Optimal sampling design for functional and spatio-temporal random fields.

Dissertation submitted in partial fulfillment for the degree of Doctor in Science - Statistics

Martha Patricia Bohorquez Castañeda

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
Diciembre de 2015
Optimal sampling design for functional and spatio-temporal random fields.

Dissertation submitted in partial fulfillment for the degree of Doctor in Science - Statistics

Martha Patricia Bohorquez Castañeda

Advisor
Ramón Giraldo Henao, Ph.D.

Coadvisor
Jorge Mateu, Ph.D.

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
Diciembre de 2015
Abstract: We extend the framework of optimal sampling designs to the spatial prediction of univariate and multivariate functional data. In both cases, we derive unbiased predictors and their variances. In the univariate case, we propose to use a simple cokriging predictor with the scalar random fields resulting from the scores associated with the representation of the functional data with the empirical functional principal components. In the multivariate case, we develop spatial prediction of a functional variable at unsampled sites, using functional covariates, that is, we present a functional cokriging method. We show that through the representation of each function in terms of its empirical functional principal components, the functional cokriging only depends on the auto-covariance and cross-covariance of the associated score vectors, which are scalar random fields. Design criteria are given for all predictors derived in this thesis. In addition, we develop a methodology for dynamic spatial sampling designs to find the optimal spatial mean estimation and the optimal spatial prediction at some future time points, based on the temporal variation of the spatial dependence structure. The methodologies are applied to the networks of air quality of Bogotá and México.

Resumen: Esta tesis extiende los diseños de muestreo óptimos a la predicción espacial univariada y multivariada de datos funcionales. En ambos casos, se presentan predictores insesgados con sus respectivas varianzas. En el caso univariado, se propone usar cokriging simple sobre el campo aleatorio escalar formado por los puntajes asociados con la representación de los datos funcionales en términos de sus componentes principales funcionales empíricos. En el caso multivariado, se desarrolla la predicción espacial de una variable funcional en sitios no muestreados, usando covariables funcionales, es decir, se presenta el cokriging funcional. Se demuestra que a través de la representación de cada función en términos de sus componentes principales funcionales empíricos, el cokriging funcional solo depende de la auto-covarianza y de la covarianza cruzada de los vectores de puntajes asociados, los cuales son campos aleatorios escalares. Se proponen criterios de diseño para todos los predictores desarrollados en esta tesis. Adicionalmente, se construye una metodología para diseños de muestreo espacial dinámicos que permitan encontrar la estimación óptima de la media espacial y la predicción espacial óptima en un tiempo futuro, basados en la variación temporal de la estructura de dependencia espacial. Las metodologías son aplicadas a las redes de calidad del aire de Bogotá y México.

Keywords: Functional data, Geostatistics, Optimal sampling, Spatio-temporal data.

Palabras clave: Datos funcionales, Geostadística, Muestreo óptimo, Datos espacio-temporales.
Acceptation Note

Thesis Work

“mention”

Jury

Jury

Jury

Advisor
Ramón Giraldo Henao

Coadvisor
Jorge Mateu

Bogotá, D.C., Diciembre de 2015
Dedicado a

A mis padres, esposo e hija.
Agradecimientos

Agradezco a Ramón Giraldo y a Jorge Mateu por su continuo apoyo y orientación. A mi familia por su paciencia y amor. A mis amigos los zoomatemáticos por su compañía y alegría. A Sotico, Sergio, Ruben, Mayito y Luisfer por su amistad y consejo.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>I</td>
</tr>
<tr>
<td>List of Tables</td>
<td>III</td>
</tr>
<tr>
<td>List of Figures</td>
<td>IV</td>
</tr>
<tr>
<td>Introduction</td>
<td>VI</td>
</tr>
<tr>
<td>1. Background</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Spatial process</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Spatio-temporal process</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Multivariate spatial prediction</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Optimal sampling</td>
<td>8</td>
</tr>
<tr>
<td>1.5 Functional data</td>
<td>11</td>
</tr>
<tr>
<td>1.5.1 Functional principal components</td>
<td>12</td>
</tr>
<tr>
<td>1.5.2 Spatial prediction of functional data</td>
<td>14</td>
</tr>
<tr>
<td>2. Optimal sampling for spatial prediction of functional data</td>
<td>16</td>
</tr>
<tr>
<td>2.1 Optimal spatial sampling for functional data</td>
<td>16</td>
</tr>
<tr>
<td>2.1.1 Optimal spatial sampling using ordinary functional kriging</td>
<td>16</td>
</tr>
<tr>
<td>2.1.2 Optimal spatial sampling using FPCA and simple kriging</td>
<td>17</td>
</tr>
<tr>
<td>2.1.3 Optimal spatial sampling for functional prediction using basis coefficients</td>
<td>18</td>
</tr>
<tr>
<td>2.1.3.1 Functional prediction using FPCA and simple cokriging</td>
<td>18</td>
</tr>
<tr>
<td>2.1.3.2 Optimal spatial sampling using FPCA and simple cokriging</td>
<td>20</td>
</tr>
<tr>
<td>2.2 Real data analysis</td>
<td>21</td>
</tr>
<tr>
<td>3. Multivariate spatial functional random field. Prediction and optimal sampling.</td>
<td>27</td>
</tr>
</tbody>
</table>
3.1 Functional cokriging .................................................. 27
   3.1.1 Cokriging with two functional random fields .......... 28
   3.1.2 Cokriging with $P$ functional random fields .......... 30
   3.1.3 Optimal spatial sampling for functional cokriging .... 31
3.2 Simulation study .................................................... 32
3.3 Real data analysis .................................................. 34

4. Optimal dynamic spatial sampling ............................... 43
   4.1 Model and covariance structure ............................... 43
   4.2 Forecasting of the spatial covariance parameter .......... 44
   4.3 Optimal spatial sampling for mean estimation .......... 48
   4.4 Optimal sampling for spatial prediction ................... 49
   4.5 Simulation study ............................................... 50
   4.6 Real data analysis ............................................ 54

Discussion and conclusions ........................................ 58

Bibliography ......................................................... 61
List of Tables

3.1 Nested variogram components of the linear model of coregionalization model using two nested Matérn structures. ............................................. 35

4.1 Median and median absolute deviation over 100 realizations of the optimal spatial mean estimation for the future time $T + 1$ when the exponential covariance model is used and the covariance parameter $\Theta_t = (\sigma_t^2, \phi_t)$ follows different time series models A-F. See the meaning of letters A to F in the text. 52

4.2 Optimal spatial prediction estimation for the future time $T + 1$ when the exponential covariance model is used and the covariance parameter $\Theta_t = (\sigma_t^2, \phi_t)$ follows the VAR(1) model F. See the meaning of letter F in the text. The data are generated from a 12-dimensional VAR(1) model. ............... 53
List of Figures


2.2 Top Left Panel: Variogram of $f_1$. Top Right Panel: Variogram of $f_2$. Bottom Left Panel: Cross-variogram of $f_1, f_2$. Bottom Right Panel: Trace-variogram. ........................................... 24

2.3 Left: Optimal sampling network assuming that all stations can be moved. Right: Optimal location for the mobile station keeping the current network fixed. .......................................................... 25


3.1 Top Left Panel: Simulated data, $\chi_1^s$. Top Right Panel: Simulated data, $\chi_2^s$. Bottom Left Panel: Grid of simulation locations. Bottom Right Panel: Cross validation residuals and residual mean. ......................... 33

3.2 Left Panel: México city. Air quality network RAMA (shown in red). The blue points are the stations that measure temperature but belong to REDMET. Right Panel: PM10, May 23 to May 30, 2015. ......................... 34


3.4 Experimental and theoretical variogram fitted according to the linear model of coregionalization. .......................................................... 36

3.5 Left Panel: Optimal location for one additional station. Right Panel: Cross validation residuals and residual mean. ......................... 37

3.6 Cross validation residuals and residual mean of the predictions of PM10 using functional kriging method proposed in [14]. ....................... 38

3.7 Density plots for the PM10 data at some stations. ................................. 39

3.8 Density plots for the NO2 data at some stations. .................................. 40

3.9 Density plots for the Temperature data at some stations. ....................... 41
3.10 Density plots for the first two scores of each variable. .......................... 42

4.1 $\Theta_{T+1}$ is the vector of forecasts based on the time series of the spatial covariance parameter estimates at time points $1, ..., T$. Note that the spatial sampling sizes $n_t$ can be different for each temporal instant. Also, an observation location $s_i$ at time $t$ can vary to $s_i$ at next time $t+1$ and so on. $s_i, s_i, s_i \in \mathbb{R}^d$. ........... 45

4.2 Spatial configuration of the unsampled interest sites $S_{501}^0$ (unfilled circles) and optimal dynamic sampling design $S_{501}$ (filled circles). .......................... 54

4.3 Some semivariograms obtained on May 7th, 2014 .......................... 55

4.4 Autocorrelation and cross-correlation plots for time series $\Theta_t = (\hat{\sigma}_t^2, \hat{\phi}_t)$ .......................... 56

4.5 Spatial locations of the mobile station for the next seven days for the optimal mean estimation. .......................... 57
A sampling design that finds the best combination estimator-design or predictor-design, according to the optimization of a criterion previously established is called optimal. Shows several design criteria based on the information matrix. In order to use these criteria with correlated data, spatial sampling designs are built using the concept of entropy, selecting the subset of sites maximizing the amount of information about the subset of the non observed sites (see, for example, [33]; [155], [5]). Develop a hierarchical Bayesian framework for the design of monitoring networks in staircase data patterns and they emphasize the importance of considering auxiliary information in the environmental network design. Illustrate the effect of the spatio-temporal points set on the information measure by using several spatio-temporal autoregressive moving average (STARMA) processes with different levels and types of covariance. [29], [154], [4] and [42] propose optimal sampling designs based on the state space model approach for dynamic spatio-temporal processes. Build static spatial sampling designs which optimize the covariance parameter estimation and prediction. Present a complete review on the theory of sampling designs to collect spatial and spatio-temporal data. All these methods require distributional assumptions, the iterative inversion of the covariance matrix with too much data or a deep scientific knowledge of the phenomenon, for example, the differential equation that describes the process ([42]). Due to the big volume of information registered in many areas such as monitoring of the weather and air pollution, public health, medicine ([83] and [79]), among others, functional data analysis has gained relevance. When the curves are time series varying spatially, spatial functional data analysis is an alternative approach to spatio-temporal modeling. Thus, spatial and spatio-temporal sampling strategies need new developments, facing these goals and features. Suggest the combination of the state space models and geostatistical methods with filtering and smoothing techniques. In fact, the spatial functional data analysis comprises these techniques. Recognizes the necessity of new developments in spatial correlated functional data and extends the spatial autoregressive model and the spatial moving average model to stochastic processes taking values in Hilbert spaces. The projection methodology based on the eigenfunctions basis of the autocovariance operator has been used in [130], [128] and [129]. Assume the existence of an underlying spatial functional Markovian model, and consider the problem of least-squares linear estimation for random fields with values in fractional Sobolev spaces of variable order. Spatial data allow the optimal prediction of the variable of interest at unsampled sites using geostatistical methods. The advantage is that here is carried out the optimal prediction of the whole curve [135]. Extensions of univariate geostatistical methods to functional data have been given by [67], [56], [77] and [58]. The ordinary kriging method to predict functions at unsampled sites is provided by [57], using non-parametric methods to build the curves. Extensions that introduce exogenous variables
for the external drift with applications to air quality monitoring and weather analysis are given by [31] and [84]. [54] develops functional prediction by using multivariate geostatistics. The existing methods of spatial and spatio-temporal sampling optimize the prediction of spatial or spatio-temporal points, and so far there is no step forward in sampling for spatially correlated curves.

Here we propose a methodology for optimal spatial sampling for functional data. This approach can be also used when data at each site of a region correspond to long time series. In this case, we can apply the methodology by carrying out in a first step of the analysis a smoothing analysis to get functional data from discrete observations. Specifically, the interest is optimal spatial sampling design in order to ensure a good quality in the spatial prediction of functions. For this purpose, we define design criteria considering ordinary kriging for functional data [57] and simple kriging [77]; we also propose, an adaptation of the functional spatial prediction problem [54, 111] using empirical functional principal components.

Additionally, we generalize the methodology to the case of multivariate functional geostatistics. The performance of the prediction of a functional variable is improved taking the spatial cross-covariance with other functional variables into account. That is, we propose optimal sampling designs that ensure a good quality of the spatial functional prediction in presence of spatially correlated covariates. For this purpose, we use again the empirical functional principal components (EFPC) representation, because it allows the use of scalar variables to find the spatial functional prediction by modeling the spatial auto-covariance and cross-covariance of the EFPC associated score vectors.

When the time series do not have too many observations, or there is no interest in functional data analysis, we propose an alternative methodology based on the temporal dynamic behavior of the covariance parameters. The interest now is the optimal spatial mean estimation and the optimal spatial prediction at some future time points. The criteria to select the new locations depends on the variance of the estimation or the prediction to optimize, which, in turn depends on the spatial covariance structure. [154] and [42] recognize the dynamic nature of the covariance matrix and emphasize that modeling spatio-temporal processes, requires extensive knowledge of the phenomena and hence its dynamical structure is usually unknown. So, we propose to focus on the time-varying spatial covariance function and model its dynamic behavior. We present dynamic spatial sampling designs using criteria that depend only on the covariance structure, thus, our approach does not need the knowledge of the spatio-temporal process model. First, we estimate the parameters of the spatial covariance at each time point observed, obtaining time series with measurement errors (uncertainty). Then, we model these time series and find their forecasting at the next time points when the sampling locations can be changed, and finally we evaluate the respective criterion. The design criteria are computed by plugging the forecasts for the unknown parameters, so, we find locally optimal designs (78). For simplicity, we do not use the word locally hereafter.

We develop dynamic spatial sampling for the optimal spatial mean estimation and the optimal spatial prediction at some future time points. For the mean estimation, we use the generalized least squares estimator (133), and for prediction we use ordinary kriging (see, for example, 39). We propose design criteria based on the dynamic behavior of the variance of the generalized mean estimation and the prediction error, respectively.

The thesis follows with some background in Chapter 1, the methodological setup for univariate and multivariate spatial functional data in Chapters 2 and 3. The dynamic spatial sampling methodology is described en Chapter 4. Real datasets are analyzed in
each chapter. We apply the proposed methodologies to the design of two environmental monitoring networks. The thesis ends with some concluding remarks.
CHAPTER 1

Background

1.1 Spatial process

A spatial process is a stochastic process \( \{Z(s), s \in D_s \subset \mathbb{R}^d\} \), where \( \mathbb{R}^d \) is the spatial index set. This approach fits models in continuous space based on a finite number \( n \) of spatial observations \( Z \equiv (Z(s_1), ..., Z(s_n)) \). The geostatistical method is based on the first two moments of \( Z(s) \). Given that \( \mu(s) = E(Z(s)) \) is known for \( s \in D_s \), the optimal prediction of the random variable \( Z(s_0) \), at unsampled location \( s_0 \in D_s \), is the following linear combination of the data \[42]:

\[
Z^*(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) = \lambda'Z
\]

where

- Let \( \mu = (\mu(s_i), i = 1, ..., n) \)
- \( \lambda \) is chosen to minimize the mean squared prediction error,

\[
E(\lambda'Z - Z(s_0))^2
\]

\( \lambda \) is a \( n \)-vector.

\[
\lambda' \equiv (\lambda(s_i); i = 1, ..., n)'
\]

\( \lambda(s_i) \) represents the contribution of the observation at the site \( s_i \) to the prediction \( Z^*(s_0) \). In this procedure, the main stage is to estimate the covariance matrix by fitting a valid function \( C \). That is, a positive definite function \( C \) which ensures that the prediction variance \( Var(Z^*(s_0)) \) be nonnegative. The optimal prediction after the optimization is

\[
Z^*(s_0) = \mu(s_0) + c'\Sigma^{-1}(Z - \mu)
\]

with minimized mean-squared prediction error

\[
E(Z^*(s_0) - Z(s_0))^2 = C(0) - c'\Sigma^{-1}c
\]
where

\[ C(0) = \text{Var}(Z(s)), \quad \Sigma \equiv \text{Cov}(Z), \quad c' = \text{cov}(Z(s_0); Z) \quad \mu \equiv E(Z); \tag{1.6} \]

and \( \Sigma \) is a positive definite matrix. The \( i, j \)-th entry of \( \Sigma \) is \( (\Sigma)_{ij} = \text{Cov}(Z(s_i), Z(s_j)) \). The kriging has variations in different cases like unknown mean or non-constant mean. In those cases, constraints over \( \lambda \) are required in order to ensure unbiasedness. The kriging is an optimal linear predictor and does not require the stationarity or normality assumptions.

**Covariance parameter estimation for the spatial process**

**Least squares methods**

Least squares methods are used to choose the model \( C \) and initial values \( \Theta_0 \) based on graphical exploratory data analysis. Often, these methods carry out the estimation of the variogram \( 2\gamma \), which under weak stationarity is related with the covariance function through the relation

\[ 2\gamma(s_i - s_j|\Theta) = \text{Var}(Z(s_i) - Z(s_j)) = 2(C(0|\Theta) - C(s_i - s_j|\Theta)) \tag{1.7} \]

for \( s_i - s_j \in \mathbb{R}^d \). In geostatistical analysis it is widely used the weighted least-squares method which makes use of the matrix \( W(\Theta) \), proposed in [38], and given by

\[ W(\Theta) \simeq \text{Diag} \left( \frac{2(2\gamma(s_i - s_j|\Theta))^2}{|N(s_i - s_j)|} \right) \]

where \( N(s_i - s_j) = \{(i, j) : s_i - s_j\} \), for \( i, j = 1, \ldots, n \) that generates the first \( q \) spatial lags \( q = 1, \ldots, Q \). These methods usually use spatial lag classes until half of maximum distance between each pair of locations.

**Composite likelihood (CL)**

[101] avoids the use of a multivariate distribution by maximizing a pseudo-likelihood function \( CL(\Theta) \), see (1.8). The marginal distributions \( f(Z(s_i), \Theta) \) are assumed known, except for the parameter \( \Theta \). Then, if \( l(Z(s_i), \Theta) = \ln(f(Z(s_i), \Theta)) \) is the log-likelihood function, the associated composite likelihood is given by

\[ CL(\Theta) = \sum_{i=1}^{n} l(Z(s_i), \Theta) \tag{1.8} \]

The gradient of (1.8), \( \nabla CL(\Theta) = CS(\Theta) \), is called the composite score function. The estimator \( \hat{\Theta} = ((\hat{\theta}_1), ..., (\hat{\theta}_K)) \) is found by solving the equations system

\[ CS((\theta)_k) = \sum_{i=1}^{n} \nabla l(Z(s_i), (\theta)_k) = 0, \quad k = 1, ..., K. \tag{1.9} \]
The adaptation of this method to estimate $\gamma$ is as follows. Let $V(i,j) = Z(s_i) - Z(s_j)$, $i,j = 1,...,n$ the increments variable. Under the assumption of marginal Gaussian distributions $Z(s_i) \sim N(\mu,\sigma^2)$ for $s \in \mathbb{R}^d$, the distribution of $V(i,j)$ is

$$V(i,j) \sim N(0, 2\gamma(s_i - s_j|\Theta)).$$  \hspace{1cm} (1.10)

After applying (1.8) and (1.9), the resulting composite score function for $V^2(i,j)$ is

$$CS(\Theta; i,j) = \frac{n-1}{n} \sum_{i=1}^{n} \sum_{j>i} \frac{\partial \gamma(s_i - s_j|\Theta)/\partial \Theta}{4\gamma^2(s_i - s_j|\Theta)} \left( V^2(i,j) - 2\gamma(s_i - s_j|\Theta) \right)$$  \hspace{1cm} (1.11)

which is a nonlinear weighted least-squares estimation of $V^2(i,j) = 2\gamma(s_i - s_j|\Theta) + \varepsilon(i,j)$, where $\varepsilon(i,j)$ are independent random variables with $E(\varepsilon(i,j)) = 0$ and $Var(\varepsilon(i,j)) = 8\gamma^2(s_i - s_j|\Theta)$

**Maximum likelihood**

The estimation of $\Theta$ for the spatial covariance model $C$ using maximum likelihood (ML) requires the assumption that the vector $Z$ follows a multivariate Gaussian distribution, $Z \sim N_n(\mu, \Sigma)$, where $\Sigma$ is defined in (1.6).

### 1.2 Spatio-temporal process

A spatio-temporal process is the stochastic process $\{Z(s,t) : (s,t) \in D_s \times D_t\}$, where $D_s \times D_t$ is the spatio-temporal index set. $D_s \times D_t \subseteq \mathbb{R}^d \times \mathbb{R}$ with $\mathbb{R}^d$ for space and $\mathbb{R}$ for the time. These approaches fit models in continuous space and time, based on a finite number $n$ of spatial observations and $T$ of temporal observations of the process: $Z \equiv (Z(s_1,t_1),...,Z(s_1,t_T),...,Z(s_n,t_1),...,Z(s_n,t_T))$.

However, sometimes the available observations of the series have different length at all locations. Then, a more general form of expressing the spatio-temporal observations of the phenomenon of interest is

$$Z \equiv (Z(s_i; t_{ij}), \hspace{0.2cm} i = 1,...,n, \hspace{0.2cm} j = 1,...,T_i)$$  \hspace{1cm} (1.12)

- \{s_1,...,s_n\} are the $n$ known spatial locations,
- $T_i$ is the length of the time series available at each of the $n$ locations.

If $\mu(s,t) = E(Z(s,t))$ is known for $(s,t) \in D_s \times D_t$, the optimal prediction of the random variable $Z(s_0,t_0)$, in one unsampled space-time coordinate $(s_0,t_0) \in D_s \times D_t$, is the following linear combination of the data [42]:

$$Z^*(s_0,t_0) = \sum_{i=1}^{n} \sum_{j=1}^{T_i} \lambda_{ij} Z(s_i; t_{ij}) = \mathbf{X}'Z$$  \hspace{1cm} (1.13)

where
Let \( \mu = (\mu(s_i; t_{ij}), i = 1, ..., n; j = 1, ..., T_i) \)

- \( \lambda \) is chosen to minimize the mean squared prediction error,
  \[ E \left( Z(s_0, t_0) - \lambda' Z \right)^2 \]  
  (1.14)

\( \lambda \) is a \( \sum_{i=1}^{n} T_i \)-vector.

\[ \lambda(s_i, t_{ij}) \] represents the contribution of the observation at the site \( s_i \) at time \( t_{ij} \) to the prediction \( Z^*(s_0, t_0) \).

\[ Z^*(s_0; t_0) = \mu(s_0; t_0) + c(s_0; t_0)' \Sigma^{-1}(Z - \mu) \]  
(1.16)

with minimized mean-squared prediction error

\[ E \left( Z^*(s_0; t_0) - Y(s_0; t_0) \right)^2 = \text{Var}(Z(s_0; t_0)) - c(s_0; t_0)' \Sigma^{-1} c(s_0; t_0) \]  
(1.17)

where

\[ \Sigma \equiv \text{Cov}(Z), \quad c(s_0; t_0)' = \text{cov}(Z(s_0; t_0); Z) \quad \text{y} \quad \mu \equiv \text{E}(Z); \]  
(1.18)

and \( \Sigma \) is a positive definite matrix.

Notice that the equations 1.16 and 1.17 involve calculating \( \Sigma^{-1} \), which can be a drawback because of the sizes of spatio-temporal datasets. Too much data can lead to problems with the iterative inversion of the covariance matrix. This is an advantage of the estimation of the spatial covariance at each time because in most of the cases, there are not too much spatial locations. Nevertheless, there are alternatives that make operative the kriging even for large spatio-temporal datasets [41].

**Dynamical models**

When enough scientific knowledge about the phenomenon of interest is available, hierarchical models provide a quite general framework. The model can be formulated like a dynamic one (based on conditional distributions), and it can be considered as a hierarchical statistical model. For example, if the process of interest \( Y(s_i; t_{ij}) \) cannot be observed directly, the data are

\[ Z(s_i; t_{ij}) = Y(s_i; t_{ij}) + \varepsilon(s_i; t_{ij}) \quad j = 1, ..., T_i \quad i = 1, ..., n \]  
(1.19)

where \( \{\varepsilon(s_i; t_{ij})\} \sim iid(0, \sigma^2_\varepsilon) \) is independent of \( Y(\cdot, \cdot) \) and represents the measurement error.

Thus, it can be formulated like a dynamic model. From model (1.19), the first level and second level respectively are

\[ [Z|Y(\cdot, \cdot), \theta_D] \]

and

\[ [Y(\cdot, \cdot)|\theta_P] \]

where \( \theta_D \) are parameters in the data model and \( \theta_P \) are parameters of the process model.

Now, with Bayes’ Theorem can be made inferences about \( Y(\cdot, \cdot) \); there are two options:
1. Empirical hierarchical model (EHM). Estimate $\Theta = \{\Theta_D, \Theta_P\}$ from the data $Z$ and replace this value in the posteriori distribution (plug-in).

$$[Y(\cdot; \cdot)|Z, \Theta]\alpha[Z|Y(\cdot; \cdot), \Theta_D][Y(\cdot; \cdot)|\Theta_P]$$

2. Bayesian hierarchical model (BHM). Add another level to structure for the parameter $\Theta$

$$[Y(\cdot; \cdot)|Z, \Theta]\alpha[Z|Y(\cdot; \cdot), \theta_D][Y(\cdot; \cdot)|\theta_P][\Theta]$$

For instance, if the spatial index is continuous, it can be a representation of a hierarchical bayesian kriging where the distribution of the covariance parameters is in the last level.

If the space is considered discrete the model could be a spatio temporal autoregressive moving average, (STARMA); for example a $VAR(1)$ model,

$$Y_t = MY_{t-1} + W_t$$

where $M$ is the propagator matrix influencing over $Y_t$ of the variable in every observed place in the previous time, and $\{W_t\}$ is a $n - variate$ white noise process, $\{W_t\} \sim iid(0, Q)$; in this case the key is the estimation of $M$; it has $n^2$ elements and needs to reduce the dimension, which makes necessary to find ways to reduce the quantity of parameters. One of the advantages of this particular model is that some stochastic differential equations can be replaced by $VAR(p)$ - models, making the statistical analysis simpler.

### 1.3 Multivariate spatial prediction

The spatial multivariate prediction is a topic of big interest. Sometimes the study of the phenomenon requires the knowledge of $p$ variables simultaneously. However, this work has an additional importance because its connection with geostatistics for functional data. The data are $n, p \times 1$-vectors $Z(s_1), ..., Z(s_n)$, and the data matrix is an $n \times p$ matrix with $(i, j)th$ element $Z_j(s_i)$:

$$Z \equiv (Z(s_1), ..., Z(s_n))'$$

and

$$Z(s_i) = (Z_1(s_i), Z_2(s_i), ..., Z_p(s_i))'$$

The interest generally is to predict

$$Z(s_0) \equiv (Z_1(s_0), ..., Z_p(s_0))'$$

The predictor for the $Z_l(s_0), 1 \leq l \leq p$, based on all $p$ variables is the following linear combination of the data:

$$Z^*_l(s_0) = \sum_{i=1}^n \sum_{q=1}^p \lambda_{lq}Z_q(s_i)$$

With this method, the prediction is carried out for each variable once at the time. Assume that this time the process can be measured without error. This method is known like cokriging, given the following assumptions

1. Constant mean $E(Z(s)) = \mu$, with $\mu = (\mu_1, ..., \mu_p)'$, para todo $s \in D_s$ and
2. The covariance structure for each pair of locations is a $p \times p$ matrix $\Sigma_{(s_i, s_{i'})}$, which only depends of locations. That is, the cross covariance structure between two variables $Z_l$ and $Z_{l'}$, $l, l' = 1, \ldots, p$ at two sites $s_i, s_{i'} \in D_s$, is a function of $s_i$ and $s_{i'}$,

$$\text{Cov}(Z_l(s_i), Z_{l'}(s_{i'})) = C_{l'l'}(s_i, s_{i'})$$  \hspace{1cm} (1.25)$$

$$\Sigma_{(s_i, s_{i'})} = \begin{pmatrix}
C_{11}(s_i, s_{i'}) & C_{12}(s_i, s_{i'}) & \cdots & C_{1p}(s_i, s_{i'}) \\
C_{21}(s_i, s_{i'}) & C_{22}(s_i, s_{i'}) & \cdots & C_{2p}(s_i, s_{i'}) \\
\vdots & \vdots & \ddots & \vdots \\
C_{p1}(s_i, s_{i'}) & C_{p2}(s_i, s_{i'}) & \cdots & C_{pp}(s_i, s_{i'})
\end{pmatrix} \hspace{1cm} (1.26)$$

which is not necessarily symmetric because in general, there is no reason for the assumption

$$\text{Cov}(Z_l(s_i), Z_{l'}(s_{i'})) = \text{Cov}(Z_{l'}(s_i), Z_l(s_{i'}))$$  \hspace{1cm} (1.27)$$

about the cross covariances. Thus in general, $[110]$,

$$\Sigma_{(s_i, s_{i'})} \neq \Sigma_{(s_{i'}, s_i)} \quad \text{and} \quad \Sigma_{(s_i, s_{i'})} \neq \Sigma'_{(s_i, s_{i'})}$$

but it always satisfies

$$\Sigma'_{(s_i, s_{i'})} = \Sigma_{(s_{i'}, s_i)}$$

Then, the complete covariance matrix $\text{Cov}(\mathbf{Z}) = \sum$ with all variables and the observed locations is

$$\sum = \begin{pmatrix}
\Sigma_{(s_1, s_1)} & \Sigma_{(s_1, s_2)} & \cdots & \Sigma_{(s_1, s_n)} \\
\Sigma_{(s_2, s_1)} & \Sigma_{(s_2, s_2)} & \cdots & \Sigma_{(s_2, s_n)} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{(s_n, s_1)} & \Sigma_{(s_n, s_2)} & \cdots & \Sigma_{(s_n, s_n)}
\end{pmatrix} \hspace{1cm} (1.28)$$

In predictor [1.24] is necessary to have measures for all variables in the same places; however, there are modifications to allow using auxiliary variables even when this is not the case. To guarantee unbiasedness using the predictor [1.24] $E(Z_l^*(s_0) - Z_l(s_0)) = 0$ that is,

$$E(Z_l^*(s_0) - Z_l(s_0)) = E \left[ \sum_{i=1}^{n} \sum_{q=1}^{p} \lambda_{qi} Z_q(s_i) - Z_l(s_0) \right] = 0 \hspace{1cm} (1.29)$$

Taking the expectation and applying the assumption of constant mean

$$\sum_{i=1}^{n} \sum_{q=1}^{p} \lambda_{qi} \mu_q = \mu_l \hspace{1cm} (1.30)$$

$$\sum_{i=1}^{n} \lambda_{l_i} \mu_1 + \sum_{i=1}^{n} \lambda_{2i} \mu_2 + \cdots + \sum_{i=1}^{n} \lambda_{l_i} \mu_l + \cdots + \sum_{i=1}^{n} \lambda_{p_i} \mu_p = \mu_l \hspace{1cm} (1.31)$$

Then, the necessary and sufficient conditions for unbiasedness are

1. $\sum_{i=1}^{n} \lambda_{l_i} = 1$
2. $\sum_{i=1}^{n} \lambda_{qi} = 0$ para $q = 1, \ldots, l - 1, l + 1, \ldots, p$
The best linear unbiased predictor is found by the minimization of

\[ E \left( Z_l(s_0) - \sum_{i=1}^{n} \sum_{q=1}^{p} \lambda_{qi} Z_q(s_i) \right)^2 \]  

subject to the constraints to guarantee unbiasedness. A shortcoming in this approach is that the correlation between the predicted values is not taken into account. Therefore, it is better to use it only to predict one variable at a time. When the study requires joint prediction, there is a generalization due to [150], that is to say, to predict at once,

\[ Z(s_0) := (Z_1(s_0), ..., Z_p(s_0))' \]

Nevertheless, in any case, the prediction depends entirely on the dependence structure. Assuming stationarity of the covariance matrix, that is, \( Cov(Z_l(s_i), Z_l'(s_{i'})) = C_{ll'}(s_i - s_{i'}) \), it can be modeled with the cross-variogram which is defined

\[ 2\nu_{ll'}(s_i, s_{i'}) = Cov[Z_l(s_i) - Z_l(s_{i'}), Z_l'(s_i) - Z_l'(s_{i'})] \]  

(1.33)

However, this definition needs the special condition of symmetry \( C_{ll'}(s_i - s_{i'}) = C_{l'l}(s_i - s_{i'}) \). A better option is the pseudo cross-variogram \( 2\gamma_{ll'}(s_i, s_{i'}) \) which allows asymmetry. The pseudo cross-variogram is given by

\[ 2\gamma_{ll'}(s_i, s_{i'}) = Var(Z_l(s_i) - Z_l'(s_{i'})) \]  

(1.34)

Generally, there is no justification to impose the restriction of symmetry and [150] show that performance of the prediction is better without this restriction and using the pseudo cross-variogram. These approaches are based in the multivariate distribution. Therefore, in the most general case, it is necessary to model the spatial dependence between all variables. That is, for \( p \) variables observed in \( n_r \) sites, \( r = 1, ..., p \), the covariance matrix needs to be found as:

\[ \sum = Var(Z_1(s_{11}), ..., Z_1(s_{1n_1}), ..., Z_p(s_{p1}), ..., Z_p(s_{pn_p})) \]

and has to be a positive-definite matrix. Its dimension is \( \sum_{r=1}^{p} n_r \times \sum_{r=1}^{p} n_r \). If the interest is the vector prediction, the optimal predictor of 1.23 is

\[ Z^*(s_0) = \sum_{i=1}^{n} Z(s_i) \Gamma_i \]  

(1.35)

where \( Z(s_i) \) is defined in 1.22 and \( \Gamma_i \) is a \( p \times p \)-matrix, formed by the weights \( \lambda_{ij} \), which represents the contribution of the \( lth \) variable at location \( s_i \) to the prediction of the \( l'th \) variable.

\[ \Gamma_i = \begin{pmatrix} \lambda_{11}^i & \lambda_{12}^i & \cdots & \lambda_{1p}^i \\ \lambda_{21}^i & \lambda_{22}^i & \cdots & \lambda_{2p}^i \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1}^i & \lambda_{p2}^i & \cdots & \lambda_{pp}^i \end{pmatrix} \]  

(1.36)
Finally, the best linear unbiased predictor is found by the minimization of 1.37 or 1.38.

\[
\max_{1 \leq l \leq n} \{ \text{Var} [Z_l(s_0) - Z_l^*(s_0)] \} \tag{1.37}
\]

\[
\sum_{i=1}^{n} \text{Var} [Z_i(s_0) - Z_i^*(s_0)] \tag{1.38}
\]

1.38 is computationally more tractable, so it will be used in this work.

1.4 Optimal sampling

An optimal sampling design is the one that finds the best combination predictor-design or estimator-design, according to the optimization of a criterion previously established. Therefore, the optimal design criterion must be defined based on the aims of the study. The goal of this study is the optimal spatial prediction of functional data.

In this study, we design or redesign the set \( S = \{ s_1, ..., s_n \} \) of observed spatial locations ensuring a minimal variance of the univariate spatial functional prediction in Chapter 2, a minimum variance of functional cokriging in Chapter 3, and a minimum variance of the mean estimation or the prediction at some future time point in Chapter 4.

An optimal design \( S_n^* \) is defined by \( \Phi(\Theta, S_n) \) as one that

\[
S_n^* = \arg \max_{S_n \in \Xi_n} \Phi(\Theta, S_n) \tag{1.39}
\]

where \( \Phi(\Theta, S_n) \) is the design criterion and any scalar measure of information obtained with the design \( S_n \) that depends on the parameter vector \( \Theta \) and \( \Xi_n \) is the set of all \( n \)-observation designs. The design criterion \( \Phi(\Theta, S_n) \) in the spatial sampling context comes from the variance of the error associated with each predictor. The set \( \Xi_n \) is the spatial domain \( D_s \). However, \( D_s \) is a continuous set, so there are infinite options for the new locations. Thus, in practice, the criterion is computed over a set \( D'_s \subset D_s \) that contains a finite number of available possibilities previously determined. In addition, it does not make sense to take sites extremely close because the spatial correlation leads to a redundant information and therefore to a waste of resources. So, \( D'_s \) must be built according to some knowledge of the region conditions, the possibility of access, and maybe economical criteria. In other cases, the best option is the evaluation of the criterion over a fine regular grid.

Entropy optimal sampling

Entropy is a way to measure the reduction in uncertainty when a random variable is observed. Then, the entropy is used as a criterion in the optimal design. The entropy \( H(Z) \) of random variable \( Z \) is

\[
H(Z) = \frac{E(- \log f_Z(Z))}{m_Z(Z)} \tag{1.40}
\]

where \( f_Z \) is the probability density function of \( Z \) and \( m_Z \) is a reference density against which uncertainty about \( Z \) is to be measured with the same units as \( f_Z \). The most common choice is \( m = 1 \), then a more known definition of entropy is given by

\[
H(Z) = E(- \log f_Z(Z)). \tag{1.41}
\]
For example, assume that the spatial process $Z$, is such that $Z \sim N(\mu(s, \beta), \Sigma(\Theta))$ for a set of locations $(s_1, s_2, ..., s_n)'$, $s_i \in D_s$, and $\beta$ and $\Theta$ are mean vector and covariance parameters, respectively. Then, the log-likelihood function is

$$l(\beta, \Theta) = -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma(\Theta)| - \frac{1}{2} (Z - \mu(s; \beta))' \Sigma^{-1}(\Theta)(Z - \mu(s; \beta))$$

(1.42)

and its entropy is

$$H(Z) = E[-\ln f_Z(Z)] = \frac{m}{2} \ln(2\pi) + \frac{1}{2} \ln |\Sigma(\Theta)| + \frac{m}{2}.$$  

The last step relies on the result,

$$E \left[ (Z - \mu(s; \beta))' \Sigma^{-1}(\Theta)(Z - \mu(s; \beta)) \right] = E \left[ Tr \left\{ (Z - \mu(s; \beta))' \Sigma^{-1}(\Theta)(Z - \mu(s; \beta)) \right\} \right] = m.$$  

(1.43)

[98] obtain the entropy criterion of predictive distribution in a hierarchical Bayesian kriging for data having a staircase structure. That is, in many applications, data from distinct networks with stations having different operational periods must be combined in spatial prediction. Sometimes, even the stations in a network, have different starting times of operations. The data are reassembled in an increasing order of operational periods; the data matrix appears to be an ascending staircase. They consider $o$ locations where levels of a random field are observed over time and a further $n$ locations where prediction is required. Let $Z_t$ be a $n + o$-dimensional random vector at time $t$. The first $n$ coordinates are those corresponding to the locations without observations and remaining $o$ those with observations. The vector $Z_t$ can be partitioned as

$$Z_t = (Z_t^{(n)}, Z_t^{(o)})$$

(1.44)

$Z_t$ is assumed to be independent and have a multivariate Gaussian distribution,

$$Z_t|x_t, B, \Sigma \overset{Ind.}{\sim} N_p(x_tB, \Sigma); \quad p = n + o$$

(1.45)

$x_t \equiv (x_t1, ..., x_{tp})$ denotes a $p$-dimensional vector of covariates and $B$ is the $p \times (n + o)$-matrix of regression coefficients, partitioned in accord with the partitioning of $Z_t$

$$B = (B^{(n)} \quad B^{(o)})$$

The covariance matrix is partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{nn} & \Sigma_{no} \\ \Sigma_{no} & \Sigma_{oo} \end{pmatrix}$$

for which the prior assigned is an inverted Wishart distribution, in order to have conjugate prior distributions. [98] obtain a predictive multivariate Student’s $t$-distribution. Besides, they find a closed expression for entropy of it. Although Gaussian distribution is a restricted case this is one very important beginning, to work later in other most realistic cases. Finally, they show some applications using transformations of the response variable for meeting the assumptions.
Optimal designs for specific objectives

In a geostatistical model, [158] proposes some criteria based on the maximum likelihood estimation of covariance parameters (Θ) under the Gaussian assumption, using the inverse of the Fisher information matrix $I^{-1}(\Theta; S)$ like an approximation to the covariance matrix of $\Theta$. This design criterion is called the $D$-optimal:

$$V_0(S; \Theta) = -\ln \det I(\Theta; S) \quad (1.46)$$

[161] and [159] propose a design criterion to optimal spatial prediction, when the covariance parameters are estimated; this criterion is based on the kriging variance and the information matrix. The criterion $(1.46)$ depends on the unknown parameter. Nevertheless, it can be used with the next modification:

$$S(\hat{\Theta}_0) = \arg \min_{S \subseteq D} V_0(S; \hat{\Theta}_0)$$

assuming the existence of an initial value $\theta_0$ and the design $S(\hat{\Theta}_0)$ is an only locally optimal design. To avoid this problem an option is to use a Bayesian approach to specify a utility function $U(\Theta, S, Z)$, which measures the effectiveness of a particular design to reach the objective. A general Bayesian criterion is

$$U(S) = \int \int U(\Theta, S, Z)p(\Theta|Z, S)p(Z|S)d\Theta dZ$$

where $p(\Theta|Z, S)$ is the posteriori distribution of $\Theta$ and $p(Z|S) = p(Z|\Theta, S)p(\theta)d\Theta$ is the marginal distribution of the data over the priori distribution of $\Theta$. Although the utility function depends on each particular problem, the Shannon entropy has been used for several authors. [29] give a formal definition and implements the entropy in a state space framework, when the variable of interest is not directly observable but is known its relation with another variable $Y$, which is observable. [34] deduces the following criterion by using Shannon entropy as utility function and a normal approximation to predictive distribution of $\Theta$:

$$U(S) = -\int_{\Theta} \ln(|I(\Theta; S)|)p(\Theta)d\Theta.$$

A different purpose would be the detection of outliers, for example, sites where a standard is failing to comply, as a zone where the maximum level of a contaminant is exceeded. [49] build a spatial design to verify ozone levels. If the covariance parameters and dynamic model are known and the interest is the minimum prediction error variance, a dynamic design of monitoring networks in general could be treated by a recursive processes using Kalman filter

$$P_{t|t-1} \equiv \text{var}(Y_t|Z_1, ..., Z_{t-1}) \quad \text{where} \quad Z_t = (Z_t(s_1), ..., Z_t(s_n)) \quad t = 0, 1, ...$$

This expression changes for each dynamical model. Either one, the data model and process model admit any structure, even nonlinear and non-Gaussian. For each $t$, the algorithm obtains the mean and covariance of the forecast distribution: $Y_{t|t-1}$ and $P_{t|t-1}$, [42].
1.5 Functional data

Let $\chi(t)$ be a function defined over some set $T$. Then:

1. A random variable $\chi(t), t \in T$, is called a functional variable if it takes values in an infinite dimensional space (or functional space), see [48] for a good exposition.

2. An observation $\chi(t)$ of $\chi(t), t \in T$ is called a functional datum. We assume the observations are in $L^2$.

3. A functional dataset $\chi_1(t), \ldots, \chi_n(t)$ are the realizations of $n$ functional variables, $\chi_1(t), \ldots, \chi_n(t)$ identically distributed as $\chi(t)$ which is square integrable.

4. Let $D_s \subset \mathbb{R}^d$ be the spatial index set, and let $\chi_s(t), s \in D_s$ be a spatial functional random variable. A spatial functional dataset $\chi_s^1(t), \ldots, \chi_s^n(t)$ is the observation of $n$ functional variables $\chi_{s_1}(t), \ldots, \chi_{s_n}(t)$ at the set of spatial sites $S = \{s_1, \ldots, s_n\}, S \subset D_s$.

This thesis considers the case when $T \subset \mathbb{R}$, that is, the functional variable is a curve. Given that the functions are observed only for a finite number of measured values, it is necessary to fit a model in order to reconstruct the whole function $\chi_i(t)$. The model can be parametric $\chi(t, \hat{\beta}_i) i = 1, \ldots, n$ as in [67], or non-parametric as in [118] using basis functions. A basis function system is a set of known functions $\phi_1, \ldots, \phi_K$ that are mathematically independent of each other and approximate arbitrarily well any curve by a linear combination of a sufficiently large number $K$ of these functions [118]. Basis functions approximate a function $\chi_i(t)$ by using a fixed truncated basis expansion as follows

$$\chi_i(t) = \sum_{j=1}^{K} a_{ij} \phi_j(t) = \Phi^T(t)a_i$$

where the $a_{ij}$ are the coefficients of the corresponding basis function $\phi_j(t), j = 1, \ldots, K$, for the datum $\chi_i(t)$ and $a_i$ is the vector that contains the $K$ coefficients. $\Phi^T(t)$ is the vector whose elements are the $K$ basis functions.

Spatial functional random vector

Let $D_s \subset \mathbb{R}^d$ be the spatial index set, and let $\chi^1_s(t), \ldots, \chi^P_s(t), s \in D_s$, be $P$ spatial functional square integrable random fields, such that, $\chi^p_s(t) \in \mathcal{H} = L^2(\mathcal{B}) p = 1, \ldots, P$. This paper considers the case when $t \in \mathcal{B} \subset \mathbb{R}$, that is, the functional variable is a curve, see [48] for a good exposition. Note that $L^2(\mathcal{B})$ is a real separable Hilbert space.

A multivariate spatial functional random field is given by

$$\{\Xi_s : s \in D_s \subset \mathbb{R}^d\}$$

where

$$\Xi_s = (\chi^1_s(t), \ldots, \chi^p_s(t)).$$

Now, let $\mathcal{H}^P = \mathcal{H} \oplus \ldots \oplus \mathcal{H}$ be the direct sum of the $P$ real separable Hilbert spaces, then $\Xi_s \in \mathcal{H}^P$ ([121] and [17]). The sum and scalar multiplication for the elements of $\mathcal{H}^P$ are
defined by
\[
[\xi, \zeta] \equiv (\xi^1 + \zeta^1, ..., \xi^p + \zeta^p) \quad b\zeta \equiv (b\zeta^1, ..., b\zeta^p) \quad \xi, \zeta \in \mathcal{H}^P, \quad b \in \mathbb{R}
\] (1.48)
and the inner product as
\[
[\xi, \zeta] = \langle \xi^1, \zeta^1 \rangle + \ldots + \langle \xi^p, \zeta^p \rangle \quad \xi, \zeta \in \mathcal{H}^P
\] (1.49)
where \( \langle \xi^p, \zeta^p \rangle \) is the \( L_2 \)-inner product.

A multivariate spatial functional dataset is an observation of \( \Xi_s \) at a particular set of spatial sites, \( S \subset D_s \). If the \( P \) functional random fields can be measured at the same set of locations \( S = \{s_1, ..., s_n\} \), we have
\[
(\chi^1_{s_i}(t), ..., \chi^p_{s_i}(t)) \quad i = 1, ..., n
\]
In other case, each spatial functional random field \( \chi^p_s(t) \) is observed at a different set \( S_p \) of \( n_p \) spatial locations \( p = 1, ..., P \) as follows
\[
(\chi^p_{s_1}(t), ..., \chi^p_{s_{n_p}}(t)) , \quad p = 1, ..., P
\]
Usually, at least some locations are common for several variables.

1.5.1 Functional principal components

Assuming the spatial functional random fields are random elements of \( L^2(B) \) and that \( E(\chi_s(t)) = 0 \), then the covariance operator \( C \) of \( \chi_s(t) \) is defined as
\[
C(y) = E[\langle \chi_s, y \rangle \chi_s] \quad y \in L^2(B)
\] (1.50)
Thus,
\[
C(y)(t) = \int c(t, r)y(r)dr, \quad \text{where} \quad c(t, r) = E[\chi_s(t)\chi_s(r)]
\]
with estimators given by
\[
\hat{C}(y) = \frac{1}{n} \sum_{i=1}^{n} (\langle \chi_{s_i}, y \rangle \chi_{s_i}) , \quad y \in L^2(B)
\]
and
\[
\hat{C}(y)(t) = \int \hat{c}(t, r)y(r)dr, \quad \text{where} \quad \hat{c}(t, r) = \frac{1}{n} \sum_{i=1}^{n} \chi_{s_i}(t)\chi_{s_i}(r)
\]
A bounded continuous linear operator \( C \) on \( \mathcal{H} \) is a covariance operator if and only if it is symmetric positive-definite and its eigenvalues \( \eta^k \) satisfy \( \sum_{k=1}^{\infty} \eta^k < \infty \).

The functional principal components (FPC) are defined as the eigenfunctions of the covariance operator \( C \). The estimators of FPC are called the empirical functional principal components (EFPC’s). Because the main interest is the reconstruction of the curve \( \chi_s \), a reasonable choice for a basis functions system is the EFPC’s formed by the eigenfunctions \( \xi_j(t) \), \( j = 1, ..., K \) of the covariance operator \( C \) of \( \chi_s(t) \) with basis coefficients.
given by the associated principal component scores \( f_j(s_i) \), defined as

\[
f_j(s_i) = \langle \chi(s_i), \xi_j \rangle, \quad j = 1, \ldots, K, \quad i = 1, \ldots, n \quad (1.51)
\]

According to [77], the approximation of this basis is uniformly optimal, in the sense of minimizing \( \hat{S}^2 \) given by

\[
\hat{S}^2 = \sum_{i=1}^{n} \left\| \chi(s_i(t)) - \sum_{j=1}^{K} f_j(s_i) \xi_j(t) \right\|^2 . \quad (1.52)
\]

It is possible to have a very good approximation using only a few empirical functional principal components. Denoting by \( \eta_j \) the corresponding eigenvalue, we choose \( K \) that ensures a minimum percentage of accumulated variability, previously established. The most frequently used value is 85% but the user makes the decision. Using the Karhunen-Loève expansion [20], we assume the model takes the form

\[
Y_s(t) = \mu(t) + \chi_s(t) = \mu(t) + \sum_{j=1}^{\infty} f_j(s_i) \xi_j(t), \quad Y_s(t) \in L^2 \quad (1.53)
\]

where \( E(Y_s(t)) = \mu(t) \). The mean function \( \mu(t) \) is estimated by the sample mean function \( \hat{\mu}(t) = \bar{Y}_s(t) = \frac{1}{n-1} \sum_{i=1}^{n} Y_s(t) \). So, \( E(\chi_s(t)) = 0 \) and from now on, we use the random variable \( \chi_s(t) \). In addition, for \( k = 1, \ldots, K \)

\[
E \left( f^k_s \right) = E \left( \chi_s, \xi^k \right) = \left\langle 0, \xi^k \right\rangle = 0 \quad (1.54)
\]

Note that for each \( k \) and \( s \in D_s \), \( f^k_s \) is a scalar spatial random field. So, the vector \( (f^1_s, \ldots, f^K_s) \) is a \( K \)-dimensional spatial random field.

According to [17], \( \Xi_s \in \mathcal{H}^P \) is a joint Gaussian \( \mathcal{H}^P \)-valued random field, if the real variable

\[
[\Xi_s, \zeta] = \langle \chi^1_s, \zeta^1 \rangle + \ldots + \langle \chi^P_s, \zeta^P \rangle \quad (1.55)
\]

is Gaussian for all \( \zeta \in \mathcal{H}^P \).

Let \( \xi^p_1, \ldots, \xi^p K_p, \quad p = 1, \ldots, P \) the first \( K_p \) eigenfunctions of the covariance operator of \( \chi^p_s \) and following the notation in (1.51) let the corresponding scores be

\[
f^p K_p_s = \langle \chi^p_s, \xi^{pK_p} \rangle , \quad (1.56)
\]

for \( b_{11}, \ldots, b_{PK_p} \) arbitrary real numbers, and let the vector \( \zeta \) be given by

\[
\zeta = (b_{11} \xi^{11} + \ldots + b_{1K_1} \xi^{1K_1}, \ldots, b_{P1} \xi^{P1} + \ldots + b_{PK_p} \xi^{PK_p})
\]

We thus have that

\[
[\Xi_s, \zeta] = b_{11} f^{11}_s + \ldots + b_{1K_1} f^{1K_1}_s + \ldots + b_{P1} f^{P1}_s + \ldots + b_{PK_p} f^{PK_p}_s \quad (1.57)
\]

is a real Gaussian variable and therefore the vector

\[
(f^{11}_s, \ldots, f^{1K_1}_s, \ldots, f^{P1}_s, \ldots, f^{PK_p}_s)
\]
is a joint Gaussian multivariate random field in \( \mathbb{R}^{K_1 + ... + K_P} \).

### 1.5.2 Spatial prediction of functional data

#### Ordinary kriging for functional data

Following [20], the predictor of the curve \( \chi_{s_0}(t) \) based on the set of functions \( \chi_{s_i}(t) \), \( i = 1, ..., n \) is given by

\[
\hat{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \chi_{s_i}(t) \quad t \in T, \quad \lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{R} \quad (1.58)
\]

The weights \( \lambda_1, \lambda_2, ..., \lambda_n \) in (1.58) are found as the solution of the minimization problem of the variance of the prediction error

\[
\min_{\lambda_1, \lambda_2, ..., \lambda_n} \int_T Var (\chi_{s_0}(t) - \hat{\chi}_{s_0}(t)) \, dt \quad (1.59)
\]

subject to the constraint \( \sum_{i=1}^{n} \lambda_i = 1 \) to ensure unbiasedness. The ordinary functional kriging, as in the scalar case, depends on the spatial dependence structure which is modeled under the second order stationarity assumption, through the trace-variogram function \( \gamma(\chi_{s_i}(t), \chi_{s_{i'}}(t)) \), defined as

\[
\gamma(\chi_{s_i}(t), \chi_{s_{i'}}(t)) = \frac{1}{2} Var (\chi_{s_i}(t) - \chi_{s_{i'}}(t)) = \gamma(\|s_i - s_{i'}\|, t) \quad (1.60)
\]

Once (1.60) has been integrated for every pair of curves, the variogram obtained, \( \gamma(\|s_i - s_{i'}\|) \), is scalar and modeled with usual spatial variogram models which allow to include geometric anisotropy.

#### Simple kriging using functional principal components

The spatial functional prediction under the assumption of a known mean function \( \mu(t) \), and using a linear combination of the observed curves

\[
\hat{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \chi_{s_i}(t)
\]

is solved for \( \lambda_1, \lambda_2, ..., \lambda_n \) by the minimum least squares method [77],

\[
E \left\| \chi_{s_0}(t) - \hat{\chi}_{s_0}(t) \right\|^2 = E (\langle \chi_{s_0}, \chi_{s_0} \rangle) - 2 \sum_{i=1}^{n} \lambda_i E (\langle \chi_{s_i}, \chi_{s_0} \rangle) + \sum_{i,i'=1}^{n} \lambda_i \lambda_{i'} E (\langle \chi_{s_i}, \chi_{s_{i'}} \rangle)
\]

where

\[
E (\langle \chi_{s_i}, \chi_{s_{i'}} \rangle) = \sum_{j=1}^{\infty} \sum_{j'=1}^{\infty} E (f_j(s_i) f_{j'}(s_{i'})) (\langle \xi_j(t), \xi_{j'}(t) \rangle) = \sum_{j=1}^{\infty} E (f_j(s_i) f_{j}(s_{i'})) \quad (1.61)
\]

Thus, due to the representation as a linear combination of the empirical functional principal components, see (1.52) and (1.53), the functional covariances between two locations
$E(\langle \chi_{s_i}, \chi_{s_{i'}} \rangle)$ are completely determined by the sum of the spatial auto-covariances of all score components $f_j(s)$ for the pair $(s_i, s_{i'})$, see (1.51). Note that this procedure does not need the cross-covariances between score vectors.

**Functional spatial prediction with basis function coefficients**

[54] and [111] approximate each function in the dataset by $K$ basis functions, and then perform ordinary cokriging at the unsampled site $s_0$ with the spatial process formed by the basis-coefficient vector. The predictor for the curve $\chi_{s_0}(t)$ is

$$\chi_{s_0}^*(t) = \sum_{j=1}^{K} a_j^*(s_0) \phi_j(t) = \Phi^T a^*(s_0)$$

where $a^*(s_0) = (a_1^*(s_0), a_2^*(s_0), \ldots, a_K^*(s_0))^T$ is the cokriging predictor of $a(s_0) = (a_1(s_0), a_2(s_0), \ldots, a_K(s_0))^T$. Thus, the basis coefficients of the $K$ basis functions at each site form the spatial variables, and can be written in matrix form as $A = (a_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, K$. The vector of coefficients to reconstruct the function $\chi_{s_i}(t)$ comes from the $i$-th row. The columns contain the coefficients for each basis function over the $n$ locations measured. Each column is a realization of a spatial process and denoted here by $a_j(s)$, $j = 1, \ldots, K$. Thus,

$$\begin{pmatrix} a_1(s) \\ \vdots \\ a_K(s) \end{pmatrix}$$

is a $K - dimensional$ spatial stochastic process. The covariance matrix for the multivariate spatial process of the basis coefficients is $\Sigma = \text{Cov}(A)$, given by $\Sigma = \Sigma(s_i, s_{i'})$, where $\Sigma(s_i, s_{i'}) = \text{Cov}(a_j(s_i), a_{j'}(s_{i'}))$, $j, j' = 1, \ldots, K$, $i, i' = 1, \ldots, n$.

Under the second order stationarity assumption, the function that determines the matrix $\Sigma$ depends only on the distances between coefficients $\text{Cov}(a_j(s_i), a_{j'}(s_{i'})) = C_{jj'}(\|s_i - s_{i'}\|)$. This approach allows to analyze the change of the coefficients of each basis function from one place to another and the cross-covariance between them. The linear model of coregionalization (LMC) is used in order to ensure that $\Sigma$ is positive definite. However, as the number of coefficients increases, so does the difficulty of the LMC, and this model can become intractable. So, to make this approach useful, it is required the use of a basis functions system that ensures a reduced number of coefficients. Thus, we present in the Section 2.1.3.1 an alternative proposal using the empirical functional principal components.
2.1 Optimal spatial sampling for functional data

In this section, the techniques for optimal spatial sampling are developed for each of the three predictors considered in Section 1.5.2. However, for the predictor presented in Section 1.5.2 we first propose a modification using simple cokriging based on functional principal components. For all spatial processes considered, we assume second order stationarity, that is, constant mean and finite variance and covariance structure depending only on the distance between locations. We now set the procedure to select the optimal spatial configuration in the sense of minimum variance of the prediction error for each predictor. The goal is the optimal prediction of \( L \) curves in a set of interest locations \( S_0 = \{s^1_0, \ldots, s^L_0\} \). Let

\[
S = \{s_1, \ldots, s_n\}
\]

be the current set of sampling locations and

\[
S_m = \{s_{n+1}, \ldots, s_{n+m}\}
\]

the set of new locations that must be determined. The enlarged network is then

\[
S' = S \cup S_m = \{s_1, \ldots, s_n, s_{n+1}, \ldots, s_{n+m}\}
\]

So, among all possible subsets of size \( m \) such that \( S_m \subset D'_s \), we must choose the one that minimizes the total variance of the prediction error.

2.1.1 Optimal spatial sampling using ordinary functional kriging

The uncertainty associated to the ordinary functional kriging prediction for an unsampled site \( s_0 \), is called the trace-variance and is given by

\[
\sigma_{OFK}^2 = \int_T Var(\chi_{s_0}(t) - \hat{\chi}_{s_0}(t)) = \sum_{i=1}^n \lambda_i \gamma(\|s_i - s_0\|) - \delta
\]
where $\delta$ is a Lagrange multiplier. The trace-variance when $m$ new locations $\{s_{n+1}, ..., s_{n+m}\}$ are added for the prediction at an unsampled site is $\sum_{i=1}^{n+m} \gamma(|s_i - s_0|) - \delta$. Now, the constraint to ensure unbiasedness turns into $\sum_{i=1}^{n+m} \lambda_i = 1$. According to the optimization in (1.59) and the trace-variogram model (1.60), the solution for the weights vector $\lambda = (\lambda_1, \lambda_2, ..., \lambda_{n+m})$ and the Lagrange multiplier $\delta$ are given by

$$\lambda = \left( \gamma + 1 - \frac{1 - \lambda^T \Gamma^{-1} \gamma}{1 - \lambda^T \Gamma^{-1} \delta} \right)^T \Gamma^{-1}$$

$$\delta = - \frac{1 - \lambda^T \Gamma^{-1} \gamma}{1 - \lambda^T \Gamma^{-1} \delta}$$

respectively. Note that $\gamma^T = (\gamma(|s_1 - s_0|), ..., \gamma(|s_{n+m} - s_0|))$ and $\Gamma$ is a $(n + m) \times (n + m)$ matrix whose $(i, i')$th element is $\gamma(|s_i - s_{i'}|)$. So the design criterion for the optimal prediction of $L$ curves in a set $S_0 = \{s_0^1, ..., s_0^L\}$ of interest locations using the total trace-variance, is given by

$$\arg \min_{S_m \subseteq D_n} \sum_{l=1}^L \left( \sum_{i=1}^{n+m} \lambda_i \gamma(|s_i - s_0^l|) - \delta_l \right)$$

(2.1)

and depends only on the distances. This predictor admits intrinsic stationarity for the trace-variogram. If the trace-variogram model has to be estimated from the data, its classical empirical estimator given by the method-of-moments as proposed in [57] takes the form

$$\hat{\gamma}(|h|) = \frac{1}{2|N(|h|)|} \left( \frac{1}{|N(|h|)|} \int_T (\chi_{s_i}(t) - \chi_{s_i'}(t))^2 dt \right)$$

where for a fixed $|h|$, $N(|h|) = \{(s_i, s_i') : |s_i - s_{i'}| = |h|, i, i' = 1, ..., n + m$ with number of elements $|N(|h|)|$. Now, the model parameters can be estimated as usual.

### 2.1.2 Optimal spatial sampling using FPCA and simple kriging

The variance of the prediction error of simple kriging for an unsampled site $s_0$ based on the enlarged network $S \cup S_m$, applying (1.61), takes the form

$$\sigma_{\text{SK}}^2 = \sum_{j=1}^{n+m} \eta_j - 2\varsigma \lambda + \lambda^T \Omega \lambda$$

where $\varsigma$ and $\Omega$ are the $(n + m)$-vector and the $(n + m) \times (n + m)$-matrix formed by the sum of the sequence of functional auto-covariances between observations and prediction site and given by

$$\varsigma = \left( \sum_{j=1}^{\infty} E(f_j(s_i) f_j(s_0)) \right), \quad \Omega = \sum_{j=1}^{\infty} \Omega_j \quad \text{and} \quad \Omega_j = E(f_j(s_i) f_j(s_{i'}))$$

for $i = 1, ..., n + m$. The solution vector with simple kriging is $\lambda = \Omega^{-1} \varsigma$, hence

$$\sigma_{\text{SK}}^2 = \sum_{j=1}^{\infty} \eta_j - \varsigma \varsigma$$
reducing the design criterion for the optimal prediction of $L$ curves to

$$\arg \max_{S_m \subset D_\ast} \sum_{l=1}^{L} \varsigma_l \Omega_{l}$$

where $\varsigma_l = \left( \sum_{j=1}^{\infty} E \left( f_j(s_i)f_j(s_0) \right) \right)$, $l = 1, ..., L$. The covariance function that determines $\Omega_j$ and $\varsigma_j$, depends only on the distances between observations and prediction sites. The value of $K$ that truncates the representation in terms of empirical functional principal components can be even more flexible and we can include more terms until the cumulative variance reaches some prefixed threshold, because this method does not use cross-covariances between score vectors, and fitting the model for auto-covariances is simpler.

2.1.3 Optimal spatial sampling for functional prediction using basis coefficients

There are only two cases where the dimension $K$ of the multivariate spatial process formed by the coefficients is not important, see [1.62]: when there is no cross-correlation between the coefficient vectors $a_j(s)$ $j = 1, ..., K$ or when all direct and cross-covariance functions are chosen proportional to a same basic function [151]. In both cases, the performance of the cokriging predictor is equivalent to applying ordinary kriging to each spatial process $a_j(s)$, $j = 1, ..., K$. In any other case it is necessary the use of the LMC. However, the coefficients involved in the reconstruction of each function $\chi_{s_i}(t)$ with the empirical functional principal components can be two or three in many practical cases, making the LMC a more feasible option.

2.1.3.1 Functional prediction using FPCA and simple cokriging

According to Section 1.5.1 we have $E(\chi(s)) = 0$ and $E(f_j(s_i)) = 0$, $i = 1, ..., n$ and $j = 1, ..., K$ and we can use simple cokriging [110] to predict the vector

$$f(s_0) = (f_1(s_0), ..., f_K(s_0))^T$$

at the unsampled location $s_0$. Now, let $\xi^T(t)$ be the vector containing the first $K$ chosen eigenfunctions. The representation of the functions is given by

$$\chi_{s_i}(t) = \xi^T(t)f(s_i), \quad i = 1, ..., n,$$

and our proposal to predict the curve $\chi_{s_0}(t)$ is

$$\chi_{s_0}^\ast(t) = \xi^T(t)f^\ast(s_0), \quad i = 1, ..., n.$$ 

The simple cokriging predictor of the score vector at $s_0$ is given by [110]

$$f^\ast(s_0) = \sum_{i=1}^{n} f^T(s_i)\Gamma_i,$$

where $f(s_i) = (f_1(s_i), ..., f_K(s_i))^T$ and $\Gamma_i$ is a $K \times K$-matrix formed by the weights $\lambda^i_{jj'}$, representing the contribution of the $j-th$ score at location $s_i$ to the prediction of the $j'-th$
score. Then the matrix \( \Gamma = (\Gamma_i) \) \( i = 1, ..., n \) is the solution of the system

\[
\begin{pmatrix}
\Sigma(s_1, s_1) & \Sigma(s_1, s_2) & \cdots & \Sigma(s_1, s_n) \\
\Sigma(s_2, s_1) & \Sigma(s_2, s_2) & \cdots & \Sigma(s_2, s_n) \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma(s_n, s_1) & \Sigma(s_n, s_2) & \cdots & \Sigma(s_n, s_n)
\end{pmatrix}
\begin{pmatrix}
\Gamma_1 \\
\Gamma_2 \\
\vdots \\
\Gamma_n
\end{pmatrix}
= 
\begin{pmatrix}
\Sigma(s_0, s_1) \\
\Sigma(s_0, s_2) \\
\vdots \\
\Sigma(s_0, s_n)
\end{pmatrix}
\tag{2.3}
\]

with

\[
\Gamma_i = \begin{pmatrix}
\lambda_{i1} & \lambda_{i2} & \cdots & \lambda_{iK} \\
\lambda_{21} & \lambda_{22} & \cdots & \lambda_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_{K1} & \lambda_{K2} & \cdots & \lambda_{KK}
\end{pmatrix}
\tag{2.4}
\]

Note that although we use an orthonormal basis, the cross-covariances between the respective score coefficients depend on the cross-covariance between the observed functions, so in general terms, there is no reason to assume independence between the score vectors. Indeed note that

\[
E(f_j(s_i)f_{j'}(s_{i'})) = E\left(\langle \chi_{s_i}, \xi_j \rangle \langle \chi_{s_{i'}}, \xi_{j'} \rangle \right)
= E\left(\int \chi_{s_i}(t)\xi_j(t)dt \int \chi_{s_{i'}}(r)\xi_{j'}(r)dr \right)
= E\left(\int \int \chi_{s_i}(t)\xi_j(t)\chi_{s_{i'}}(r)\xi_{j'}(r)dtdr \right)
= \int \xi_{j'}(r)\left(\int E(\chi_{s_i}(t)\chi_{s_{i'}}(r))\xi_j(t)dt \right)dr
= \int \xi_{j'}(r)\left(\int c_{s_i, s_{i'}}(t, r)\xi_j(t)dt \right)dr
= \int \xi_{j'}(r)C_{s_i, s_{i'}}(\xi_j)dr
= \langle C_{s_i, s_{i'}}(\xi_j), \xi_{j'} \rangle
\tag{2.5}
\]

As a particular case, when \( i = i' \) the covariance for different scores is 0 as in the multivariate case,

\[
E(f_j(s_i)f_{j'}(s_i)) = \int \xi_{j'}(r)\left(\int c_{s_i, s_i}(t, r)\xi_j(t)dt \right)dr
= \int \xi_{j'}(r)\eta_j\xi_j(r)dr
= \eta_j \langle \xi_{j'}, \xi_j \rangle
= \begin{cases}
\eta_j & \text{if } j = j' \\
0 & \text{if } j \neq j'
\end{cases}
\tag{2.6}
\]
So, the spatial auto-covariance function for each score vector \( j \) is given by
\[
E(f_j(s_i)f_j(s_{i'})) = \begin{cases} 
\eta_j, & \text{if } i = i' \\
\eta_j \rho_j(\|s_i - s_{i'}\|; \Theta), & \text{if } i \neq i'.
\end{cases}
\tag{2.7}
\]
where \( \rho_j(.) \) is the correlation function of the spatial scalar field determined by the score vector \( f_j \). Consequently, (2.7) shows that the covariance structure is second order stationary such that the variance of each score vector is the corresponding eigenvalue, and so the covariance model for each \( f_j \) has finite and known variance (sill). The matrices in the main diagonal of (2.3) can be denoted in more general form as \( \Sigma_0 \), that is, \( \Sigma_0 = \Sigma_{(s_i,s_i)} = (\text{Cov}(f_j(s_i),f_{j'}(s_i)))\), \( j,j' = 1,...,K \). Therefore, its trace is constant and given by
\[
\text{Tr}(\Sigma_0) = \sum_{j=1}^{K} \eta_j
\tag{2.8}
\]
The variance of the prediction error can be obtained as
\[
\text{Var}(\chi_{s_0}(t) - \chi_{s_0}^*(t)) = \text{Var}(\xi_t(f(s_0)) - \xi_t(f^*(s_0)))
\]
\[
= \xi_t(\text{Var}(f(s_0) - f^*(s_0)) \xi(t)
\]
\[
= \xi_t((\text{Tr}(\Sigma_0) - \text{Tr}(\sum_{i=1}^{n} (\Sigma_{(s_0,s_i)} \Gamma_i))) \xi(t)
\]
\[
= \sigma_{\text{cok}}^2 \xi_t(\xi(t)
\tag{2.9}
\]
where
\[
\sigma_{\text{cok}}^2 = \text{Tr}(\Sigma_0) - \text{Tr}(\sum_{i=1}^{n} (\Sigma_{(s_0,s_i)} \Gamma_i))
\]
is the accumulated variance of the cokriging predictor of the vector \( f(s_0) \), where \( \text{Tr}(\Sigma_0) \) is constant, see (2.8).

### 2.1.3.2 Optimal spatial sampling using FPCA and simple cokriging

Given that \( \xi(t) \) is known, the variance of the prediction error for an unsampled site based on the enlarged network \( S \cup S_m \) only depends on \( \sigma_{\text{cok}}^2 \), see (2.9).
\[
\sigma_{\text{cok}}^2 = \text{Tr}(\Sigma_0) - \text{Tr}(\sum_{i=1}^{n+m} (\Sigma_{(s_0,s_i)} \Gamma_i)) = \sum_{j=1}^{K} \eta_j - \text{Tr}(\sum_{i=1}^{n+m} (\Sigma_{(s_0,s_i)} \Gamma_i))
\]
The design criterion for the optimal prediction of \( L \) curves in a set \( S_0 = \{s_0^1,...,s_0^L\} \) of interest locations using the total prediction error variance, goes through the calculation of
\[
\arg \max_{S_m \subset B_s} \sum_{l=1}^{L} \left( \text{Tr}(\sum_{i=1}^{n+m} (\Sigma_{(s_0^l,s_i)} \Gamma_i)) \right)
\tag{2.10}
\]
Denoting by \( \Delta_s^0(\Sigma_{(s_0^i,s_i)}) = 1,...,n+m \), see (2.3), the cokriging solution is \( \Gamma = \Sigma^{-1} \Delta_s^0 \). So, the LMC and criterion (2.10) depend only on the distance between observations and prediction sites due to the assumption of second order stationarity. The number \( K \) of the
chosen principal components is usually small; therefore, the iterative computation of the matrix \( \sum^{-1} \) does not represent a high computational cost, if the spatial points are not too dense.

Finally, these criteria allow to determine the performance of the whole set, so that all sampling locations could be changed. Some networks contain \( p \) mobile stations, \( S = \{s_1, ..., s_p, s_{p+1}, ..., s_n\} \), \( p < n \), then the criterion is computed for all possible sets \( S_p = \{s_1, ..., s_p\} \subset D_s^\prime \). Also, our statistical criteria can be used to decide the number of observation sites \( n \). A maximum prediction variance is previously determined and the optimization is carried out for \( n = 1 \), then keeping this location fixed, the second location is optimized and so on until finding the \( n \) that reaches the established threshold.

All design criteria shown in this section depend on the parameter \( \Theta \). If this parameter is unknown and has to be estimated, the design is not longer optimum, it is only locally optimum because the optimization process is based also on \( \Theta \) and not only on the design \( S_n \). Thus (1.39) turns into \( S_n^* = \arg \max_{S_n \in \Xi_n} \Phi(\hat{\Theta}, S_n) \). For the covariance structures necessary for the predictors given in Sections 1.5.2 and 2.1.3.1, the classical empirical variogram and cross-variogram can be used, and the model parameters can be fitted by ordinary or weighted least squares to avoid distributional assumptions, or by methods based on the likelihood function. Note that the ordinary kriging based on the trace-variogram admits intrinsic stationarity unlike the predictors based on FPCA. We use the plug-in estimators in the linear model of coregionalization and trace-variogram models, to carry on the optimization of the sampling criteria for each case. That is, in every place where the terms \( \eta_j \rho_j (\|s_i - s_i'\|; \Theta) \) or \( \gamma(\|s_i - s_i'\|; \Theta) \) appear we replace them with \( \eta_j \rho_j (\|s_i - s_i'\|; \hat{\Theta}) \) and \( \gamma(\|s_i - s_i'\|; \hat{\Theta}) \). [73] propose a correction of the kriging variance to incorporate the uncertainty due to the lack of knowledge of \( \Theta \). [159] find that this correction could be important only for weak spatial autocorrelation cases. Nevertheless, this correction is based on the Gaussian assumption, on the use of the maximum likelihood or Restricted Maximum Likelihood estimation and it depends on \( \Theta \). So, the best option is to use the plug-in method as long as the spatial auto-correlation is moderate or strong, see [133].

2.2 Real data analysis

We analyze network data for air quality in Bogotá, Colombia. The air quality network monitors hourly the concentration of several air pollutants and meteorological parameters, although not all are measured at all stations. Particulate matter up to 10\( \mu g/m^3 \) in size (PM10) is an important component of air pollution, consisting of a mixture of solid and liquid particles suspended in the air. The data correspond to consecutive hours from May 13, 2013 at 1:00 a.m. to May 13, 2014 at 12:00 a.m., at 10 environmental stations (see Figure 2.1 left panel). Bogota’s environment agency manages the network of air quality monitoring in order to obtain, process and disclose air quality to assess compliance with standards and the basis for the definition of policies pollution control (137 and 51). The city is mixed-use, each area has industrial sites, commercial activity, housing and heavy traffic. However, the areas of the stations 2, 5 and 6 have more industrial facilities, and the area of the station 2 exceeds the rest of stations in volume of vehicle traffic. So, the station 2 is classified as traffic-industrial, the stations 5 and 6 as industrial and the rest of stations as background stations. One of the most important physical characteristics of the PM10 is its diameter, since it may enter the respiratory tract and cause damage to tissues and organs or serve as a vehicle
for bacteria and viruses. Several studies have found evidence of positive association between exposure to PM10 with the morbidity and mortality due to respiratory and cardiovascular disease, cancer, influenza and asthma (see for example, \[68\]; \[114\]; \[115\] and \[52\]). In addition, PM10 can damage human-made materials and is a major cause of reduced visibility \[113\]. Therefore, its physicochemical characteristics and its adverse effects on human health make it necessary to monitor and control this component of air pollution. The air quality local standards for PM10 are $50\mu g/m^3$ measured as an annual mean, and $100\mu g/m^3$ measured as a daily concentration \[137\]. To convert the dataset to curves, we use a set of 191 $B-splines$ basis functions of order four with equally spaced knots and a smoothing parameter $0.00001$, chosen by cross-validation. Figure 2.1 (right panel) shows the curves for the first week in the dataset. The first principal component accumulates 94.1% of the variability and the second only 2.5%. Thus, it is enough to include only the first score vector and use simple kriging. However, for illustrative purposes, we analyze the spatial auto-covariance and cross-covariance between the first two score vectors (see Figure 2.2). From the empirical variograms in Figure 2.2, we consider that there is no reason to assume discontinuity at the origin, since there is no jump in $\|s_i - s_i\| = 0$. So, we kept the nugget parameter fixed and equal to zero. Also, according to Figure 2.2 (bottom left), there is no significant cross-correlation between $f_1(s)$ and $f_2(s)$. Regarding the selection of the model, we chose the Gneiting-Matérn covariance model, with parameter vector $\Theta = (\sigma^2, \phi, \kappa_1, \kappa_2)$ that is highly flexible and shows a parabolic behavior at the origin without the differentiability issues of the Gaussian model \[15\]. This

![Figure 2.1](image-url)
model takes the form

\[
C(\|h\|; \Theta) = \frac{1}{2^{\kappa_1-1}} \Gamma(\kappa_1) \left( \frac{\|h\|}{\phi} \right)^{\kappa_1} K_{\kappa_1} \left( \frac{\|h\|}{\phi} \right) \\
\sigma^2 \left( 1 + 8 \left( \frac{\|h\|}{\phi \kappa_2} \right) + 25 \left( \frac{\|h\|}{\phi \kappa_2} \right)^2 + 32 \left( \frac{\|h\|}{\phi \kappa_2} \right)^3 \left( 1 - \left( \frac{\|h\|}{\phi \kappa_2} \right)^8 \right) \right)
\]

(2.11)

where \( h = s_i - s_j \), \( i \neq j \), \( \frac{\|h\|}{\phi \kappa_2} \leq 1 \) and \( C(\|h\|; \Theta) = 0 \) otherwise. Following [60], \( \lambda = 10/47 \) and \( K_{\kappa_1}(.) \) is the modified Bessel function of the third kind of order \( \kappa_1 \). The weighted least squares estimator for \( \Theta \) in both cases are shown in Figure 2.2. The weights used are \( \frac{N(||s_i - s_j||)}{||s_i - s_j||^2} \), with \( N(||s_i - s_j||) \) the number of point pairs at distance \( ||s_i - s_j|| \). In this estimation procedure \( \sigma_j^2 \) is assumed fixed and equal to \( \tilde{\eta}_j, j = 1, 2 \). In figure 2.2 (bottom right) note that the sill of the trace-variogram estimates the total variance and therefore is greater than the sum of the first two eigenvalues that correspond to the sills of the variograms of \( f_1 \) and \( f_2 \).

In order to illustrate the methodology, we first assume that all stations can be moved. We chose as the interest locations for prediction, \( s_0^1 \) and \( s_0^2 \). These points are located in zones with high population density, so the pollution has a strong impact. As for the set \( D' \), we took the sampling grid of 300 spatial points separated by 1 km, 10 points west to east and 30 south to north. Figure 2.3 (left) shows the interest locations \( s_0^1 \) and \( s_0^2 \) and the optimal sampling grid obtained using the global criterion (2.10) that is equivalent in this particular case to (2.2) truncated in the second score vector, \( \Omega = \Omega_1 + \Omega_2 \). Currently, this network has a mobile air quality monitoring station to improve the information obtained with the network, and given that this network cannot be moved frequently, our methodology can be used to optimize the location of this mobile station. Figure 2.3 (right) shows the optimal sampling location for this mobile station keeping fixed the current network, and using (2.10). For the optimization procedure, we use simulated annealing [28] with state given by the spatial sampling design applied at each iteration. The energy function is given by the corresponding criteria.

The community near \( s_0^2 \) need solutions for heavy traffic. The stations 7 and 8 are very near the point \( s_0^2 \) while station 6 practically determines the curve in \( s_0^1 \). The location obtained for station 6 is the beginning of another zone of industrial facilities. This is the border where housing turns into industrial. The network was moved to the center and to the left, but closer than \( s_0^2 \) than \( s_0^1 \). Thus, the stations 6, 5, 9, and 4 were relocated precisely on the 3 industrial zones in Bogotá. The location for station 0 could be justified because there are brick industries, and it could increase the PM10. About the mobile station \( M \), we expected the results equidistant at both interest points but a little closer to the station 1 because it is a source of PM10 due to the industry facilities.

Finally, to assess the quality of the spatial prediction based on the FPCA, we use the leave-one-out functional cross validation method, i.e., removing each functional observation and using the rest of functional observations to predict a smoothed function at the removed location [107]. The sets of observations and predictions are shown in Figures 2.4 (top left panel) and 2.4 (top right panel), respectively. For illustrative purposes, Figure 2.4 (bottom left panel) shows the graphical comparison between observation and prediction for station 4. The performance is good; the residual mean function varies close to zero, from -10 to 10 in most of the cases, see Figure 2.4 (bottom right panel). However, there are some large residuals, in particular for station 2 and station 8. These two functions have the largest variation range. By now, this is an univariate functional prediction and auxiliary information.
CHAPTER 2. OPTIMAL SAMPLING FOR SPATIAL PREDICTION OF FUNCTIONAL DATA

Figure 2.2. Top Left Panel: Variogram of $f_1$. Top Right Panel: Variogram of $f_2$. Bottom Left Panel: Cross-variogram of $f_1, f_2$. Bottom Right Panel: Trace-variogram.

should improve the performance. In addition, detection and analysis of functional outliers can be very helpful. [79] develop these tools for the multivariate case.
Figure 2.3. Left: Optimal sampling network assuming that all stations can be moved. Right: Optimal location for the mobile station keeping the current network fixed.
Figure 2.4. Top Left Panel: PM10, May 13, 2013 to May 13, 2014. Top Right Panel: Spatial functional prediction for the 10 original sites by cross validation. Bottom Left Panel: Observation vs prediction for station 4 (Guaymaral), Bottom Right Panel: Cross validation residuals and residual mean.
Multivariate spatial functional random field. Prediction and optimal sampling.

3.1 Functional cokriging

We first develop the cokriging method for the case of two spatial functional random fields, then we give the generalization to the case of \( P \) spatial functional random fields and finally we propose the optimal sampling designs.

Let \( \mathbf{\xi}^T(t) \) be the vector containing the first \( K \) eigenfunctions chosen according to some threshold of percentage of accumulated variability previously established. We use here a more compact notation. The representation of the functions in terms of its functional principal components is given by

\[
\chi_s(t) = \mathbf{\xi}^T(t)f_{s_i}, \quad i = 1, ..., n
\]

where

\[
f_{s_i} = (f_{s_1}^1, ..., f_{s_K}^K)^T
\]

\[
E(\chi_s(t)) = 0 \quad \text{and} \quad E(f_{s_i}^k) = 0, \quad i = 1, ..., n \quad k = 1, ..., K \quad \text{according to Section 1.5.1}
\]

Thus, \( f_{s_i}^k \) for each \( k \) and \( s \in D_s \) is the scalar spatial random field of the scores associated with the representation of the process \( \chi_s(t) \) in terms of its empirical functional principal components \( \xi^k(t) \) and the vector \( (f_{s_1}^1, ..., f_{s_K}^K) \) is a \( K \)-dimensional spatial scalar random field. Also, \( (2.5) \) shows that in general terms the score vectors are not independent. The spatial auto-covariance function for each score vector is given by

\[
E(f_{s_i}^k f_{s_i'}^k) = \begin{cases} 
\eta^k, & \text{if } i = i' \\
\eta^k \rho^k (||s_i - s_{i'}||; \Theta) & \text{if } i \neq i'
\end{cases}
\]

where \( \rho^k(.) \) is the correlation function of the spatial scalar random field determined by the score vector \( f_{s_i}^k \). Consequently, \( (3.1) \) shows that the covariance structure is second order stationary such that the variance of each score vector is the corresponding eigenvalue, and so the covariance model for each \( f_{s_i}^k \) has finite and known variance (sill).
3.1.1 Cokriging with two functional random fields

Let \( \chi_s^1(t) \) and \( \chi_s^2(t) \) be two spatial functional random fields such that \( E(\chi_s^1(t)) = 0 \) and \( E(\chi_s^2(t)) = 0 \). The cokriging predictor of the \( \chi_{s_0}^1(t) \) for an unsampled site \( s_0 \), using \( \chi_s^2(t) \) as a spatial covariate, is given by

\[
\hat{\chi}_{s_0}^1(t) = \sum_{i=1}^{n_1} \lambda_i^{11} \chi_{s_i}^1(t) + \sum_{j=1}^{n_2} \lambda_j^{12} \chi_{s_j}^2(t)
\]

where \( \lambda_i^{11} i = 1, \ldots, n_1 \) are the weights of the \( n_1 \) observations of \( \chi_s^1(t) \) and \( \lambda_j^{12} j = 1, \ldots, n_2 \) are the weights of the \( n_2 \) observations of \( \chi_s^2(t) \). Note that it is not required that both processes are measured at the same places. In addition, the unbiasedness of the predictor is ensured given that the mean is known,

\[
E(\hat{\chi}_{s_0}^1(t) - \chi_{s_0}^1(t)) = E\left(\sum_{i=1}^{n_1} \lambda_i^{11} \chi_{s_i}^1(t) + \sum_{j=1}^{n_2} \lambda_j^{12} \chi_{s_j}^2(t)\right) = 0
\]

and \( \lambda = (\lambda_i^{11}) \ i = 1, \ldots, n_1 \) and \( \beta = (\lambda_j^{12}) \ j = 1, \ldots, n_2 \) are constants that minimize

\[
Q = E\|\chi_{s_0}^1(t) - \hat{\chi}_{s_0}^1(t)\|^2
\]

Now, we carry out the minimization of \( Q \) to obtain the system of cokriging equations

\[
\frac{\partial dQ}{\partial \lambda_i^{11}} = -2E(\chi_{s_i}^1, \chi_{s_0}^1) + 2 \sum_{i'=1}^{n_1} \lambda_i^{11} E(\chi_{s_i}^1, \chi_{s_{i'}}^1) + 2 \sum_{j=1}^{n_2} \lambda_j^{12} E(\chi_{s_j}^2, \chi_{s_0}^1)
\]

and

\[
\frac{\partial dQ}{\partial \lambda_j^{12}} = -2E(\chi_{s_j}^2, \chi_{s_0}^1) + 2 \sum_{i=1}^{n_1} \lambda_i^{11} E(\chi_{s_i}^1, \chi_{s_j}^1) + 2 \sum_{j'=1}^{n_2} \lambda_j^{12} E(\chi_{s_j}^2, \chi_{s_{j'}}^2)
\]

Thus, for \( i = 1, \ldots, n_1 \) and \( j = 1, \ldots, n_2 \) the partial derivatives are given by

\[
\sum_{i'=1}^{n_1} \lambda_i^{11} E(\chi_{s_i}^1, \chi_{s_{i'}}^1) = E(\chi_{s_i}^1, \chi_{s_0}^1) - \sum_{j=1}^{n_2} \lambda_j^{12} E(\chi_{s_j}^2, \chi_{s_i}^1)
\]

Therefore, the cokriging equations are

\[
\sum_{i'=1}^{n_1} \lambda_i^{11} E(\chi_{s_i}^1, \chi_{s_{i'}}^1) = E(\chi_{s_i}^1, \chi_{s_0}^1) - \sum_{j=1}^{n_2} \lambda_j^{12} E(\chi_{s_j}^2, \chi_{s_i}^1)
\]
and
\[
\sum_{j=1}^{n_2} \lambda_{12}^{12} E(\chi_{s,j}^2, \chi_{s,j'}^2) = E(\chi_{s,j}^2, \chi_{s,j}^1) - \sum_{i=1}^{n_1} \lambda_{11}^{1} E(\chi_{s,j}^1, \chi_{s,i}^1)
\]  
(3.4)

Replacing (3.3) and (3.4) in (3.2) we obtain
\[
E\|\chi_{s,0}^1(t) - \chi_{s,0}^1(t)\|^2 = E(\chi_{s,0}^1, \chi_{s,0}^1) - \sum_{i=1}^{n_1} \lambda_{11}^{1} E(\chi_{s,i}^1, \chi_{s,0}^1) - \sum_{j=1}^{n_2} \lambda_{12}^{12} E(\chi_{s,j}^2, \chi_{s,0}^1)
\]  
(3.5)

Let \( f_{s}^{1k} k = 1, ..., K \) and \( f_{s}^{2l} l = 1, ..., L \) be the spatial scalar random fields formed by the scores of the functions \( \chi_{s}^1(t) \) and \( \chi_{s}^2(t) \), respectively. Expressing each function in (3.5) in terms of its principal components, we obtain
\[
E(\chi_{s,0}^1, \chi_{s,0}^1) = \sum_{k=1}^{K} \sum_{k'=1}^{K} E\left(f_{s,0}^{1k}, f_{s,0}^{1k'}\right) \langle \xi_{1k}, \xi_{1k'} \rangle
\]
and
\[
\sum_{i=1}^{n_1} \lambda_{11}^{1} E(\chi_{s,i}^1, \chi_{s,0}^1) = \sum_{i=1}^{n_1} \sum_{k=1}^{K} \sum_{k'=1}^{K} \lambda_{i1}^{11} E\left(f_{s,i}^{1k}, f_{s,0}^{1k'}\right) \langle \xi_{1k}, \xi_{1k'} \rangle
\]
\[
\sum_{j=1}^{n_2} \lambda_{12}^{12} E(\chi_{s,j}^2, \chi_{s,0}^1) = \sum_{j=1}^{n_2} \sum_{l=1}^{L} \sum_{k=1}^{K} \lambda_{j2}^{12} E\left(f_{s,j}^{2l}, f_{s,0}^{1k}\right) \langle \xi_{2l}, \xi_{1k} \rangle
\]
where \( \xi_{1k}, k = 1, ..., K \) are the eigenfunctions of the covariance operator of \( \chi_{s}^1(t) \) and \( \xi_{2l}, l = 1, ..., L \) are the eigenfunctions of the covariance operator of \( \chi_{s}^2(t) \). Due to the orthonormality of the empirical functional principal components, we have that
\[
\langle \xi_{1k}, \xi_{1k'} \rangle = \begin{cases} 
0 & \text{if } k \neq k' \\
1 & \text{if } k = k'
\end{cases}
\]
Also note that
\[
\sum_{k=1}^{K} E\left(f_{s,0}^{1k}, f_{s,0}^{1k}\right) = \sum_{k=1}^{K} \eta_{1k}
\]
Thus, \( E\|\chi_{s,0}^1(t) - \chi_{s,0}^1(t)\|^2 \) can be simplified as follows
\[
\sum_{k=1}^{K} E\left(f_{s,0}^{1k}, f_{s,0}^{1k}\right) - \sum_{i=1}^{n_1} \sum_{k=1}^{K} \lambda_{i1}^{11} E\left(f_{s,i}^{1k}, f_{s,0}^{1k}\right) - \sum_{j=1}^{n_2} \sum_{l=1}^{L} \sum_{k=1}^{K} \lambda_{j2}^{12} E\left(f_{s,j}^{2l}, f_{s,0}^{1k}\right)
\]  
(3.6)
\[
= \sum_{k=1}^{K} \eta_{1k} - \sum_{i=1}^{n_1} \sum_{k=1}^{K} \lambda_{i1}^{11} E\left(f_{s,i}^{1k}, f_{s,0}^{1k}\right) - \sum_{j=1}^{n_2} \sum_{l=1}^{L} \sum_{k=1}^{K} \lambda_{j2}^{12} E\left(f_{s,j}^{2l}, f_{s,0}^{1k}\right)
\]
where \( c_{12}^{lk} = \langle \xi_{2l}, \xi_{1k} \rangle \). Hence, once the representation with the functional principal components of the functional variables involved is used, the variance and the equation system of functional cokriging depend only on the auto-covariances and cross-covariances of the scores vectors, which are scalar processes.
3.1.2 Cokriging with $P$ functional random fields

The more general goal is the optimization of the spatial functional prediction of $\chi_{s_0}^r(t) \ 1 \leq r \leq P$ at the unsampled site $s_0$ based on the $P$ spatial functional variables,

$$\hat{\chi}_{s_0}^r(t) = \sum_{p=1}^{P} \sum_{i=1}^{n_p} \lambda_{i}^{rp} \chi_{s_i}^p(t)$$

The interest is the minimization of the squared norm of the prediction error given by

$$Q = E \| \chi_{s_0}^r(t) - \hat{\chi}_{s_0}^r(t) \|^2$$

For $m = 1, ..., P$, the derivatives and the cokriging equations take the form

$$\frac{\partial dQ}{d\lambda_{i}^{rm}} = -2E \langle \chi_{s_j}^m, \chi_{s_0}^r \rangle + 2 \sum_{p=1}^{P} \sum_{i=1}^{n_p} \lambda_{i}^{rp} E \langle \chi_{s_j}^m, \chi_{s_i}^p \rangle, \quad j = 1, ..., n_m \quad (3.8)$$

and

$$E \langle \chi_{s_j}^m, \chi_{s_0}^r \rangle = \sum_{p=1}^{P} \sum_{i=1}^{n_p} \lambda_{i}^{rp} E \langle \chi_{s_j}^m, \chi_{s_i}^p \rangle, \quad j = 1, ..., n_m \quad (3.9)$$

respectively. Replacing (3.8) and (3.9) in the squared norm of the prediction error (3.7), we obtain

$$E \| \chi_{s_0}^r(t) - \hat{\chi}_{s_0}^r(t) \|^2 = E \chi_{s_0}^r \chi_{s_0}^r - \sum_{p=1}^{P} \sum_{i=1}^{n_p} \lambda_{i}^{rp} E \chi_{s_i}^p \chi_{s_0}^r \quad (3.10)$$

Now, using the functional principal components representation we show that (3.10) only depends on the auto-covariances and cross-covariances between the score vectors chosen for each random field, that is,

$$E \| \chi_{s_0}^r(t) - \hat{\chi}_{s_0}^r(t) \|^2 = \sum_{k=1}^{K} E \left( f_{s_0}^{rk} f_{s_0}^{rk} \right) - \sum_{p=1}^{P} \sum_{i=1}^{n_p} \sum_{k=1}^{K} \sum_{l=1}^{L} \lambda_{i}^{lp} f_{s_i}^{lk} E \left( f_{s_0}^{rk} f_{s_l}^{kl} \right) \quad (3.11)$$

where, as before, denoting by $\eta_{rk}$, $k = 1, ..., K$ the eigenvalues of the observation of $\chi_{s_0}^r$, we have that

$$\sum_{k=1}^{K} E \left( f_{s_0}^{rk} f_{s_0}^{rk} \right) = \sum_{k=1}^{K} \eta_{rk}$$

and

$$c_{lp} = \begin{cases} 1 & \text{If } p = r \text{ and } k = l \\ 0 & \text{If } p = r \text{ and } k \neq l \\ \langle \xi_{k}^p, \xi_{l}^p \rangle & \text{If } p \neq r \end{cases}$$

In most of the cases it is sufficient with a few principal components, maybe 1 or 2 for each functional random field, due to the fact that the eigenvalue of the first principal component is too much larger than the rest of them $\eta_{1r} \gg \eta_{2r}$. So, the use of the linear model of coregionalization is a feasible option. Note that all spatial processes of scores considered have constant mean, finite variance and covariance structure depending only on the distance.
between locations, see (3.1), that is, all processes of score vectors are second order stationary processes.

Finally, as a global measure for the quality of the optimal prediction of the functional random field \( \chi_{\mathbf{r}s}^r(t) \) at \( B \) unsampled sites, we can use

\[
\sum_{b=1}^{B} \left( E \| \chi_{\mathbf{r}s}^r(t) - \hat{\chi}_{\mathbf{r}s}^r(t) \|^2 \right). \tag{3.12}
\]

### 3.1.3 Optimal spatial sampling for functional cokriging

We need to design or redesign the \( p \) sets \( S_p = \{\mathbf{s}_1, \ldots, \mathbf{s}_{n_p}\} \) \( p = 1, \ldots, P \), or at least those that can be changed, of observed spatial locations ensuring an optimal spatial functional prediction of \( \chi_{\mathbf{r}s}^r(t) \) in a set of interest locations \( S_0 = \{\mathbf{s}_0^1, \ldots, \mathbf{s}_0^B\} \) based on the \( P \) spatially correlated functional random fields. We now set the procedure to select the optimal spatial configuration in the sense of the total minimum square norm of the prediction error for functional cokriging, see (3.12). Suppose first that \( m_p \) stations can be added for the observation of each random field \( \chi_p^r(s) \), \( p = 1, \ldots, P \). So, the enlarged network for each case is then

\[
S_p' = S_p \cup S_{mp} = \{\mathbf{s}_1, \ldots, \mathbf{s}_{n_p}, \mathbf{s}_{n_p+1}, \ldots, \mathbf{s}_{n_p+m_p}\}, \ p = 1, \ldots, P
\]

Let \( \bigcup_{p=1}^{P} S_{mp} = S_{m_1} \cup \ldots \cup S_{mp} \subset D_{\mathbf{s}}' \) the set of new locations that must be determined. So, among all possible subsets \( \bigcup_{p=1}^{P} S_{mp} \), we must choose the one that minimizes the total minimum square norm of the prediction error for functional cokriging. So, according to (3.11) and (3.12) the design criterion is given by

\[
\text{arg min}_{\bigcup_{p=1}^{P} S_{mp} \subset D_{\mathbf{s}}'} \sum_{b=1}^{B} E \| \chi_{\mathbf{r}s}^r(t) - \hat{\chi}_{\mathbf{r}s}^r(t) \|^2 \tag{3.13}
\]

where \( \sum_{k=1}^{K} \eta^{rk} \) is constant and therefore the criterion (3.13) turns into

\[
\text{arg max}_{\bigcup_{p=1}^{P} S_{mp} \subset D_{\mathbf{s}}'} \sum_{b=1}^{B} \sum_{i=1}^{n_p+m_p} \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{p=1}^{P} \chi_{\mathbf{r}s}^{rp,i} \eta^{lk} E \left( f_{\mathbf{s}_i}^p f_{\mathbf{s}_0}^{r'l} \right). \tag{3.14}
\]

The criterion (3.14) establishes the general case, but frequently all random fields are measured at the same set of places \( S = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \). Here, the optimization is over the possible sets \( S_m \subset D_{\mathbf{s}}' \) and there is only one enlarged network \( S \cup S_m \subset D_{\mathbf{s}}' \) for all random fields. Finally, this criterion allows to determine the performance of the whole set so that all sampling locations could be changed, or to update the locations when there are mobile stations available.

So, the LMC and criterion (3.14) depend only on the distance between observations and prediction sites due to the second order stationarity. The numbers \( K \) and \( L \) of the chosen principal components are usually small, 1 or 2; so, the iterative computation of the inverse of covariance matrix does not represent a high computational cost, if the spatial points are not too dense.
Again, the design criterion shown in this section depends on parameter $\Theta$. So, we use the plug-in method, see the end of the Section 2.1.3.2.

### 3.2 Simulation study

We conducted a simulation study to illustrate and evaluate the performance of our proposal. We generated samples of zero mean processes

$$\chi_s^1 = \sum_{k=1}^{2} f_{s,1}^{1k}\xi_1^k(t); \quad \chi_s^2 = \sum_{k=1}^{3} f_{s,2}^{2k}\xi_2^k(t), \quad s \in \mathbb{R}^2$$

\{\xi_1^1(t) = \sin(\pi t), \xi_1^2(t) = \cos(\pi t)\} \text{ and } \{\xi_2^1(t) = \sqrt{\frac{3}{2}}t, \xi_2^2(t) = \frac{3(3t^2-1)}{2\sqrt{2}}, \xi_2^3(t) = \frac{5}{2}\sqrt{\frac{3}{2}}(t^3 - \frac{3}{4}t^2)\} \text{ are sets of orthonormal functions in } L^2[-1,1]. \text{ These functions could be selected from different basis or one common basis. Let } \gamma_1(h) \text{ stand for an isotropic exponential model with range } 1, \gamma_2(h) \text{ stand for an isotropic spherical model with range } 0.5, \text{ and } \gamma_3(h) \text{ stand for an isotropic Gaussian model with range } 0.7. \text{ The vector of scores } (f_{s,1}^{11}, f_{s,2}^{12}, f_{s,1}^{13}, f_{s,2}^{21}, f_{s,2}^{22}) \text{ is a realization of a multivariate Gaussian distribution of zero mean and dependence structure given by the linear model of corregionalization}

$$
\begin{pmatrix}
\gamma_{f_{11}^{11}}(h) & \gamma_{f_{12}^{11}}(h) & \gamma_{f_{13}^{11}}(h) & \gamma_{f_{21}^{11}}(h) & \gamma_{f_{22}^{11}}(h) \\
\gamma_{f_{11}^{12}}(h) & \gamma_{f_{12}^{12}}(h) & \gamma_{f_{13}^{12}}(h) & \gamma_{f_{21}^{12}}(h) & \gamma_{f_{22}^{12}}(h) \\
\gamma_{f_{11}^{13}}(h) & \gamma_{f_{12}^{13}}(h) & \gamma_{f_{13}^{13}}(h) & \gamma_{f_{21}^{13}}(h) & \gamma_{f_{22}^{13}}(h) \\
\gamma_{f_{21}^{21}}(h) & \gamma_{f_{22}^{11}}(h) & \gamma_{f_{22}^{12}}(h) & \gamma_{f_{22}^{13}}(h) & \gamma_{f_{22}^{22}}(h)
\end{pmatrix} = (3.16)
$$

\begin{pmatrix}
25.0 & 2.0 & 1.1 & 0.6 & 2.8 \\
2.0 & 0.4 & 0.5 & 0.5 & 0.7 \\
1.1 & 0.5 & 1.2 & 1.7 & 1.8 \\
0.6 & 0.5 & 1.7 & 2.5 & 2.6 \\
2.8 & 0.7 & 1.8 & 2.6 & 2.9
\end{pmatrix}
\begin{pmatrix}
\gamma_1(h) + \\
\gamma_2(h) + \end{pmatrix}
\begin{pmatrix}
2.7 & 1.4 & 1.4 & 0.7 & 2.1 \\
1.4 & 0.7 & 0.7 & 0.3 & 1.0 \\
1.4 & 0.7 & 0.8 & 1.6 & 1.4 \\
0.7 & 0.3 & 1.6 & 16.0 & 4.0 \\
2.1 & 1.1 & 1.4 & 4.0 & 2.4
\end{pmatrix}
\begin{pmatrix}
15.0 & 1.0 & 1.3 & 10.0 & 0.9 \\
1.0 & 1.3 & 0.9 & 1.0 & 0.9 \\
1.3 & 0.9 & 1.1 & 0.9 & 0.8 \\
1.0 & 1.0 & 0.9 & 0.9 & 0.8 \\
0.9 & 0.9 & 0.8 & 0.8 & 0.7
\end{pmatrix}
\gamma_3(h)

The elements of the matrices in (3.16) are rounded. The simulated data are shown in top panels of Figure 3.1. The simulation grid is shown in the bottom left panel of Figure 3.1. The goal is to carry out the functional cokring method to predict $\chi_s^1$ using $\chi_s^2$ as explanatory variable. The first functional principal component of the generated data accumulates 87.6% and 86.1% of the variability for $\chi_s^1$ and $\chi_s^2$ respectively. Thus, using 85% as a threshold, the prediction can be carried out based on the first principal component of each variable. Thus, we have included only the first score vector of each variable in the fitted LMC. In the bivariate case, ensuring that the LMC is positive definite is a simple procedure (66). The fitted LMC for the bivariate spatial scalar random field $(f_{s,1}^{11}, f_{s,2}^{21})$ is given by

$$
\gamma_{f_{11}^{11}}(h) = 26.3\gamma_1(h) + 2.1\gamma_2(h) \\
\gamma_{f_{21}^{21}}(h) = 2.1\gamma_1(h) + 17.5\gamma_2(h) \\
\gamma_{f_{11}^{11}}(h) = 0.3\gamma_1(h) + 1.3\gamma_2(h)
$$

(3.17)
Using the results in (3.17), we obtain the expression of the variance and the equation system of the functional cokriging in terms of the auto-covariances and cross-covariances of the scores vectors, see equation (3.6). In order to assess the quality of the spatial prediction based on functional cokriging, we used the leave-one-out functional cross validation method, i.e., removing each functional observation at each data location in the simulation grid from the dataset and further predicting it based on the remaining functions ([107]). Note in bottom right panel of Figure 1. that the residual mean function varies close to zero. The functional residuals are varying from -3 to 3 in most of the cases, showing that the predictions are good in all spatial locations. Although there are a few large residuals that take values outside the range (-3, 3), these are at some specific time points more than for the whole domain of the functions.
3.3 Real data analysis

We analyze network data for air quality in México city during the dry season because in the rainy season all air pollutants diminish. The data correspond to consecutive hours from January 01, 2015 at 1:00 a.m. to May 30, 2015 at 12:00 a.m., at 23 environmental stations (see Figure 3.2 left panel). The stations in the air quality network RAMA, monitor hourly particulate matter up to 10 micrometers in size (PM10) and Nitrogen dioxide (NO2) among others. The particulate matter (PM) is an important component of air pollution, see Section 2.2. NO2 is a gaseous air pollutant produced by the road traffic and other fossil fuel combustion processes and it contributes to the formation and modification of other air pollutants such as particulate matter. In addition, PM10 and NO2 damage human-made materials and are the major causes of reduced visibility [136]. Therefore, its physicochemical characteristics and its adverse effects on human health make it necessary to monitor and control these components of air pollution. Also, we explore how the temperature affect the pollutant concentrations. The temperature (Temp) data are taken from the meteorological network REDMET. RAMA and REDMET have 15 stations in common. The Secretariat of Environment of México currently operates the network of air quality monitoring in order to obtain, process and disclose air quality to assess compliance with standards and the basis for the definition of policies pollution control. The data are obtained from the Automatic Monitoring System [136]. To convert the datasets to curves, we use $B-splines$ basis functions of order four with equally spaced knots and a smoothing parameter 0.00001. For the dataset of PM10 we use a set of 163 $B-splines$ basis functions, for the dataset of NO2 we use a set of 157 $B-splines$ basis functions and for the data set of Temp we use 121 $B-splines$ basis functions. Right panel of Figure 3.2 and Figure 3.3 show the curves for the last week in the dataset.

![Figure 3.2](image-url)  

**Figure 3.2.** Left Panel: México city. Air quality network RAMA (shown in red). The blue points are the stations that measure temperature but belong to REDMET. Right Panel: PM10, May 23 to May 30, 2015.
The first principal component explains 75%, 84.6% and 85.7% of the variability for PM10, NO2 and Temp respectively, and the second principal component only accounts for 13.9%, 13.8% and 13.1. Thus, using 85% as a threshold, we include two score vectors for PM10, while only the first score vector is included for NO2 and Temp.

Our interest is the spatial functional prediction of PM10 using NO2 and Temp as functional covariates. So according to the notation in Section 3.1, PM10 is \( \chi_1^s(t) \), NO2 is \( \chi_2^s(t) \) and Temp is \( \chi_3^s(t) \). Figure 3.4 shows experimental and theoretical variogram fitted according to the linear model of coregionalization. We use two nested Matérn structures linearly combined, with smoothing parameters 0.1 and 5, and ranges 3000 and 13000. Thus \( \gamma_{f1}^{11} \) and \( \gamma_{f1}^{12} \) are the variograms for the first principal component of PM10, \( \gamma_{f2}^{21} \) and \( \gamma_{f3}^{21} \) are the variograms for the score vectors corresponding to the first principal component of NO2 and Temp, respectively, and the rest of variograms in Figure 3.4 and in Table 3.1 are the cross variograms between each pair of score vectors. From the empirical variogram there is no reason to assume discontinuity at the origin, since there is no jump in \( \|s_i - s_i'\| = 0 \). So, we kept the nugget parameter fixed and equal to zero.

To illustrate the methodology for optimal sampling designs, we chose two interests locations for prediction, \( s_1^0 \) and \( s_2^0 \), see Figure 3.5 (left panel). As for the set \( D_s' \), we took the sampling grid of 375 spatial points separated by 2 km, 25 points west to east and 15 south to north,
Figure 3.4. Experimental and theoretical variogram fitted according to the linear model of coregionalization.
CHAPTER 3. MULTIVARIATE SPATIAL FUNCTIONAL RANDOM FIELD. PREDICTION AND OPTIMAL SAMPLING.

restricted to the area with stations. Figure 3.5 (left panel) shows the locations of interest \( s_0 \) and \( s_2 \) and the optimal sampling location (OL) to add a new station keeping fixed the current network, and using (3.14). For the optimization procedure, we use simulated annealing \[28\] with state given by the spatial sampling design criterion applied at each iteration. The energy function is given by the criterion (3.14).

In order to assess the quality of the spatial prediction based on the functional cokriging, we use the leave-one-out functional cross validation method \[107\]. Although there are some large residuals at the beginning of the season due to the variation of pollutants in this period, the performance is good; the residual mean function varies close to zero, from -20 to 20 in most of the cases, see Figure 3.5 (right panel).

Finally, Figure 3.6 shows the cross validation results when univariate functional prediction is carried out. These predictions are obtained using the simple cokriging predictor only with the scores resulting from the representations of the PM10 with the empirical functional principal components \[14\]. The interval of variation of the residual curves goes from -30 to 30 in most of the cases. It makes clear the gain in accuracy obtained with the multivariate functional prediction if NO2 and Temp are included.

As a further exercise, we perform some exploratory analysis of data and scores. Figures 3.7, 3.8 and 3.9 show some density plots of PM10, NO2 and Temp, and Figure 3.10 shows the density plots of the first two scores of each variable. These data are a finite dimensional realization of the infinite dimensional process that we attempt to model. To check the joint Gaussian distribution hypothesis, we use the tests of Mardia, Royston and Henze-Zirkler \[94\]. Any of the tests lead to reject the hypothesis of a joint Gaussian distribution of the data. However, the tests for the score vector do not reject the joint Gaussian distribution hypothesis. The p-values are 0.5694, 0.1467 and 0.2374, respectively. Note that, the joint Gaussian distribution of functional variables implies the joint Gaussian distribution of the score vectors but the converse is not necessarily true. The use of transformations
Figure 3.6. Cross validation residuals and residual mean of the predictions of PM10 using functional kriging method proposed in [14].

of the observed data and its implications in the empirical functional principal components representation is a possibility that must be investigated further in this context. For the purposes of this paper, the optimal linear predictor proposed here has a good performance for this application.
Figure 3.7. Density plots for the PM10 data at some stations.
Figure 3.8. Density plots for the NO2 data at some stations.
Figure 3.9. Density plots for the Temperature data at some stations.
Figure 3.10. Density plots for the first two scores of each variable.
In this section, we outline the procedure to find optimal dynamic spatial sampling designs. We propose design criteria for the optimal spatial mean estimation and the optimal spatial prediction at some future time points. These criteria depend on the spatial covariance parameter forecast, so we first present the spatial data model and the time series forecasting methods that take into account the uncertainty caused by the estimation of the covariance parameter. Our proposal considers discrete and continuous temporal processes. At the end, we show a five-step summary of the proposed methodology.

### 4.1 Model and covariance structure

Let $Z_t(s)$ be the spatial process at time $t$, where $t \in D_t \subset \mathbb{R}$, $s \in D_s \subset \mathbb{R}^2$, with $\mathbb{R}^2$ standing for the spatial component and $\mathbb{R}$ for the temporal one. Consider the model

$$Z_t = 1 \mu_t + \delta_t, \quad t = 1, ..., T \quad (4.1)$$

where $Z_t = (Z_t(s_1), ..., Z_t(s_{n_t}))'$, $t = 1, ..., T$, is the random vector at $n_t$ spatial locations $S_t = \{s_1, ..., s_{n_t}\}$ at each time point $t$, $\mu_t = E(Z_t(s))$, $1 = (1, ..., 1)'$, $\delta_t = (\delta_t(s_1), ..., \delta_t(s_{n_t}))'$ is a correlated second-order stationary process with $E(\delta_t) = 0$ and $\text{Var}(\delta_t) = \text{Var}(Z_t) = \Sigma_t$ is the covariance matrix of the spatial process at time $t$ at the set of locations $S_t$, i.e., the $i,j$-th entry of $\Sigma_t$ is $(\Sigma_t)_{ij} = \text{Cov}(Z_t(s_i), Z_t(s_j))$ for $i, j = 1, ..., n_t$, $t = 1, ..., T$ and $\text{Cov}(Z_t(s_i), Z_t(s_j))$ is given by a known positive-definite covariance function $C(Z_t(s_i), Z_t(s_j)|\Theta_t)$ with parameter vector $\Theta_t$. We have the following possibilities:

a. $C(Z_t(s_i), Z_t(s_j)|\Theta)$; the function and parameter are constant for all $t \in D_t$.

b. $C(Z_t(s_i), Z_t(s_j)|\Theta_t)$; the function is constant but the parameter is time-varying. So, $\Theta_t$ $t = 1, ..., T$, comes from the time series $\{\Theta_t, \ t \in D_t\}$.

Under spatial second-order stationarity, we have (a) $C(s_i - s_j|\Theta)$ and (b) $C(s_i - s_j|\Theta_t)$ respectively. The case $C_t(s_i - s_j|\Theta_t)$ when besides of the parameter vector, the spatial covariance function also varies at each $t$ is not taken into account here. This assumption implies a very unstable process, and we do not consider it a realistic assumption.
To ensure positive definiteness, frequently a known valid parametric model is used, for example, Matern, Exponential, Spherical among others or some of these nested. However, the selection of the model requires graphical exploratory data analysis. Taking advantage of the relation between covariance and semivariance functions, the most-used graph is the empirical semivariogram that allows to visualize the possible existence of a sill and hence the range to decide about the assumption of second order stationarity. The behavior of the semivariogram near the origin gives information about the smoothness and the mean square continuity of the process, how fast it increases allows to distinguish among several options of theoretical semivariograms. In addition, here it is necessary to assess these characteristics for available times and decide the model to proceed with the estimation of the covariance parameters for each $t$.

The dependence of the vector of the estimators of the spatial covariance parameters on $t$ can be either stochastic or deterministic. Adjacent observations are frequently correlated, so a reasonable assumption for the structure of any element of $\hat{\Theta}_t$ is that of an autoregressive process of order $p$ (Jones, 1993). For instance, if the parameter vector of some covariance model $C$ has two components, such as an exponential model with variance and range, $\hat{\Theta}_t = ((\hat{\theta}_t)_1, (\hat{\theta}_t)_2) = (\hat{\sigma}_t^2, \hat{\phi}_t)$, each element may depend on $t$ in a different way, for example $(\hat{\theta}_t)_1 = \theta_1$ for all $t \in D_t$, but $(\hat{\theta}_t)_2$ follows an $AR(p)$ process. The parameter estimates of the covariance model may be correlated, so $\hat{\Theta}_t$ could follow a $VAR(p)$ process (vector autoregressive process of order $p$). Another option is that the parameter follows a deterministic function of $t$, for example, a polynomial model.

In practical cases $\Theta_t$, $t = 1,...,T$ must be estimated based on the observations of the process $Z_t$ at the set of sampling spatial locations $S_t$, which we denote as $z_t = (z_t(s_1), ..., z_t(s_n))^\prime$. There are several options to estimate spatial covariance parameters: least squares methods, composite likelihood, maximum likelihood among others, see for example, [39] and [133]. The spatial covariance parameter at $t$ is denoted by $\Theta_t = ((\theta_t)_1, ..., (\theta_t)_K)$ and its estimate by $\hat{\Theta}_t = \left((\hat{\theta}_t)_1, ..., (\hat{\theta}_t)_K\right)$. The spatial covariance parameter $\Theta_t$ is estimated for the observed spatial processes at time points $1,...,T$, and so, a time series of $\hat{\Theta}_t$ is obtained. Some forecasting method is applied to the series $\hat{\Theta}_t$, $t = 1,...,T$, to find one-step ahead forecast value, that is denoted as $\hat{\Theta}_{T+1} = \left((\hat{\theta}_{T+1})_1, ..., (\hat{\theta}_{T+1})_K\right)$. The design criterion that allows to redesign the spatial sampling scheme for the time point $T + 1$ is a function of this forecast value. See Figure 4.1 for a graphical summary.

### 4.2 Forecasting of the spatial covariance parameter

In this Section we review the methods to find forecasts of the spatial covariance parameters at the future time $T + 1$. These time series are estimations of covariance parameters and therefore we must use models for data with measurement errors. The more general approach is using state space models to extract the signal and generate forecasts. The error term is assumed an additive white noise process, and the bias of the estimators $(\hat{\theta}_t)_k$, $k = 1,...,K$ negligible. For example, the local level model ([46]) for some covariance parameter $((\theta_t)_k)$ for discrete and continuous time takes the form, respectively,
Forecast: \[ \tilde{\Theta}_{T+1} \]

Criterion: \[ \arg \min \Phi(\tilde{\Theta}_{T+1}, S_{T+1}) \]

Optimal design \[ S_{T+1} \sim \]

Figure 4.1. \( \tilde{\Theta}_{T+1} \) is the vector of forecasts based on the time series of the spatial covariance parameter estimates at time points 1,...,T. Note that the spatial sampling sizes \( n_t \) can be different for each temporal instant. Also, an observation location \( s_i \) at time \( t \) can vary to \( s_i \) at next time \( t+1 \) and so on. \( s_i, s_i, s_i \in \mathbb{R}^d \)

Discrete time \( t = 1, \ldots, T \)

\[
\begin{align*}
(\hat{\theta}_t)_k &= (\theta_t)_k + (\epsilon_t)_k, & (\epsilon_t)_k &\sim N(0, \sigma^2_{\epsilon_k}) \\
(\theta_{t+1})_k &= (\theta_t)_k + (\eta_t)_k, & (\eta_t)_k &\sim N(0, \sigma^2_{\eta_k})
\end{align*}
\] (4.2)

\( (\epsilon_t)_k \) and \( (\eta_t)_k \) are mutually independent, and independent of \( (\theta_t)_k \).

Continuous time \( 0 \leq t \leq T \)

\[
\begin{align*}
(\hat{\theta}(t))_k &= (\theta(t))_k + (\epsilon(t))_k, & t = t_1, \ldots, t_T, & (\epsilon(t_i))_k &\sim N(0, \sigma^2_{\epsilon_k}(t_i)) \\
(\theta(t))_k &= (\theta(0))_k + \sigma_{\eta_k} w(t)
\end{align*}
\] (4.3)
Consider the estimation of parameters in ARMA near to the unit circle; in other cases, the impact of the measurement error is very small. The range between these bounds increases significantly only when the parameters are ARMA with some stationary and invertible established bounds for the difference between the mean squared error of forecast obtained may be reduced to the additive one by taking logarithms ([72] and [46]). Nevertheless, [93] assure the identifiability. Although the model could have a multiplicative form, this case shows an estimation procedure under the required constraints to assure identifiability. We assume the measurement errors as an additive white noise process, to ensure the identifiability. Another approach to the estimation problem with equally spaced times is given to CARMA to obtain parameters of the CARMA identification of a CARMA process, making the inference easier than for any other CARMA process at times.

$w(t)$ is the Brownian motion process, and $\sigma_{nk} > 0$ is a scale parameter. $\sigma_{nk}^2(t)$ is a non-stochastic function that may depend on unknown parameters and it is bounded significantly away from zero. For estimation of parameters we use the reduced models

\[
(\hat{\theta}(t_i))_k = (\theta(t_i))_k + (\epsilon(t_i))_k, \quad t = t_1, ..., t_T, \quad (\epsilon(t_i))_k \sim N(0, \sigma_{nk}^2(t_i)),
\]

\[
(\theta(t_i + 1))_k = (\theta(t_i))_k + (\eta_{ti})_k
\]

where the $(\epsilon(t))_k$’s are assumed to be independent and $(\eta_{ti})_k = [\sigma_{nk}]_k [w(t_i + 1) - w(t_i)]$. This is a discrete level model that allows unequal variance of the $(\epsilon(t))_k$’s. This is the only difference with [4.2].

Also, there is a huge number of developments about the use of autoregressive integrated moving average ARIMA models considering the effect of these measurement errors. [24] show that for a general ARIMA$(p, d, q)$ process and added white noise, the resulting model is ARIMA$(p, d, \max(q, p + d))$. If $p + d \leq q$, the order of the process with error does not change. As a particular case, if the parameter $(\theta(t_i)_r$ follows a stationary AR(1) process, and the uncertainty added by the estimation is denoted with $\epsilon_t$ and is a white noise process, then $(\theta_i)_r$ follows an ARMA(1, 1) process. In general, if $Y_t$ is an AR$(p)$ process and $\epsilon_t$ is white noise, the process $Y_t + \epsilon_t$ follows an ARMA$(p, p)$ process ([30]). Note that $(\theta_i)_r$ is supposed to be a zero-mean process.

[24] and [30] provide several methods to involve the measurement errors in parameter estimation of ARMA and ARIMA models, even allowing for additive correlated noise. [99] consider the estimation of parameters in ARMA models when the errors follow ARMA models, and they show an estimation procedure under the required constraints to assure identifiability. We assume the measurement errors as an additive white noise process, to assure the identifiability. Although the model could have a multiplicative form, this case may be reduced to the additive one by taking logarithms ([72] and [16]). Nevertheless, [93] established bounds for the difference between the mean squared error of forecast obtained with some stationary and invertible ARMA models, including and ignoring the measurement error. The range between these bounds increases significantly only when the parameters are near to the unit circle; in other cases, the impact of the measurement error is very small.

Although some authors ([97] and [117]) use discrete-time ARMA$(p, q)$ processes to forecast in continuous time, this is not always right because there are some ARMA$(p, q)$ processes for which there are not continuous-time autoregressive moving average CARMA$(p, q)$ processes whose autocovariance functions match over the integer values of $t$ ([24]). Moreover, according to [22], there are ARMA$(p, q)$ processes that match with more than one CARMA$(p, q)$ process on the discrete observations. So, it is necessary to check if the identification of a CARMA$(p, q)$ process from the observations at discrete times is possible. Using the fact that the state and observation equations of a CARMA$(p, q)$ process at times $t_1, ..., t_T$ satisfy discrete-time state and observation equations ([24] it is possible to apply the Kalman recursions and the likelihood function ([23] and [89]) in order to estimate the parameters. Another approach to the estimation problem with equally spaced times is given by [116], that uses the techniques of classical discrete time series applying a transformation to obtain parameters of the CARMA$(p, q)$ process. For continuous autoregressive CAR$(p)$ processes, [31] finds approximations based on the autocovariance function of the continuous time estimator for discrete data. A CAR$(1)$ process is more frequently used because it is a Markovian process and has the advantage that its discretization is a discrete-time AR$(1)$ process, making the inference easier than for any other CARMA$(p, q)$ process. The
computational implementation for simulation and estimation of the CARMA\((p, q)\) processes in the \(R\) package \texttt{ctarma} is discussed in [145, 146]. We show below an example of an \(AR(1)\) and \(CAR(1)\) measured with error and written in state space form ([72]):

**AR(1) processes observed with error.** If some component of the parameter vector follows a discrete autoregressive process \(AR(1)\), the model that takes into account the measured error is given by
\[
\begin{align*}
(\hat{\theta}_t)_k &= (\theta_t)_k + (\epsilon_t)_k, \quad t = 1, \ldots, T, \quad (\epsilon_t)_k \sim N(0, \sigma^2_{\epsilon_k}) \\
(\theta_t)_k &= \varphi_{0k} + \varphi_{1k}(\theta_{t-1})_k + (\eta_t)_k
\end{align*}
\]  
\[(4.5)\]
where \(\varphi_{0k}\) and \(\varphi_{1k}\) are parameters, and \((\epsilon_t)_k\) and \((\eta_t)_k\) are mutually uncorrelated white noise disturbances with variances \(\sigma^2_{\epsilon_k}\) and \(\sigma^2_{\eta_k}\) respectively. \(|\varphi_{1k}| < 1\) is the condition for stationarity. The variance of the initial state \((\theta_1)_k\) is \(\frac{\sigma^2_{\theta_k}}{1-\varphi^2_{1k}}\).

**CAR(1) processes observed with error.** If the autoregressive process \(CAR(1)\) is measured with error, it is represented by the following state space model
\[
\begin{align*}
(\hat{\theta}(t))_k &= (\theta(t))_k + (\epsilon(t))_k, \quad t = t_1, \ldots, t_T, \quad (\epsilon(t))_k \sim N(0, \sigma^2_{\epsilon_k}(t)), \\
\frac{d(\theta(t))_k}{dt} &= (\rho_{0k} - \rho_{1k}(\theta(t))_k)dt + \sigma_{\eta_k}dw(t)
\end{align*}
\]  
\[(4.6)\]
where \(w(t)\) is the Brownian motion process, and \(\sigma_{\eta_k} > 0\) is a scale parameter. \(\rho_{0k}\) and \(\rho_{1k}\) are parameters and the condition \(\rho_{1k} > 0\) assures that \((4.6)\) is a stationary process.

The analysis of time series observed at irregular points and the missing data can be handled very efficiently with continuous time models even for phenomena that are not in continuous time (Harvey, 1990 and Durbin and Koopman, 2012). Note that \((3), (4)\) and \((6)\) are defined in a discrete set of time points not necessarily equally spaced.

Since the estimates of the spatial covariance parameters may be correlated, it is necessary to evaluate the interdependence between \((\theta_t)_k\) and \((\theta_t)_{k'}\), \(k, k' = 1, \ldots, K\) and to consider a multivariate time-series analysis. Note that in this context, the dimension of the time series vector is small because the spatial covariance parameter dimension is usually two or three. For example, the vector autoregressive process \(VAR(1)\) taking into account the measured error is given by
\[
\begin{align*}
\hat{\theta}_t &= \theta_t + \epsilon_t, \quad t = 1, \ldots, T, \quad \epsilon_t \sim N_K(0, \mathbf{Q}) \\
\theta_t &= \varphi_0 + \varphi_1\theta_{t-1} + \eta_t
\end{align*}
\]  
\[(4.7)\]
where \(\hat{\theta}_t\) is the \(K\)-vector with the estimates of the spatial covariance parameter \(\theta_t\) at time point \(t\). The \(K\)-vector \(\varphi_0\) and the \(K \times K\) real matrix \(\varphi_1\) are parameters and \(\eta_t \sim N_K(0, \mathbf{V})\) ([76]). The necessary and sufficient condition for stationarity is that all the eigenvalues of \(\varphi_1\) are less than 1 in absolute value. Under stationarity \(\mu = (I_k - \varphi_1)^{-1} \varphi_0\). The vector autoregressive models have the advantage that the method of ordinary least squares applied to each equation involved is asymptotically efficient. Finally, for the continuous time multivariate processes it is no longer possible obtain the estimations of the parameters from the corresponding discrete multivariate case. [72] shows an estimation method for this type of models.
4.3 Optimal spatial sampling for mean estimation

Estimation To consider the spatial covariance structure for the spatial mean estimation at some temporal instant \( t \), we use the generalized least squares (GLS) estimator \( \hat{\mu}_t \) (4.8),

\[
\hat{\mu}_t = \frac{1}{\hat{\Sigma}_t^{-1}} z_t, \quad t \in D_t, \quad t = 1, \ldots, T
\]

with \( \hat{\Sigma}_t = \left( C(s_i - s_j | \hat{\Theta}_t) \right) \), \( i, j = 1, \ldots, n_t \). The estimator of the variance \( \text{Var}(\hat{\mu}_t) \) is given by

\[
\hat{\text{Var}}(\hat{\mu}_t) = \left( \frac{1}{\hat{\Sigma}_t^{-1}} \right)^{-1} = \left( \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} C_{s_i, s_j} \right)^{-1}
\]

where \( C_{s_i, s_j} \) is the \( ij \)th entry of the matrix \( \hat{\Sigma}_t^{-1} \), \( i, j = 1, \ldots, n_t \). We have \( T \) spatial vectors \( z_t = (z(s_1), \ldots, z(s_{n_t})) \), \( t = 1, \ldots, T \) observed at each set of spatial locations \( S_t = \{ s_1, \ldots, s_{n_t} \} \), \( t = 1, \ldots, T \). Based on these datasets we need \( S_{T+1} = \{ s_1, \ldots, s_{n_{T+1}} \} \) that optimizes the estimation of \( \hat{\mu}_{T+1} \) for the spatial process at the time point \( T + 1 \). Note that an observation location \( s_i \) at time \( T \) can vary to \( s_i \) at next time \( T + 1 \) and so on. In addition, the spatial sampling sizes can be different for each temporal instant. \( s_i, s_i \in \mathbb{R}^d \).

Design criteria

Optimal sampling design at time \( T + 1 \). The design at time \( T + 1 \) goes through the calculation of the variance forecast \( \hat{\text{Var}}(\hat{\mu}_{T+1}) \) which is given by

\[
\hat{\text{Var}}(\hat{\mu}_{T+1}) = \left( \frac{1}{\hat{\Sigma}_{T+1}^{-1}} \right)^{-1} = \left( \sum_{i=1}^{n_{T+1}} \sum_{j=1}^{n_{T+1}} C_{s_i, s_j} \right)^{-1}
\]

with \( \hat{\Sigma}_{T+1} = \left( C(s_i - s_j | \hat{\Theta}_{T+1}) \right) \), \( i, j = 1, \ldots, n_{T+1} \) and \( C_{s_i, s_j} \) is the \( ij \)th entry of the matrix \( \hat{\Sigma}_{T+1}^{-1} \). Then, the optimal spatial configuration \( S_{T+1} \) to estimate \( \mu_{T+1} \) is given by the \( n_{T+1} \) points design which minimizes

\[
\hat{\text{Var}}(\hat{\mu}_{T+1})
\]

Due to the continuity of \( D_s \), it is not possible to evaluate (4.10) in all its subsets, see Section 1.4.

Optimal sampling design at time \( T + m + 1 \). A time series model or a deterministic function, \( \theta_t = f(t) \) allows to carry out forecasts \( m \) steps ahead of \( \Theta_{T+1}, \ldots, \Theta_{T+m+1} \). If the sampling design only can be changed at the future time point \( T + m + 1 \), the criterion based on the observations at time points \( t = 1, \ldots, T \), is such that optimizes the mean estimation at \( T + m + 1 \) using the covariance parameter forecast for this time point. That is,

\[
\arg \min_{S_{T+m+1} \subset D_s} \hat{\text{Var}}(\hat{\mu}_{T+m+1})
\]
**Optimal sampling design for the next m time points.** If the observation locations change at time \( T + 1 \) and remain fixed for the next \( m \) time points, the criterion can be modified to a global measure such as the total variance, \( \text{Tr}(\text{Var}(\hat{\mu}_{T+1}, \ldots, \hat{\mu}_{T+m+1})) \). In this case, we should use the set of locations that minimizes

\[
\sum_{t=T+1}^{t=T+m+1} \left( \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} C(s_i, s_j) \right)^{-1}.
\]

(4.13)

If not all future temporal points are equally interesting for the study, the criteria could be modified by including only those temporal points whose estimations need to be optimized.

**Optimal sampling design for mobile stations.** This approach can be used when it is not necessary or possible to move all sampling locations but only a few. Some networks contain \( r \) mobile stations, then the criterion is computed for all possible sets \( \{s_1, \ldots, s_r\} \subset D'_s \). For example, suppose there is a mobile station, and this station is at location \( s_1 \) at time \( T \). The design criterion is the variance forecast of the mean estimation for the next time point \( T + 1 \) with the new observation location \( s_1 \), but keeping fixed \( s_2, \ldots, s_{n_T} \). So \( S_{T+1} = \{s_1, s_2, \ldots, s_{n_T}\} \) and the optimal spatial location for the mobile station at time point \( T + 1 \) is \( s_1 \) that minimizes

\[
\arg\min_{s_1 \in D_{s_1}} \widehat{\text{Var}}(\hat{\mu}_{T+1})
\]

(4.14)

where \( D_{s_1} \subset D_s \) is the set that contains the \( N \) possible locations for \( s_1 \).

### 4.4 Optimal sampling for spatial prediction

Now we want to optimize the prediction of the random vector

\[
Z_{T+1} = (Z_{T+1}(s_1^0), \ldots, Z_{T+1}(s_L^0))
\]

in the set of locations of interest \( S_{T+1}^0 = \{s_1^0, \ldots, s_L^0\} \), based on the observation locations \( S_{T+1} = \{s_1, \ldots, s_{n_{T+1}}\} \). The most common approach is to use a kriging predictor, which has several cases according to the knowledge of the mean and the underlying distributional assumptions (39). We show the methodology to obtain the optimal spatial sampling design, when we use the ordinary kriging for predicting at the time point \( T + 1 \). As we use the variance of the prediction error as the design criterion, it can be adapted to any other kriging alternative.

The updated prediction \( Z^*(s_l^0) \) at each location \( s_l^0 \in S_{T+1}^0 \) with \( l = 1, \ldots, L \), is obtained by applying ordinary kriging at time point \( T + 1 \),

\[
Z^*(s_l^0) = \hat{\mu}_{T+1} + c_{T+1}^T \Sigma_{T+1}^{-1} (z_{T+1} - 1 \hat{\mu}_{T+1}),
\]

(4.15)

where \( c_{T+1} = (C(s_i - s_l^0) | \Theta_{T+1}) \), with \( s_i \in S_{T+1} \) \( i = 1, \ldots, n_{T+1} \) the vector of covariances between the variable at the unsampled location \( s_l^0 \) and the variables at the observed locations \( S_{T+1} \). \( \hat{\mu}_{T+1} \) is as defined in (4.8), and the covariance model \( C \) and the matrix \( \Sigma_{T+1} \) are as defined in Section 4.1.
The variance forecast of the prediction error \( Z^*(s^0_l) - Z(s^0_l) \) for each location \( l = 1, \ldots, L \), based on the set of sampling locations \( S_{T+1} \), is given by

\[
\tilde{\text{Var}} \left( Z^*(s^0_l) - Z(s^0_l) \right) = C(0|\tilde{\Theta}_{T+1}) - c^T_{T+1} \tilde{\Sigma}_{T+1}^{-1} c_{T+1} + (1 - 1^T \tilde{\Sigma}_{T+1}^{-1} 1)^2 \left( 1^T \tilde{\Sigma}_{T+1}^{-1} 1 \right)^{-1},
\]

where \( C(0|\tilde{\Theta}_{T+1}) \) is the variance forecast of the spatial process at time point \( T + 1 \). It is assumed that each spatial process is a second-order stationary process. Let \( Z_{T+1}^* = (Z_{T+1}^*(s^0_1), \ldots, Z_{T+1}^*(s^0_L)) \) be the vector predictor. The criterion design at time \( T + 1 \) is the total variance forecast of the prediction error \( Tr \left( \tilde{\text{Var}}(Z_{T+1}^* - Z_{T+1}) \right) \), and the optimal spatial configuration at time \( T + 1 \) is given by the \( n_{T+1} \) points design that minimizes the sum of the prediction error of ordinary kriging variance forecasts. Then the optimal spatial configuration \( S_{T+1} \) for predicting at the \( L \) sites in the set \( S^0_{T+1} \) is given by the \( n_{T+1} \) points design which minimizes \( Tr \left( \tilde{\text{Var}}(Z_{T+1}^* - Z_{T+1}) \right) \).

Both, if there are mobile stations or the sampling design can not be changed for the next \( m \) times, we obtain the same case considered for the optimal estimation of the mean given in (4.13), according to the covariance model and the parameter forecasts. If the spatial mean at the time point \( T + 1 \) is known, the process is the same but with simple kriging.

The steps that we should follow to put in practice our methodological proposal are the following:

1. Determine the spatial covariance model \( C \).
2. Estimate the spatial covariance parameter \( \Theta_t \) at each time point \( t = 1, \ldots, T \) based on \( z_t \).
3. Model the time series formed by the spatial covariance parameter estimates \( \hat{\Theta}_1, \ldots, \hat{\Theta}_T \).
4. Find the forecast \( \tilde{\Theta}_{T+m+1} \) of the spatial covariance parameter for the future time \( T + m + 1 \), when the observation locations can be changed, \( m \in \mathbb{N}, m \geq 0 \).
5. Determine the new set \( S_{T+m+1} \) of spatial observation locations at the time point \( T + m + 1 \) optimizing the respective criterion.

Finally, our statistical criteria can be used to decide the number of observation sites \( n_{T+m+1} \). A maximum prediction variance is previously determined and the optimization is carried out for \( n_{T+m+1} = 1 \), then keeping this location fixed, the second location is optimized and so on until finding the \( n_{T+m+1} \) that reaches the established threshold.

### 4.5 Simulation study

We conduct a simulation study to illustrate and evaluate the performance of our proposal under a variety of scenarios of dynamic spatial dependence structures. Although the methodology proposed does not require the knowledge of the dynamic model, we perform the simulation using vector autoregressive moving average (VARMA) models to ensure the temporal dependence and determine the dynamical behavior of the data.

We simulate Gaussian multivariate time series

\[
Z_t = (Z_t(s_1), \ldots, Z_t(s_{n_t})) \quad t = 1, \ldots, T
\]

(4.17)
or a stochastic way. The process is simulated on the unit square $[0, 1] \times [0, 1]$ in Table 4.1. The capital letters A-F refer to the following time series models for the parameter estimation $\hat{\mu}_t$ and covariance models. For each case we carried out 100 realizations and we obtained the median as long as the number of unknown parameters in the model is small in relation to the number of observations. The semivariogram has a very simple form and the bias of the least squares estimator is negligible asymptotically Gaussian. In addition, the bias of the composite likelihood estimator of the covariance parameters in order to find the forecast of $\Theta_{T+1}$ (see [26, 27]; [76] and [145, 146] for the computational aspects of simulation and estimation we follow here). We use $C(s_i - s_j|\Theta_t)$ to calculate the criteria given in Sections 4.3 and 4.4 according to the particular scenario. We use simulated annealing with state $i$ given by the sampling design applied at the $i$th-iteration ($\Theta_t$) to calculate the criteria given in Sections 4.3 and 4.4 according to the particular scenario. We use simulated annealing with state $i$ given by the sampling design applied at the $i$th-iteration ($\Theta_t$). The energy function is given by the sampling criterion. Basically, under certain regularity conditions, ($\Theta_t$), the maximum likelihood spatial covariance parameter estimators are consistent, unbiased and asymptotically Gaussian. In addition, the bias of the composite likelihood estimator of the semivariogram has a very simple form and the bias of the least squares estimator is negligible as long as the number of unknown parameters in the model is small in relation to the number of observations, [9], so these methods are also good options.

The data are generated varying spatial sample sizes, time series lengths and isotropic covariance models. For each case we carried out 100 realizations and we obtained the median $\hat{\mu}_{T+1}$ and the median absolute deviation (MAD) $\hat{\mu}_{T+1}^{mad}$. The median and MAD of the mean estimation $\hat{\mu}_{T+1}$ obtained from the optimal spatial sampling designs at time $T + 1$ for each $n_{T+1}$-dimensional $VAR(1)$ process, under the exponential spatial covariance model are shown in Table 4.1. The capital letters A-F refer to the following time series models for the parameter $\Theta_t$.

A. $C(\cdot | \sigma^2(t), \phi)$ with $\sigma^2(t) = \ln(t)$, $\phi = 0.6$.

B. $C(\cdot | \sigma^2, \phi)$ where $\sigma^2$ comes from an $AR(1)$ process with autoregressive parameter $\varphi_1 = 0.7$, $\eta_t \sim N(0, 0.1)$ and $E(\sigma^2) = 10$ $t = 1, ..., T$, see (4.5). $\phi = 0.6$.

C. $C(\cdot | \sigma^2, \phi(t))$ where $\sigma^2 = 1$ and $\phi(t)$ comes from a $CAR(1)$ process with autoregressive parameter 0.5. $\sigma^2 = 1$ and $E(\phi(t)) = 1$ $t = 1, ..., T$, see (4.6).

D. $C(\cdot | \sigma^2, \phi_t)$ where $\sigma^2 = 1$ and $\phi_t$ comes from the discrete random walk with and $\sigma^2 = 1$, see (4.2).

E. $C(\cdot | \sigma^2(t), \phi)$ where $\sigma^2(t)$ comes from the continuous random walk, and for $t = 1, ..., T + 1$. $\sigma^2 = 1$ see (4.3). $\phi = 0.6$. 

\[ Z_t \sim N_{n_t}(1\mu_t, \Sigma_t), \quad t = 1, ..., T + 1 \] (4.18)
F. \( C(\cdot, \sigma_t^2, \phi_t) \) where \((\sigma_t^2, \phi_t)\) follows a \( VAR(1) \) model with parameters \( \varphi, V \) and \( \mu \\

\phi = \begin{pmatrix} 0.9 & 0 \\ 0.4 & 0.5 \end{pmatrix} \quad V = \begin{pmatrix} 10 & 0.2 \\ 0.2 & 0.1 \end{pmatrix} \quad Q = \begin{pmatrix} 1 & 0 \\ 0 & 0.03 \end{pmatrix} \quad \mu = \begin{pmatrix} 115 \\ 1 \end{pmatrix} \) (4.19)

\( \mu_{T+1} \) in Table 4.1 is the spatial mean estimation obtained from the simulation of a 30-dimensional \( VAR(1) \) process, according to the A-F models. In each case, the optimization process to obtain the optimal spatial sampling designs is carried out among the \( n_{T+1} \)-subsets of this set of 30 locations. Note that we assume \( n_T = n_{T+1} \). We use this procedure due to the computational limitations of simulate from high-dimensional VARMA models.

<table>
<thead>
<tr>
<th>( T )</th>
<th>( n_{T+1} )</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>49.57</td>
<td>50.07</td>
<td>1.80</td>
<td>49.87</td>
<td>50.05</td>
<td>2.17</td>
<td>50.17</td>
</tr>
<tr>
<td>10</td>
<td>50.70</td>
<td>49.95</td>
<td>3.97</td>
<td>49.73</td>
<td>50.16</td>
<td>2.19</td>
<td>50.18</td>
</tr>
<tr>
<td>15</td>
<td>50.45</td>
<td>50.55</td>
<td>7.80</td>
<td>50.01</td>
<td>49.77</td>
<td>7.48</td>
<td>49.88</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>49.65</td>
<td>49.31</td>
<td>4.61</td>
<td>49.55</td>
<td>50.38</td>
<td>1.53</td>
<td>49.29</td>
</tr>
<tr>
<td>10</td>
<td>49.29</td>
<td>49.07</td>
<td>4.48</td>
<td>49.99</td>
<td>50.19</td>
<td>1.17</td>
<td>87.09</td>
</tr>
<tr>
<td>15</td>
<td>50.60</td>
<td>49.22</td>
<td>6.45</td>
<td>50.17</td>
<td>49.89</td>
<td>1.38</td>
<td>49.85</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>50.15</td>
<td>51.06</td>
<td>3.69</td>
<td>51.05</td>
<td>46.73</td>
<td>1.28</td>
<td>49.95</td>
</tr>
<tr>
<td>10</td>
<td>49.67</td>
<td>50.28</td>
<td>4.02</td>
<td>49.33</td>
<td>49.97</td>
<td>3.38</td>
<td>49.65</td>
</tr>
<tr>
<td>15</td>
<td>50.48</td>
<td>50.38</td>
<td>3.94</td>
<td>50.27</td>
<td>49.73</td>
<td>4.70</td>
<td>49.90</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>48.03</td>
<td>49.4</td>
<td>1.08</td>
<td>50.02</td>
<td>50.96</td>
<td>1.15</td>
<td>49.74</td>
</tr>
<tr>
<td>10</td>
<td>49.74</td>
<td>48.63</td>
<td>1.42</td>
<td>50.05</td>
<td>50.16</td>
<td>1.67</td>
<td>50.25</td>
</tr>
<tr>
<td>15</td>
<td>50.13</td>
<td>50.02</td>
<td>2.41</td>
<td>50.10</td>
<td>50.18</td>
<td>0.88</td>
<td>49.50</td>
</tr>
<tr>
<td>20</td>
<td>51.71</td>
<td>49.83</td>
<td>2.50</td>
<td>50.14</td>
<td>50.02</td>
<td>1.52</td>
<td>46.22</td>
</tr>
</tbody>
</table>

Table 4.1. Median and median absolute deviation over 100 realizations of the optimal spatial mean estimation for the future time \( T + 1 \) when the exponential covariance model is used and the covariance parameter \( \Theta_t = (\sigma_t^2, \phi_t) \) follows different time series models A-F. See the meaning of letters A to F in the text.

Now, the goal is to optimize the prediction of the random vector \( Z_{501} \) in a set of locations of interest \( S_0^{501} \), based on the set of observation locations \( S_{501} \). The coordinates of the set \( S_0^{501} = \{ s_1^{501}, ..., s_{30}^{501} \} \) are in the columns 2-3 of Table 4.2. Again, due to the computational issues, we simulate from a 50-dimensional \( VAR(1) \) process, with dynamic spatial exponential covariance model, which parameter \( \Theta_t = (\sigma_t^2, \phi_t) \) follows the \( VAR(1) \) model given in (4.19). In practice, the interest could be the prediction in some places or the visualization of the phenomenon in a map. Here, we select randomly a set of 30 points of interest and the optimal spatial sampling design of size 12 is selected from the other 20 points in the simulation. The spatial configurations of the sets \( S_0^{501} \) and \( S_{501} \) are shown in Figure 4.2. The last two columns in Table 4.2 show the ordinary kriging predictions (see (4.15)) based on the optimal spatial sampling design at the time point 501, and the simulation results, respectively. The optimal sampling design is found according to the criterion given in Section 4.4, i.e., the set of 12 observation locations that minimize the total variance of prediction error at time 501.

In all cases above considered, both the optimal mean estimation and the optimal prediction perform reasonably well. The estimations are close to the real value parameter, and predictions are close to the simulated value. The performance over 100 realizations is evaluated with the median absolute deviation (\( \hat{\mu}_{T+1}^{mad} \)). There are some large values of the \( \hat{\mu}_{T+1}^{mad} \) in particular for \( n_{T+1} = 15 \) and \( T = 25 \), which is the shorter time series used
<table>
<thead>
<tr>
<th>$s^0_l$, $l = 1, \ldots, 30$</th>
<th>$x$</th>
<th>$y$</th>
<th>$Z^*_t$, $n_T = 12$</th>
<th>$Z_{501}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s^0_1$</td>
<td>0.41</td>
<td>0.69</td>
<td>56.80</td>
<td>53.37</td>
</tr>
<tr>
<td>$s^0_2$</td>
<td>0.73</td>
<td>0.48</td>
<td>51.06</td>
<td>53.58</td>
</tr>
<tr>
<td>$s^0_3$</td>
<td>0.28</td>
<td>0.36</td>
<td>44.86</td>
<td>47.25</td>
</tr>
<tr>
<td>$s^0_4$</td>
<td>0.42</td>
<td>0.54</td>
<td>51.59</td>
<td>53.98</td>
</tr>
<tr>
<td>$s^0_5$</td>
<td>0.30</td>
<td>0.45</td>
<td>58.00</td>
<td>42.63</td>
</tr>
<tr>
<td>$s^0_6$</td>
<td>0.45</td>
<td>0.17</td>
<td>47.31</td>
<td>50.09</td>
</tr>
<tr>
<td>$s^0_7$</td>
<td>0.15</td>
<td>0.80</td>
<td>50.21</td>
<td>42.36</td>
</tr>
<tr>
<td>$s^0_8$</td>
<td>0.80</td>
<td>0.25</td>
<td>45.47</td>
<td>50.29</td>
</tr>
<tr>
<td>$s^0_9$</td>
<td>0.65</td>
<td>0.25</td>
<td>51.38</td>
<td>56.49</td>
</tr>
<tr>
<td>$s^0_{10}$</td>
<td>0.79</td>
<td>0.29</td>
<td>56.55</td>
<td>50.05</td>
</tr>
<tr>
<td>$s^0_{11}$</td>
<td>0.37</td>
<td>0.68</td>
<td>46.99</td>
<td>50.44</td>
</tr>
<tr>
<td>$s^0_{12}$</td>
<td>0.29</td>
<td>0.11</td>
<td>48.72</td>
<td>45.86</td>
</tr>
<tr>
<td>$s^0_{13}$</td>
<td>0.64</td>
<td>0.26</td>
<td>49.94</td>
<td>50.96</td>
</tr>
<tr>
<td>$s^0_{14}$</td>
<td>0.62</td>
<td>0.43</td>
<td>46.26</td>
<td>51.47</td>
</tr>
<tr>
<td>$s^0_{15}$</td>
<td>0.48</td>
<td>0.72</td>
<td>55.05</td>
<td>51.09</td>
</tr>
<tr>
<td>$s^0_{16}$</td>
<td>0.27</td>
<td>0.19</td>
<td>51.81</td>
<td>52.06</td>
</tr>
<tr>
<td>$s^0_{17}$</td>
<td>0.68</td>
<td>0.46</td>
<td>47.09</td>
<td>51.84</td>
</tr>
<tr>
<td>$s^0_{18}$</td>
<td>0.54</td>
<td>0.53</td>
<td>49.27</td>
<td>47.63</td>
</tr>
<tr>
<td>$s^0_{19}$</td>
<td>0.73</td>
<td>0.85</td>
<td>49.81</td>
<td>54.16</td>
</tr>
<tr>
<td>$s^0_{20}$</td>
<td>0.38</td>
<td>0.34</td>
<td>49.51</td>
<td>47.73</td>
</tr>
<tr>
<td>$s^0_{21}$</td>
<td>0.61</td>
<td>0.34</td>
<td>44.63</td>
<td>40.89</td>
</tr>
<tr>
<td>$s^0_{22}$</td>
<td>0.45</td>
<td>0.59</td>
<td>55.22</td>
<td>48.87</td>
</tr>
<tr>
<td>$s^0_{23}$</td>
<td>0.28</td>
<td>0.12</td>
<td>47.87</td>
<td>53.07</td>
</tr>
<tr>
<td>$s^0_{24}$</td>
<td>0.79</td>
<td>0.81</td>
<td>48.57</td>
<td>52.68</td>
</tr>
<tr>
<td>$s^0_{25}$</td>
<td>0.72</td>
<td>0.40</td>
<td>54.57</td>
<td>53.24</td>
</tr>
<tr>
<td>$s^0_{26}$</td>
<td>0.69</td>
<td>0.56</td>
<td>46.75</td>
<td>47.29</td>
</tr>
<tr>
<td>$s^0_{27}$</td>
<td>0.14</td>
<td>0.26</td>
<td>47.35</td>
<td>53.14</td>
</tr>
<tr>
<td>$s^0_{28}$</td>
<td>0.36</td>
<td>0.35</td>
<td>52.41</td>
<td>44.58</td>
</tr>
<tr>
<td>$s^0_{29}$</td>
<td>0.55</td>
<td>0.77</td>
<td>54.75</td>
<td>51.78</td>
</tr>
<tr>
<td>$s^0_{30}$</td>
<td>0.13</td>
<td>0.33</td>
<td>47.14</td>
<td>45.50</td>
</tr>
</tbody>
</table>

**Table 4.2.** Optimal spatial prediction estimation for the future time $T + 1$ when the exponential covariance model is used and the covariance parameter $\Theta_t = (\sigma_t^2, \phi_t)$ follows the VAR(1) model F. See the meaning of letter F in the text. The data are generated from a 12-dimensional VAR(1) model.
There is evidence that the behavior of the optimal spatial sampling designs is affected by the length of the time series because forecasts improve along with the increase of the length of the time series. It is known that for small spatial sample sizes, the estimated parameters are less precise, making the inclusion of the uncertainty in the models, more relevant. This combination is frequently found in environmental studies, long time series but a few spatial locations. However, the forecasts are better when the length of the time series is considerably larger than the number of spatial observation locations. So, the inclusion in the modeling steps of the error process has an important and positive role in the quality of forecasts. Note that the computational limitations in this simulation study, are not shared with the methodology proposed. Also, we carried out simulations from $VARMA(1,1)$ processes with similar results. Additional simulation results can be found at the website http://ciencias.bogota.unal.edu.co/fileadmin/content/gruposdeinvestigacion/estadisticaespacial/.

### 4.6 Real data analysis

We analyze network data for air quality in Bogotá, Colombia described in Section 2.2. The Figure 4.3 shows the empirical variograms and the fitted models by composite likelihood, for some temporal instants of May 7th, 2014, 5 am, 6 am, 7 am and 8 am. From the empirical variograms, we consider that there is no reason to assume discontinuity at the origin, since there is no jump in $|s_i - s_j| = 0$. We fit a generalized Cauchy covariance, \[ C_{\mathbf{\Theta}_t}(s_i - s_j) = \sigma_t^2 \left( 1 + \left( \frac{|s_i - s_j|}{\phi_t} \right)^\gamma \right)^{-\nu} \quad \phi_t > 0, \quad 0 < \gamma \leq 2, \quad \nu > 0, \quad t = 1, \ldots, T. \] (4.20)
where $\phi_t$ is a scale parameter, $\gamma$ is a shape parameter and $\nu$ parameterizes the long-memory dependence. So, we fit models with the nugget parameter fixed and equal to zero. Regarding the selection of the covariance model, we first chose the parameter model $\gamma = 3/2$ because of the shape of the short-lag terms of the empirical variograms. Then this parameter is kept fixed and we run several maximizations on a grid of $\nu$ values, and the estimation is restricted to the parameters $\sigma^2_t$ and $\phi_t$ at each time point. We follow this procedure because the estimation of $\gamma$ and $\nu$ leads some numerical instability issues. The fitted model for the spatial covariance at each time $t$, is a particular case of (4.20) which is

$$C(s_i - s_j \mid \Theta_t) = \sigma^2_t \left( 1 + \left( \frac{\|s_i - s_j\|}{\phi_t} \right)^{3/2} \right)^{-1/3}$$

$t = 1, \ldots, T$ \hspace{1cm} (4.21)

![Figure 4.3. Some semivariograms obtained on May 7th, 2014](image)

The fitted model for the time series $\hat{\sigma}^2_t$ is a $CAR(1)$. As the process is sampled at equally spaced intervals of length one hour, the process obtained is an $AR(1)$ with autoregression
coefficient 0.894. The time series of $\hat{\phi}_t$ has a constant mean and there is no significant autocorrelation, so the forecasting for every $T + 1$ is 4729.37. Based on these results, we test the independence between the two processes by using the sample cross-correlation function $\sigma^2$, see Figure 4.4. There is no evidence of significant cross-correlation.

The air quality network cannot be moved frequently, but it has a mobile station to improve the information obtained. Then, our methodology can be used to optimize the next location of this mobile station. Unfortunately, we do not have the data of the mobile station. So, in order to illustrate the methodology, we use the data until May 7th to optimize the mobile station location for May 8th. Then we use the information obtained by the network for May 8th and we simulate the data at mobile station with conditional simulation, and so on until May 14th. As for the set $D_t$, we take the allowable sampling grid of 300 spatial points separated 1km, 10 points west to east and 30 south to north. So, if only this mobile station can be moved, its optimal location to estimate the spatial mean on the next day is
the one that minimizes expression (4.14) but applied as in (4.13) with $m = 24$. Figure 4.5 shows the optimal spatial locations of the mobile station for the days May 8th to May 14th.

Figure 4.5. Spatial locations of the mobile station for the next seven days for the optimal mean estimation.
Discussion and conclusions

In this dissertation we have studied the spatio-temporal sampling from several perspectives. In the first part, we have proposed a methodology to find optimal sampling designs for spatio-temporal processes combining classical geostatistics and functional data analysis. The criteria used here are useful when moving only one location or even lots of them. The use of functional data analysis brings several advantages. There is no restriction of the temporal index sets, it could be discrete or continuous and it is not necessary that the point times are equally spaced. The ordinary functional kriging has the advantage of using the function directly instead of its representation in terms of basis functions, but it implies more effort in the estimation of the trace-variogram model while the other two predictors carry out all the procedures with the score vectors based on the approximation of empirical functional principal components. In addition, in the functional principal component analysis, these temporal points have the same role of the variables in the multivariate principal components approach, taking into account the temporal correlation in an alternative way. The performance of these methodologies could be improved, taking other variables into account to explain the features of the observation locations. So, in the second part, we have developed the functional cokriging predictor using the representation of the functions in terms of the empirical functional principal components and of the associated scores. We show that the coefficients of the system of equations for the cokriging weights depend only on the auto-covariances and cross-covariances of the score vectors. Thus, once the number of eigenfunctions according to some predetermined percentage threshold are chosen, and hence the respective scores are found, the multivariate functional random field prediction can be found using only the multivariate scalar random field formed by the score vectors. As an additional advantage of the method presented here is that it only uses the functional principal component representation of each random field. The functional cokriging method does not require multivariate functional principal component analysis. Thus, the advantages and drawbacks of the functional cokriging are the same of the scalar cokriging. The limitation of the dimension for the LMC is the most critical issue of the cokriging method. However, this difficulty is solved by the fact that the EFPC representation does not usually need too many eigenfunctions, even with one or two could be sufficient in most of the cases, making feasible to use this type of covariance model.

The spatial optimal sampling design for spatial functional data is a natural extension of its counterpart with scalar variables. Once the covariance between curves has been modeled, the optimization process to find an optimal design has the same computational effort as in the scalar case. Its performance depends on the quality of the covariance parameter estimators and on the optimization algorithm used. A functional datum is not observed entirely but in a finite number of points. The first step of the procedure is to reconstruct the functions
from observed data. Here we have not investigated the effect of the uncertainty introduced in the selection of the basis functions and by the estimation of the coefficients vector. The functional principal components are computed with the centered data using \( \hat{\mu}(t) = \bar{Y}_s(t) \).

We are aware that the spatial covariance between curves can modify the true mean function. Another option could be to estimate the spatial covariance structure \( \Sigma_t \) at each time point \( t \) and use the generalized least squares mean estimator to center data before building the curves. However, it is difficult to apply this option when \( T \) is big, unless there is a known structure and the parameter estimation can be carried out in an automatic way.

In chapter 4 we propose a methodology to find optimal sampling designs for non-stationary spatio-temporal processes. Our approach is very useful when we do not know the dynamic behavior of the spatio-temporal process. Instead, we model the time-varying spatial covariance parameter using multivariate time series. The design criteria proposed in this work only depend on the forecast of the spatial covariance parameter using time series analysis. Although we use ARIMA models, any other alternative of modeling that improves the forecast of the spatial covariance parameter could be used. For example, a good choice could be the multiplicative error model for nonnegative time-series data. This analysis is carried out with the data obtained from the spatial covariance parameter estimation at each time point. The methodology is straightforward by combining time series models with geostatistical analysis. The temporal index set could be discrete or continuous, which is very useful because the time series observed at irregular points can be handled very efficiently with continuous time models.

We have presented criteria for optimal spatial mean estimation and optimal spatial prediction either when all sampling locations or only some of them can be moved. Also, if the optimal spatial configuration cannot be moved in a long time, the criteria can only include those temporal points that are more relevant. The methodology is evaluated through a simulation study. According to the results, the performance of the design is good when the parameter estimation and forecasting of the dynamic covariance structure are good enough. The uncertainty of the estimated parameters is taken into account through state space models or time series models adjusted with measurement error. We did not find a significant cross-correlation between the time series of spatial covariance parameter estimators for the data set used here. However, this is a possibility that has to be evaluated, because its existence implies the use of multivariate time series. The computational efficiency depends on the behavior of the simulated annealing. We have reviewed here some literature useful to forecast, but there are other possibilities. The goal is to use a method that ensures good forecasts including uncertainty and possible biases. Finally, if there is scientific knowledge about a spatio-temporal process, such as physical statistical models, it has to be used because in some occasions, these models determine the dynamic behavior of the covariance matrix.

We have not taken into account that the longer-term forecasts are less accurate. We have focused on dynamic designs, thinking that the new measured data have to be added to the existing records, so that when the new designs have to be found, time series models and forecasts are updated and are not too far from the time point at which the network is going to be moved. We must involve in the methodology, an option to adjust the criteria, allowing to take into account the increasing uncertainty of the forecasts. For example, the criteria can be penalized depending on the variance of forecast error. We have to explore and evaluate several options.

The computations are done using the free software R and the packages \textit{fda}, \textit{fda.usc}, \textit{GenSA}, \textit{intamapInteractive}, \textit{gstat}, \textit{RandomFields}, \textit{ctarma} and \textit{maptools}. Here we used
simulated annealing because it allows to define discrepancy functions in the iterations, but others such as genetic algorithms can also be applied \cite{126}. When there is not a big number of available options for new sites, it is possible to compute the criteria in all cases.
Bibliography


[56] _____, *Continuous time-varying kriging for spatial prediction of functional data: An environmental application*, Journal of Agricultural, Biological, and Environmental Statistics **15** (2010), no. 1, 66–82.


[144] , *ctarma: Estimation and simulation of carma(p,q)*, 2013b, R package version 0.1.4.


[146] , *ctarma: Estimation and simulation of carma(p,q)*, 2013b, R package version 0.1.4.


