Stoch Environ Res Risk Assess
DOI 10.1007/s00477-016-1266-y

ORIGINAL PAPER

Multivariate functional random fields: prediction and optimal sampling

M. Bohorquez[1] · R. Giraldo[1] · J. Mateu[2]

Abstract This paper develops 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 scores vectors, which are scalar random fields. In addition, we propose the methodology to find optimal sampling designs in this context. The proposal is applied to the network of air quality in México city.

Keywords Functional data · Multivariate geostatistics · Optimal sampling

1 Introduction

In recent years, functional data analysis has seen a large development in a wide variety of scientific fields such as ecology (Yen et al. 2014), medicine (Srensen et al. 2013) or environmental sciences (Giraldo et al. 2011; Torres et al. 2011). In particular, functional data analysis has largely treated problems related to monitoring networks of the weather and pollutants (see Escabias et al. 2005; Ignaccolo et al. 2013, among others). Ruiz-Medina (2011) 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 Ruiz-Medina and Espejo (2012), Ruiz-Medina (2012) and Ruiz-Medina et al. (2015). Ruiz-Medina and Espejo (2012) and Ruiz-Medina (2012) assume the existence of an underlying spatial functional Markovian model, and Ruiz-Medina et al. (2015) consider the problem of least-squares linear estimation for random fields with values in fractional Sobolev spaces of variable order. The geostatistical methods allow to predict the variable of interest at unsampled sites. When the curves are time series varying spatially, spatial functional data analysis is an alternative approach to spatio-temporal modeling, with the advantage that it allows the optimal prediction of the whole curve. There are several extensions of univariate geostatistical methods to functional data. The ordinary kriging method to predict functions at unsampled sites is provided by Giraldo et al. (2011) and Giraldo et al. (2012), using non-parametric methods to build the curves. Goulard and Voltz (1993), Nerini et al. (2010), Giraldo (2014) and Bohorquez et al. (2015) develop univariate functional prediction by using scalar multivariate geostatistics. Horvath and Kokoszka (2012) and Bohorquez et al. (2015) carry out spatial functional prediction using the scores associated with the functional principal components representation. Extensions that introduce exogenous variables for the external drift with applications to air quality monitoring and weather analysis are given by Caballero et al. (2013) and Ignaccolo

✉ M. Bohorquez
mpbohorquezc@unal.edu.co

R. Giraldo
rgiraldoh@unal.edu.co

J. Mateu
mateu@mat.uji.es

[1] Department of Statistics, National University of Colombia, Bogotá, Colombia

[2] Department of Mathematics, Universitat Jaume I, Castellón, Spain

Published online: 01 June 2016

et al. (2014). See Giraldo et al. (2010) and Giraldo and Mateu (2013) for other cases. However, so far there are no advances in multivariate functional geostatistics. Our interest in this paper is to carry out the prediction of a functional variable taking the spatial cross-covariance with other functional variables into account to improve the performance of the prediction. For this purpose, we use 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. In addition, we want to take into account the impact of the observation locations in the statistical spatial methods (Le and Zidek 2006; Angulo et al. 2000). Zhu and Stein (2006) and Zimmerman (2006) develop optimal sampling designs in the scalar case to optimize the spatial prediction. Recently, Bohorquez et al. (2015) generalize the optimal sampling designs for several univariate spatial functional predictors. Now we propose optimal sampling designs that ensure a good quality of the spatial functional prediction in presence of spatially correlated covariates.

The paper follows with some background in Sect. 2, and the methodological setup in Sect. 3. A simulation study is performed in Sect. 4 and a real dataset is analyzed in Sect. 5. The paper ends with some concluding remarks in Sect. 6.

2 Background

2.1 Spatial functional random vector

Let $D_s \subset \mathbb{R}^d$ be the spatial index set, and let $\boldsymbol{\chi}_s^1(t), \ldots, \boldsymbol{\chi}_s^P(t)$, $\boldsymbol{s} \in D_s$, be P spatial functional square integrable random fields, such that, $\boldsymbol{\chi}_s^p(t) \in \mathcal{H} = L^2(\mathcal{B})$ $p = 1, \ldots, P$. This paper considers the case when $t \in \mathcal{B} \subset \mathbb{R}$, that is, the functional variable is a curve, see Ferraty and Vieu (2006) 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

$$\{\boldsymbol{\Xi}_s : \boldsymbol{s} \in D_s \subset \mathbb{R}^d\}$$

where

$$\boldsymbol{\Xi}_s = \left(\boldsymbol{\chi}_s^1(t), \ldots, \boldsymbol{\chi}_s^P(t)\right)$$

Now, let $\mathcal{H}^P = \mathcal{H} \oplus \ldots \oplus \mathcal{H}$ be the direct sum of the P real separable Hilbert spaces, then $\boldsymbol{\Xi}_s \in \mathcal{H}^P$ (Reed and Simon 1980; Bongiorno et al. 2014). The sum and scalar multiplication for the elements of $\mathcal{H}^P$ are defined by

$$\begin{aligned} &[\boldsymbol{\xi}, \boldsymbol{\zeta}] \equiv \left(\xi^1 + \zeta^1, \ldots, \xi^P + \zeta^P\right) \\ &b\boldsymbol{\zeta} \equiv \left(b\zeta^1, \ldots, b\zeta^P\right) \\ &\boldsymbol{\xi}, \boldsymbol{\zeta} \in \mathcal{H}^P, \quad b \in \mathbb{R} \end{aligned} \tag{1}$$

and the inner product as

$$[\boldsymbol{\xi}, \boldsymbol{\zeta}] = \langle \xi^1, \zeta^1 \rangle + \cdots + \langle \xi^P, \zeta^P \rangle \qquad \boldsymbol{\xi}, \boldsymbol{\zeta} \in \mathcal{H}^P \tag{2}$$

where $\langle \xi^p, \zeta^p \rangle$ is the L^2-inner product. A multivariate spatial functional dataset is an observation of $\boldsymbol{\Xi}_s$ 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 $S = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, we have

$$\left(\chi_{s_i}^1(t), \ldots, \chi_{s_i}^P(t)\right) \quad i = 1, \ldots, n$$

In other case, each spatial functional random field $\boldsymbol{\chi}_s^p(t)$ is observed at a different set S_p of n_p spatial locations $p = 1, \ldots, P$ as follows

$$\left(\chi_{s_1}^p(t), \ldots, \chi_{s_{np}}^p(t)\right), \quad p = 1, \ldots, P$$

Usually, at least some locations are common for several variables.

Given that the functions are known only for a finite number of measured values, it is necessary to fit a model in order to reconstruct the whole function $\chi_{s_i}(t)$. The model can be parametric $\chi_{s_i}(t, \hat{\boldsymbol{\beta}}_i)$ $i = 1, \ldots, n$ as in Goulard and Voltz (1993), or non-parametric as in Ramsay and Silverman (2005) 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 that approximate arbitrarily well any curve by a linear combination of a sufficiently large number K of these functions (Ramsay and Silverman 2005). Basis functions approximate a function $\chi_{s_i}(t)$ by using a fixed truncated basis expansion as follows

$$\chi_{s_i}(t) = \sum_{k=1}^{K} a_{s_i}^k \phi^k(t) = \boldsymbol{\Phi}^T(t)\boldsymbol{a}_{s_i} \tag{3}$$

where the $a_{s_i}^k$ are the coefficients of the corresponding basis function $\phi^k(t)$, $k = 1, \ldots, K$, for the datum $\chi_{s_i}(t)$ and $\boldsymbol{a}_{s_i}$ is the vector that contains the K coefficients. $\boldsymbol{\Phi}^T(t)$ is the vector whose elements are the K basis functions.

2.2 Functional principal components

Assuming the spatial functional random fields are random elements of $L^2(\mathcal{B})$ and that $E(\boldsymbol{\chi}_s(t)) = 0$, then the covariance operator C of $\boldsymbol{\chi}_s(t)$ is defined as

$$C(y) = E[\langle \chi_s, y \rangle \chi_s] \quad y \in L^2(\mathcal{B}) \tag{4}$$

Thus,

$$C(y)(t) = \int c(t,r)y(r)dr, \quad \text{where} \quad \mathrm{c(t,r)} = \mathrm{E}[\chi_s(\mathrm{t})\chi_s(\mathrm{r})]$$

with estimators given by

$$\hat{C}(y) = \frac{1}{n}\sum_{i=1}^{n}\left(\langle \chi_{s_i}, y\rangle \chi_{s_i}\right), \qquad y \in L^2(\mathcal{B})$$

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 η^k satisfy $\sum_{k=1}^{\infty}\eta^k < \infty$. The functional principal components (FPC) are defined as the eigenfunctions of the covariance operator (4), Horvath and Kokoszka (2012). The estimators of FPC are called the *empirical functional principal components* (EFPC's).

Because the main interest is the reconstruction of the curve χ_{s_0}, a reasonable choice for a basis functions system is the EFPC's formed by the eigenfunctions $\xi^k(t)$, $k = 1,\ldots,K$ of the covariance operator C of $\boldsymbol{\chi}_s(t)$ with basis coefficients given by the associated principal component scores $f_{s_i}^k$, defined as

$$f_{s_i}^k = \left\langle \chi_{s_i}, \xi^k \right\rangle, \quad k = 1,\ldots,K, \quad i = 1,\ldots,n \tag{5}$$

According to Horvath and Kokoszka (2012), the approximation of this basis is uniformly optimal, in the sense of minimizing $\hat{\mathcal{S}}^2$ given by

$$\hat{\mathcal{S}}^2 = \sum_{i=1}^{n}\left\|\chi_{s_i}(t) - \sum_{k=1}^{K} f_{s_i}^k \xi^k(t)\right\|^2. \tag{6}$$

It is possible to have a very good approximation using only a few *empirical functional principal components*. Denoting by η^k 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 (Bosq 2000), we assume the model takes the form

$$\mathcal{Z}_s(t) = \mu(t) + \boldsymbol{\chi}_s(t) = \mu(t) + \sum_{k=1}^{\infty} f_{s_i}^k \xi^k(t)$$

where $E(\mathcal{Z}_s(t)) = \mu(t)$. The mean function $\mu(t)$ is estimated by the sample mean function $\hat{\mu}(t) = \mathcal{Z}_s(t)$ with $\mathcal{Z}_s(t) = n^{-1}\sum_{i=1}^{n}\mathcal{Z}_{s_i}(t)$. So $E(\boldsymbol{\chi}_s(t)) = 0$ and from now on, we use the random variable $\boldsymbol{\chi}_s(t)$. In addition, for $k = 1,\ldots,K$

$$E\left(f_{s_i}^k\right) = E\left\langle \chi_{s_i}, \xi^k \right\rangle = \left\langle 0, \xi^k \right\rangle = 0 \tag{7}$$

Note that for each k and $\boldsymbol{s} \in D_s$, $\boldsymbol{f}_s^k$ is a scalar spatial random field. So, the vector $\left(\boldsymbol{f}_s^1,\ldots,\boldsymbol{f}_s^K\right)$ is a K-dimensional spatial random field.

According to Bongiorno et al. (2014), $\boldsymbol{\Xi}_s \in \mathcal{H}^P$ is a joint Gaussian $\mathcal{H}^P$-valued random field, if the real variable

$$[\boldsymbol{\Xi}_s, \boldsymbol{\zeta}] = \langle \boldsymbol{\chi}_s^1, \zeta^1\rangle + \cdots + \langle \boldsymbol{\chi}_s^P, \zeta^P\rangle \tag{8}$$

is Gaussian for all $\boldsymbol{\zeta} \in \mathcal{H}^P$.

Let $\xi^{p1},\ldots,\xi^{pK_p}$, $p = 1,\ldots,P$ the first K_p eigenfunctions of the covariance operator of χ_s^p and following the notation in (5) let the corresponding scores be

$$f_s^{pKp} = \left\langle \chi_s^p, \xi^{pK_p}\right\rangle, \tag{9}$$

for $b_{11},\ldots,b_{PK_P}$ arbitrary real numbers, and let the vector $\boldsymbol{\zeta}$ be given by

$$\boldsymbol{\zeta} = \left(b_{11}\xi^{11} + \cdots + b_{1K_1}\xi^{1K_1},\ldots,b_{P1}\xi^{P1} + \cdots + b_{PK_p}\xi^{PK_P}\right)$$

We thus have that

$$[\boldsymbol{\Xi}_s,\boldsymbol{\zeta}] = b_{11}f_s^{11} + \cdots + b_{1K_1}f_s^{1K_1} + \cdots + b_{P1}f_s^{P1} + \cdots + b_{PK_P}f_s^{PK_P} \tag{10}$$

is a real Gaussian variable and therefore the vector

$$\left(\boldsymbol{f}_s^{11},\ldots,f_s^{1K_1},\ldots,f_s^{P1},\ldots,f_s^{PK_P}\right)$$

is a joint Gaussian multivariate random field in $\mathbb{R}^{K_1+\cdots+K_P}$.

2.3 Geostatistics for functional data

2.3.1 Ordinary kriging for functional data

Following Giraldo et al. (2011), the predictor of the curve $\chi_{s_0}(t)$ based on the set of functions $\chi_{s_i}(t)$, $i = 1,\ldots,n$ is given by

$$\overset{*}{\chi}_{s_0}(t) = \sum_{i=1}^{n}\lambda_i\chi_{s_i}(t) \qquad t \in [0,T], \quad \lambda_1,\lambda_2,\ldots,\lambda_n \in \mathbb{R}$$

The weights $\lambda_1,\lambda_2,\ldots,\lambda_n$ in (2.3.1) are found as the solution of the minimization problem of the variance of the prediction error

$$\min_{\lambda_1,\lambda_2,\ldots,\lambda_n}\int_T Var\left(\boldsymbol{\chi}_{s_0}(t) - \overset{*}{\chi}_{s_0}(t)\right)dt$$

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 $\gamma(\boldsymbol{\chi}_{s_i}(t), \boldsymbol{\chi}_{s_{i'}}(t))$, defined as

$$\gamma(\boldsymbol{\chi}_{\boldsymbol{s}_i}(t), \boldsymbol{\chi}_{\boldsymbol{s}_{i'}}(t)) = \frac{1}{2} Var\left(\boldsymbol{\chi}_{\boldsymbol{s}_i}(t) - \boldsymbol{\chi}_{\boldsymbol{s}_{i'}}(t)\right) = \gamma(\|\boldsymbol{s}_i - \boldsymbol{s}_{i'}\|, t) \tag{11}$$

Once (11) has been integrated for every pair of curves, the variogram obtained, $\gamma(\|\boldsymbol{s}_i - \boldsymbol{s}_{i'}\|)$, is scalar and modeled with usual spatial variogram models which allow to include geometric anisotropy.

2.3.2 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

$$\breve{\chi}_{\boldsymbol{s}_0}(t) = \sum_{i=1}^{n} \lambda_i \chi_{\boldsymbol{s}_i}(t)$$

is solved for $\lambda_1, \lambda_2, \ldots, \lambda_n$ by the minimum least squares method (Horvath and Kokoszka 2012),

$$E\left\|\boldsymbol{\chi}_{\boldsymbol{s}_0}(t) - \breve{\chi}_{\boldsymbol{s}_0}(t)\right\|^2 = E\langle \boldsymbol{\chi}_{\boldsymbol{s}_0}, \boldsymbol{\chi}_{\boldsymbol{s}_0}\rangle - 2\sum_{i=1}^{n} \lambda_i E\langle \chi_{\boldsymbol{s}_i}, \boldsymbol{\chi}_{\boldsymbol{s}_0}\rangle + \sum_{i,i'=1}^{n} \lambda_i \lambda_{i'} E\left\langle \chi_{\boldsymbol{s}_i}, \chi_{\boldsymbol{s}_{i'}}\right\rangle$$

where

$$E\left(\langle \chi_{\boldsymbol{s}_i}, \chi_{\boldsymbol{s}_{i'}}\rangle\right) = \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} E\left(f_{\boldsymbol{s}_i}^k f_{\boldsymbol{s}_{i'}}^{k'}\right)\left\langle \xi^k(t)\xi^{k'}(t)\right\rangle = \sum_{k=1}^{\infty} E\left(f_{\boldsymbol{s}_i}^k f_{\boldsymbol{s}_{i'}}^k\right) \tag{12}$$

Thus, due to the representation as a linear combination of the *empirical functional principal components*, see (6) and (7), the functional covariances between two locations $E\left(\langle \chi_{\boldsymbol{s}_i}, \chi_{\boldsymbol{s}_{i'}}\rangle\right)$ are completely determined by the sum of the spatial auto-covariances of all score components $f_{\boldsymbol{s}}^k$ for the pair $(\boldsymbol{s}_i, \boldsymbol{s}_{i'})$, see (5). Note that this procedure does not need the cross-covariances between score vectors.

2.3.3 Functional spatial prediction with basis function coefficients

Nerini et al. (2010) and Giraldo (2014) approximate each function in the dataset by K basis functions, see (3). The basis coefficients

$$\left(\boldsymbol{a}_{\boldsymbol{s}}^1, \ldots, \boldsymbol{a}_{\boldsymbol{s}}^K\right)$$

form a *K-dimensional* spatial random field. Thus, the proposed predictor for the curve $\boldsymbol{\chi}_{\boldsymbol{s}_0}(t)$ is

$$\widetilde{\chi}_{\boldsymbol{s}_0}(t) = \sum_{k=1}^{K} \widetilde{a}_{\boldsymbol{s}_0}^k \phi^k(t) = \boldsymbol{\Phi}^T \widetilde{\boldsymbol{a}}_{\boldsymbol{s}_0}^k$$

where $\widetilde{\boldsymbol{a}}_{\boldsymbol{s}_0} = \left(\widetilde{a}_{\boldsymbol{s}_0}^1, \ldots, \widetilde{a}_{\boldsymbol{s}_0}^K\right)^T$ is the cokriging predictor of $\boldsymbol{a}_{\boldsymbol{s}_0} = \left(a_{\boldsymbol{s}_0}^1, \ldots, a_{\boldsymbol{s}_0}^K\right)^T$.

The linear model of coregionalization (LMC) is used in order to ensure that the covariance matrix for the multivariate spatial process of the basis coefficients 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. An an alternative proposal using the *empirical functional principal components* is given by Bohorquez et al. (2015), see Sect. 2.3.4.

2.3.4 Functional prediction using FPCA and simple cokriging

Bohorquez et al. (2015) propose to represent the observed functions using as basis functions the first K eigenfunctions. Let $\boldsymbol{\xi}^T(t)$ be the vector containing the first K eigenfunctions chosen according to some threshold of percentage of accumulated variability previously established. The representation of the functions is given by

$$\chi_{\boldsymbol{s}_i}(t) = \boldsymbol{\xi}^T(t)\boldsymbol{f}_{\boldsymbol{s}_i},\ i = 1, \ldots, n$$

According to Sect. 2.2, $E(\chi_{\boldsymbol{s}}(t)) = 0$ and $E\left(\boldsymbol{f}_{\boldsymbol{s}_i}^k\right) = 0$, $i = 1, \ldots, n$ $k = 1, \ldots, K$ so simple cokriging (Myers 1982) can be used to predict the vector

$$\boldsymbol{f}_{\boldsymbol{s}_0} = \left(f_{\boldsymbol{s}_0}^1, \ldots, f_{\boldsymbol{s}_0}^K\right)^T$$

and the proposed predictor for the curve $\chi_{\boldsymbol{s}_0}(t)$ is

$$\widetilde{\chi}_{\boldsymbol{s}_0}(t) = \boldsymbol{\xi}^T(t)\widetilde{\boldsymbol{f}}_{\boldsymbol{s}_0}, \qquad i = 1, \ldots, n$$

where $\widetilde{\boldsymbol{f}}_{\boldsymbol{s}_0}$ is the simple cokriging predictor of the score vector $\boldsymbol{f}_{\boldsymbol{s}_0}$ at the unsampled location $\boldsymbol{s}_0$. Bohorquez et al. (2015) show that in general terms the score vectors are not independent. The spatial auto-covariance function for each score vector j is given by

$$E\left(f_{\boldsymbol{s}_i}^k f_{\boldsymbol{s}_{i'}}^k\right) = \begin{cases} \eta^k, & if \quad i = i' \\ \eta^k \rho^k(\|\boldsymbol{s}_i - \boldsymbol{s}_{i'}\|; \boldsymbol{\Theta}) & if \quad i \neq i' \end{cases} \tag{13}$$

where $\rho^k(.)$ is the correlation function of the spatial scalar random field determined by the score vector $\boldsymbol{f}^k$. Consequently, (13) 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 $\boldsymbol{f}^k$ has finite and known variance (sill). The variance of the prediction error can be obtained as

$$Var\left(\chi_{\boldsymbol{s}_0}(t) - \widetilde{\chi}_{\boldsymbol{s}_0}(t)\right) = \sigma_{cok}^2 \boldsymbol{\xi}^T(t)\boldsymbol{\xi}(t)$$

where σ^2_{cok} is the accumulated variance of the cokriging predictor for the vector $\boldsymbol{f}_{s_0}$.

3 Functional cokriging

We first develop the cokriging method for the case of two spatial functional random fields and then we give the generalization to the case of P spatial functional random fields.

3.1 Cokriging with two functional random fields

Let $\boldsymbol{\chi}^1_{\boldsymbol{s}}(t)$ and $\boldsymbol{\chi}^2_{\boldsymbol{s}}(t)$ be two spatial functional random fields such that $E\left(\boldsymbol{\chi}^1_{\boldsymbol{s}}(t)\right) = 0$ and $E\left(\boldsymbol{\chi}^2_{\boldsymbol{s}}(t)\right) = 0$. The cokriging predictor of the $\boldsymbol{\chi}^1_{\boldsymbol{s}_0}(t)$ for an unsampled site $\boldsymbol{s}_0$, using $\boldsymbol{\chi}^2_{\boldsymbol{s}}(t)$ as a spatial covariate, is given by

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

where $\lambda_i^{11}\ i = 1, \ldots, n_1$ are the weights of the n_1 observations of $\chi^1_{\boldsymbol{s}}(t)$ and $\lambda_j^{12}\ j = 1, \ldots, n_2$ are the weights of the n_2 observations of $\chi^2_{\boldsymbol{s}}(t)$. 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,

$$E\left(\check{\chi}^1_{s_0}(t) - \boldsymbol{\chi}^1_{\boldsymbol{s}_0}(t)\right) = E\left(\sum_{i=1}^{n_1} \lambda_i^{11} \chi^1_{s_i}(t) + \sum_{j=1}^{n_2} \lambda_j^{12} \chi^2_{s_j}(t)\right) = 0$$

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

$$\mathcal{Q} = E\|\chi^1_{s_0}(t) - \check{\chi}^1_{s_0}(t)\|^2$$

Now, we carry out the minimization of $\mathcal{Q}$ to obtain the system of cokriging equations

$$\begin{aligned}
\mathcal{Q} &= E\|\chi^1_{s_0}(t) - \check{\chi}^1_{s_0}(t)\|^2 = E\langle\chi^1_{s_0}, \chi^1_{s_0}\rangle - 2E\langle\chi^1_{s_0}, \check{\chi}^1_{s_0}\rangle \\
&\quad + E\langle\check{\chi}^1_{s_0}, \check{\chi}^1_{s_0}\rangle \\
&= E\langle\chi^1_{s_0}, \chi^1_{s_0}\rangle - 2\sum_{i=1}^{n_1} \lambda_i^{11} E\langle\chi^1_{s_i}, \chi^1_{s_0}\rangle - 2\sum_{j=1}^{n_2} \lambda_j^{12} E\langle\chi^2_{s_j}, \chi^1_{s_0}\rangle \\
&\quad + \sum_{i=1}^{n_1}\sum_{i\prime=1}^{n_1} \lambda_i^{11}\lambda_{i\prime}^{11} E\langle\chi^1_{s_i}, \chi^1_{s_{i\prime}}\rangle + 2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \lambda_i^{11}\lambda_j^{12} E\langle\chi^2_{s_j}, \chi^1_{s_i}\rangle \\
&\quad + \sum_{j=1}^{n_2}\sum_{j\prime=1}^{n_2} \lambda_j^{12}\lambda_{j\prime}^{12} E\langle\chi^2_{s_j}, \chi^2_{s_{j\prime}}\rangle
\end{aligned} \tag{14}$$

Thus, for $i = 1, \ldots, n_1$ and $\mathrm{j} = 1, \ldots, \mathrm{n}_2$ the partial derivatives are given by

$$\begin{aligned}
\frac{\partial dQ}{d\lambda_i^{11}} &= -2E\langle\chi^1_{s_i}, \chi^1_{s_0}\rangle + 2\sum_{i\prime=1}^{n_1} \lambda_{i\prime}^{11} E\langle\chi^1_{s_i}, \chi^1_{s_{i\prime}}\rangle \\
&\quad + 2\sum_{j=1}^{n_2} \lambda_j^{12} E\langle\chi^2_{s_j}, \chi^1_{s_i}\rangle
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial dQ}{d\lambda_j^{12}} &= -2E\langle\chi^2_{s_j}, \chi^1_{s_0}\rangle + 2\sum_{i=1}^{n_1} \lambda_i^{11} E\langle\chi^2_{s_j}, \chi^1_{s_i}\rangle \\
&\quad + 2\sum_{j\prime=1}^{n_2} \lambda_{j\prime}^{12} E\langle\chi^2_{s_j}, \chi^2_{s_{j\prime}}\rangle
\end{aligned}$$

Therefore, the cokriging equations are

$$\sum_{i\prime=1}^{n_1} \lambda_{i\prime}^{11} E\langle\chi^1_{s_i}, \chi^1_{s_{i\prime}}\rangle = E\langle\chi^1_{s_i}, \chi^1_{s_0}\rangle - \sum_{j=1}^{n_2} \lambda_j^{12} E\langle\chi^2_{s_j}, \chi^1_{s_i}\rangle \tag{15}$$

and

$$\sum_{j\prime=1}^{n_2} \lambda_{j\prime}^{12} E\langle\chi^2_{s_j}, \chi^2_{s_{j\prime}}\rangle = E\langle\chi^2_{s_j}, \chi^1_{s_0}\rangle - \sum_{i=1}^{n_1} \lambda_i^{11} E\langle\chi^2_{s_j}, \chi^1_{s_i}\rangle \tag{16}$$

Replacing (15) and (16) in (14) we obtain

$$\begin{aligned}
E\|\chi^1_{s_0}(t) - \check{\chi}^1_{s_0}(t)\|^2 &= E\langle\chi^1_{s_0}, \chi^1_{s_0}\rangle - \sum_{i=1}^{n_1} \lambda_i^{11} E\langle\chi^1_{s_i}, \chi^1_{s_0}\rangle \\
&\quad - \sum_{j=1}^{n_2} \lambda_j^{12} E\langle\chi^2_{s_j}, \chi^1_{s_0}\rangle
\end{aligned} \tag{17}$$

Let $f_{\boldsymbol{s}}^{1k}\ k = 1, \ldots, K$ and $f_{\boldsymbol{s}}^{2l}\ l = 1, \ldots, L$ be the spatial scalar random fields formed by the scores of the functions $\chi^1_{\boldsymbol{s}}(t)$ and $\chi^2_{\boldsymbol{s}}(t)$, respectively. Expressing each function in (17) in terms of its principal components, we obtain

$$\begin{aligned}
E\langle\chi^1_{s_0}, \chi^1_{s_0}\rangle &= \sum_{k=1}^{K}\sum_{k'=1}^{K} E\left(f_{s_0}^{1k} f_{s_0}^{1k'}\right)\langle\xi^{1k}, \xi^{1k'}\rangle \\
\sum_{i=1}^{n_1} \lambda_i^{11} E\langle\chi^1_{s_i}, \chi^1_{s_0}\rangle &= \sum_{i=1}^{n_1}\sum_{k=1}^{K}\sum_{k'=1}^{K} \lambda_i^{11} E\left(f_{s_i}^{1k} f_{s_0}^{1k'}\right)\langle\xi^{1k}, \xi^{1k'}\rangle \\
\sum_{j=1}^{n_2} \lambda_j^{12} E\langle\chi^2_{s_j}, \chi^1_{s_0}\rangle &= \sum_{j=1}^{n_2}\sum_{l=1}^{L}\sum_{k=1}^{K} \lambda_j^{12} E\left(f_{s_i}^{2l} f_{s_0}^{1k}\right)\langle\xi^{2l}, \xi^{1k}\rangle
\end{aligned}$$

where $\xi^{1k},\ k = 1, \ldots, K$ are the eigenfunctions of the covariance operator of $\chi^1_{\boldsymbol{s}}(t)$ and $\xi^{2l},\ l = 1, \ldots, L$ are the eigenfunctions of the covariance operator of $\chi^2_{\boldsymbol{s}}(t)$. 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, (17) can be simplified as follows

$$\begin{aligned} E\|\chi_{s_0}^1(t) - \breve{\chi}_{s_0}^1(t)\|^2 &= \sum_{k=1}^{K} E\left(f_{s_0}^{1k} f_{s_0}^{1k}\right) \\ &- \sum_{i=1}^{n_1}\sum_{k=1}^{K} \lambda_i^{11} E\left(f_{s_i}^{1k} f_{s_0}^{1k}\right) - \sum_{j=1}^{n_2}\sum_{l=1}^{L}\sum_{k=1}^{K} \lambda_j^{12} c_{12}^{lk} E\left(f_{s_i}^{2l} f_{s_0}^{1k}\right) \\ &= \sum_{k=1}^{K} \eta^{1k} - \sum_{i=1}^{n_1}\sum_{k=1}^{K} \lambda_i^{11} E\left(f_{s_i}^{1k} f_{s_0}^{1k}\right) \\ &- \sum_{j=1}^{n_2}\sum_{l=1}^{L}\sum_{k=1}^{K} \lambda_j^{12} c_{12}^{lk} E\left(f_{s_i}^{2l} f_{s_0}^{1k}\right) \end{aligned} \tag{18}$$

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.

3.2 Cokriging with P functional random fields

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

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

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

$$Q = E\|\chi_{s_0}^r(t) - \breve{\chi}_{s_0}^r(t)\|^2 \tag{19}$$

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

$$\frac{\partial dQ}{d\lambda_i^{rm}} = -2E\langle \chi_{s_j}^m, \chi_{s_0}^r \rangle + 2\sum_{p=1}^{P}\sum_{i=1}^{n_p} \lambda_i^{rp} E\langle \chi_{s_j}^m, \chi_{s_i}^p \rangle, \quad j = 1, \ldots, n_m \tag{20}$$

and

$$E\langle \chi_{s_j}^m, \chi_{s_0}^r \rangle = \sum_{p=1}^{P}\sum_{i=1}^{n_p} \lambda_i^{rp} E\langle \chi_{s_j}^m, \chi_{s_i}^p \rangle, \quad j = 1, \ldots, n_m \tag{21}$$

respectively. Replacing (20) and (21) in the squared norm of the prediction error (19), we obtain

$$E\|\chi_{s_0}^r(t) - \breve{\chi}_{s_0}^r(t)\|^2 = E\langle \chi_{s_0}^r, \chi_{s_0}^r \rangle - \sum_{p=1}^{P}\sum_{i=1}^{n_p} \lambda_i^{rp} E\langle \chi_{s_i}^p, \chi_{s_0}^r \rangle \tag{22}$$

Now, using the functional principal components representation we show that (22) only depends on the auto-covariances and cross-covariances between the score vectors chosen for each random field, that is,

$$\begin{aligned} E\|\chi_{s_0}^r(t) - \breve{\chi}_{s_0}^r(t)\|^2 &= \sum_{k=1}^{K} E\left(f_{s_0}^{rk} f_{s_0}^{rk}\right) \\ &- \sum_{i=1}^{n_p}\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{p=1}^{P} \lambda_i^{rp} c_{rp}^{lk} E\left(f_{s_i}^{pk} f_{s_0}^{rl}\right) \end{aligned} \tag{23}$$

where, as before, denoting by η^{rk}, $k = 1, \ldots, K$ the eigenvalues of the observation of χ_s^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

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

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 and finite variance and covariance structure depending only on the distance between locations, see (13), 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 $\chi_{s_0}^r(t)$ at B unsampled sites, we can use

$$\sum_{b=1}^{B}\left(E\|\chi_{s_0^b}^r(t) - \breve{\chi}_{s_0^b}^r(t)\|^2\right) \tag{24}$$

The functional cokriging predictor presented in this paper is the optimal linear predictor, that is, the linear unbiased predictor of minimum prediction error variance. In the scalar case, under the Gaussianity assumption and an square-error loss function, this predictor and its error variance coincides with the optimal predictor. If the data are not Gaussian and it is not possible to reach the Gaussianity with some transformation, it is necessary to use a non-Gaussian model because in this case the optimal predictor is generally a nonlinear function of the observed data. The definition of joint Gaussian distribution in a Hilbert space is similar to that in the multivariate case, and it is characterized completely by the mean function and the covariance operator. Nevertheless, extensions of functional

geostatistics under Gaussianity or goodness-of-fit tests for a joint Gaussian distribution in Hilbert spaces have not been studied. This is an interesting subject because even multivariate normality tests lose power for large dimensions.

3.3 Spatial 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. An optimal design S_n^* is defined by Müller (2007) as one that

$$S_n^* = \arg\max_{S_n \in \Xi_n} \Phi(\boldsymbol{\Theta}, S_n) \tag{25}$$

where $\Phi(\boldsymbol{\Theta}, S_n)$ is the design criterion and any scalar measure of information obtained with the design S_n that depends on the parameter vector $\boldsymbol{\Theta}$ and Ξ_n is the set of all *n-observation* designs. The design criterion $\Phi(\boldsymbol{\Theta}, S_n)$ in the spatial sampling context comes from the variance of the error associated with each predictor. The set Ξ_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.

3.4 Spatial optimal sampling for functional cokriging

We need to design or redesign the p sets $S_p = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n_p}\}$ $p = 1, \ldots, P$, or at least those that can be changed, of observed spatial locations ensuring an optimal spatial functional prediction of $\chi_{\boldsymbol{s}}^r(t)$ in a set of interest locations $S_0 = \{\boldsymbol{s}_0^1, \ldots, \boldsymbol{s}_0^B\}$ 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 (24). Suppose first that m_p stations can be added for the observation of each random field $\chi_{\boldsymbol{s}}^p(t)$, $p = 1, \ldots, P$. So, the enlarged network for each case is then

$$S_p' = S_p \cup S_{m_p} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n_p}, \boldsymbol{s}_{n_p+1}, \ldots, \boldsymbol{s}_{n_p+m_p}\},\ p = 1, \ldots, P$$

Let $\bigcup_{p=1}^P S_{m_p} = S_{m_1} \cup \cdots \cup S_{m_P} \subset D_s'$ the set of new locations that must be determined. So, among all possible subsets $\bigcup_{p=1}^P S_{m_p}$, we must choose the one that minimizes the total minimum square norm of the prediction error for functional cokriging. So, according to (23) and (24) the design criterion is given by

$$\arg\min_{\bigcup_{p=1}^P S_{m_p} \subset D_s'} \sum_{b=1}^B E\|\chi_{\boldsymbol{s}_0^b}^r(t) - \breve{\chi}_{\boldsymbol{s}_0^b}^r(t)\|^2 \tag{26}$$

where $\sum_{k=1}^K \eta^{rk}$ is constant and therefore the criterion (26) turns into

$$\arg\max_{\bigcup_{p=1}^P S_{m_p} \subset D_s'} \sum_{b=1}^B \sum_{i=1}^{n_p+m_p} \sum_{k=1}^K \sum_{l=1}^L \sum_{p=1}^P \lambda_i^{rp} c_{rp}^{lk} E\left(f_{\boldsymbol{s}_i}^{pk} f_{\boldsymbol{s}_0^b}^{rl}\right) \tag{27}$$

The criterion (27) establishes the general case, but frequently all random fields are measured at the same set of places $S = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$. Here, the optimization is over the possible sets $S_m \subset D_s'$ and there is only one enlarged network $S \cup S_m \subset D_s'$ 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 (27) depend only on the distance between observations and prediction sites due to the second order stationarity. The numbers K, 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.

The design criterion shown in this section depends on parameter $\boldsymbol{\Theta}$. 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 $\boldsymbol{\Theta}$ and not only on the design. Thus (25) turns into the optimization of $\Phi(\hat{\boldsymbol{\Theta}})$. For the covariance structures necessary for the functional cokriging predictor, 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. We use the plug-in estimators in the linear model of coregionalization to carry on the optimization of the sampling criteria for each case. That is, in every place where the terms $\eta_j \rho_j(\|\cdot\|; \boldsymbol{\Theta})$ or $\gamma(\|\cdot\|; \boldsymbol{\Theta})$ appear, we replace them with $\eta_j \rho_j\left(\|\cdot\|; \hat{\boldsymbol{\Theta}}\right)$ and $\gamma(\|\cdot\|; \hat{\boldsymbol{\Theta}})$. Harville and Jeske (1992) propose a correction to incorporate the uncertainty due to the lack of knowledge of $\boldsymbol{\Theta}$. Zhu and Stein (2006) 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 $\boldsymbol{\Theta}$. So, the best option is to use the plug-in method as long as the spatial auto-correlation is moderate or strong, see Schabenberger and Gotway (2004).

4 Simulation

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

$$\boldsymbol{\chi}_s^1 = \sum_{k=1}^{2} f_{s_i}^{1k}\xi^{1k}(t); \qquad \boldsymbol{\chi}_s^2 = \sum_{k=1}^{3} f_{s_i}^{2k}\zeta^{2k}(t), \quad \boldsymbol{s} \in \mathbb{R}^2 \tag{28}$$

$\left\{\xi^{11}(t) = sin(\pi t),\ \xi^{12}(t) = cos(\pi t)\right\}$ and $\left\{\zeta^{21}(t) = \sqrt{\frac{3}{2}}t,\ \zeta^{22}(t) = \frac{\sqrt{5}(3t^2-1)}{2\sqrt{2}},\ \zeta^{23}(t) = \frac{5}{2}\sqrt{\frac{7}{2}}(t^3 - \frac{3}{5}t)\right\}$ are sets of orthonormal functions in $L^2[-1,1]$. These functions could be selected from different basis or one common basis. Let $\gamma_1(h)$ stand for an isotropic exponential model with range 1, $\gamma_2(h)$ stand for an isotropic spherical model with range 0.5, and $\gamma_3(h)$ stand for an isotropic Gaussian model with range 0.7. The vector of scores $(f_s^{11}, f_s^{12}, f_s^{13}, f_s^{21}, f_s^{22})$ is a realization of a multivariate Gaussian distribution of zero mean and dependence structure given by the linear model of coregionalization

$$\begin{pmatrix} \gamma_{f^{11}}(h) & \gamma_{f^{11}f^{12}}(h) & \gamma_{f^{11}f^{13}}(h) & \gamma_{f^{11}f^{21}}(h) & \gamma_{f^{11}f^{22}}(h) \\ \gamma_{f^{12}f^{11}}(h) & \gamma_{f^{12}}(h) & \gamma_{f^{12}f^{13}}(h) & \gamma_{f^{12}f^{21}}(h) & \gamma_{f^{12}f^{22}}(h) \\ \gamma_{f^{13}f^{11}}(h) & \gamma_{f^{13}f^{12}}(h) & \gamma_{f^{13}}(h) & \gamma_{f^{13}f^{21}}(h) & \gamma_{f^{13}f^{22}}(h) \\ \gamma_{f^{21}f^{11}}(h) & \gamma_{f^{21}f^{12}}(h) & \gamma_{f^{21}f^{13}}(h) & \gamma_{f^{21}}(h) & \gamma_{f^{21}f^{22}}(h) \\ \gamma_{f^{22}f^{11}}(h) & \gamma_{f^{22}f^{12}}(h) & \gamma_{f^{22}f^{13}}(h) & \gamma_{f^{22}f^{21}}(h) & \gamma_{f^{22}}(h) \end{pmatrix} = \tag{29}$$

$$\begin{pmatrix} 25.0 & 2.0 & 1.1 & 0.6 & 2.8 \\ 2.0 & 0.4 & 0.5 & 0.5 & 0.7 \\ 1.1 & 0.5 & 1.2 & 1.7 & 1.8 \\ 0.6 & 0.5 & 1.7 & 2.5 & 2.6 \\ 2.8 & 0.7 & 1.8 & 2.6 & 2.9 \end{pmatrix}\gamma_1(h)$$

$$+\begin{pmatrix} 2.7 & 1.4 & 1.4 & 0.7 & 2.1 \\ 1.4 & 0.7 & 0.7 & 0.3 & 1.0 \\ 1.4 & 0.7 & 0.8 & 1.6 & 1.4 \\ 0.7 & 0.3 & 1.6 & 16.0 & 4.0 \\ 2.1 & 1.1 & 1.4 & 4.0 & 2.4 \end{pmatrix}\gamma_2(h)$$

$$+\begin{pmatrix} 1.5 & 1.0 & 1.3 & 1.0 & 0.9 \\ 1.0 & 1.3 & 0.9 & 1.0 & 0.9 \\ 1.3 & 0.9 & 1.1 & 0.9 & 0.8 \\ 1.0 & 1.0 & 0.9 & 0.9 & 0.8 \\ 0.9 & 0.9 & 0.8 & 0.8 & 0.7 \end{pmatrix}\gamma_3(h)$$

The elements of the matrices in (29) are rounded. The simulated data are shown in top panels of Fig. 1.

The simulation grid is shown in the bottom left panel of Fig. 1. The goal is to carry out the functional cokriging method to predict $\boldsymbol{\chi}_s^1$ using $\boldsymbol{\chi}_s^2$ as explanatory variable. The first functional principal component of the generated data accumulates 87.6% and 86.1% of the variability for $\boldsymbol{\chi}_s^1$ and $\boldsymbol{\chi}_s^2$ 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 (Goovaerts, 1997). The fitted LMC for the bivariate spatial scalar random field (f_s^{11}, f_s^{21}) is given by

$$\begin{aligned} \gamma_{f^{11}}(h) &= 26.3\gamma_1(h) + 2.1\gamma_2(h) \\ \gamma_{f^{21}}(h) &= 2.1\gamma_1(h) + 17.5\gamma_2(h) \\ \gamma_{f^{11}f^{21}}(h) &= 0.3\gamma_1(h) + 1.3\gamma_2(h) \end{aligned} \tag{30}$$

Using the results in (30), 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 Eq. (18). 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 (Montero et al. 2015). Note in bottom right panel of Fig. 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.

5 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 Fig. 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 (NO_2) among others. The particulate matter (PM) is an important component of air pollution. One of the most important physical characteristics of this material is the particle diameter, since it

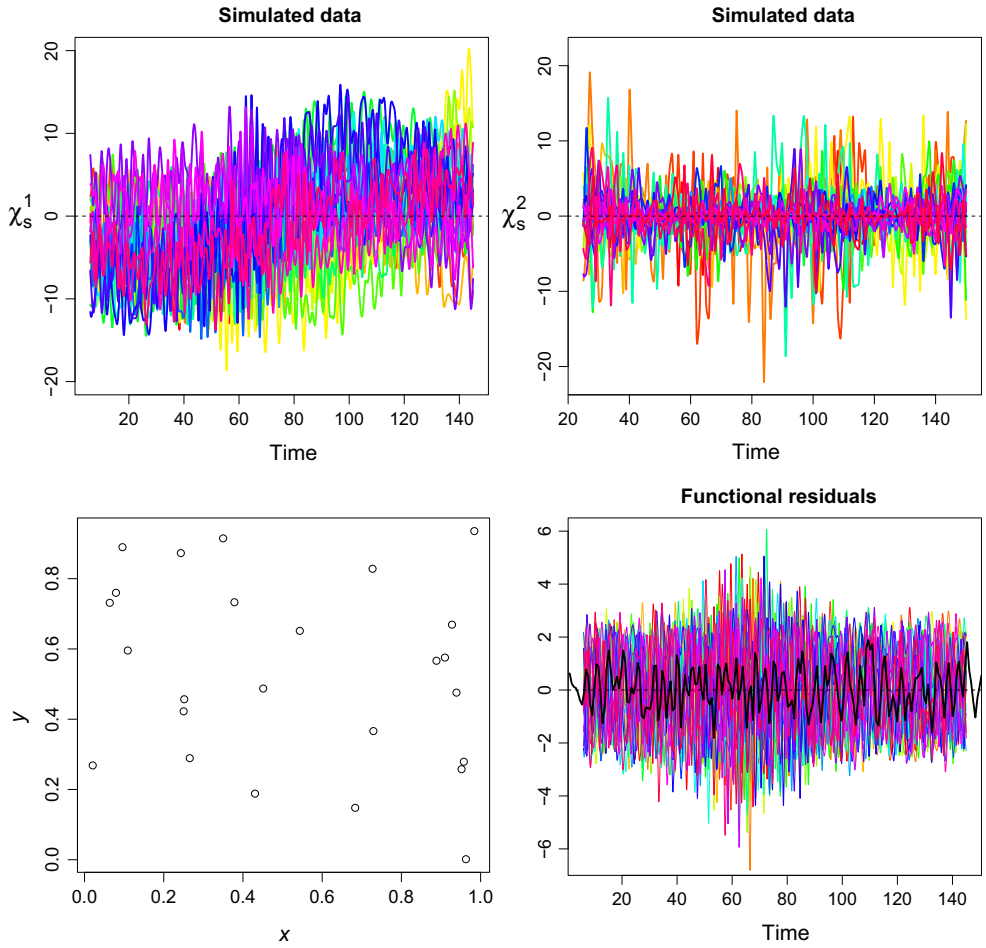

Fig. 1 *Top Left Panel* Simulated data, χ_s^1. *Top Right Panel* Simulated data, χ_s^2. *Bottom Left Panel* Grid of simulation locations. *Bottom Right Panel* Cross validation residuals and residual mean

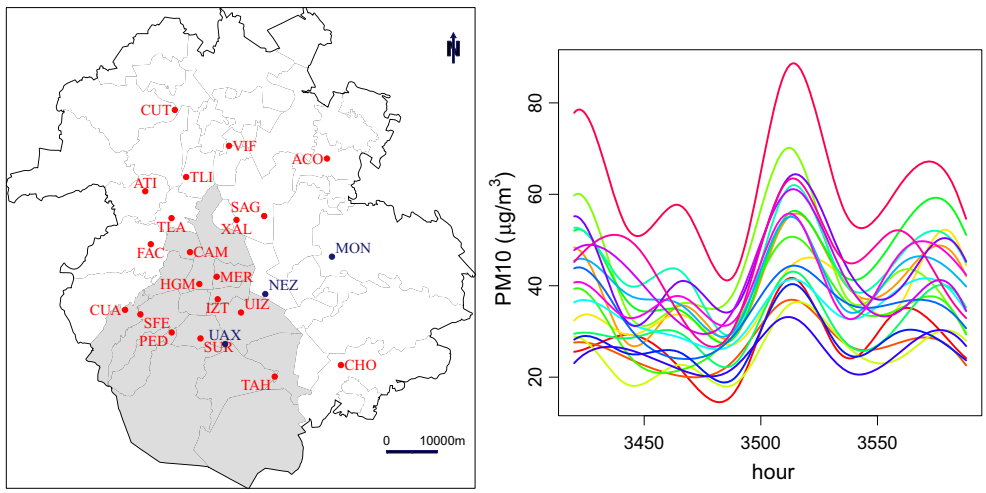

Fig. 2 *Left Panel* México city. Air quality network RAMA (shown in *red*). The *blue* points are the stations that measure temperature but belong to REDMET. *Right Panel* PM10, May 23 to May 30, 2015

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, Greenbaum et al. 2001; Paldy et al. 2006). NO_2 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 NO_2 damage human-made materials and are the major causes of reduced visibility (Secretaría del Medio Ambiente México 2015). 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 (Secretaría del Medio Ambiente México 2015).

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 NO_2 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 Figs. 2 and 3 show the curves for the last week in the dataset.

The first principal component accumulates 75%, 84.6% and 85.7% of the variability for PM10, NO_2 and Temp respectively, and the second principal component only 13.9%, 13.8% and 13.1. Thus, using 85 % as a threshold, we include two score vectors for PM10, and for NO_2 and Temp it is enough to include only the first score vector.

Our interest is the spatial functional prediction of PM10 using as functional covariates NO_2 and Temp. So according to the notation in Sect. 3, PM10 is $\chi_s^1(t)$, NO_2 is $\chi_s^2(t)$ and Temp is $\chi_s^3(t)$. Figure 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 $\gamma_{f^{11}}$ and $\gamma_{f^{12}}$ are the variograms for the first principal component of PM10, $\gamma_{f^{21}}$ and $\gamma_{f^{31}}$ are the variograms for the score vectors corresponding to the first principal component of NO_2 and Temp, respectively, and the rest of variograms in Fig. 4 and in Table 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 $\|\boldsymbol{s}_i - \boldsymbol{s}_{i'}\| = 0$. 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, $\boldsymbol{s}_0^1$ and $\boldsymbol{s}_0^2$, see Fig. 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, restricted to the area with stations. Figure 5 (left panel) shows the locations of interest $\boldsymbol{s}_0^1$ and $\boldsymbol{s}_0^2$ and the optimal sampling location (OL) to add a new station keeping fixed the current network, and using (27). For the optimization procedure, we use simulated annealing (Brooks and Morgan 1995) with state given by the spatial sampling design criterion applied at each iteration. The energy function is given by the criterion (27).

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, i.e., removing each

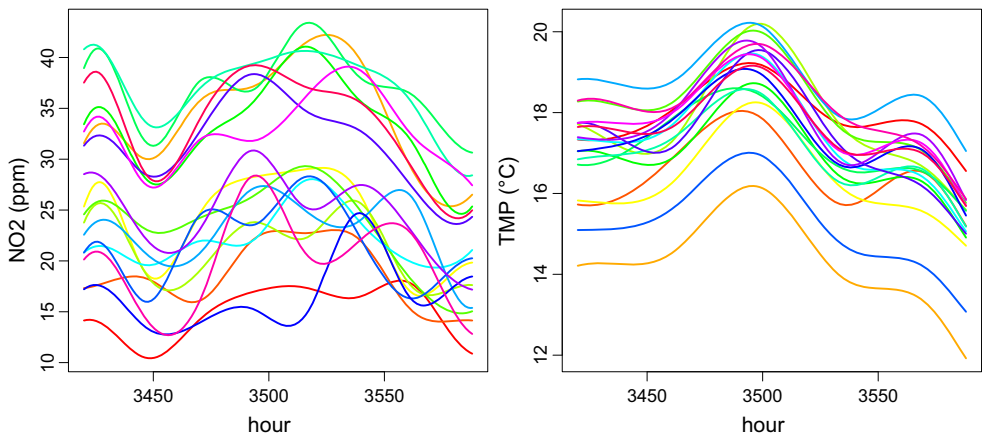

Fig. 3 *Left panel* NO_2, May 23 to May 30, 2015. *Right panel* Temperature, May 23 to May 30, 2015

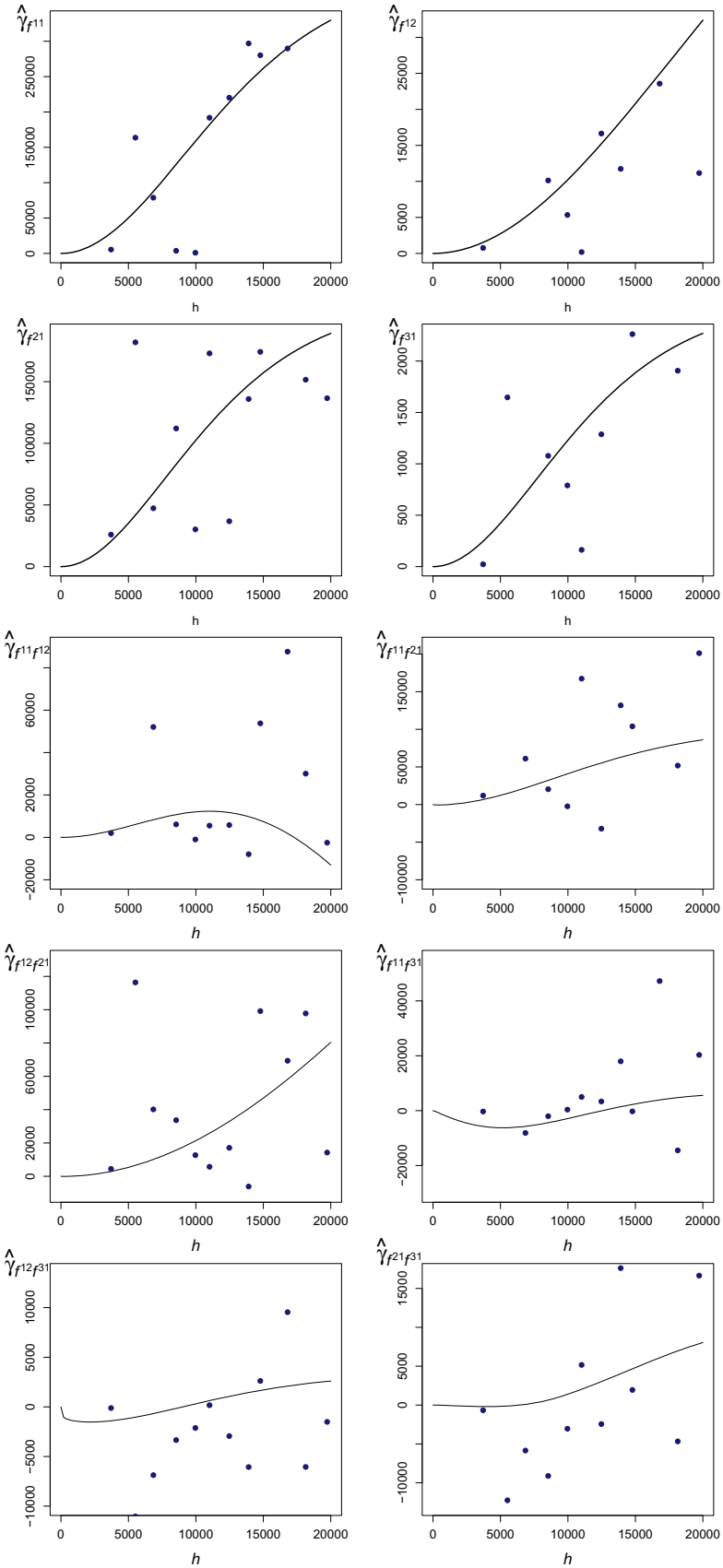

Fig. 4 Experimental and theoretical variogram fitted according to the linear model of coregionalization

Table 1 Nested variogram components of the linear model of coregionalization model using two nested Matérn structures

γ_f^{pk}	$\hat{\gamma}_{f^{11}}$	$\hat{\gamma}_{f^{21}}$	$\hat{\gamma}_{f^{31}}$	$\hat{\gamma}_{f^{11}f^{12}}$	$\hat{\gamma}_{f^{11}f^{21}}$	$\hat{\gamma}_{f^{11}f^{31}}$	$\hat{\gamma}_{f^{12}}$	$\hat{\gamma}_{f^{12}f^{21}}$	$\hat{\gamma}_{f^{12}f^{31}}$	$\hat{\gamma}_{f^{21}f^{31}}$
$v = 0.1$	216778.0	24415.3	56381.7	−11931	71160.1	104764.3	−11931.2	31932.8	−3508.696	−6856.4
$v = 5$	1923456.9	386064.8	1181738.7	202568	−680077.6	−92372.9	202568.5	388685.0	3251.8	164083.5

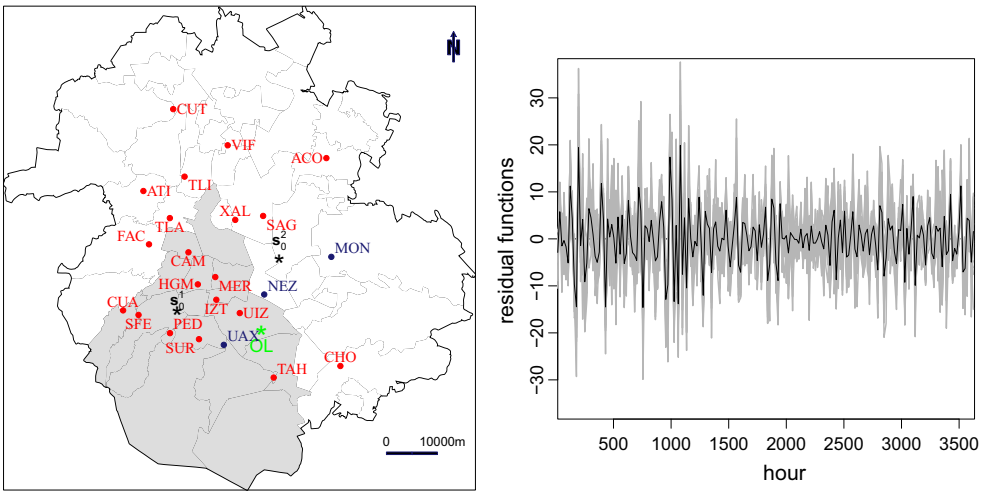

Fig. 5 *Left Panel* Optimal location for one additional station. *Right Panel* Cross validation residuals and residual mean of the predictions of PM10 using functional cokriging method

functional observation and using the rest of functional observations to predict a smoothed function at the removed location (Montero et al. 2015). 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 Fig. 5 (right panel).

Finally, Fig. 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 (Bohorquez et al. 2015). 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 NO_2 and Temp are included.

As a further exercise, we perform some exploratory analysis of data and scores. Figures 7, 8 and 9 show some density plots of PM10, NO_2 and Temp, and Fig. 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 (Korkmaz et al. 2014). Any of the tests lead to reject the hypothesis of

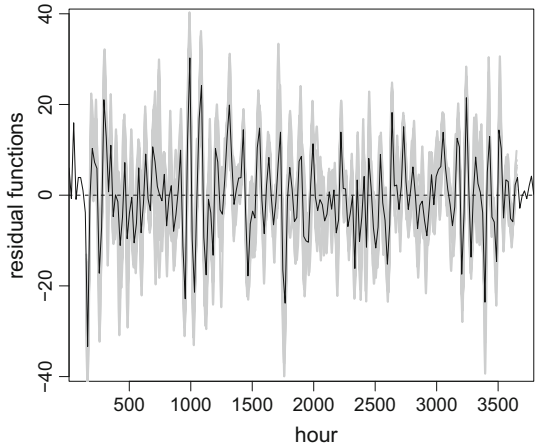

Fig. 6 Cross validation residuals and residual mean of the predictions of PM10 using functional kriging method proposed in Bohorquez et al. (2015)

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 of the observed

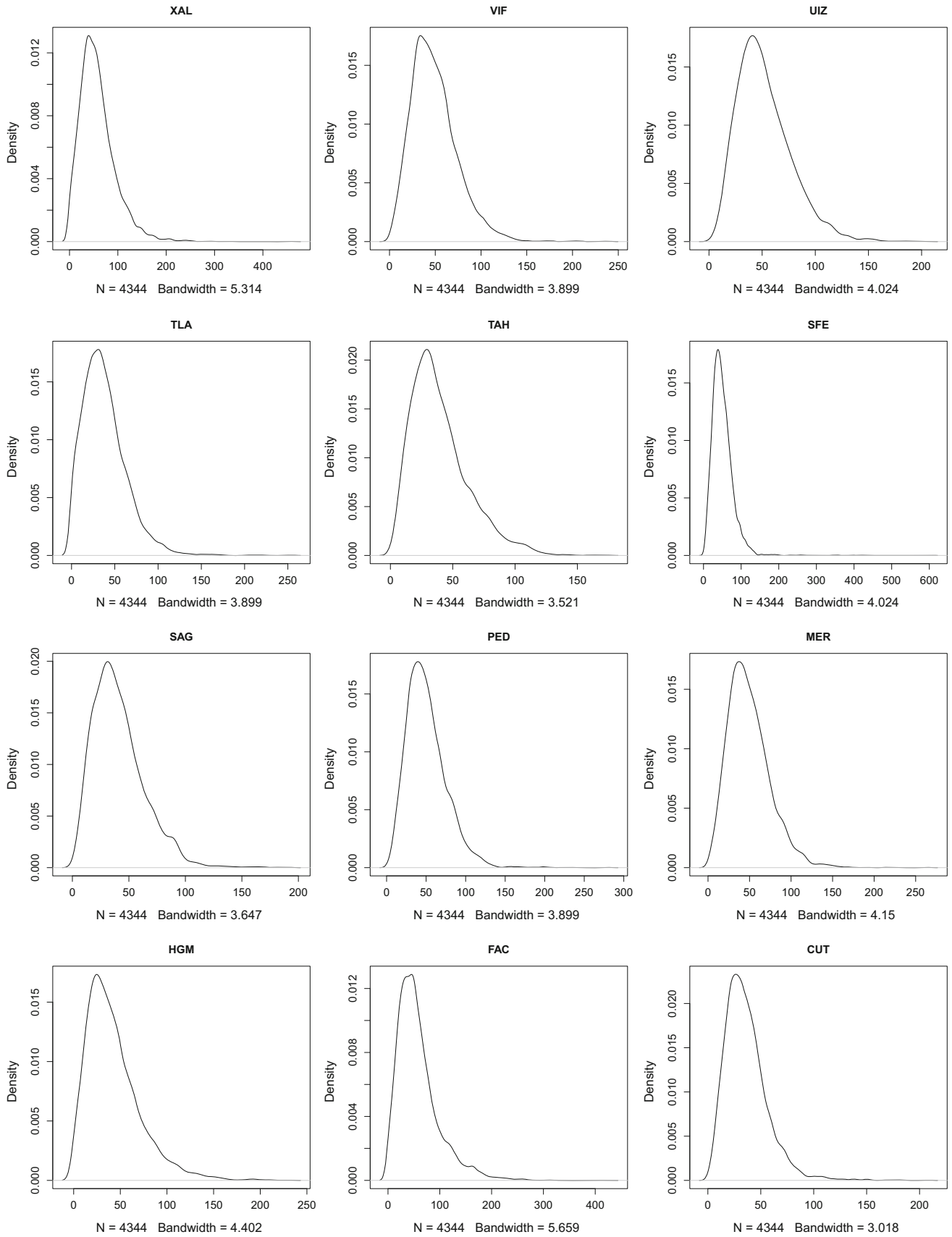

Fig. 7 Density plots for the PM10 data at some stations

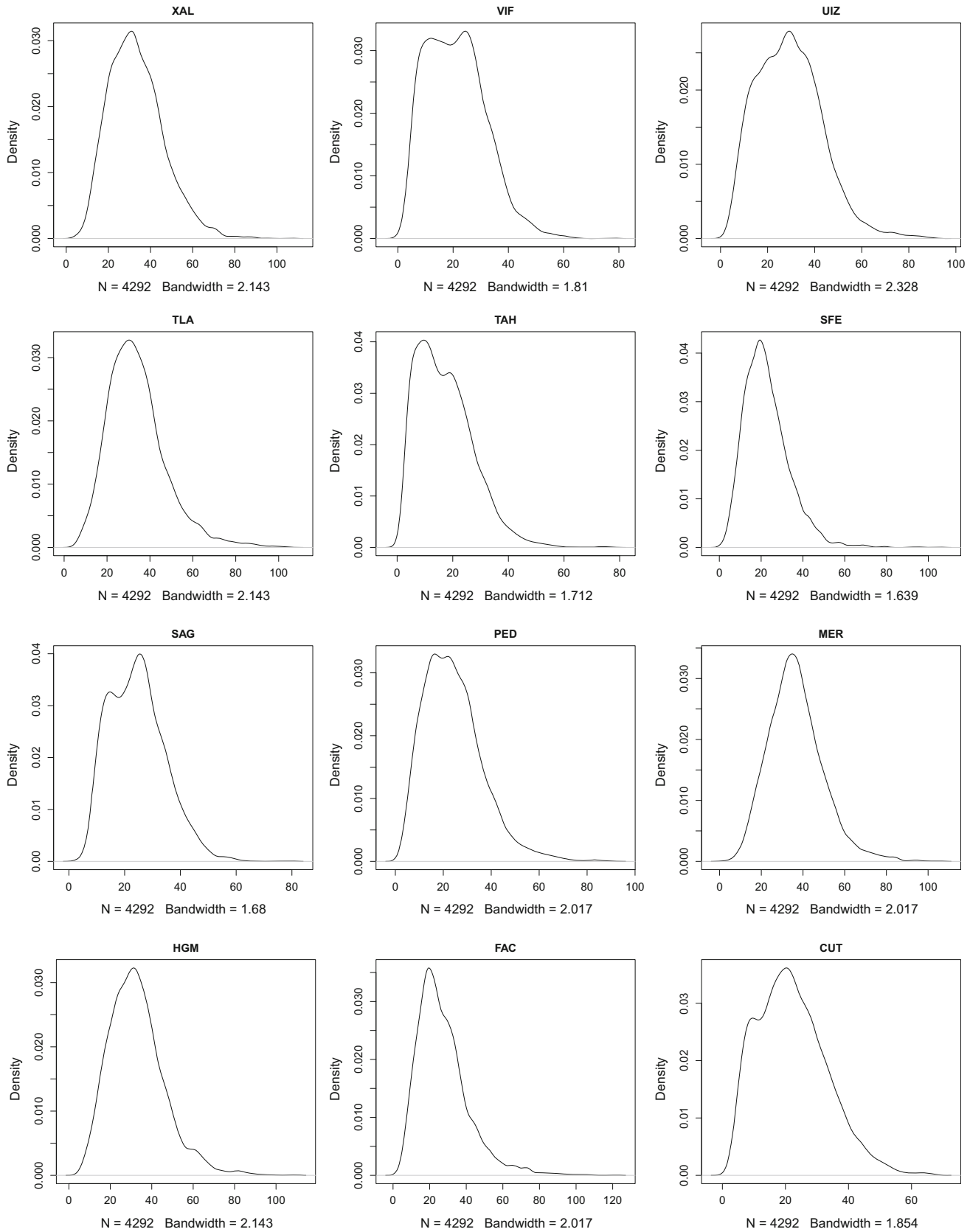

Fig. 8 Density plots for the NO_2 data at some stations

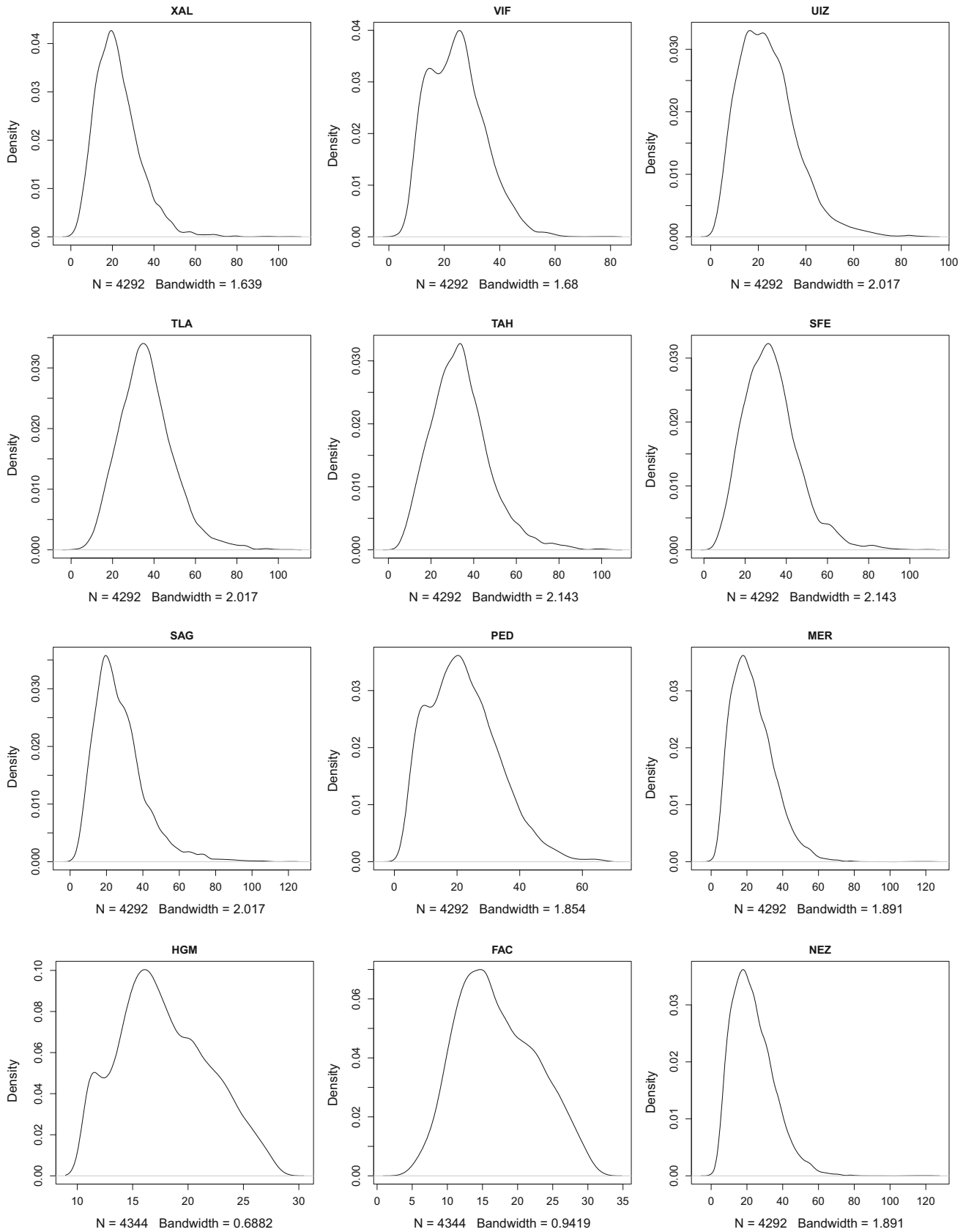

Fig. 9 Density plots for the Temperature data at some stations

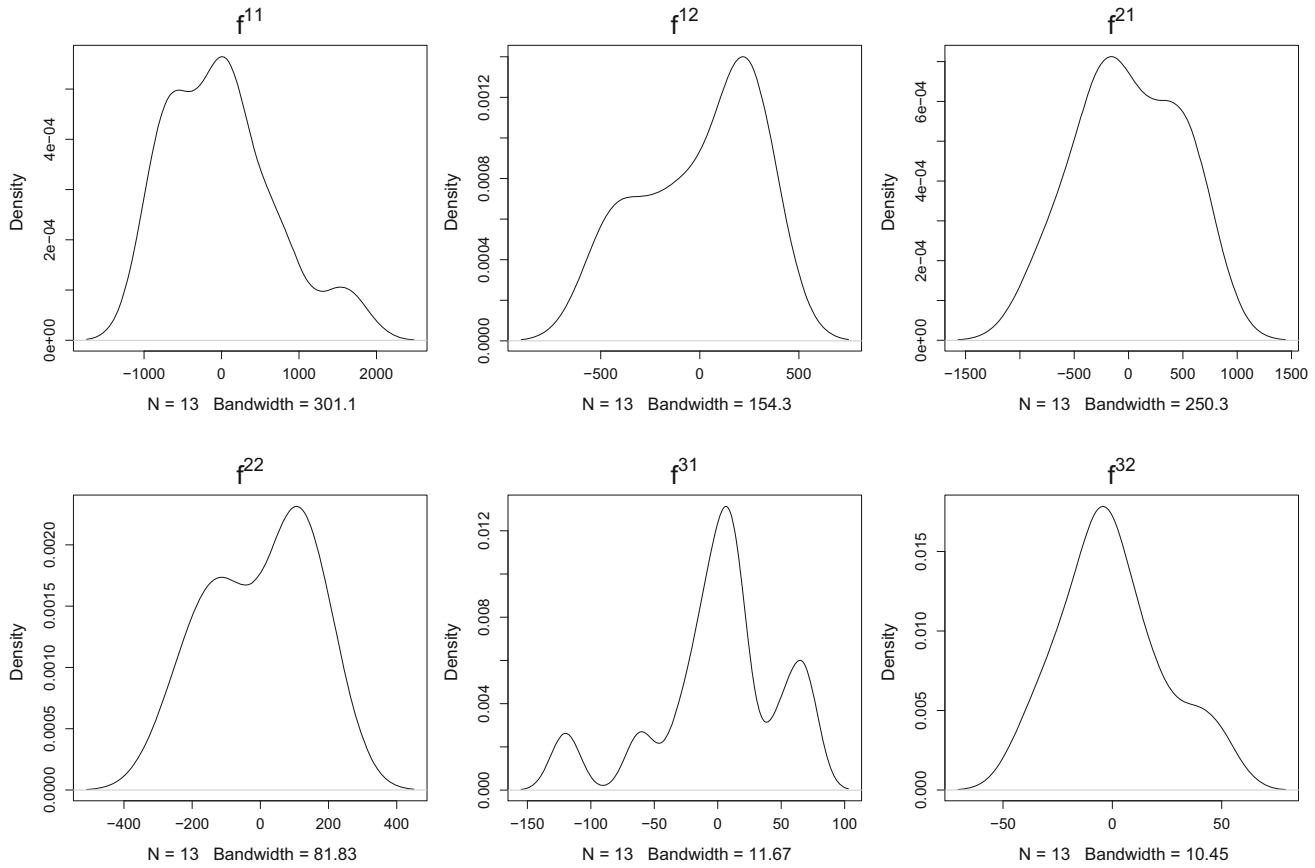

Fig. 10 Density plots for the first two scores of each variable

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.

6 Discussion and conclusions

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 $\hat{\mu}(t) = \mathcal{Z}_s(t)$. We are aware that the spatial

covariance between curves can modify the true mean function. A more formal option could be to estimate the spatial covariance structure Σ_t at each time point t and use the generalized least squares mean estimator to center data before building the curves. However, this is a hard option to be applied when T is big, unless there is a known structure and the parameter estimation can be carried out in an automatic way.

The computations in this paper are done using the free software R and the packages *fda*, *fda.usc*, *GenSA*, *gstat*, *RandomFields*, *MVN*, *MSBVAR* and *maptools*. Here we used simulated annealing because it allows to define discrepancy functions in the iterations, but others such as genetic algorithms can also been applied (Ruiz-Cárdenas et al. 2010). When there is not a big number of available options for new sites, it is possible to compute the criteria in all cases.

Acknowledgments The authors are grateful to the positive comments made by the reviewers and the Associate Editor that have improved the quality of the manuscript.

References

Angulo J, Bueso M, Alonso F (2000) A study on sampling design for optimal prediction of space-time stochastic processes. Stoch Environ Res Risk Assess 14(6):412–427

Bohorquez M, Giraldo R, Mateu J (2015) Optimal sampling for spatial prediction of functional data. Stat Methods Appl 25:39–54. doi:10.1007/s10260-015-0340-9

Bongiorno EG, Salinelli E, Goia A, Vieu P (2014) Contributions in infinite-dimensional statistics and related topics. Societa Editrice Esculapio, Bologna

Bosq D (2000) Linear processes in function spaces: theory and applications, 149th edn. Springer, New York

Brooks S, Morgan B (1995) Optimization using simulated annealing. J R Stat Soc Ser D 44(2):241–257

Caballero W, Giraldo R, Mateu J (2013) A universal kriging approach for spatial functional data. Stoch Environ Res Risk Assess 27(7):1553–1563

Escabias M, Aguilera AM, Valderrama MJ (2005) Modeling environmental data by functional principal component logistic regression. Environmetrics 16(1):95–107

Ferraty F, Vieu P (2006) Nonparametric functional data analysis: theory and practice. Springer, New York

Giraldo R (2014) Cokriging based on curves, prediction and estimation of the prediction variance. InterStat 2:1–30

Giraldo R, Delicado P, Mateu J (2010) Continuous time-varying kriging for spatial prediction of functional data: an environmental application. J Agric Biol Environ Stat 15(1):66–82

Giraldo R, Delicado P, Mateu J (2011) Ordinary kriging for function-valued spatial data. Environ Ecol Stat 18(3):411–426

Giraldo R, Mateu J (2013) Kriging for functional data. Wiley, New York

Giraldo RA, Mateu J, Delicado P (2012) geofd: an R package for function-valued geostatistical prediction. Revista Colombiana de Estadística 35:385–407

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press on Demand

Goulard M, Voltz M (1993) Geostatistical interpolation of curves: a case study in soil science. In: Soares A (ed) Geostatistics Tróia'92. Kluwer Academic Press, Boston

Greenbaum DS, Bachmann JD, Krewski D, Samet JM, White R, Wyzga RE (2001) Particulate air pollution standards and morbidity and mortality: case study. Am J Epidemiol 154(12):78–90

Harville DA, Jeske DR (1992) Mean squared error of estimation or prediction under a general linear model. J Am Stat Assoc 87(419):724–731

Horvath L, Kokoszka P (2012) Inference for functional data with applications. Springer, New York

Ignaccolo R, Ghigo S, Bande S (2013) Functional zoning for air quality. Environ Ecol Stat 20(1):109–127

Ignaccolo R, Mateu J, Giraldo R (2014) Kriging with external drift for functional data for air quality monitoring. Stoch Environ Res Risk Assess 28(5):1171–1186

Korkmaz S, Goksuluk D, Zararsiz G (2014) Mvn: An r package for assessing multivariate normality. R J 6(2):151–162

Le N, Zidek J (2006) Statistical analysis of environmental space-time processes. Springer, New York

Montero JM, Fernandez-Aviles G, Mateu J (2015) Spatial and spatio-temporal geostatistical modeling and kriging. Wiley, New York

Müller W (2007) Collecting spatial data: optimum design of experiments for random fields. Springer, Berlin

Myers D (1982) Matrix formulation of co-kriging. Math Geol 14(3):249–257

Nerini D, Monestiez P, Manté C (2010) Cokriging for spatial functional data. J Multiv Anal 101(2):409–418

Paldy A, Bobvos J, Lustigova M, Moshammer H, Niciu EM, Otorepec P, Puklova V, Szafraniec K, Zagargale T, Neuberger M et al (2006) Health impact assessment of pm10 on mortality and morbidity in children in central-eastern european cities. Epidemiology 17(6):S131

Ramsay J, Silverman B (2005) Functional data analysis. Springer, New York

Reed M, Simon B (1980) Methods of modern mathematical physics I: functional analysis. Academic Press, Inc., San Diego

Ruiz-Cárdenas R, Ferreira M, Schmidt A (2010) Stochastic search algorithms for optimal design of monitoring networks. Environmetrics 21(1):102–112

Ruiz-Medina MD (2011) Spatial autoregressive and moving average hilbertian processes. J Multiv Anal 102(2):292–305

Ruiz-Medina M (2012) Spatial functional prediction from spatial autoregressive hilbertian processes. Environmetrics 23(1):119–128

Ruiz-Medina M, Espejo R (2012) Spatial autoregressive functional plug-in prediction of ocean surface temperature. Stoch Environ Res Risk Assess 26(3):335–344

Ruiz-Medina M, Anh V, Espejo R, Angulo J, Frías MP (2015) Least-squares estimation of multifractional random fields in a hilbert-valued context. J Optim Theory Appl 167(3):888–911

Schabenberger O, Gotway C (2004) Statistical methods for spatial data analysis. CRC Press, London

Secretaría del Medio Ambiente México S (2015) http://www.aire.df.gob.mx/default.php

Srensen H, Goldsmith J, Sangalli LM (2013) An introduction with medical applications to functional data analysis. Stat Med 32(30):5222–5240

Torres JM, Nieto P, Alejano L, Reyes A (2011) Detection of outliers in gas emissions from urban areas using functional data analysis. J Hazard Mater 186(1):144–149

Yen JDL, Thomson JR, Paganin DM, Keith JM, Mac Nally R (2014) Function regression in ecology and evolution: free. Methods Ecol Evol 6:17–26

Zhu Z, Stein M (2006) Spatial sampling design for prediction with estimated parameters. J Agric Biol Environ Stat 11(1):24–44

Zimmerman D (2006) Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics 17(6):635–652