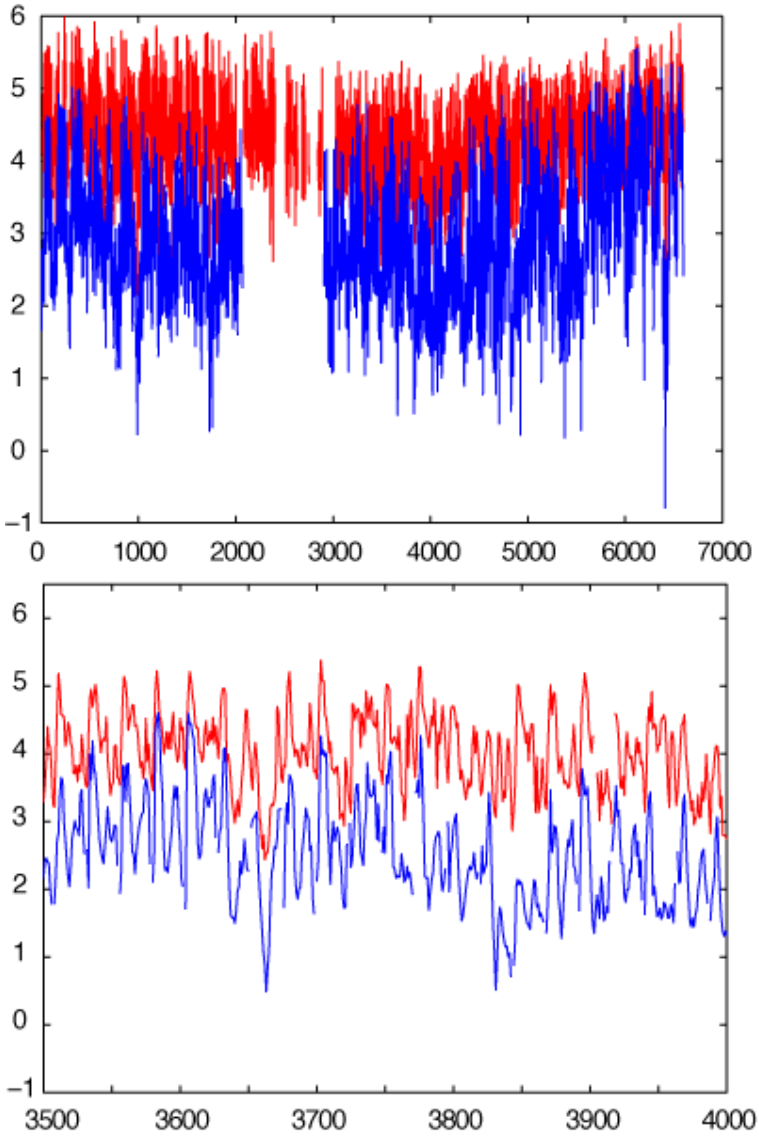


Figure 4.10: Temporal evolution of $NO_x$ (log-scale) as measured in two different monitoring stations, background (blue) and traffic (red), for the entire period (March-November 2005, 6600 hours) (top) and for a sub period (bottom).

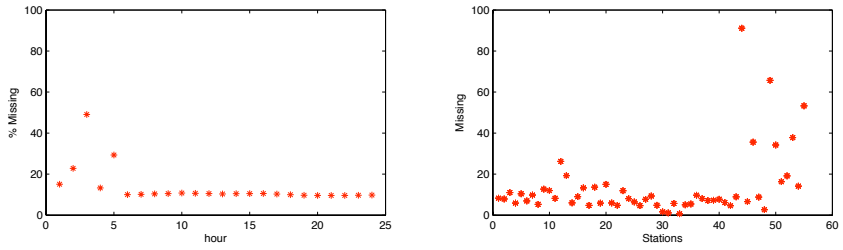


Figure 4.11: Percentage of missing data at different stations (left) and at different hours (right)

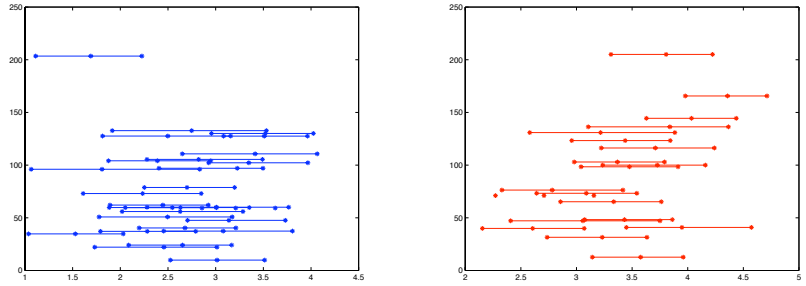

Figure 4.12: Concentrations versus emissions (Mg per year) falling into a radius of 1 km from the monitoring station in background (left) and traffic (red) stations (IRSE).

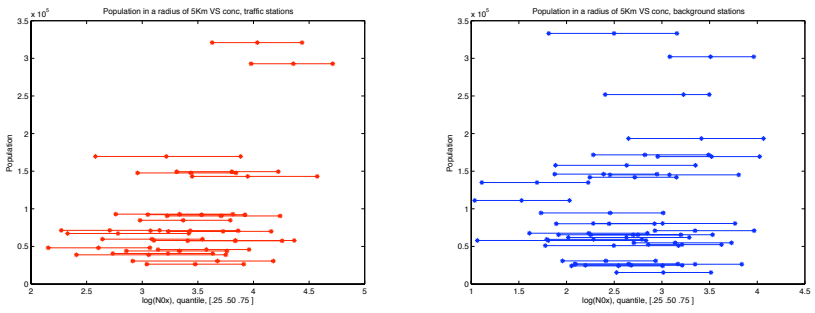


Figure 4.13: Concentrations versus resident population in a 5 km radius area, for background (left) and traffic (red) stations.

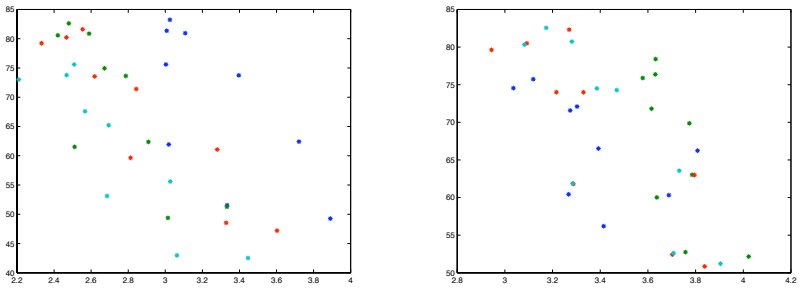

Figure 4.14: Monthly average concentrations versus monthly average temperature. Different colors indicate different stations

able from station to station with a minimum of near zero percent (0.6) and a maximum value close to 90. Moreover the most of missing data appear in the early morning (3, 4 and 5 am), corresponding to hours of low concentrations and low exposure risk, and thus chosen for calibration procedures. Monitoring stations are classified as *background* or *traffic* with respect to distance from emission sources. While traffic stations are located very close to roads, background stations are located in sites as far as possible from emissions sources, trying to measure the value of the pollutant over a large representative area. Of course traffic monitoring stations measure higher concentrations values (see Fig. 4.8 and 4.9), with a daily cycle reflecting daily traffic pattern. The log value of time series for two sample stations are shown in Fig. 4.10.
We have meteorological variables for the whole period, without missing data, over a 4x4km grid . The domain does not cover the entire Tuscany region, but all the monitoring stations fall into the coverage area. Variables corresponding to the nearest grid point are assigned to each station. The data set includes, among others, temperature, wind speed and some variables of the similarity theory, like friction velocity $u^*$, the velocity scale, or sensible heat flux. Figure 4.14 shows monthly average of temperature with respect to monthly average of $NO_x$ concentrations for a subset of monitoring stations. For what concerns yearly emissions we obtain a value for each station by adding all the emissions (linear, point and diffuse, expressed in Mg per year) falling into a radius of 1 km from the monitoring station. This approach has some drawbacks (for example the emissions could be not smooth enough due to the rough inclusion of point emissions), but it is the simplest way to proceed. Figure 4.12 shows concentration values with respect to emission values for each station, and indicates a positive dependence, although less strong than expected, probably due to the estimation procedure in the database of emissions. The strong relationship between emissions and popu-

lation density yields to a very similar behavior of these variables with respect to concentrations (see Fig.4.12 and 4.13).

In the following we report a short summary of the preliminary work on the data we have done to provide as the tools for a proper choice of model variables and structure. A series of nested regression equations (with least square estimations) make us able to test some hypothesis regarding the main variables involved and the spatiotemporal behavior and correlation structure.
The first step we made is to link concentration values to emissions and temperature covariates, considering the regression

$$Y_{it} = \mu_i + \beta_i^T T_{it} + \beta^E E_i + \epsilon_{it}^a \tag{4.1}$$

where $Y_{it}$ is (log) concentration of $NO_x$ at site $i$ at time $t$, $\mu_i$ is a spatial varying mean, $T_{it}$ is temperature at the RAMS point nearest to station $i$ at time t, and $E_i$ is the total $NO_x$ yearly emission value within a distance less than $1km$ from the station, and $\beta_i^T$ and $\beta_E$ the corresponding regression coefficients. We found both emissions and temperature to be significative.
We performed a spectral analysis of the residuals that showed a periodicity of period 24 hours and 7 days, corresponding to daily and weekly cycles.
We can thus model these residuals with the corresponding sinusoidal components

$$\epsilon_{it}^a = \alpha_i^1 \cos(2\pi t/24) + \beta_i^1 \sin(2\pi t/24) + \alpha_i^7 \cos(2\pi t/168) + \beta_i^7 \sin(2\pi t/168) + \epsilon_{it}^b \tag{4.2}$$

with a different coefficient for each station $i$. Once estimated all these coefficients can be viewed as a spatial process, and so we checked the spatial structure with both visual inspection and variogram.

The autoregressive structure of the residuals is supported by the estimation of

$$\epsilon_{it}^b = \Phi_i \epsilon_{i,t-1} + \eta_{i,t} \tag{4.3}$$

Now we introduce a (gaussian and univariate) kernel in order to reduce the dimensionality of the spatial process, by writing

$$\epsilon_t^b = \mathbf{K}\mathbf{a}_t + \nu_t \tag{4.4}$$

where $\epsilon_t$ is the vector of $n = 55$ dimensions and $\nu_t \sim N(0, \sigma_\nu^2 I_{nxn})$.
The element of the $nxp$ kernel matrix, where $p$ is an arbitrary number of reference points, is $k_{ij}(\theta, s_i - r_j)$, with $i = 1, \ldots, n$ and $j = 1, \ldots, p$, and the kernel depending on the $\theta$ parameter and the distance between measurements locations and reference points. The $\theta$ parameter can be spatially varying but we assume now a constant (and known) value.

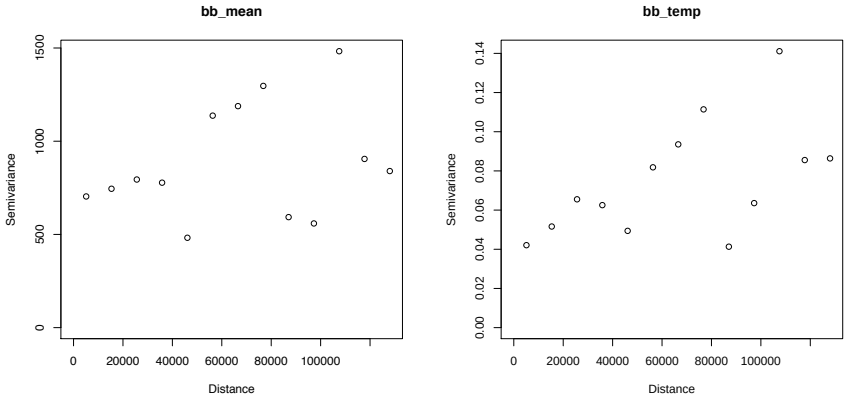


Figure 4.15: Estimated variograms for (left) estimated mean ($\mu_i$ in equation 4.1) and (right) temperature coefficients ($\beta_i^T$ in equation 4.1).

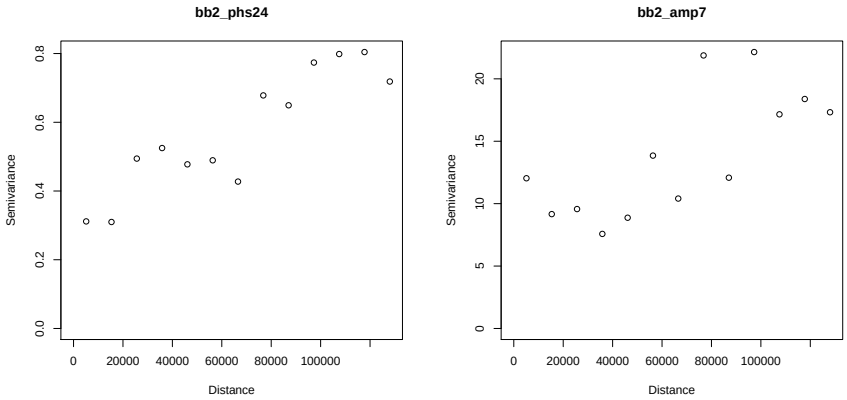


Figure 4.16: Estimated variograms for (left) phase of the 24 hour periodicity and (right) amplitude of 7 days periodicity (see equation 4.2)

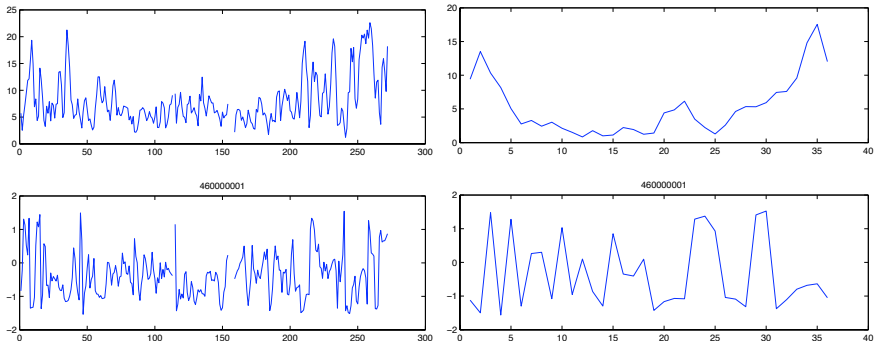


Figure 4.17: Temporal evolution of amplitude (top) and phase (bottom) of 24 hours (left) and 7 days (right) cycles as estimated by moving average estimation for a sample station (Lucca-Capannori).

Once $\mathbf{a}_t$ has been estimated (as MSE) for each time $t$, we fit an autoregressive model similar to that of equation 4.3 but in a reduced dimension setting

$$\mathbf{a}_t = M\mathbf{a}_{t-1} + \gamma_t, \qquad \gamma_t \sim N(0, Q) \tag{4.5}$$

The temporal evolution of the coefficients associated to 24 hours and 7 days cycles has also been checked, by a moving average estimation of $\alpha_i^1$, $\beta_i^1$, $\alpha_i^7$, $\beta_i^7$ of equation 4.2, with a temporal window of three days with one overlapping day for coefficients of 24hour cycle, and a temporal window of three weeks, with one overlapping, for the 7 days cycle.
This analysis outlines the temporal evolution for these coefficients (see figure 4.17), that will be included in our modeling strategy.

## 4.4 Tools for model choice

As we are dealing with a complex model, where a lot of different choices can be made with regard to both modeling strategies (kind and parameters of the kernel, number and locations of reference points ecc) and to prior assumptions, it is important to have appropriated tools for model validation and selection. Here we focus on a Bayesian framework and on the methods developed in this context.
Although, or maybe for, the crucial role of model selection, different approaches can be found in literature. Before describing a short selection of these methods, it is worth noting that none of them are free of drawbacks and that there not exists general agreement in literature about which would turn to be the right approach.

Given a model selection problem in which we have to choose between two models, M1 with priors on parameters $\pi(\theta_1)$ and M2, with priors on parameters $\pi(\theta_2)$, and data vector $Y$, the Bayes factor is given by the ratio of the observed marginal densities for the two models

$$p(y|M_i) = \int f(y|\theta_i, M_i)\pi_i(\theta_i)d\theta_i, \qquad i = 1, 2 \tag{4.6}$$

This ratio becomes a likelihood ratio test when the two models share the same parameterization and the hypothesis are both simple. However when the priors are improper, Bayes factor is not well defined, while numerical instabilities also arise when the priors are proper but diffuse.
Likelihood ratio test statistics is the base for the AIC and BIC criterions (Akaike and Bayesian information criterion respectively). Both these methods account for the number of parameters, including a term that acts as a penalty for complexity.
A generalization of the AIC is the so called deviance information criterion (DIC, [39]), that has become very popular in recent years. A term accounting for the goodness of fit is added to a model complexity indicator. The key quantity is the deviance,

$$D(\theta) = -2\log f(y|\theta) + 2\log h(y) \tag{4.7}$$

where $h(y)$ is a normalizing function of data alone, and $f(y|\theta)$ is the likelihood. The effective number of parameters is defined as

$$p_D = \overline{D(\theta)} - D(\bar{\theta}) \tag{4.8}$$

where $\overline{D(\theta)} = E_\theta[-2\log f(y|\theta)|y] + 2\log h(y)$ is the posterior mean deviance that can be regarded as a measure of fit, and $\bar{\theta}$ is an estimate of $\theta$, for example the posterior mean or mode or median of $\theta$. Thus $DIC$ is defined as

$$DIC = \overline{D(\theta)} + p_D \tag{4.9}$$

with the smallest value of DIC indicating a better fit, and only differences in DIC values being significant for model comparison.
The DIC criterion, well suitable for gaussian models, can be computed easily within a MCMC computation, but there are some limitations and criticism about this method. DIC is not invariant to parameterization, and it is dependent on which part of the model is considered as part of the likelihood. This *focus* issue can be problematic in some cases.
In missing data models different focus yields to different definition of DIC, depending on the choice of the likelihood function ([83]). Given the observed data $Z$ and a latent stage $Y$, one can consider the observed likelihood $f(Z|\theta)$, the complete likelihood $f(Y, Z|\theta)$ or the conditional likelihood $f(Z|Y, \theta)$. Celeux et al.([83]) explored this field defining eight variation of

DIC.
Recently Plummer ([84]) suggests some limitations on DIC fields of application and provides improved asymptotic approximations. Moreover the asymptotic justification of DIC seems to be inappropriate for models with more than two hierarchical levels.

Another approach to Bayesian model choice is based on hypothetical replicates from the same process that generated the data, the so called *posterior predictive approach*. In this context replicate data sets are simulated from the posterior distribution of model parameters and a *distance* between these replications and the original data is evaluated. A model choice criterion belonging to this framework is the *posterior predictive loss*.
In [85] the authors used the standard utility idea, and, replacing experiments with models, minimize the loss over the models. Denoting $L(y_{rep}, a|y_{obs})$ the loss for guessing the action $a$ when a replicate of the data $y_{rep}$ has been obtained and $y_{obs}$ was observed, then the minimization of the expected value of this loss is performed over $a$, where the expectation is taken with respect to the posterior predictive distribution for $y_{rep}$ under model $m$.
Focusing on one of the component of $y_{obs}$, say the $l^{th}$, the corresponding replication $y_{l,rep}$ is assumed to have the same distribution as $y_{l,obs}$.
For squared error loss the resulting criterion is

$$D_k = \frac{k}{k+1}G + P \tag{4.10}$$

$$G = \sum_{l=1}^{n}(\mu_l - y_{l,obs})^2 \tag{4.11}$$

$$P = \sum_{l=1}^{n}\sigma_l^2 \tag{4.12}$$

where $\mu_l = E[Y_{l,\text{rep}}|y]$ and $\sigma_l^2 = Var[Y_{l,\text{rep}}|y]$, that is the mean and variance of the predictive distribution of $Y_{l,\text{rep}}$ given the observed data. This criterion accounts for closeness both to the observed data and to the replication, considering both the fit and the smoothness of the estimation, corresponding to the $G$ and $P$ term respectively. This last term tends to increase when the variance is inflated by over-fitting.
For what concerns $k$, it indicates the relative regret for departure from the observation compared with departure from the replications, but in practice model ranking is often insensitive to the choice of $k$.
The posterior distribution can be written as

$$p(y_{l,\text{rep}}) = \int p(y_{l,\text{rep}}|\theta)p(\theta|y)d\theta \tag{4.13}$$

In a MCMC algorithm it is possible to obtain $y_{l,\text{rep}}$ by drawing from $p(y_{l,\text{rep}}|\theta = \theta^m)$, where $\theta^m$ is the $m$th posterior realization of $\theta$. Thus, with this extra

level of simulation, one can obtain $\mu_l$ and $\sigma_l$ as the mean and variance of the drawn samples.

In the case of spatiotemporal data, the posterior predictive loss criterion becomes

$$G = \sum_{i=1}^{n} \sum_{i=1}^{T} (\mu_t(s_i) - y_{t,obs}(s_i))^2 \tag{4.14}$$

$$P = \sum_{i=1}^{n} \sum_{i=1}^{T} \sigma_t^2(s_i) \tag{4.15}$$

where $\mu_t(s_i)$ and $\sigma_t^2(s_i)$ are the mean and the variance of the posterior predictive distribution at location $s_i$ and time $t$.

## 4.5 The model

To describe the spatiotemporal behavior of $NO_x$ over the Tuscany region we propose a dynamical hierarchical spatiotemporal model. A reduced dimension framework is needed due to the high dimensionality of the process, and this issue is achieved by using a discretized version of convolution kernel method. Bayesian approach is adopted to estimation purposes. The hierarchical Bayesian structure is very useful in presence of complicated processes as in our case and in the most of environmental problems: the presence of a not negligible number of missing data is another issue that makes this choice the more appropriate.
To take into account all the features envisaged in the data set we have to consider the role of the emissions and temperature, the time periodicity and the residual dynamics. Spatial structure, also non stationary, and dimension reduction is addressed by using kernel convolution processes, while the dynamical behavior of the coefficients and of the residual term is introduced with an additional level of hierarchy.
In detail we developed the model described below. Normality is approached using the log transformation of the data.

The first stage of the model is designed to account for measurement errors and missing data.
Let $\mathbf{Z}_t$ be a vector of variable length of measurements at time $t$ over the $m_t$ not missing stations at time $t$, and let $\mathbf{Y}_t$ be the true process at time $t$ over all the $n$ stations. We can define

$$\mathbf{Z}_t = M_t \mathbf{Y}_t + \epsilon_{zt} \qquad \epsilon_{zt} \sim N(0, \sigma_z^2 I_{m_t X m_t}) \tag{4.16}$$

where $M_t$ is a $m_t$x$n$ mapping matrix whose $i$th row is the $i$th row of the identity matrix of size $n$ if and only if the $i$th station is not missing.

The $\epsilon_{zt}$ term is a mean zero gaussian error, representing the measurement error. We assume a diagonal structure for the covariance matrix of this error (it is reasonable that the measurement error is independent between stations) and with a common variance, $\Sigma_z = \sigma_z^2 I_{m_t X m_t}$.
At the second stage we model the latent process as

$$\begin{aligned}\mathbf{Y}_t &= K_m\mathbf{m} + \beta^E\mathbf{E} + T_t K_T \mathbf{b}^T + K_{a1}\mathbf{a1}_t \cos(2\pi t/24) + K_{b1}\mathbf{b1}_t \sin(2\pi t/24) + \\ &\quad + K_{a7}\mathbf{a7}_t \cos(2\pi t/168) + K_{b7}\mathbf{b7}_t \sin(2\pi t/168) + K\mathbf{a}_t + \epsilon_t \\ &\quad \epsilon_t \sim N(0, \sigma_\epsilon^2 I_{nXn}) \end{aligned} \tag{4.17}$$

where $Y_t$ is a $n$x1 vector representing the process at time $t$ over the $n$ stations, $E$ is a $n$x1 vector with emissions, $T_t = diag(\text{Temp}_t)$ is a $nxn$ matrix obtained from diagonalization of the vector of temperatures in the $n$ locations at time $t$. The first term represents a spatially correlated mean process expressed by a discrete equivalent of a convolution of the process $\mathbf{m}$, a vector of dimension $p_m$ defined over a reduced dimension grid. The $\mathbf{m}$ process is assumed to be a mean zero process distributed as multivariate normal with variance covariance matrix $\Sigma_m$. This matrix can be assumed diagonal or not, that is the reduced process can be assumed spatially independent or dependent. The element of the $nxp_m$ kernel matrix is defined as $k_{ij}(\theta, s_i - r_j)$, using one of the valid kernel functions described in section 1.3, for example gaussian or Matern form, and it is assumed to be known.
The second term gives the contribution of the emissions to the concentrations: as we have only one value for each station we assume the coefficient $\beta_E$ constant in space and time.
The third term models the temperature effect. Again we use the process convolution approach with the process $\mathbf{b}^T \sim N(0, \Sigma_T)$ defined over $p_T$ points and a corresponding $nxp_T$ kernel matrix.
Then we have the periodical terms, each defined on a reduced dimension grid with appropriate kernel matrix, but now we assume a temporal correlation for each of the processes $\mathbf{a1}_t$, $\mathbf{b1}_t$, $\mathbf{a7}_t$, $\mathbf{b7}_t$, with an autoregressive structure defined at next stage:

$$\mathbf{a1}_t = \mathbf{a1}_{t-1} + \eta_{a1}, \quad \eta_{a1} \sim N(0, \Sigma_{a1}) \tag{4.18}$$
$$\mathbf{b1}_t = \mathbf{b1}_{t-1} + \eta_{b1}, \quad \eta_{b1} \sim N(0, \Sigma_{b1}) \tag{4.19}$$
$$\mathbf{a7}_t = \mathbf{a7}_{t-1} + \eta_{a7}, \quad \eta_{a7} \sim N(0, \Sigma_{a7}) \tag{4.20}$$
$$\mathbf{b7}_t = \mathbf{b7}_{t-1} + \eta_{b1}, \quad \eta_{b7} \sim N(0, \Sigma_{b7}) \tag{4.21}$$
$$\tag{4.22}$$

This approach is similar to the dynamical approach described in section 3.3. A term accounting for spatiotemporal dynamics is then added: a reduced dimension process $\mathbf{a}_t$ is assumed to evolve with a transitional equation, like

$$\mathbf{a}_t = H\mathbf{a}_{t-1} + \eta_a, \quad \eta_a \sim N(0, Q) \tag{4.23}$$

where the $H$ matrix governs the spatiotemporal evolution of the process. No assumption about the structure of this matrix (that is diagonal or tridiagonal ecc) can be made here because the structure of this model is not easily comparable with a physical dynamics.
Finally we add a residual error assumed to be zero mean gaussian with diagonal and homogeneous variance matrix, and constant in time, $\epsilon_t = \epsilon \sim N(0, \sigma_\epsilon^2 I_{nXn})$.
Thus we have the following priors over the parameters:

$$\mathbf{m} \sim N(0, \Sigma_m) \tag{4.24}$$
$$\beta_E \sim N(0, \sigma^2_{\beta_E}) \tag{4.25}$$
$$\mathbf{b}^T \sim N(0, \Sigma_T) \tag{4.26}$$

We define the remaining priors using conjugate distributions:

$$\begin{aligned}
\Sigma_{a1} &\sim IW((\nu_{a1} C_{a1}), C_{a1}) && (4.27)\\
\Sigma_{b1} &\sim IW((\nu_{b1} C_{b1}), C_{b1}) && (4.28)\\
\Sigma_{a7} &\sim IW((\nu_{a7} C_{a7}), C_{a7}) && (4.29)\\
\Sigma_{b7} &\sim IW((\nu_{b7} C_{b7}), C_{b7}) && (4.30)\\
Q &\sim IW((\nu_q C_q), C_q) && (4.31)\\
h = vec(H) &\sim N(\mu_h, \Sigma_h) && (4.32)\\
\sigma_\epsilon^2 &\sim IG(q_\epsilon, r_\epsilon) && (4.33)\\
\sigma_z^2 &\sim IG(q_z, r_z) && (4.34)\\
& && (4.35)
\end{aligned}$$

We also need to specify the initial value of time varying coefficients

$$\mathbf{b1}_0 \sim N(0, \sigma^2_{b_0^1}) \tag{4.36}$$
$$\mathbf{a1}_0 \sim N(0, \sigma^2_{a_0^1}) \tag{4.37}$$
$$\mathbf{b7}_0 \sim N(0, \sigma^2_{b_0^7}) \tag{4.38}$$
$$\mathbf{a7}_0 \sim N(0, \sigma^2_{a_0^7}) \tag{4.39}$$
$$\mathbf{a}_0 \sim N(0, \sigma^2_{a0}) \tag{4.40}$$

We sample the parameters $\mathbf{m}, \beta_E, \mathbf{b}^T, \mathbf{Y}_t, \mathbf{a1}_t, \mathbf{b1}_t, \mathbf{a7}_t, \mathbf{b7}_t$, for $t = 0, 1, \ldots, T$, $\sigma_\epsilon^2, \sigma_z^2, H, Q, \Sigma_{a1}, \Sigma_{b1}, \Sigma_{a7}, \Sigma_{b7}$ with Gibbs sampler, with full conditional specified in Appendix.
The required hyperparameters are $\Sigma_m$, $\sigma^2_{beta_E}$, $\Sigma_T, q_\epsilon$, $r_\epsilon$, $\sigma^2_{b_0^1}$, $\sigma^2_{a_0^1}$, $\sigma^2_{b_0^7}$, $\sigma^2_{a_0^7}$, $\sigma^2_{a0}$, $\mu_h$,$\Sigma_h$, $\nu_{a1}$, $C_{a1}$, $\nu_{b1}$, $C_{b1}$, $\nu_{a7}$, $C_{a7}$, $\nu_{b7}$, $C_{b7}$. All these hyperparameters and the initial values required have been set from data analysis and simulation steps.

### 4.5.1 Kernel choice

As first temptative we choose an univariate gaussian kernel, that is

$$K_{ij}(\theta) = \frac{1}{\sqrt{2\pi}\theta_j} \exp\left(-\frac{1}{2\theta_j}|s_i - r_j|^2\right) \tag{4.41}$$

The locations of reference points chosen are shown in Fig.4.18 The reference points are chosen following an ideal main grid with spacing of 60 km and a nested smaller grid (30 km) in the north of the region. This fact is motivated by the non homogeneity of the number of monitoring stations across the region, in order to increase accuracy where we have more detailed information.
We checked model sensitivity to the value chosen for the variance of the kernel, finding a reasonable behavior in the range from 2500 to 10000. As it is reccomended that grid spacing is no more than the standard deviation of the kernel (for example see [8]), the minimum acceptable value in our case is a variance of 4000. Thus the corresponding spatial process have a range parameter from about 90 to 140 km.
In the following we want to compare different choices of kernel variance and shape, and thus we need to choose a tool that make us able to do that in a sensible way.

### 4.5.2 Model selection

We choose an univariate kernel with variance 10000 over the whole domain as a first choice. In this way we observe a different behavior for different kind of stations (that is traffic or background) where an underestimation for the former coexists with an overestimation for the latter, and so we decided to take into account the hypothesis of using different kernel variances for the two different situations. As traffic stations tend to reflect local situations it seems natural to consider a narrower range for these stations with respect to background stations that are representative of a larger scale dynamics. We choose a value of 4000 for traffic stations and 9000 for background.
Now we proceed with an analysis of the residuals of this model.

We observe an overall good performance of the model particularly for high concentrations, while for very low concentrations the model tends to an overestimation (see figures 4.21, 4.20). Residuals do not show spatial trend (see figure 4.23). For what concerns meteorological variables we checked the distribution of the residuals with respect to temperature, finding no more trend as expected (see figure 4.24). To test if it is possible to improve model performances by introducing other variables, we checked the influence of wind speed. A rough inspection of the correlation between residuals and

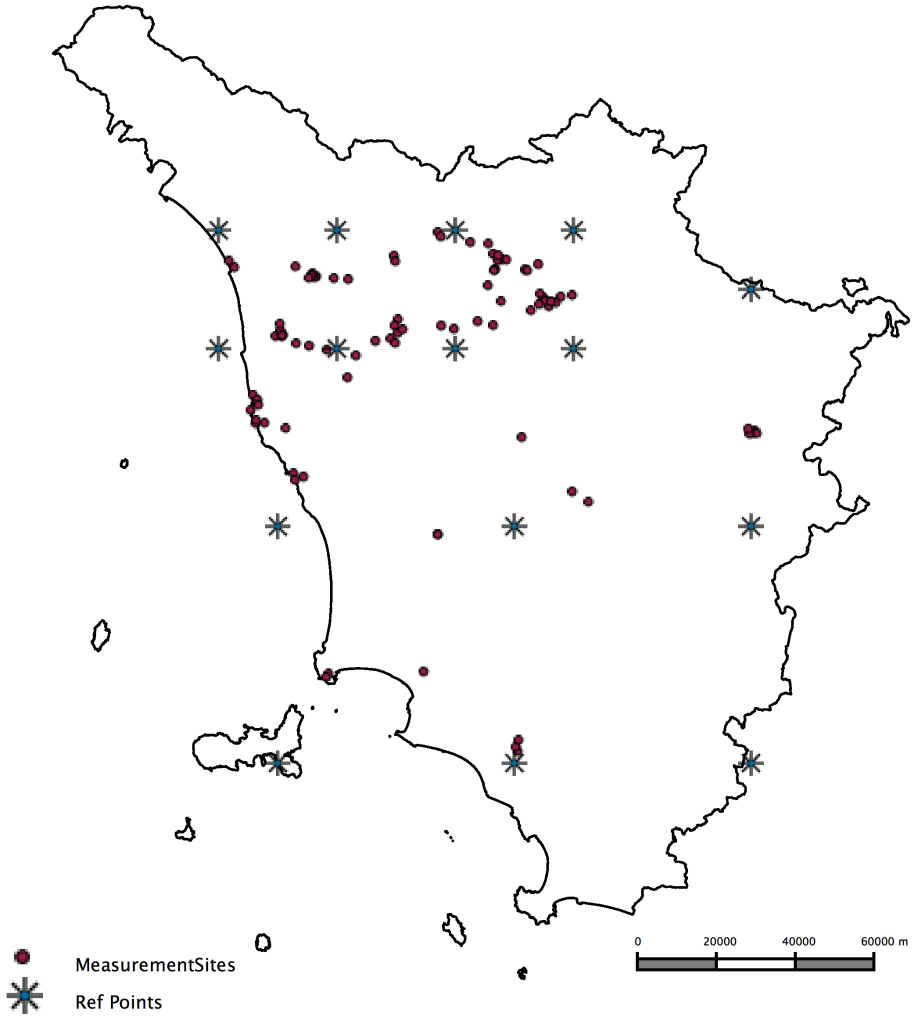


Figure 4.18: Lattice locations for kernel convolution latent processes.

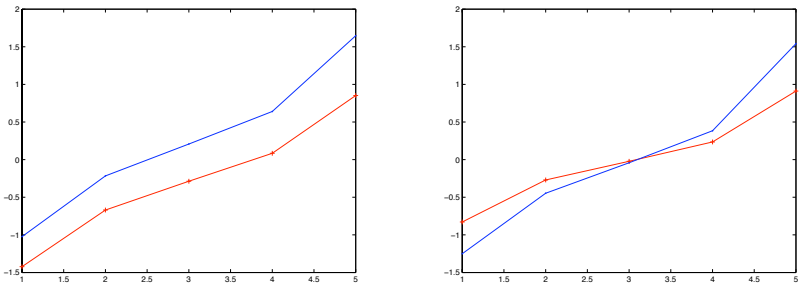

Figure 4.19: (left) Quantiles of the residuals [.025 .25 .50 .75 .975] for the model with costant kernel for traffic (red) and background (blue) stations. (right) Quantiles of the residuals [.025 .25 .50 .75 .975] for the model with different kernel for traffic (red) and background stations (blue).

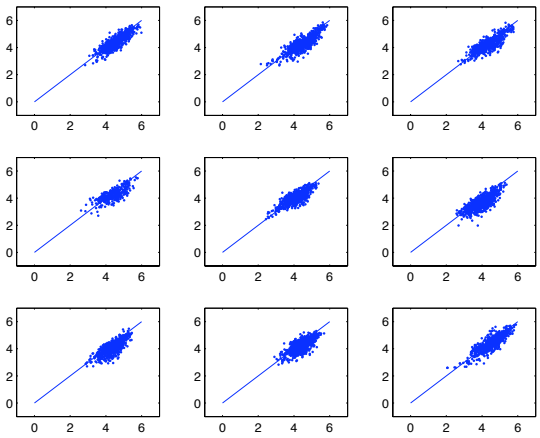

Figure 4.20: Estimated values vs. measured values for each of the nine months of simulation, March-November 2005, Firenze, Gramsci traffic station.

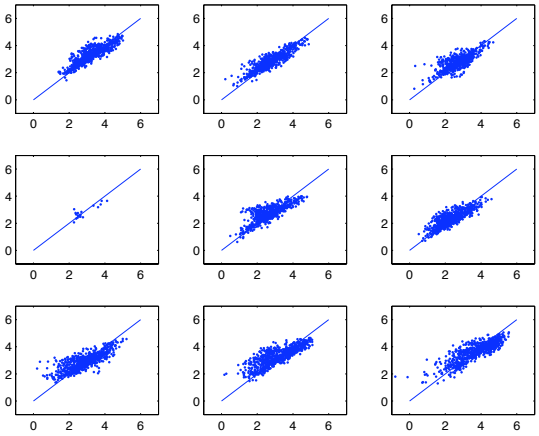

Figure 4.21: Estimated values vs. measured values for each of the nine months of simulation, March-November 2005, Prato, via Roma background station.

wind speed does not support this hypothesis. To check a possible influence of wind speed and direction in shape and variance of the kernel, we divided the data in eight classes (the four quadrants for wind direction and two classes of wind speed, below or above 3.5 m/s) and calculated empirical semivariogram contour plots for each class. This can be an useful tool to assess anisotropy. The method uses the separation distance between each axis. Rectangular bins are then formed and the empirical semivariogram in each bin is calculated. The value of each semivariogram is assigned to the center of the bins. Contour plot of this map can show departure from isotropy as long as the isolines are far from circular contours. We found very similar contour plots for different wind classes, but these plots are not circular, showing a shorter range in the direction NE-SW with respect to the opposite direction (NW-SE). For this reason we choose a bivariate kernel with the major axis of the corresponding ellipse oriented in direction NW-SE. This behavior can be motivated by the specific morphological shape of Tuscany (see figure 4.25), where both the mountains and the coastline are oriented along this direction.

We compared the three models described above, stationary univariate kernel (case 1), nonstationary univariate kernel (case 2), nonstationary bivariate and asymmetric kernel (case 3) by posterior predictive loss. In table 4.26 we report the D (with k=1), G and P terms described in chapter 4.4 for the three cases for the complete model (model 1) and time reduced model (model 2, that we will describe in the next chapter). We see that the preferred model

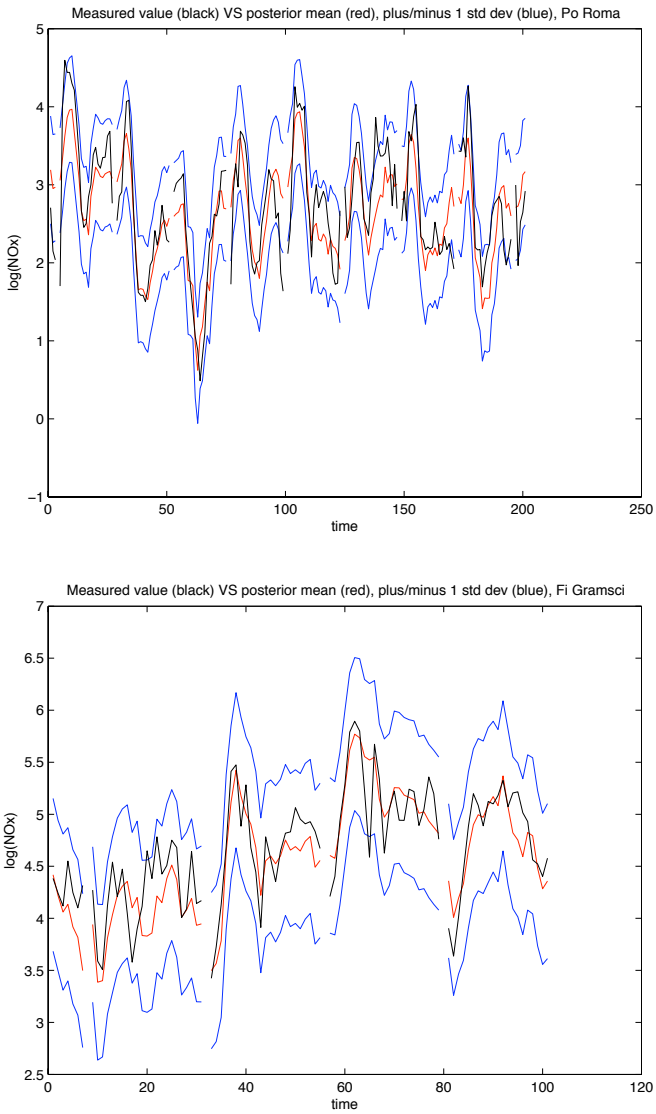


Figure 4.22: Posterior mean (red), posterior mean plus/minus one posterior standard deviation (blue), measured concentration (black) for a background station (top), Prato, via Roma, 1000-1300 hours, (bottom), and for the traffic station Firenze-via Gramsci 6500-6600 hours (bottom).

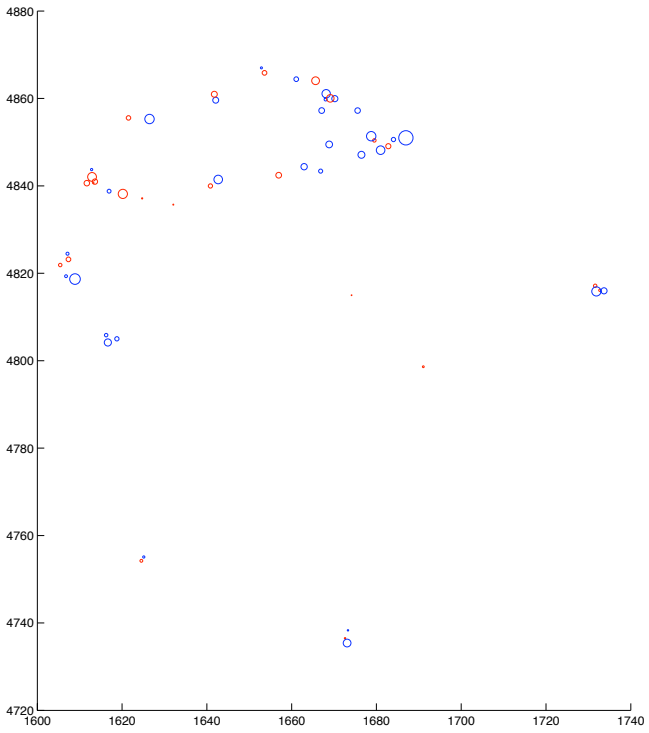


Figure 4.23: Spatial distribution of the residuals. Circles are proportional to the absolute value of the mean of the residual for each station (red for traffic and blue for background).

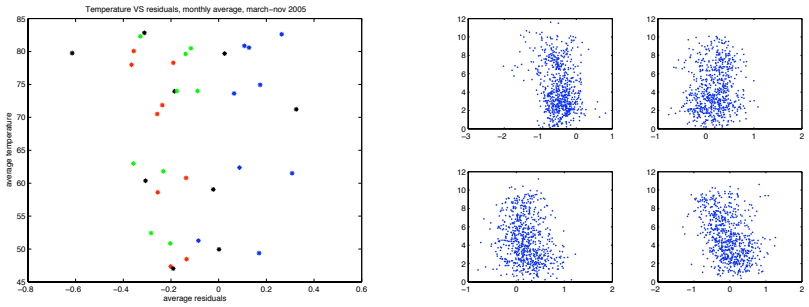


Figure 4.24: (left) Monthly average temperature with respect to monthly average of the concentrations for four different stations (different colors). (right) Hourly wind speed with respect to hourly concentrations for four stations in May.

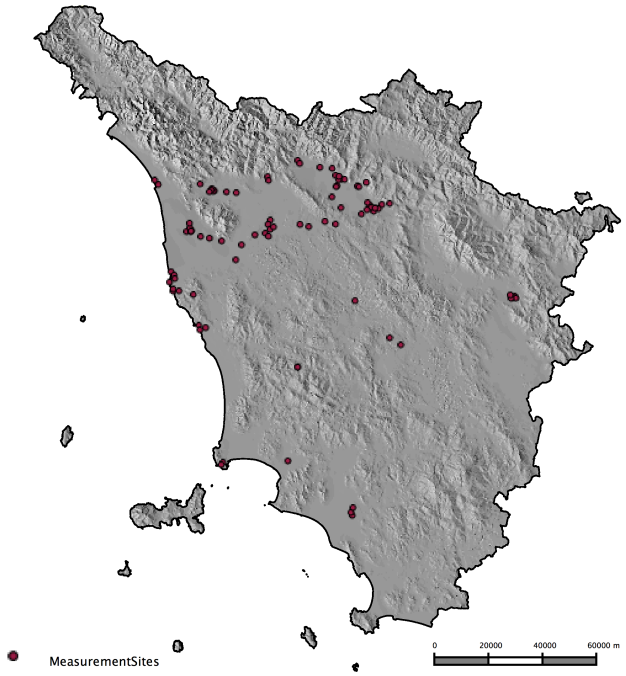


Figure 4.25: Morphology of the Tuscany region.

| | D | G | P |
|---|---|---|---|
| M1-1 | 311786 | 205460 | 209056 |
| **M1-2** | **214835** | **114870** | **157400** |
| M1-3 | 223405 | 127810 | 159500 |
| **M2-2** | **214915** | **141870** | **143980** |
| M2-3 | 281005 | 188610 | 186700 |

Figure 4.26: Posterior predictive loss comparisons between different models. Stationary univariate kernel (case 1), nonstationary univariate kernel (case 2), nonstationary bivariate and asymmetric kernel (case 3) for complete model (Model 1) and time reduced model (Model 2), see section 4.6.

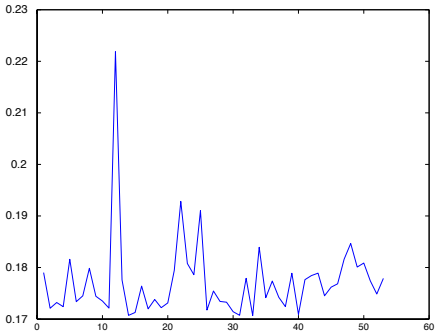


Figure 4.27: Posterior predictive loss (Model 1, case 2) for different stations.

is the nonstationary univariate kernel case, although the case with bivariate kernel is very similar for model 1.
If we consider the single component in space or time that is summed up in equation (4.15), we can obtain information about the model performance at each time or site. In figure 4.27 the D term for each station is reported: the peak in this graph corresponds to a background station near Florence. Looking closely we find out that emission database provides for this station a very high emission value. Presumably this is not a fair value, because this station is located far from main roads and other industrial emissions.

Before checking the model predictive ability in next section we introduce a time-reduced model by using the EOF approach.

## 4.6 Time reduced version of the model: an EOF approach

The empirical orthogonal functions technique that we described in section 3.6.1 is designed to find spatial structure (maps) explaining most of the variance, and the time series evolution of the principal components.
The same technique can be used to find the principal components of the temporal structure and the spatial behavior of these, by simply arranging the matrix in a $n$x$T$ fashion, that is with rows corresponding to time series and columns representing a map for each time. So let $F^* = F'$, with $F$ the matrix defined as in section 3.6.1[1], and the corresponding $T$x$T$ covariance matrix $C^* = F'^*F^*/(n-1)$. Solving the associated eigenvalue problem $C^* = \Psi\Lambda\Psi'$ we find in the columns of $\Psi$ the eigenvectors ($\in R^T$ space), and for each eigenvector the corresponding projected map $\mathbf{a}_i = F^*\psi_i$.
In our case the number of spatial locations ($n = 55$) is much smaller than the time period of observations ($T = 6600$), with the corresponding covariance matrix having high dimension, 6600x6600. Since the rank of $F^*$ is at most $n$, the rank of $C^*$ is at most $n = 55$, and the number of zero eigenvalues of $C^*$ is at least $T - n$. Thus it is possible to use a more efficient strategy to find eigenvectors (Von Storch method). Let $L^* = F^*F'^*$, with size $n$x$n$, it can be proven that the eigenvalue problem $C^*\Psi = \Psi\Lambda$ is equivalent to $L^*B^* = B^*\Lambda$, where $B^* = F^*\Psi$. Equivalence means that both the equations yield to the same eigenvalues, while the eigenvectors are not the same. Just projecting $F'^*$ on the vectors from $B^*$ we obtain vectors proportional to the original EOFs, with a proportionality factor of $1/\sqrt{\lambda_i}$. Thus, once calculated, the $T$x$n$ matrix $D^* = F'^*B^*$, $\psi_i = d_i/\sqrt{\lambda_i}$, where $d_i$ are the column vectors of $D^*$, gives the EOFs we are looking for. The advantage of this procedure is that we have to manage only matrices of dimensions $n$x$n$ instead of the huge matrices as before.
We apply this decomposition to our data (after fitting the missing values with splines), both with the original data and the residuals after regression with temperature and emissions. As we found that the first 20 eigenvalues explain the 80% of the variability, we envisage in this method a way to achieve reduction of time dimension.
In the following we propose a model based on the EOF technique in order to reduce dimension.
Arranging the spatiotemporal process (after removing the mean) in a $n$x$T$ matrix (called $F^*$), where each row corresponds to the time series of each location, we have the ($T$x$T$) covariance matrix $C^* = F'^*F^*/(n-1)$. The solution of the eigenvalue problem gives the desired matrix of eigenvectors, of which only $n$ corresponding to non null eigenvalues (because the rank of

[1] F is the $T$x$n$ matrix with the i-th row corresponding to a map of the process at time i and the $j$th column corresponding to a time series at location $j$.

$F^*$ is at most $n$). We can gain a grater reduction in the dimensionality of the problem choosing only the first $p*$ eigenvectors (from data analysis we saw that with $p^* = 20$ we can explain most of the variance). We call $\Phi$ the $T$x$p^*$ matrix formed by the eigenvectors corresponding to the first $p^*$ eigenvalues and use this matrix to project the $n$x$T$ data matrix in a $n$x$p^*$ spatiotemporal process. Moreover we have $p^*$ time independent components, each of these can be viewed again as a spatial process, and thus projected to a $p$ space trough the kernel matrix. In the end we obtain $p^*$ processes over $p$ points, independent in time and space. According to this we can model the process in matrix form as:

$$Y = K_m\mathbf{m}*ones(1,T)+\beta^E\mathbf{E}*ones(1,T)+T_tK_T\mathbf{b}^T*ones(1,T)+(K\Delta)\Phi'+\epsilon \tag{4.42}$$

where $Y$ is the $n$x$T$ matrix of the spatiotemporal process and *ones* represents a matrix of ones of the specified dimension, $\Phi$ is the $T$x$p^*$ EOFs matrix, and K is the $n$x$p$ kernel matrix. So $\Delta$ is the $p$x$p^*$ matrix of the reduced dimension process. Rewriting the whole model, we have:

$$\begin{aligned} Z_t &= M_tY_t + \epsilon_{zt} \qquad \epsilon_{zt} \sim N(0, \sigma_z^2 I_{m_tXm_t}) && (4.43)\\ \mathbf{Y}_t &= K_m\mathbf{m} + \beta^E\mathbf{E} + T_tK_T\mathbf{b}^T + K\Delta\Phi'_{.t} + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2 I_{nXn}) && (4.44)\end{aligned}$$

where $\Phi_{.t}$ is the $t$th column of the $\Phi$ matrix. If we want to use different kernels for each of the $p^*$ processes we can arrange the last term of the equation above as $\tilde{K}\tilde{\Delta}\Phi'_{.t}$. $\tilde{\Delta}$ is defined as a $(pp^*)$x$p^*$ matrix

$$\tilde{\Delta} = \begin{pmatrix} \delta_{\mathbf{1}} & & & \\ & \delta_{\mathbf{2}} & & \\ & & \ddots & \\ & & & \delta_{\mathbf{p^*}} \end{pmatrix} \tag{4.45}$$

with each $\delta_i$ being a p-dimensional vector, while $\tilde{K}$ is a $n$X$pxp^*$ matrix formed by the $p^*$ kernel $\tilde{K} = [K_1, \ldots, K_{p*}]$, each $K_i$ with dimension $n$x$p$. At the lower stage we define

$$\mathbf{m} \sim N(0, \Sigma_m) \tag{4.46}$$
$$\beta_E \sim N(0, \sigma^2_{\beta_E}) \tag{4.47}$$
$$\mathbf{b}^T \sim N(0, \Sigma_T) \tag{4.48}$$

and the remaining hyperpriors, with $\underline{\delta} = vec(\Delta)$

$$\underline{\delta} \sim N(0, \Sigma_\delta) \tag{4.49}$$
$$\sigma_\epsilon^2 \sim IG(q_\epsilon, r_\epsilon) \tag{4.50}$$
$$\sigma_z^2 \sim IG(q_z, r_z) \tag{4.51}$$
$$\tag{4.52}$$

This formulation reduces computational efforts (the time required is less than half the time required for model 1) without losing much in terms of posterior predictive loss (see table 4.26). In the next section the predictive ability of the model both in complete and time reduced version will be checked.

## 4.7 Predictive analysis

To check the predictive ability of the models described above we left out two randomly chosen stations (one classified as traffic and the other one classified as background station, see fig 4.28) and calculated the posterior predictive distribution for each time in that point in space. We chofse univariate and nonstationary (different for traffic and background stations) gaussian kernel for both the models. Figures 4.29 - 4.33 compare the posterior predictive distribution for each time with respect to measured concentrations: the posterior predictive 0.025 and 0.975 quantiles are represented in green, the median is represented as a yellow star, while measured concentrations are in blue. For both model 1 and model 2 time series predictions are quite good, with a good behavior for the median, although the uncertainty bounds are pretty high.
To better evaluate the predictive behavior of the models we try to construct a synthetic measure of the prediction ability. We proceeded as follows. Let $y_t$ a random variable that follows the posterior distribution at time t for a given station and let $y_t^*$ the measured concentration at time t in that station and consider the probability that $y_t$ is less then $y_t^*$, $p_t = P(y_t < y_t^*)$. By doing this we obtain 6600 values (one for each time) of the variable $p = [p_1, \ldots, p_T]$: the histogram of this variable gives an overall representation of the predictive ability as long as it is peaked around the value of 0.5. These histograms are shown in figures 4.34 in four cases. The first case (top left) is the result for background station by using model 1: the peak is on 0.5, although there is a skewness with a higher tail on the left side. This is the indication of an overestimation for certain time points: these time points correspond to the case of low concentrations during nighttime when the model does not decrease as much as the measured values. This fact is made clear if we notice that the histogram calculated considering only daytime hours (figure 4.34, bottom left) displays the desired symmetry, and can thus be motivated by the high number of missing data registered during nighttime hours. The histogram referred to model 2 seems to indicate a lower predictive performance, as expected: the histogram is again centered on 0.5 but is is less peaked (figure 4.34, top right). Finally (bottom right) the histogram for the traffic station regarding model 1 is reported: for this station the histogram is more peaked than for the corresponding case for the background station, but the skewness is still present.

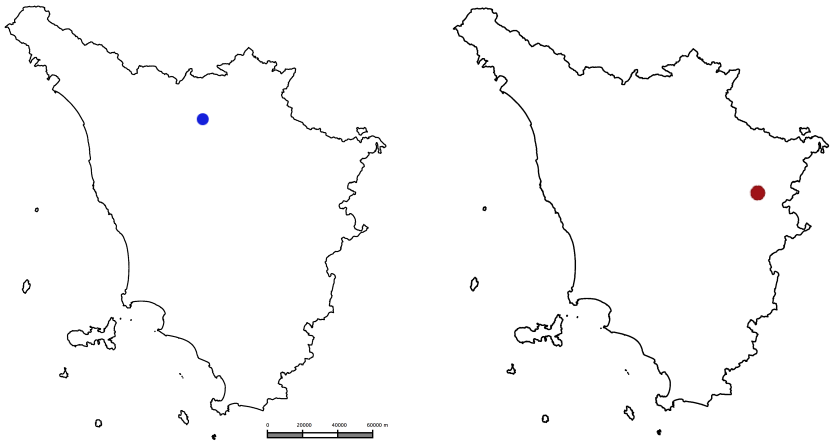

Figure 4.28: Test stations: background, Prato, via Roma (left) and traffic, Arezzo, via fiorentina (right).

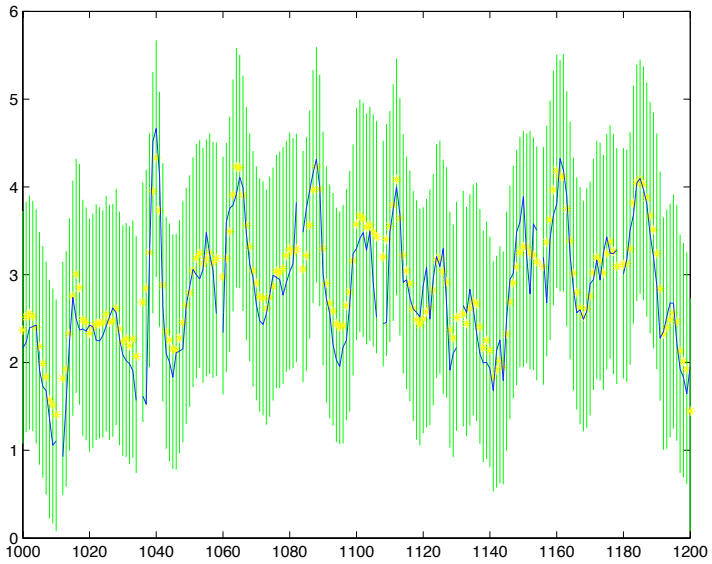


Figure 4.29: Model 1: Posterior predictive distribution versus measured concentrations for the background station (Prato, via Roma) for hours from 1000-1200 of considered period (0.025- 0.975 quantiles in green, median as a yellow star, measured concentrations in blue).

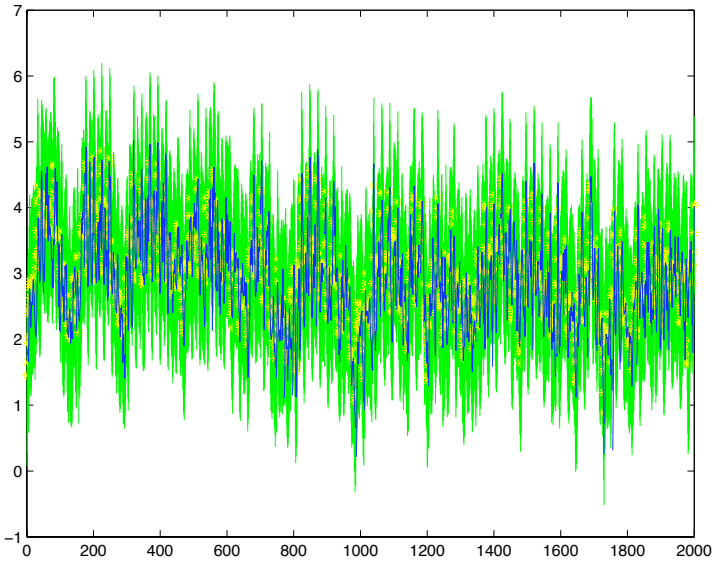


Figure 4.30: Model 1: Posterior predictive distribution versus measured concentrations for the background station (Prato, via Roma) for hours from 1000-2000 of considered period (0.025- 0.975 quantiles in green, median as a yellow star, measured concentrations in blue).

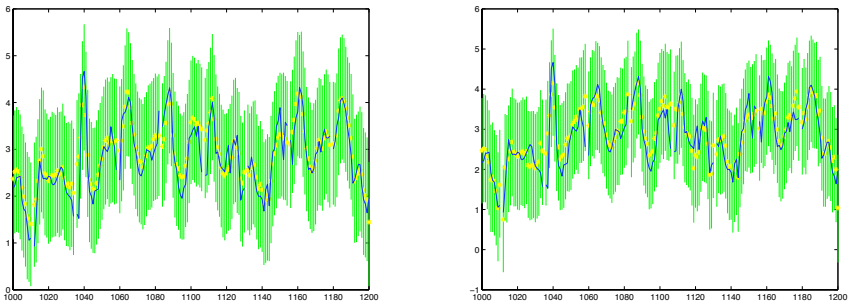

Figure 4.31: Comparison between model 1 and model 2: Posterior predictive distribution versus measured concentrations for the background station (Prato, via Roma) for hours from 1000-1200 of considered period (symbols as in previous figures).

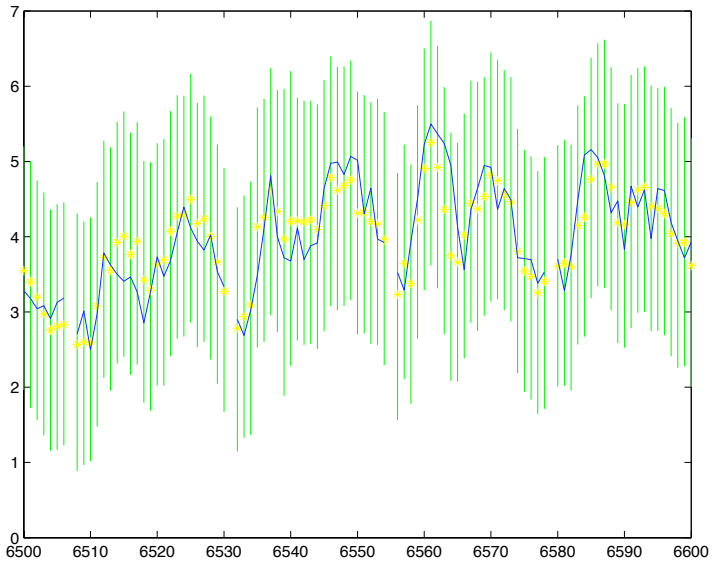


Figure 4.32: Model 1: Posterior predictive distribution versus measured concentrations for the traffic station (Arezzo, via fiorentina) for hours from 6000-6600 of considered period (symbols as in previous figures).

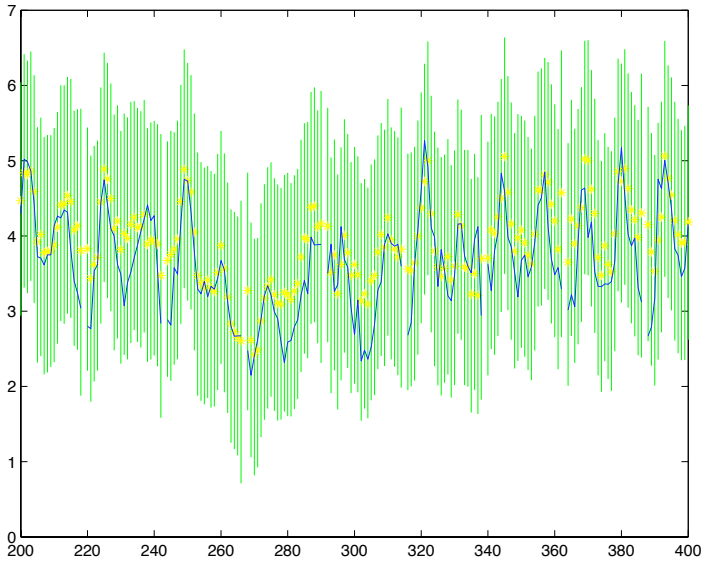


Figure 4.33: Model 1: Posterior predictive distribution versus measured concentrations for the traffic station (Arezzo, via fiorentina) for hours from 200-400 of considered period (symbols as in previous figures).

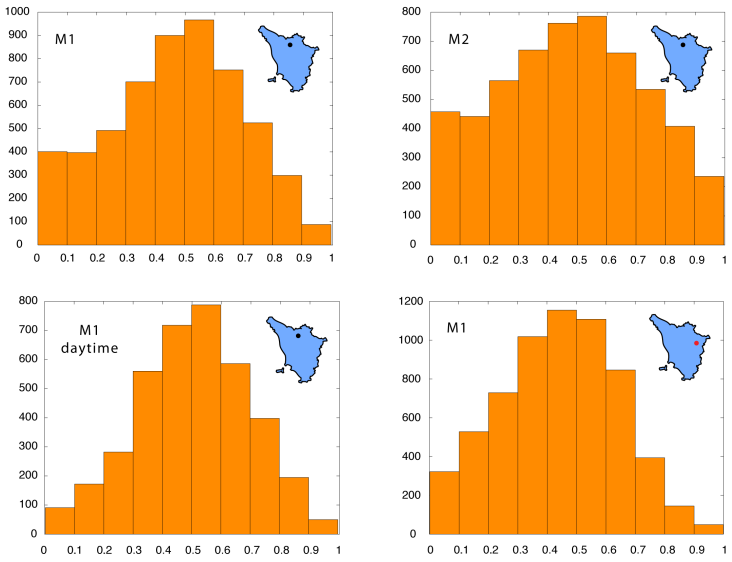


Figure 4.34: Synthetic measure of predictive behavior of the proposed models, see text for details.

In summary the model shows good prediction ability and makes possible to estimate the concentrations in unobserved spatial locations with a good prediction ability, requiring only emissions, temperature and distance from road (traffic/background). The model introduced has a very flexible structure: many choices of the kernel functions accounting for different correlation structures are possible. Moreover the kernel can be space and time varying allowing for nonstationarity in space and time and each subprocess could be modeled with a different kernel function or set of reference points. We also obtained estimations of missing data in a straight manner inside the model, and no additional requirement or methods (as imputation) is needed. Moreover, anomalous behavior of monitoring stations and outliers in the dataset could be easily detected analyzing model results.

Although the dimension reduction strategy adopted made the MCMC estimation feasible, this is still computationally demanding. The tradeoff between the prediction ability of the model and computational tractability is apparent when comparing the complete model (model 1) with the EOF model (model 2). Another drawback of this model is the quite high variability in predictions that could be the subject of a further improvement. Finally, as in the proposed models the estimated parameters cannot be directly related to physical dynamics, in the following chapter we introduce a new model that could overcome this limit.

## 4.8 Future developments

In this final subsection we propose a model suitable for future developments. Here we use the concept of dimension reduction via kernel convolution approach, but now we try to link the dynamics of the process with the physical mechanism of the advection diffusion process.
As we saw in section 3.5.1 the time continuous advection diffusion equation can be viewed in discrete time as an integro-differential equation of the form

$$y_{t+1}(s) = \gamma \int k_s(r;\theta_s) y_t(r) dr + \tilde{\eta}_{t+1}(s) \tag{4.53}$$

where $\eta$ is a spatially colored noise process and $\gamma$ is a parameter that controls explosive growth. This model is nonseparable and, by using spatially varying kernel, also nonstationary.
In this context the advection term is represented by the mean of the kernel (i.e. the translation) and the diffusion term is related to the variance covariance matrix of the kernel. In this way it is possible to mimic the physical behavior of the process we would like to model.
We saw in section 3.5.1 that this model can be written in spectral representation by using Fourier transform. This method is well suited for a regular

grid of points, but unfortunately this is not our case.
However it is possible to envisage a feasible modeling strategy by using the spatial dimension reduction we used in previous sections and the framework of integro-differential models. We can imagine a model like this (assuming as a first guess a unitary growth/depletion term, $\gamma = 1$)

$$\begin{aligned} Z_t &= M_t Y_t + \beta^E E + \epsilon_{zt} && (4.54)\\ Y_t &= K_{sp}\Phi\alpha_t && (4.55)\\ \alpha_t &= \Phi' B'_\theta \alpha_{t-1} + \eta_t && (4.56)\\ & && (4.57) \end{aligned}$$

where $Z_t$ and $Y_t$ are defined as in the previous models, and $B_\theta = [\mathbf{b}(s_1;\theta_{s_1}), \ldots, \mathbf{b}(s_n;\theta_{s_n})]$ is defined as in section 3.5, that is, in the gaussian case

$$b_j(s;\theta_1(s),\theta_2(s)) = \exp[i\omega_j(\theta_1(s)+s) - 0.5\omega_j^2\theta_2(s)] \qquad (4.58)$$

where $\omega_j$ is the spatial frequency. $\theta_1(s)$ and $\theta_2(s)$ are respectively the translation and scale parameters and can be directly linked to the transport effect and the diffusion process, that is with the meteorological fields of wind and diffusivity. This dynamics is applied now to a regular grid, defined by using a (purely spatial) kernel $K_{sp}$ analogous to that used in previous models, and assumed known. The key issues now are the choice of lattice spacing, that needs to be chosen in a coherent manner with respect to the spatial scale of dynamics, and properly accounting for modeling errors due to discretization in time and space.

# Conclusions

In this thesis we have dealt with spatiotemporal modeling for environmental applications. Throughout the text we discussed many challenging issues involved in statistical modeling of environmental systems, for what concerns both their theoretical and applied aspects.
In particular, in the present work we focused on the application of spatiotemporal modeling to the study of nitrogen oxides concentrations in Tuscany. Nitrogen dioxides are toxic by inhalation and a not negligible percentage of the urban population in last years was potentially exposed to ambient air nitrogen dioxide concentrations above the health protection limit value. Moreover $NO_2$ and other nitrogen oxides are precursors for a number of harmful secondary air pollutants like ozone and particulate matter.
Due to the collaboration with Environmental Protection Agency - Tuscany region -, that has the role of air quality monitoring and assessment, we developed models suited for application, with particular concern on the ability of describing and predicting the spatiotemporal behavior of these pollutants over the area of interest by using the minimal set of available informations. To achieve this goal we developed a hierarchical Bayesian spatiotemporal reduced rank model in a complete and temporal reduced version able to predict $NO_x$ concentrations, with a good average behavior and a quite high, but known, uncertainty, in any arbitrary unobserved locations in our region, as long as emissions, temperature and distance from main streets are known for that point. The ability to achieve this knowledge about spatiotemporal distribution of pollutant concentrations is of outmost importance for Environmental Protection Agency for many reasons ranging from risk and exposure assessment to authorization procedures for air emissions, environmental preservation, recovery and planning.
Actual needs in regulatory activity made us also concentrate on some specific questions as the multivariate aspect or a physically based modeling approach, that are now methodologically relevant issues for environmental problems. For example the correlation between nitrogen dioxide and nitrogen oxide over the area of interest is crucial to evaluate authorization procedures for new emission plants, as they are based on nitrogen oxides emissions, while only nitrogen dioxide is limited by law. We investigated

the bivariate relationship between nitrogen oxides and nitrogen dioxides and their spatial and nonspatial correlation structure by using the coregionalization approach.

Very complex spatiotemporal processes, occurring on a wide variety of scales and arising from the interactions of many subprocesses involving physical and chemical mechanisms, take part in determining the value of concentration of a pollutant ($NO_x$ in this case) in a given point in space and time. Statistical methods able to deal with this complexity and suited for spatial and spatiotemporal processes have been reviewed in the thesis.
A hierarchical Bayesian framework has been introduced to deal with the complexity of the underlying spatiotemporal process. This framework made us able to solve the problem of missing data and to use the different stages of the structure to address different terms involved in the process. Time evolution of the process has been modeled through a space time dynamical setting: a nonseparable spatiotemporal dynamic process being added to the dynamical evolution of covariates coefficients.
Environmental and physical processes are often neither separable nor stationary: recently developed methods to overcome these assumptions have been reviewed and examined in order to find the proper way to account for these issues in our application.
Due to the high dimensionality of the involved variables, reduced rank methods were required to make computation manageable. Thus a modeling framework and methods to achieve dimension reduction in spatial and spatiotemporal contexts have been provided in the text, and these proved of valuable help in the application of our case study. In fact we developed and applied a reduced rank hierarchical Bayesian dynamical model, potentially nonstationary and nonseparable, to estimate concentrations at unobserved space and time points of nitrogen oxides concentrations over the Tuscany region. Kernel convolution in space and empirical orthogonal functions in a temporal reduced setting made the MCMC estimation feasible, even if still computationally demanding.
Comparison between different modeling strategies has been performed by using posterior predictive loss. The selected model showed good prediction ability and made possible to estimate the concentrations (with the associated uncertainty) in unobserved spatial locations, requiring only emissions, temperature and distance from road (traffic/background). The model introduced has a very flexible structure: many choices of the kernel functions accounting for different correlation structures are possible. Moreover kernel can be space and time varying allowing for nonstationarity in space and time and each subprocess could be modeled with a different kernel function or set of reference points. We obtained estimations of missing data in a straight manner inside the model, with no additional requirement or methods being needed and anomalous behavior of monitoring stations and outliers in

the dataset have been detected analyzing model results. The model showed good prediction ability but a quite high associated variability.
An important characteristic of environmental processes is represented by the a priori knowledge we could have from many different sources, like experts opinion and physical or empirical relationships between the variables involved. In the case of atmospheric pollution the underlying dynamics driving the advection and diffusion of a pollutant in air is known. The concern of including this dynamics inside the statistical model has been addressed and a theoretical framework suited to account for the advection diffusion equation directly into the statistical model introduced. Thanks to this model we are able to relate in a physically based structure emissions, meteorological variables and concentrations. The limits in the application of this model, due to the non regularly spaced data set we are dealing with, could be overcome by a modeling strategy sketched out at the end of this thesis, that uses a combination of two projection bases allowing for the proper dimension reduction. The introduction of this modeling framework is a promising way to take advantages of both statistical and deterministic approaches in spatiotemporal modeling of pollutant concentrations.

# Appendix

## A.1 Full-conditionals

### A.1.1 Model 1

$$
\begin{aligned}
\mathbf{Z}_t &= M_t\mathbf{Y}_t + \epsilon_{zt} \qquad \epsilon_{zt} \sim N(0, \sigma_z^2 I_{m_t X m_t}) \\
\mathbf{Y}_t &= K_m\mathbf{m} + \beta^E\mathbf{E} + T_t K_T \mathbf{b}^T + K_{a1}\mathbf{a1}_t \cos(2\pi t/24) + K_{b1}\mathbf{b1}_t \sin(2\pi t/24) \\
&\quad + K_{a7}\mathbf{a7}_t \cos(2\pi t/168) + K_{b7}\mathbf{b7}_t \sin(2\pi t/168) + K\mathbf{a}_t + \epsilon_t \\
\epsilon_t &\sim N(0, \sigma_\epsilon^2 I_{nXn})
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{a1}_t &= \mathbf{a1}_{t-1} + \eta_{a1}, \quad \eta_{a1} \sim N(0, \Sigma_{a1}) \\
\mathbf{b1}_t &= \mathbf{b1}_{t-1} + \eta_{b1}, \quad \eta_{b1} \sim N(0, \Sigma_{b1}) \\
\mathbf{a7}_t &= \mathbf{a7}_{t-1} + \eta_{a7}, \quad \eta_{a7} \sim N(0, \Sigma_{a7}) \\
\mathbf{b7}_t &= \mathbf{b7}_{t-1} + \eta_{b1}, \quad \eta_{b7} \sim N(0, \Sigma_{b7}) \\
\mathbf{a}_t &= H\mathbf{a}_{t-1} + \eta_a, \quad \eta_a \sim N(0, Q)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{m} &\sim N(0, \Sigma_m) \\
\beta_E &\sim N(0, \sigma^2_{\beta_E}) \\
\mathbf{b}^T &\sim N(0, \Sigma_T) \\
\Sigma_{a1} &\sim IW((\nu_{a1} C_{a1}), C_{a1}) \\
\Sigma_{b1} &\sim IW((\nu_{b1} C_{b1}), C_{b1}) \\
\Sigma_{a7} &\sim IW((\nu_{a7} C_{a7}), C_{a7}) \\
\Sigma_{b7} &\sim IW((\nu_{b7} C_{b7}), C_{b7}) \\
Q &\sim IW((\nu_q C_q), C_q) \\
h = vec(H) &\sim N(\mu_h, \Sigma_h) \\
\sigma^2_\epsilon &\sim IG(q_\epsilon, r_\epsilon) \\
\sigma^2_z &\sim IG(q_z, r_z)
\end{aligned}
$$

The Gibbs sampler for this model proceeds sampling from the following full conditional distributions. We omit bold text to denote vector to simplify notation. Let

$$\begin{aligned}\theta &= K_m m + \beta^E E + T_t K_T b^T + K_{a1} a1_t \cos(2\pi t/24) + K_{b1} b1_t \sin(2\pi t/24) \\ &\quad + K_{a7} a7_t \cos(2\pi t/168) + K_{b7} b7_t \sin(2\pi t/168) + K a_t\end{aligned}$$

where $T_t = \text{diag} T_t$ and $\theta_{-x}$ denotes the equation for $\theta$ when $x$ is subtracted.

- $[Y_t|.] \propto [Y_t|\theta, \sigma_\epsilon][Z_t|\sigma_z^2, M_t, Y_t] \propto N(Ab, A)$

$$\begin{aligned}A &= \left[\frac{I}{\sigma_\epsilon^2} + \frac{M_t' M_t}{\sigma_z^2}\right]^{-1} \\ b &= \left[\frac{\theta^t}{\sigma_\epsilon^2} + \frac{Z_t' M_t}{\sigma_z^2}\right]'\end{aligned}$$

- $[\sigma_z^2|.] \propto [\sigma_z^2] \prod_{t=1}^T [Z_t|\sigma_z^2, M_t, Y_t] \propto IG(q_z^*, r_z^*)$

$$\begin{aligned}q_z^* &= q_z + \sum_{t=1}^T m_t/2 \\ r_z^* &= r_z + 0.5 \sum_{t=1}^T (Z_t - M_t Y_t)'(Y_t - M_t Y_t)\end{aligned}$$

- $[m|.] \propto [m] \prod_{t=1}^T [Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned}A &= [\Sigma_m^{-1} + \frac{1}{\sigma_\epsilon^2} \sum_{t=1}^T K_m' K_m]^{-1} \\ b &= [\frac{1}{\sigma_\epsilon^2} \sum_{t=1}^T (Y_t - \theta^t_{-K_m m})' K_m]'\end{aligned}$$

- $[\beta_E|.] \propto [\beta_E] \prod_{t=1}^T [Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned}A &= [\frac{1}{\sigma^2_{\beta_E}} + \frac{1}{\sigma_\epsilon^2} \sum_{t=1}^T E'E]^{-1} \\ b &= [\frac{1}{\sigma_\epsilon^2} \sum_{t=1}^T (Y_t - \theta^t_{-\beta_E E})' E]'\end{aligned}$$

- $[b^T|.] \propto [b^T] \prod_{t=1}^T [Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned}A &= [\Sigma_T^{-1} + \frac{1}{\sigma_\epsilon^2} \sum_{t=1}^T K_T' T_t' T_t K_T]^{-1} \\ b &= [\frac{1}{\sigma_\epsilon^2} \sum_{t=1}^T (Y_t - \theta^t_{-T_t K_T b^T})' T_t K_T]'\end{aligned}$$

- $[b1_t|.] \propto [b1_t|b1_{t-1}, \Sigma_{b1}][b1_{t+1}|b1_t, \Sigma_{b1}][Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [2\Sigma_{b1}^{-1} + \frac{1}{\sigma_\epsilon^2} K_{b1}' K_{b1} sin^2(2\pi t/24)]^{-1} \\ b &= [b1_{t-1}'\Sigma_{b1}^{-1} + b1_{t+1}'\Sigma_{b1}^{-1} + \frac{1}{\sigma_\epsilon^2}(Y_t - \theta^t_{-K_{b1}b1sin()})' K_{b1} sin(2\pi t/24)]' \end{aligned}$$

- $[a1_t|.] \propto [a1_t|a1_{t-1}, \Sigma_{a1}][a1_{t+1}|a1_t, \Sigma_{a1}][Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [2\Sigma_{a1}^{-1} + \frac{1}{\sigma_\epsilon^2} K_{a1}' K_{a1} cos^2(2\pi t/24)]^{-1} \\ b &= [a1_{t-1}'\Sigma_{a1}^{-1} + a1_{t+1}'\Sigma_{a1}^{-1} + \frac{1}{\sigma_\epsilon^2}(Y_t - \theta^t_{-K_{a1}a1cos()})' K_{a1} cos(2\pi t/24)]' \end{aligned}$$

- $[b7_t|.] \propto [b7_t|b7_{t-1}, \Sigma_{b7}][b7_{t+1}|b7_t, \Sigma_{b7}][Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [2\Sigma_{b7}^{-1} + \frac{1}{\sigma_\epsilon^2} K_{b7}' K_{b7} sin^2(2\pi t/24)]^{-1} \\ b &= [b7_{t-1}'\Sigma_{b7}^{-1} + b7_{t+1}'\Sigma_{b7}^{-1} + \frac{1}{\sigma_\epsilon^2}(Y_t - \theta^t_{-K_{b7}b7sin()})' K_{b7} sin(2\pi t/24)]' \end{aligned}$$

- $[a7_t|.] \propto [a7_t|a7_{t-1}, \Sigma_{a7}][a7_{t+1}|a7_t, \Sigma_{a7}][Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [2\Sigma_{a7}^{-1} + \frac{1}{\sigma_\epsilon^2} K_{a7}' K_{a7} cos^2(2\pi t/24)]^{-1} \\ b &= [a7_{t-1}'\Sigma_{a7}^{-1} + a7_{t+1}'\Sigma_{a7}^{-1} + \frac{1}{\sigma_\epsilon^2}(Y_t - \theta^t_{-K_{a7}a7cos()})' K_{a7} cos(2\pi t/24)]' \end{aligned}$$

- $[a_t|.] \propto [a_t|a_{t-1}, H, Q][a_{t+1}|a_t, H, Q][Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [Q^{-1} + H'Q^{-1}H + \frac{1}{\sigma_\epsilon^2} K'K]^{-1} \\ b &= [(Ha_{t-1})'Q^{-1} + a_{t+1}'Q^{-1}H + \frac{1}{\sigma_\epsilon^2}(Y_t - \theta^t_{-Ka_t})'K]' \end{aligned}$$

- $[\sigma_\epsilon^2|.] \propto [\sigma_\epsilon^2] \prod_{t=1}^T [Y_t|\theta, \sigma_\epsilon] \propto IG(q_\epsilon^*, r_\epsilon^*)$

$$\begin{aligned} q_\epsilon^* &= q_\epsilon + nT/2 \\ r_\epsilon^* &= r_\epsilon + 0.5 \sum_{t=1}^T (Y_t - \theta^t)'(Y_t - \theta^t) \end{aligned}$$

- $h = vec(H), [h|.] \propto [h][a_t|a_{t-1}, H, Q] \propto N(Ab, A)$.
  Let $A_T = [a_1, \ldots, a_T]$ and $A_{T-1} = [a_0, \ldots, a_{T-1}]$, both $n$x$T$ matrices and $\tilde{Q} = I_t \otimes Q$

$$\begin{aligned} A &= [\Sigma_h^{-1} + (A_{T-1}' \otimes I_n)'\tilde{Q}^{-1}(A_{T-1}' \otimes I_n)]^{-1} \\ b &= (A_{T-1}' \otimes I_n)'\tilde{Q}^{-1} vec(A_T) + \Sigma_h^{-1}\mu_h \end{aligned}$$

Unfortunately this full conditional distribution involves matrices of huge dimensions. However, after some calculation we obtain an equivalent and manageable form:

$$\begin{aligned} A &= [\Sigma_h^{-1} + \sum_{t=1}^{T} (\underline{A}'_{T-1} \tilde{Q}^{-1} \underline{A}_{T-1}]^{-1} \\ b &= [\sum_{t=1}^{T} a'_t \tilde{Q}^{-1} \underline{A}_{T-1} + \mu'_h \Sigma_h^{-1}]' \end{aligned}$$

- $[Q|.] \propto [Q] \prod_{t=1}^{T} [a_t | a_{t-1}, H, Q], [Q^{-1}|.] \propto W(C^*, \nu^*)$

$$\begin{aligned} C^* &= [\sum_{t=1}^{T} (a_t - H a_{t-1})(a_t - H a_{t-1})' + \nu_q C_q]^{-1} \\ \nu^* &= \nu_q + T \end{aligned}$$

- $[\Sigma_{b1}|.] \propto [\Sigma_{b1}] \prod_{t=1}^{T} [b1_t | b1_{t-1}, \Sigma_{b1}], [\Sigma_{b1}^{-1}|.] \propto W(C^*, \nu^*)$

$$\begin{aligned} C^* &= [\sum_{t=1}^{T} (b1_t - b1_{t-1})(b1_t - b1_{t-1})' + \nu_{b1} C_{b1}]^{-1} \\ \nu^* &= \nu_{b1} + T \end{aligned}$$

- $[\Sigma_{a1}|.] \propto [\Sigma_{a1}] \prod_{t=1}^{T} [a1_t | a1_{t-1}, \Sigma_{a1}], [\Sigma_{a1}^{-1}|.] \propto W(C^*, \nu^*)$

$$\begin{aligned} C^* &= [\sum_{t=1}^{T} (a1_t - a1_{t-1})(a1_t - a1_{t-1})' + \nu_{a1} C_{a1}]^{-1} \\ \nu^* &= \nu_{a1} + T \end{aligned}$$

- $[\Sigma_{b7}|.] \propto [\Sigma_{b7}] \prod_{t=1}^{T} [b7_t | b7_{t-1}, \Sigma_{b7}], [\Sigma_{b7}^{-1}|.] \propto W(C^*, \nu^*)$

$$\begin{aligned} C^* &= [\sum_{t=1}^{T} (b7_t - b7_{t-1})(b7_t - b7_{t-1})' + \nu_{b7} C_{b7}]^{-1} \\ \nu^* &= \nu_{b7} + T \end{aligned}$$

- $[\Sigma_{a7}|.] \propto [\Sigma_{a7}] \prod_{t=1}^{T} [a7_t | a7_{t-1}, \Sigma_{a7}], [\Sigma_{a7}^{-1}|.] \propto W(C^*, \nu^*)$

$$\begin{aligned} C^* &= [\sum_{t=1}^{T} (a7_t - a7_{t-1})(a7_t - a7_{t-1})' + \nu_{a7} C_{a7}]^{-1} \\ \nu^* &= \nu_{a7} + T \end{aligned}$$

- $[b1_0|.] \propto [b1_0][b1_1|b1_0, \Sigma_{b1}] \propto N(Ab, A)$

$$\begin{aligned} A &= [\frac{I}{\sigma^2_{b10}} + \Sigma^{-1}_{b1}]^{-1} \\ b &= [b1'_1 \Sigma^{-1}_{b1}]' \end{aligned}$$

- $[a1_0|.] \propto [a1_0][a1_1|a1_0, \Sigma_{a1}] \propto N(Ab, A)$

$$\begin{aligned} A &= [\frac{I}{\sigma^2_{a10}} + \Sigma^{-1}_{a1}]^{-1} \\ b &= [a1'_1 \Sigma^{-1}_{a1}]' \end{aligned}$$

- $[b7_0|.] \propto [b7_0][b7_1|b7_0, \Sigma_{b7}] \propto N(Ab, A)$

$$\begin{aligned} A &= [\frac{1}{\sigma^2_{b70}} + \Sigma^{-1}_{b7}]^{-1} \\ b &= [b7'_I \Sigma^{-1}_{b7}]' \end{aligned}$$

- $[a7_0|.] \propto [a7_0][a7_1|a7_0, \Sigma_{a7}] \propto N(Ab, A)$

$$\begin{aligned} A &= [\frac{I}{\sigma^2_{a70}} + \Sigma^{-1}_{a7}]^{-1} \\ b &= [a7'_1 \Sigma^{-1}_{a7}]' \end{aligned}$$

- $[a_0|.] \propto [a_0][a_1|a_0, H, Q] \propto N(Ab, A)$

$$\begin{aligned} A &= [\frac{I}{\sigma^2_{a0}} + H'Q^{-1}H]^{-1} \\ b &= [a'_1 Q^{-1} H]' \end{aligned}$$

- $[b1_T|.] \propto [b1_T|b1_{T-1}, \Sigma_{b1}][y_T|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [\Sigma^{-1}_{b1} + \frac{1}{\sigma^2_\epsilon} K'_{b1} K_{b1} sin^2(2\pi T/24)]^{-1} \\ b &= [b1'_{T-1} \Sigma^{-1}_{b1} + \frac{1}{\sigma^2_\epsilon}(y_T - \theta^T_{-K_{b1} b1 sin})' K_{b1} sin(2\pi T/24)]' \end{aligned}$$

- $[a1_T|.] \propto [a1_T|a1_{T-1}, \Sigma_{a1}][y_T|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [\Sigma^{-1}_{a1} + \frac{1}{\sigma^2_\epsilon} K'_{a1} K_{a1} sin^2(2\pi T/24)]^{-1} \\ b &= [a1'_{T-1} \Sigma^{-1}_{a1} + \frac{1}{\sigma^2_\epsilon}(y_T - \theta^T_{-K_{a1} a1 sin})' K_{a1} sin(2\pi T/24)]' \end{aligned}$$

- $[b7_T|.] \propto [b7_T|b7_{T-1}, \Sigma_{b7}][y_T|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [\Sigma_{b7}^{-1} + \frac{1}{\sigma_\epsilon^2} K'_{b7} K_{b7} sin^2(2\pi T/24)]^{-1} \\ b &= [b7'_{T-1}\Sigma_{b7}^{-1} + \frac{1}{\sigma_\epsilon^2}(y_T - \theta^T_{-K_{b7}b7sin(.)})' K_{b7} sin(2\pi T/24)]' \end{aligned}$$

- $[a7_T|.] \propto [a7_T|a7_{T-1}, \Sigma_{a7}][y_T|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [\Sigma_{a7}^{-1} + \frac{1}{\sigma_\epsilon^2} K'_{a7} K_{a7} sin^2(2\pi T/24)]^{-1} \\ b &= [a7'_{T-1}\Sigma_{a7}^{-1} + \frac{1}{\sigma_\epsilon^2}(y_T - \theta^T_{-K_{a7}a7sin(.)})' K_{a7} sin(2\pi T/24)]' \end{aligned}$$

- $[a_T|.] \propto [a_T|a_{T-1}, H, Q][y_T|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [Q^{-1} + \frac{K'K}{\sigma_\epsilon^2}]^{-1} \\ b &= [(Ha_{T-1})'Q^{-1} + \frac{1}{\sigma_\epsilon^2}(y_T - \theta^T_{-Ka_T})'K]' \end{aligned}$$

### A.1.2 Model 2

$$\begin{aligned} Z_t &= M_t Y_t + \epsilon_{zt} \qquad \epsilon_{zt} \sim N(0, \sigma_z^2 I_{m_t X m_t}) \\ \mathbf{Y}_t &= K_m \mathbf{m} + \beta^E \mathbf{E} + T_t K_T \mathbf{b}^T + \tilde{K}\tilde{\Delta}\Phi'_{.t} + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2 I_{nXn}) \end{aligned}$$

$$\tilde{\Delta} = \begin{pmatrix} \delta_{\mathbf{1}} & & & \\ & \delta_{\mathbf{2}} & & \\ & & \ddots & \\ & & & \delta_{\mathbf{p^*}} \end{pmatrix}$$

with each $\delta_i$ being a p-dimensional vector, while $\tilde{K}$ is a $nxpp^*$ matrix formed by the $p^*$ kernel $\tilde{K} = [K_1, \ldots, K_{p*}]$, each $K_i$ with dimension $nxp$.

$$\begin{aligned} \mathbf{m} &\sim N(0, \Sigma_m) \\ \beta_E &\sim N(0, \sigma^2_{\beta_E}) \\ \mathbf{b}^T &\sim N(0, \Sigma_T) \\ \underline{\delta} &\sim N(0, \Sigma_\delta) \\ \sigma^2_\epsilon &\sim IG(q_\epsilon, r_\epsilon) \\ \sigma^2_z &\sim IG(q_z, r_z) \end{aligned}$$

where $\underline{\delta} = vec(\Delta)$.
The Gibbs sampler for this model proceeds sampling from the following full conditional distributions. We omit bold text to denote vector to simplify notation. Let

$$\begin{aligned} \theta &= K_m m + \beta^E E + T_t K_T b^T + K_{a1} a1_t \cos(2\pi t/24) + K_{b1} b1_t \sin(2\pi t/24) \\ &\quad + K_{a7} a7_t \cos(2\pi t/168) + K_{b7} b7_t \sin(2\pi t/168) + K a_t \end{aligned}$$

where $T_t = \text{diag}T_t$ and $\theta_{-x}$ denotes the equation for $\theta$ when $x$ is subtracted.

The Gibbs sampler for this model proceeds sampling from the following full conditional distributions. We omit the bold text to denote vector to simplify notation. Let

$$\theta_t = K_m m + \beta^E E + T_t K_T b^T + \tilde{K}\tilde{\Delta}\Phi'_{.t}$$

where $T_t = \text{diag}T_t$ and $\theta_{-x}$ denotes the equation for $\theta$ when $x$ is subtracted.

- $[Y_t|.] \propto [Y_t|\theta, \sigma_\epsilon][Z_t|\sigma_z^2, M_t, Y_t] \propto N(Ab, A)$

$$\begin{aligned} A &= \left[\frac{I}{\sigma_\epsilon^2} + \frac{M_t' M_t}{\sigma_z^2}\right]^{-1} \\ b &= \left[\frac{\theta_t'}{\sigma_\epsilon^2} + \frac{Z_t' M_t}{\sigma_z^2}\right]' \end{aligned}$$

- $[\sigma_z^2|.] \propto [\sigma_z^2] \prod_{t=1}^{T} [Z_t|\sigma_z^2, M_t, Y_t] \propto IG(q_z^*, r_z^*)$

$$\begin{aligned} q_z^* &= q_z + \sum_{t=1}^{T} m_t/2 \\ r_z^* &= r_z + 0.5 \sum_{t=1}^{T} (Z_t - M_t Y_t)'(Y_t - M_t Y_t) \end{aligned}$$

- $[m|.] \propto [m] \prod_{t=1}^{T} [Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [\Sigma_m^{-1} + \frac{1}{\sigma_\epsilon^2} \sum_{t=1}^{T} K_m' K_m]^{-1} \\ b &= [\frac{1}{\sigma_\epsilon^2} \sum_{t=1}^{T} (Y_t - \theta^t_{-K_m m})' K_m]' \end{aligned}$$

- $[\beta_E|.] \propto [\beta_E] \prod_{t=1}^{T} [Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [\frac{1}{\sigma^2_{\beta_E}} + \frac{1}{\sigma_\epsilon^2} \sum_{t=1}^{T} E'E]^{-1} \\ b &= [\frac{1}{\sigma_\epsilon^2} \sum_{t=1}^{T} (Y_t - \theta^t_{-\beta_E E})' E]' \end{aligned}$$

- $[b^T|.] \propto [b^T] \prod_{t=1}^{T} [Y_t|\theta, \sigma_\epsilon] \propto N(Ab, A)$

$$\begin{aligned} A &= [\Sigma_T^{-1} + \frac{1}{\sigma_\epsilon^2} \sum_{t=1}^{T} K_T' T_t' T_t K_T]^{-1} \\ b &= [\frac{1}{\sigma_\epsilon^2} \sum_{t=1}^{T} (Y_t - \theta^t_{-T_t K_T b^T})' T_t K_T]' \end{aligned}$$

- $[\sigma_\epsilon^2|.] \propto [\sigma_\epsilon^2] \prod_{t=1}^{T} [Y_t|\theta, \sigma_\epsilon] \propto IG(q_\epsilon^*, r_\epsilon^*)$

$$\begin{aligned} q_\epsilon^* &= q_\epsilon + nT/2 \\ r_\epsilon^* &= r_\epsilon + 0.5 \sum_{t=1}^{T} (Y_t - \theta^t)'(Y_t - \theta^t) \end{aligned}$$

- for $\underline{\delta} = vec(\Delta)$ we write the model in matrix form

$$\begin{aligned} \underline{Y} &= \Lambda + \tilde{K} \tilde{\Delta} \Phi' + U \\ \Lambda &= K_m m \cdot ones(1, T) + \beta^E E \cdot ones(1, T) + diag(K_T b^T) \underline{T} \end{aligned}$$

where $\underline{Y}$ is the $n$x$T$ matrix of the process $Y$ with the different stations in rows and different hours in columns, $\underline{\mathrm{T}}$ the corresponding $n$x$T$ matrix of temperatures, $\Phi$ is a $T$x$p^*$ matrix of the first $p^*$ eigenvectors and $U = [\epsilon_1, \ldots, \epsilon_T]$. The variance associated to this last term is $\tilde{\Sigma}_\epsilon = var(vec(U)) = I_{TxT} \otimes \Sigma_\epsilon = I_{TxT} \otimes \sigma_\epsilon I_{nxn}$. Thus we have

$$vec(\underline{Y}) = (\Phi \otimes \tilde{K}) vec(\tilde{\Delta}) + vec(\Lambda) + vec(U)$$

Let $G$ a matrix formed by zero and one such that $vec(\tilde{\Delta}) = Gvec(\Delta)$, thus

$$vec(\underline{Y}) = (\Phi \otimes \tilde{K}) G\underline{\delta} + vec(\Lambda) + vec(U)$$

It is possible now to write the full conditional distribution for $\underline{\delta}$.

$[\underline{\delta}|.] \propto [\underline{\delta}][\underline{Y}|\Delta, G, \Lambda, \sigma_\epsilon, \Phi, \tilde{K}] \propto N(Ab, A)$

$$\begin{aligned} A &= [[(\Phi \otimes \tilde{K})G]' \tilde{\Sigma}_\epsilon [(\Phi \otimes \tilde{K})G] + \Sigma_\delta^{-1}]^{-1} \\ b &= [[vec(\underline{Y}) - vec(\Lambda)]' \tilde{\Sigma}_\epsilon [(\Phi \otimes \tilde{K})G] + \mu_\delta' \Sigma_\delta^{-1}]' \end{aligned}$$

Unfortunately this full conditional involves matrices of huge dimensions (for example $(\Phi \otimes \tilde{K})$ is a $(Tn)$x$p^*p^*p$=363000x4800). Thus we rewrite model in vector form and obtain

$$\begin{aligned} A &= [\sum_{t=1}^{T} [\Phi_{dt}(I_{p^*xp^*} \otimes \tilde{K})G]' \Sigma_\epsilon^{-1} [\Phi_{dt}(I_{p^*xp^*} \otimes \tilde{K})G] + \Sigma_\delta^{-1}]^{-1} \\ b &= [\sum_{t=1}^{T} (Y_t - \lambda_t)' \Sigma_\epsilon^{-1} \Phi_{dt}(I_{p^*xp^*} \otimes \tilde{K})G]' \end{aligned}$$

where $\Phi_{dt} = \Phi_t' \otimes I_{nxn}$

# Bibliography


[1] Whittle, P. (1954). On stationary processes in the plane. Biometrika 41, 434-49.

[2] Heine, V. (1955). Models for two-dimensional stationary stochastic processes. Biometrika 42, 170-178.

[3] Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX: On the Matern correlation family. Biometrika, 93, 989-995.

[4] Handcock, M. S., Stein M. L. (1993) A Bayesian analysis of kriging. Technometrics 35, 403-10.

[5] Wikle, C.K., (2003). Hierarchical models in environmental science. International Statistical Review, 71 , 181-199.

[6] Higdon, D. (1998). A process-convolution approach to modeling temperatures in the North Atlantic Ocean. Journal of Environmental and Ecological Statistics, 5, 173-190.

[7] Higdon, D., Swall, J., and Kern, J. (1998). Non-stationary spatial modeling. In Bayesian Statistics 6, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith. Oxford: Oxford University Press, 761-768.

[8] Calder, C.A. and Cressie, N. Some Topics in Convolution-Based Spatial Modeling. Proceedings of the 56th Session of the International Statistics Institute. Lisbon, Portugal. August 22-29, 2007.

[9] Yaglom, A. M. (1987). Correlation Theory of Stationary and Related Random Functions. Vol. II: Supplementary Notes and References. Berlin: Springer-Verlag.

[10] Xia, G., Gelfand, A.E., 2006. Stationary process approximation for the analysis of large spatial datasets. Technical Report, ISDS, Duke University.

[11] Kern, J. (2000). Bayesian process-convolution approaches to specifying spatial dependence structure.Ó Ph.D. thesis, Institute of Statistics and Decision Sciences, Duke University, Durham, NC.

[12] Higdon (2002). Space and space-time modeling using process convolutions. In Quantitative Methods for Current Environmental Issues, eds. C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi, 37-56. New York: Springer-Verlag.

[13] Lee, H. K. H., Higdon, D. M., Calder, C. A., and Holloman, C. H. (2005). Efficient models for correlated data via convolutions of intrinsic processes.Ó Statistical Modelling, 5, 53-74.

[14] Bruno Sanso, Alexandra M. Schmidt and Aline A. Nobre. (2008) Bayesian Spatio-temporal models based on discrete convolutions Canadian Journal of Statistics, 36, 239-258.

[15] Zimmerman D. L. (1993) Another look at anisotropy in geostatistics. Mathematical Geology, 25, 453-470.

[16] Erikson M, Siska P (2000) Understanding anisotropy computations. Math. Geol. 32(6): 683-700

[17] Fuentes M. (2005) A formal test for nonstationarity of spatial stochastic processes. Journal of Multivariate Analysis 96, 30-55.

[18] Calder, C.A. (2008). A Bayesian Dynamic Process Convolution Approach to Modeling PM2.5 and PM10 Concentration Levels. Environmetrics, 19, 39-48.

[19] Paciorek, C.J., and M. Schervish. 2006. Spatial modelling using a new class of nonstationary. covariance functions. Environmetrics, 17:483-506

[20] Paciorek, C. J. (2003). Nonstationary Gaussian processes for regression and spatial modelling. Ph.D. thesis, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA.

[21] Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. Biometrika, 89, 197-210.

[22] Nychka, D., Wikle, C.K., and Royle, J.A. (2002). Multiresolution models for nonstationary spatial covariance functions. Statistical Modelling: An International Journal, 2, 315-331.

[23] Pintore A, Holmes C. 2007. Spatially adaptive non-stationary covariance functions via spatially adaptive spectra. Technical report at the Statistics Department, Oxford University.

[24] Stein, M. (2005) Nonstationary spatial covariance functions. Technical Report 21, University of Chicago, Center for Integrating Statistical and Environmental Science.

[25] Royle, JA; Wikle, CK. Efficient Statistical Mapping of Avian Count Data. Environmental and Ecological Statistics. 2005;12(2):225-243.

[26] Paciorek, C.J. 2007. Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP package. Journal of Statistical Software 19(2).

[27] Wikle, C.K., (2002). Spatial modeling of count data: A case study in modelling breeding bird survey data on large spatial domains. In Spatial Cluster Modelling, A. Lawson and D. Denison, eds. Chapman and Hall, 199-209.

[28] Royle, J. A., and L. M. Berliner (1999) A hierarchical approach to multivariate spatial modeling and prediction. Journal of Agricultural, Biological, and Environmental Statistics, 4, 29-56.

[29] Royle, J. A., Berliner, L. M., Wikle, C. K. and Milliff, R. (1999). A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea. Case Studies in Bayesian Statistics, C. Gatsonis et al., Eds., Springer-Verlag, 367-382.

[30] Matern, B. (1986). Spatial Variation, 2nd ed. Berlin: Springer-Verlag.

[31] Gelfand, A.E., Schmidt, A.M. 2003.A Bayesian Coregionalization Approach for Multivariate Pollutant Data, Journal of Geophysical Research - Atmosphere 108, D24, 8783

[32] Gelfand A, Schmidt A, Banerjee S and Sirmans C (2004). Nonstionary multivariate process modeling through spatially varying coregionalization. Test 13(2): 263-312

[33] A. Gelfand, A. Schmidt, C.F. Sirmans, Multivariate Spatial Process Models: Conditional and Unconditional Bayesian Approaches Using Coregionalization, Center for Real Estate and Urban Economic Studies, University of Connecticut

[34] Ver Hoef, J. and Barry, R. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. Journal of Statistical Planning and Inference, 69, 275-294.

[35] Ver Hoef, J., Cressie, N., and Barry, R. (2004). Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT). Journal of Computational and Graphical Statistics, 13, 265-282.

[36] Majumdar, A. and Gelfand, A. (2007). Multivariate spatial process modeling using convolved covariance functions. Mathematical Geology. Volume 39, 2,225-245

[37] Finley, O.A., Banerjee, S., Ek, A.R., McRoberts, R.E. (2008) Bayesian Multivariate Process Modeling for Prediction of Forest Attributes Journal of Agriculural, Biological, and Environmental Statistics, 13, 1.

[38] Finley, A.O., Banerjee, S., Carlin, B.P. (2006) spBayes:an R package for Univariate and Multivariate Hierarchical Point referenced Spatial Models, http://cran.r-project.org/web/packages/spBayes/index.html

[39] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., adn van der Linde, A,(2002). Bayesian Measures of Model Complexity and Fit, Journal of the Royal Statistical Society, Series B, 64, 583-639.

[40] Fuentes M. (2005) Testing for separability of spatial temporal covariance functions. Journal of Statistical Planning and Inference 136, 447-466.

[41] Kolovos, A., Christakos, G., Hristopulos, D. T. and Serre, M. L. (2004) Methods for generating non-separable spatiotemporal covariance models with potential environmental applications, Advances in Water Resources, 27, 815-830.

[42] Cressie N, Huang HC. 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of American Statistical Association 94, 1330-1340.

[43] Gneiting T. 2002. Nonseparable, stationary covariance functions for space-time data. Journal of American Statistical Association ,97, 590-600.

[44] Stein, M. L. (2005), Space-time covariance functions, Journal of the American Statistical Association, 100, 310-321.

[45] Jones, R. and Zhang, Y. (1997), Models for continuous stationary space-time processes, in Modelling Longitudinal and Spatially Correlated Data, Lecture Notes in Statistics, 122, Gregoire, T. G., Brillinger, D. R., Diggle, P. J., Russek-Cohen, E., Warren, W. G. and Wolfinger, R. D., eds., Springer, New York, pp. 289-298.

[46] Fuentes, M.,Chen, L., Davis, J. and Lackmann, G. (2005). A new class of nonseparable and nonstationary covariance models for wind fields. Environmetrics, 16, 449-464.

[47] Huang and Hsu, 2004. Modeling transport effects on ground-level ozone using a non-stationary space-time model. Environmetrics. v15. 251-268.

[48] Wikle, C.K., Berliner, L.M., and Cressie, N. (1998). Hierarchical Bayesian space-time models, Environmental and Ecological Statistics, 5, 117-154.

[49] Mardia, K. V., Goodall, C. R., Redfern, E., Alonso, F.J. (1998). The kriged Kalman filter (with Discussion). Test 7, 217-85

[50] Wikle, C. and Cressie, N. (1999), A dimension reduction approach to space time Kalman filtering. Biometrika 86, 815-829.

[51] Cressie, N. and C.K. Wikle, (2002). Space-time Kalman filter. Entry in Encyclopedia of Environmetrics, vol.4, eds. A.H. El-Shaarawi and W.W. Piegorsch. Wiley, New York, pp.2045-2049.

[52] Xu, Ke; Wikle, Christopher K. (2007) Estimation of parameterized spatio-temporal dynamic models. J. Stat. Plann. Inference 137, No. 2, 567-588

[53] Stroud, J.R., Muller, P., and Sanso, B. (2001). Dynamic models for spatio-temporal data. Journal of the Royal Statistical Society, Series B, 63, 673-689.

[54] Huerta G, Sanso B, Stroud JR. 2004. A spatiotemporal model for Mexico City ozone levels. Journal of the Royal Statistical Society, Series C 53: 231-248.

[55] Gelfand, A.E., S. Banerjee, D. Gamerman (2005) Spatial Process Modelling for Univariate and Multivariate Dynamic spatial Data. Environmetrics 16, 1-15.

[56] Higdon, D., 2001. Space and space time modeling using process convolutions. Technical Report, ISDS, Duke University.

[57] Calder, C.A. (2007). Dynamic Factor Process Convolution Models for Multivariate Space-Time Data with Application to Air Quality Assessment. Environmental and Ecological Statistics, 14, 229-247,

[58] Cressie N., Huang H.C. (1999). Classes of Nonseparable, Spatiotemporal Stationary Covariance Functions. Journal of the American Statistical Association, 94, 1330-1340.

[59] Wikle, C.K., Milliff, R.F., Nychka, D., and L.M. Berliner, (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. JASA , 96 , 382-397.

[60] Berliner, L.M., Milliff, R.F., and C.K. Wikle, (2003). Bayesian hierarchical modeling of air-sea interaction. Journal of Geophysical Research - Oceans, , 108(C4)

[61] Berliner, L.M., Wikle, C.K. and N. Cressie, (2000). Long-lead prediction of Pacific SSTs via Bayesian Dynamic Modeling. Journal of climate, 13, 3953-3968.

[62] Wikle, C.K., (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology , 84, 1382-1394

[63] Wikle, C.K. and Hooten, M.B. (2006). "Hierarchical Bayesian Spatio-Temporal Models for Population Spread". In: "Applications of Computational Statistics in the Environmental Sciences: Hierarchical Bayes and MCMC Methods". J.S. Clark and A. Gelfand (eds). Oxford University Press.

[64] Wikle, C.K., L.M. Berliner, and R.F. Milliff, (2003). Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. Monthly Weather Review , 131 , 1051-1062.

[65] Wikle, C.K. (2001). A kernel-based spectral approach for spatio-temporal dynamic models. Proceedings of the 1st Spanish Workshop on Spatio-Temporal Modelling of Environmental Processes (METMA), Benicassim, Castellon (Spain), 28-31 October 2001, pp. 167-180.

[66] Wikle, C.K. (2003). A kernel-based spectral model for non-Gaussian spatio-temporal processes. Statistical Modelling: An International Journal, 2, 299-314.

[67] Brown, P.E., Karesen, K.F., Roberts, G.O., and Tonellato, S. (2000). Blur-generated non-separable space-time models. J.R. Statistical Society, Series B, 62, 847-860.

[68] Brown, P.E., Diggle, P.J., Lord, M.E., and Young, P.C. (2001). Space-time calibration of radar rainfall data. Applied Statistics, 50, 221-241.

[69] Whittle, P. (1962). Topographic correlation, power-law covariance functions, and diffusion. Biometrika 49, 305-314.

[70] Xu, Ke, Wikle, Christopher; Fox, Neil (2005). A Kernel-Based Spatio-Temporal Dynamical Model for Nowcasting Weather Radar Reflectivities. Journal of the American Statistical Association, Volume 100, Number 472, December 2005 , pp. 1133-1144.

[71] S. Sahu, Gelfand, A.E. and D.M. Holland, 2006. Spatio-temporal modeling of fine particulate matter. Journal of Agricultural, Biological, and Environmental Statistics, 11, (1), 61-86.

[72] Shaddick, G. and Wakefield, J. (2002) Modelling daily multivariate pollutant data at multiple sites. Appl. Statist., 51, 351-372.

[73] D. Cocchi, F. P. Greco, C. Trivisano (2007). Hierarchical space-time modelling of PM10 pollution. Atmospheric Environment, 41,3, pp. 532-541.

[74] S. Sahu, Gelfand, A.E. and D.M. Holland, 2007. High resolution space-time ozone modeling for assessing trends. Journal of the American Statistical Association, 102, (480), 1221-1234.

[75] M. Fuentes, L. Chen, J. Davis 2008 A class of nonseparable and nonstationary spatial temporal covariance functions, Environmetrics vol: 19 , Issue: 5 , Pages: 487 - 507

[76] Choi, J., Reich, B., Fuentes, M., and Davis, J.M. (2008). Multivariate spatial-temporal modeling and prediction of speciated fine particles. Journal of Statistical Theory and Practice, accepted.

[77] 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.

[78] Alessandro Fassò, Michela Cameletti (2007). A general spatio-temporal model for environmental data: Working paper of GRASPA n 27

[79] Raftery, Adrian, Fuentes, Montserrat, Gneiting, Tilmann, and Gel, Yulia (2002), Assessing Uncertainty in Numerical Weather Prediction, Computing Science and Statistics, 34,

[80] Fuentes, M, Guttorp, and Challenor, P. (2003). Statistical assessment of numerical models. International Statistical Review. 71, 2, 201-221.

[81] Fuentes, M. and Raftery, A. (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with output from numerical models. Biometrics, 61, 36-45.

[82] A Hierarchical Bayesian Approach to the Spatio-Temporal Modeling of Air Quality Data, by A. Riccio et al., Atm. Env., 40, 554-566, 2006

[83] Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). Deviance information criteria for missing data models.Ó Bayesian Analysis, Number 4, pp. 651-674

[84] Plummer, M., 2008, Penalized loss functions for Bayesian model comparison. Biostatistics (Oxford, England), 9(3),523-39.

[85] Gelfand and Ghosh, 1998. Model choice: a minimum posterior predictive loss approach. Biometrika. v85. 1-11.

[86] Cressie, N. (1993). Statistics for Spatial Data (Revised Edition). John Wiley and Sons, Inc., New York.

[87] Banerjee, S; Carlin, B; Gelfand, A. 2004 Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall; 2004.

[88] Chen, L., Fuentes, M., and Davis, J., (2006) Spatial-Temporal Statistical Modeling and Prediction of Environmental Processes in Applications of Computational Statistics in the Environmental Sciences: Hierarchical Bayes and MCMC Methods edited by James S Clark and Alan Gelfand, Oxford University Press.

[89] Whittle, P. (1963). Stochastic processes in several dimensions, Bull. Int. Statist. Inst. 40, 974-94.

[90] Sahu, Sujit K., Challenor, Peter (2007). A space-time model for joint modeling of ocean temperature and salinity levels as measured by Argo floats Environmetrics, 19, (5), 509-528.

[91] Kyriakidis P.C., Journel A.G. (1999) Geostatistical Space-time models: a review, Mathematical Geology, 31, 6, 651-684.

[92] Gneiting, T., Genton, M. G. and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In Finkenstadt, B., Held, L. and Isham, V. (eds.), Statistical Methods for Spatio-Temporal Systems, Chapman & Hall/CRC, Boca Raton, pp. 151-175.

[93] Handcock, M. S. & Wallis, J. R. (1994) An approach to statistical spatio-temporal modeling of meteorological fields. J. Am. Statist. Assoc. 89, 368-78.

[94] Wikle, C.K. and L.M. Berliner, (2005). Combining information across spatial scales. Technometrics , 47, 80-91.