JIRSS (2019)
Vol. 18, No. 02, pp 115-137
DOI:10.29252/jirss.18.2.115

On a Linear Functional Mixed Effect Model for Spatial Data

Roya Nasirzadeh [1], Jorge Mateu [2], and Ahmad Reza Soltani [3,4]

[1] Department of Statistics, Faculty of Science, Fasa University, Fasa, Iran.
[2] Department of Mathematics, University Jaume I of Castellon, Spain.
[3] Department of Statistics and Operations Research, Kuwait University, Kuwait.
[4] Department of Statistics, School of Science, Shiraz University, Shiraz, Iran.

Received: 11/01/2018, Revision received: 31/05/2019, Published online: 29/08/2019

Abstract. This paper introduces a functional mixed effect random model to model spatial data. In this model, the spatial locations form the index set, while the contributing effects to the response variable are set as a linear mixture of fixed and random effects. These fixed and random effects are linear combinations of L^2 functions and random elements, respectively. However, the corresponding linear factors depend on the spatial location variable. Therefore, we develop estimation procedures to estimate the fixed and random coefficients, using spatial functional principal component analysis. Then, we perform prediction by adapting the functional universal kriging method to our model.

Keywords. Functional Principal Components, Karhunen-Loeve Expansions, Spatial Functional Mixed Effect Models, Spatial Functional Random Variable, Universal Kriging.

MSC: 62H11; 86A32.

Corresponding Author: Roya Nasirzadeh (nasirzadeh_roya@yahoo.com)
Jorge Mateu (mateu@mat.uji.es)
Ahmad Reza Soltani (asoltanir@yahoo.com)

1 Introduction

Functional data analysis dealing with random curves has attained a great deal of attention in a wide range of scientific fields. For instance, excellent works on infinite-dimensional data, such as curves, images, and surfaces can be found in textbooks and papers such as Ramsay and Silverman (2002), Ramsay and Silverman (2005), Ferraty and *et al.* (2006) or Cardot *et al.* (2007). Moreover, we can find a wealth of papers dealing with functional random effect models. In this line, Abramovich and Angelini (2006) and Krafty *et al.* (2011) developed a functional random intercept model, while Greven *et al.* (2010) used a functional random effects model (with intercept and slope) for longitudinal random functions. In their papers, they used functional principal component tools and a regression model to estimate the covariance functions. However, these standard functional data techniques are often based on the condition that the functional observations are independent, which is not realistic in many contexts.

Furthermore, in recent years, many scientific studies have collected functional data that exhibit correlation structures amenable to explicit modeling. Such structures may arise from a longitudinal study design (e.g. Morris and Carroll (2006); Greven *et al.* (2010)), crossed designs (e.g. Aston *et al.* (2010)), or spatial sampling of curves (e.g. Delicado *et al.* (2010); Giraldo *et al.* (2010); Giraldo *et al.* (2011); Gromenko *et al.* (2012); Nerini *et al.* (2010); Staicu *et al.* (2010)).

In the same line, Bosq (2000) provided a comprehensive introduction to the theory of functional time series. The introduction contains facts from the theory of probability in Hilbert and Banach spaces and also some basic theory of autoregressive and linear processes in these spaces. One of the basic works on functional time series is to split the time record into natural intervals and use the curve in each interval as a whole observed unit (see, for example, Antoniadis *et al.* (2006)). In a related work, Horvath *et al.* (2010) considered the change point in the covariance operator of the functional autoregressive process. Alternatively, we find some relevant references dealing with regression models for dependent functional data that worth mentioning. Morris and Carroll (2006) proposed wavelet-based functional mixed models, and Scheipl *et al.* (2015) considered a particular class of functional additive mixed models.

Spatial functional data analysis (SFDA) is a new branch of functional statistics that handles spatially correlated functional data. Functional kriging and cokriging are two main objectives in this context. As an application, we can refer to brain scans which are functions over a spatial domain. See Reiss and Ogden (2010), Zipunnikov *et al.* (2011)

and Guillas and Lai (2010).

Our work is motivated by neuroimaging studies containing repeated measurements of spatially correlated functional data. In particular, we develop a linear mixed effect model for spatial data when both the response and explanatory variables are functional. We introduce a functional mixed effect random model for modeling such spatial data.

To highlight the distinctiveness of our model from the preceding models such as Greven *et al.* (2010), two points are noteworthy. Firstly, Greven *et al.*'s model is a random effect one while the model we propose is a mixed effect random model. Secondly, the former model is introduced to handle longitudinal data sets while our model considers spatial data analysis. To this aim, the spatial locations form the index set, while the contributing effects to the response variable are set as a linear mixture of fixed and random effects. Specifically, let

$$\left\{\mathcal{X}_s : \quad s \in \mathbb{D} \subset \mathbb{R}^d\right\}$$

be a spatial functional process. Every $\mathcal{X}_s$ is a random element whose trajectories are real (or complex) functions in $L^2([0,1], dt)$ (Ferraty and *et al.* (2006)). We say that $\mathcal{X}_s$ is second-order integrable if $E\|\mathcal{X}_s\|^2 < \infty$, where the norm is the $L^2([0,1], dt)$ norm.

In practice, a subset of the spatial functional process is observed at a finite set T in $[0,1]$. Precisely, we assume that $\{\mathcal{X}_{s_i}(t_k),\ t_k \in T\}$ with $k = 1, \ldots, m$ and $i = 1, \ldots, n$ are the observations of the spatial functional process observed at the s_i spatial locations and times t_j. We assume that the spatial functional process $\mathcal{X}_s(t)$ satisfies

$$\begin{aligned} \mathcal{X}_s(t) &= \sum_{i=1}^{I} \eta_i(t) g_i(s) + U_1(t) f_1(s) + U_2(t) f_2(s) + \varepsilon_s(t) \qquad (1.1) \\ &= \boldsymbol{\eta}'(t)\mathbf{g}_s + \boldsymbol{U}'(t)\mathbf{f}_s + \varepsilon_s(t), \quad t \in T,\ s \in \mathbb{D}, \qquad (1.2) \end{aligned}$$

where for every i, $\eta_i \in L^2([0,1], dt)$ stands for the ith functional regression coefficient, and $U_j, j = 1,2$ are random elements in $L^2([0,1], dt)$; the factors $g_i(s), i = 1, \ldots, I$ and $f_j(s), j = 1,2$ are real functions on $\mathbb{D}$. It should be noted that $g_i(s), i = 1, \ldots, I$ and $f_j(s), j = 1,2$ are functions of the spatial locations. This highlights that our model is a spatially-dependent model. We refer to $\boldsymbol{\eta}'(t)\mathbf{g}_s$ as the fixed effects term, and to $\boldsymbol{U}'(t)\mathbf{f}_s$ as the random effect term, with $\boldsymbol{\eta}'(t) = \{\eta_1(t), \ldots, \eta_I(t)\}$ and $\boldsymbol{U}(t) = \{U_1(t), U_2(t)\}$. Also, $\mathbf{f}_s = (f_1(s), f_2(s))$ and $\mathbf{g}_s = (g_1(s), \ldots, g_I(s))$. Finally, $\varepsilon_s(t)$ is a white noise measurement error. Note that without loss of generality, we can extend the bivariate random effect part of the model to a multivariate random effect, but for simplicity of notation we use this bivariate case.

Throughout the paper, we assume that $\mathbf{U}$ and $\boldsymbol{\varepsilon}_s$ are uncorrelated zero-mean random elements in $L^2([0,1],dt)$. In addition, we assume that the white noise $\boldsymbol{\varepsilon}_s \in L^2([0,1],dt)$ is homoscedastic, i.e. its variance σ^2 is constant on $[0,1]$. Let $K_{j,j}(.,.)$ denote the auto-covariance function of U_j, $j=1,2$, and $K_{1,2}(.,.)$ be the cross-covariance function between U_1 and U_2. Therefore, $\varepsilon_s(t)$ is both white noise and homoscedastic. Thus

$$cov\left(\varepsilon_{s_i}(t), \varepsilon_{s_j}(t')\right) = \sigma^2 \delta_{i,j}\delta_{t,t'},$$

where $\delta_{.,.}$ stands for the kronecker's delta. We will show that the covariance between the two sites s_i and s_j at points t and t' depends on their spatial locations. In other words, the random part of the model (1.1) causes this model to be a non-stationary spatial model.

In this paper, the aim is to fit a linear mixed-effect model as in (1.1) to spatial functional data with repeated measurements. To estimate the covariance functions we follow Greven *et al.* (2010). On the other side, to predict a functional data at a particular unsampled site, we use a universal kriging approach for non-stationary data as in Caballero *et al.* (2013). The paper is organized as follows. Section 2 is devoted to some essential preliminaries. The estimation technique is given in Section 3, while Section 4 develops the corresponding prediction techniques. A simulation study is presented in Section 5. Finally, the data set of Canadian Temperature is analyzed in Section 6. The paper ends with some final conclusions.

2 Preliminaries

We consider the general expansion technique of the terms in (1.1) and of the associated covariances. Note that the covariance operator $\mathbb{K}(.,.)$ of $\mathbf{U}(.) = (U_1(.), U_2(.))'$ belongs to $L^2([0,1],dt) \times L^2([0,1],dt)$. For $t, t' \in T$, this operator can be written in matrix form as $\mathbb{K}(t,t') = \left[K_{i,j}(t,t')\right]$. The Mercer expansion of the covariance operator at times $t, t' \in T$ of the process $\mathbf{U}(t)$ is then given by

$$\begin{aligned}
\mathbb{K}(t,t') &= \sum_{k=1}^{\infty} \lambda_k \boldsymbol{\phi}_k(t)\boldsymbol{\phi}_k^*(t') \\
&= \begin{bmatrix} \sum_{k=1}^{\infty} \lambda_k \phi_k^1(t)\phi_k^{1^*}(t') & \sum_{k=1}^{\infty} \lambda_k \phi_k^1(t)\phi_k^{2^*}(t') \\ \sum_{k=1}^{\infty} \lambda_k \phi_k^2(t)\phi_k^{1^*}(t') & \sum_{k=1}^{\infty} \lambda_k \phi_k^2(t)\phi_k^{2^*}(t') \end{bmatrix} \\
&= \begin{bmatrix} K_{1,1}(t,t') & K_{1,2}(t,t') \\ K_{2,1}(t,t') & K_{2,2}(t,t') \end{bmatrix}, \qquad (2.1)
\end{aligned}$$

as mentioned by Greven *et al.* (2010) and Ramsay and Silverman (2005), and where $\boldsymbol{\phi}_k(t) = \{\phi_k^1(t), \phi_k^2(t)\}'$ is the k-th eigenfunction of $\mathbb{K}(t, t')$ corresponding to the k-th eigenvalue λ_k, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, and $\boldsymbol{\phi}^*$ is the conjugate transpose. Note that $\left\{\boldsymbol{\phi}_k,\ k \in \mathbb{N}\right\}$ forms an orthonormal basis of $\mathcal{H}^2$, where $\mathcal{H} = L^2([0,1])$, with respect to the additive scalar product

$$\left\langle (h_1, h_2), (h_1^*, h_2^*) \right\rangle = \int_0^1 h_1(u) h_1^*(u) du + \int_0^1 h_2(u) h_2^*(u) du.$$

Now, the Karhunen-Loeve expansion of the random process is given by

$$\boldsymbol{U}(t) = \sum_{k=1}^{\infty} \xi_k \boldsymbol{\phi}_k(t),$$

where the principal component scores (PC scores)

$$\xi_k = \left\langle \mathbf{U}, \boldsymbol{\phi}_k \right\rangle = \int_0^1 U_1(u) \phi_k^1(u) du + \int_0^1 U_2(u) \phi_k^2(u) du,$$

are uncorrelated zero-mean random variables with variances λ_k. In practice, it is often considered the finite-dimensional approximations of the $\mathbf{U}(t)$ process, namely,

$$\mathbf{U}(t) = \sum_{k=1}^{N} \xi_k \boldsymbol{\phi}_k(t).$$

In Section 3, we discuss practical methods to find a suitable N. Now, considering N fixed, the model (1.1) can be written as

$$\mathcal{X}_s(t) = \boldsymbol{\eta}'(t)\mathbf{g}_s + \sum_{k=1}^{N} \xi_k \boldsymbol{\phi}_k'(t)\mathbf{f}_s + \varepsilon_s(t), \quad t \in T, s \in \mathbb{D}. \tag{2.2}$$

3 Estimation

This section presents estimation methods of the parameters of the model in (1.1) or (2.2). In particular, the terms and associated functionals of (1.1) that we estimate in this section are

(i) the unknown functions $\eta_i(t)$, $i = 1, \ldots, I$;

(ii) the bivariate covariance function $\mathbb{K}$;

(iii) the variance σ^2 of the white noise measurement.

We assume that all the Hilbertian random elements and random processes are Gaussian. Noting that $\boldsymbol{\mathcal{Y}} = (\mathcal{X}_{s_1}(t_1), \ldots, \mathcal{X}_{s_1}(t_m), \mathcal{X}_{s_2}(t_1), \ldots, \mathcal{X}_{s_n}(t_m))$ has a Gaussian distribution, we let $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ denote the corresponding mean and variance, and $\mathbb{L}$ the corresponding log-likelihood function. In Theorem 3.1, we show that the information matrix has a block-diagonal form for parameters $\boldsymbol{\eta}$ and $vech(\boldsymbol{\Gamma})$, the latter being the vector of unique elements of matrix $\boldsymbol{\Gamma}$. This fact has a direct effect: we can estimate the mean and the variance parameters separately.

Theorem 3.1. *Under the above assumptions,*

$$\begin{aligned} \frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}} &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right) \boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu}), \\ \frac{\partial \mathbb{L}}{\partial \boldsymbol{\Gamma}} &= -0.5\boldsymbol{\Gamma}^{-1} + 0.5\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu})(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu})^{'}\boldsymbol{\Gamma}^{-1}, \end{aligned}$$

and the information matrix for parameters $\boldsymbol{\eta}$ and $vech(\boldsymbol{\Gamma})$ is given by

$$\begin{aligned} \mathcal{I} &= \begin{bmatrix} cov(\frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}, \frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}) & cov(\frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}, \frac{\partial \mathbb{L}}{\partial vech(\boldsymbol{\Gamma})}) \\ cov(\frac{\partial \mathbb{L}}{\partial vech(\boldsymbol{\Gamma})}, \frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}) & cov(\frac{\partial \mathbb{L}}{\partial vech(\boldsymbol{\Gamma})}, \frac{\partial \mathbb{L}}{\partial vech(\boldsymbol{\Gamma})}) \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)^{'} & 0 \\ 0 & 0.5\mathcal{D}^{+}(\boldsymbol{\Gamma}^{-1} \otimes \boldsymbol{\Gamma}^{-1})\mathcal{D}^{+'} \end{bmatrix}, \end{aligned}$$

where $\otimes$ denotes the Kronecker product of matrices, $vec(\boldsymbol{\Gamma})$ is a column vector obtained by stacking the column of the matrix $\boldsymbol{\Gamma}$ on top of each other, $vech(\boldsymbol{\Gamma})$ denotes the vector of unique elements of the symmetric matrix $\boldsymbol{\Gamma}$, and $\mathcal{D}$ is a full rank matrix that is called duplication matrix, where $vec(\boldsymbol{\Gamma}) = \mathcal{D}vech(\boldsymbol{\Gamma})$. Based on $\mathcal{D}$, $vech(\boldsymbol{\Gamma}) = \mathcal{D}^{+}vec(\boldsymbol{\Gamma})$ where $\mathcal{D}^{+}$ is the generalized inverse matrix $\mathcal{D}^{+} = (\mathcal{D}^{'}\mathcal{D})^{-1}\mathcal{D}^{'}$. (For more details about vec and vech, see Demidenko (2004).)

Proof. The vector of all realizations $\boldsymbol{\mathcal{Y}}$ has Gaussian distribution, thus

$$\mathbb{L} = -\frac{nm}{2}log(2\pi) - \frac{1}{2}log|\boldsymbol{\Gamma}| - \frac{1}{2}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu}),$$

and using matrix differentiation properties

$$\begin{aligned}\frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}} &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu}),\\ \frac{\partial \mathbb{L}}{\partial \boldsymbol{\Gamma}} &= -0.5\boldsymbol{\Gamma}^{-1}+0.5\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu})(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu})^{'}\boldsymbol{\Gamma}^{-1}.\end{aligned}$$

The information matrix for parameters $\boldsymbol{\eta}$ and $vech(\boldsymbol{\Gamma})$ is also given by

$$\begin{aligned}\mathcal{I} &= \begin{bmatrix} cov(\frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}) & cov(\frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}, \frac{\partial \mathbb{L}}{\partial vech(\boldsymbol{\Gamma})}) \\ cov(\frac{\partial \mathbb{L}}{\partial vech(\boldsymbol{\Gamma})}, \frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}) & cov(\frac{\partial \mathbb{L}}{\partial vech(\boldsymbol{\Gamma})}) \end{bmatrix}\\ &= \begin{bmatrix} \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)^{'} & 0 \\ 0 & 0.5\mathcal{D}^{+}(\boldsymbol{\Gamma}^{-1}\otimes\boldsymbol{\Gamma}^{-1})\mathcal{D}^{+'} \end{bmatrix},\end{aligned}$$

since

$$\begin{aligned}cov(\frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}) &= cov(\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu}), \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu}))\\ &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}cov(\boldsymbol{\mathcal{Y}},\boldsymbol{\mathcal{Y}})\boldsymbol{\Gamma}^{-1}\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)^{'}\\ &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-1}\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)^{'} = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)^{'},\end{aligned}$$

$$\begin{aligned}cov(\frac{\partial \mathbb{L}}{\partial \boldsymbol{\eta}}, \frac{\partial \mathbb{L}}{\partial \boldsymbol{\Gamma}}) &= cov(\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu}), 0.5\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu})(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu})^{'}\boldsymbol{\Gamma}^{-1})\\ &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right)\boldsymbol{\Gamma}^{-1}cov(\boldsymbol{\mathcal{Y}}, 0.5\boldsymbol{\Gamma}^{-1}(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu})(\boldsymbol{\mathcal{Y}}-\boldsymbol{\mu})^{'}\boldsymbol{\Gamma}^{-1}) = 0,\end{aligned}$$

since the third moment of the normal distribution is zero. Let $\mathbf{T} = \boldsymbol{\Gamma}^{-\frac{1}{2}}$ and $\mathbf{e} =$

$\mathbf{\Gamma}^{-\frac{1}{2}}(\mathbf{X}-\mu)$. Then,

$$\begin{aligned}
cov(vech(0.5\mathbf{\Gamma}^{-1}(\mathcal{Y}-\mu)(\mathcal{Y}-\mu)'\mathbf{\Gamma}^{-1})) &= cov(vech(0.5(\mathbf{Te})(\mathbf{Te})'))\\
&= \mathcal{D}^{+}cov(0.5(\mathbf{Te})(\mathbf{Te})')\mathcal{D}^{+'}\\
&= 0.25\mathcal{D}^{+}cov((\mathbf{Te})\otimes(\mathbf{Te}))\mathcal{D}^{+'}\\
&= 0.25\mathcal{D}^{+}cov((\mathbf{T}\otimes\mathbf{T})(\mathbf{e}\otimes\mathbf{e}))\mathcal{D}^{+'}\\
&= 0.25\mathcal{D}^{+}(\mathbf{T}\otimes\mathbf{T})cov(\mathbf{e}\otimes\mathbf{e})(\mathbf{T}'\otimes\mathbf{T}')\mathcal{D}^{+'}\\
&= 0.5\mathcal{D}^{+}(\mathbf{T}\otimes\mathbf{T})\mathbf{N}_{nm}(\mathbf{T}'\otimes\mathbf{T}')\mathcal{D}^{+'}\\
&= 0.5\mathcal{D}^{+}(\mathbf{T}\otimes\mathbf{T})(\mathbf{T}'\otimes\mathbf{T}')\mathbf{N}_{nm}\mathcal{D}^{+'}\\
&= 0.5\mathcal{D}^{+}(\mathbf{T}\otimes\mathbf{T})(\mathbf{T}'\otimes\mathbf{T}')\mathcal{D}^{+'}\\
&= 0.5\mathcal{D}^{+}(\mathbf{TT}'\otimes\mathbf{TT}')\mathcal{D}^{+'}\\
&= 0.5\mathcal{D}^{+}(\mathbf{\Gamma}^{-1}\otimes\mathbf{\Gamma}^{-1})\mathcal{D}^{+'}.
\end{aligned}$$

□

As the information matrix has a block-diagonal form, we can maximize the likelihood function over the mean and the variance parameters separately. Our estimation procedure can be outlined as follows

Step 1 Use the ordinary least square method to estimate the vector of functional regression coefficients $\boldsymbol{\eta}(t) = (\eta_1(t), \ldots, \eta_I(t))'$ as it was discussed earlier.

Step 2 Safely assume that the fixed effects part of the model is zero, otherwise it suffices to consider

$$\mathcal{X}_s(t) := \mathcal{X}_s(t) - \hat{\boldsymbol{\eta}}'(t)\mathbf{g}_s,$$

and then estimate the covariance function $\mathbb{K}$ (including auto-covariances and cross-covariances) and σ^2, as explained in Section 3.2. In the process we need to estimate the eigenvalues and eigenfunctions and the principal component scores (as detailed in Section 3.3).

Step 3 Re-estimate the fixed effects part of the model implementing a generalized least square method, where the weight is $(\hat{\mathbb{K}} + \hat{\sigma}^2\mathbf{I})^{-1}$.

Step 4 Return to step 2 if the convergence criterion is not met.

Step 5 Perform prediction at the site s_0 by evaluating $\mathcal{X}_{s_0}$ (see details in Section 4).

3.1 Estimation of the Covariance Function and σ^2

To estimate σ^2 and $\mathbb{K}$, Greven *et al.* (2010) suggested a method for finding the bivariate covariance function in a longitudinal random effect model when the subjects or clusters are independent. We extend their idea to our spatial model. Let $C_{i,j}(t,t') = Cov(\mathcal{X}_{s_i}(t), \mathcal{X}_{s_j}(t'))$ be the covariance between two sites s_i and s_j at points t and t'. Now, for all sites s_i and s_j at points t and t'

$$\begin{aligned} C_{i,j}(t,t') &= Cov(\mathcal{X}_{s_i}(t), \mathcal{X}_{s_j}(t')) \\ &= Cov(\mathbf{f}'_{s_i}\mathbf{U}(t) + \varepsilon_{s_i}(t), \mathbf{f}'_{s_j}\mathbf{U}(t') + \varepsilon_{s_j}(t')) \\ &= \mathbf{f}'_{s_i} Cov(\mathbf{U}(t), \mathbf{U}(t'))\mathbf{f}_{s_j} + Cov(\varepsilon_{s_i}(t), \varepsilon_{s_j}(t')) \\ &= \mathbf{f}'_{s_i}\mathbb{K}(t,t')\mathbf{f}_{s_j} + \sigma^2\delta_{t,t'}\delta_{s_i,s_j}, \end{aligned} \tag{3.1}$$

where δ_{s_i,s_j} is again the Kronecker's delta.

We can assume, for the moment, that $\mathcal{X}_s(t) := \mathcal{X}_s(t) - \hat{\boldsymbol{\eta}}'(t)\mathbf{g}_s$, therefore $\mathcal{X}_s(t)$ has mean zero. An estimation for the covariance function between the two sites s_i and s_j at points t and t' is given by $\hat{C}_{i,j}(t,t') = \mathcal{X}_{s_i}(t)\mathcal{X}_{s_j}(t')$. Under the settings given in the introduction and (1.1), a straightforward solution to estimate the covariance function is by using linear regression with the inputs of $\{k_{1,1}(t,t'), k_{1,2}(t,t'), k_{1,2}(t',t), k_{2,2}(t,t')\}$, $\{\mathcal{X}_{s_i}(t)\mathcal{X}_{s_j}(t'), i,j = 1,\ldots,n\}$ and $\{\mathbf{f}(s_i)\mathbf{f}(s_i), i,j = 1,\ldots,n\}$. We can also estimate the corresponding diagonal elements of the covariance function through linear regression over $\{k_{1,1}(t_k,t_k), k_{1,2}(t_k,t_k), k_{1,2}(t_k,t_k), k_{2,2}(t_k,t_k), k = 1,\ldots,m,\sigma^2\}$, $\{\mathcal{X}_{s_i}(t_k)\ \mathcal{X}_{s_j}(t_k), i,j = 1,\ldots,n, k = 1,\ldots,m\}$ and the block diagonal of $\{\mathbf{f}(s_i)\otimes\mathbf{f}(s_j), \delta_{s_i,s_j}, i,j = 1,\ldots,n\}$. The covariance functions are symmetric, so $k_{1,1}(t,t') = k_{1,1}(t',t)$, $k_{2,2}(t,t') = k_{2,2}(t',t)$ and $k_{1,2}(t,t') = k_{2,1}(t',t)$, for $t' > t$. Finally, we use a bivariate smoothing technique in t and t' over $\{\hat{k}_{1,1}(t,t'), \hat{k}_{1,2}(t,t'), \hat{k}_{1,2}(t',t), \hat{k}_{2,2}(t,t')\}$ to estimate the covariance functions (see Green and Silverman (1994) and Krivobokova *et al.* (2007)).

3.2 Eigenvalue, Eigenvector, and PC Scores Estimation

Having obtained an estimation of the covariance matrices $\hat{K}_{ij} = \{\hat{K}_{i,j}(t,t')\}$, $t,t' \in \{t_1,...,t_m\}, i,j = 1,2$, we now proceed to estimate the eigenvalues and eigenfunctions by spectral decomposition of the covariance functions. Following the expansion given in (2.1), and considering the finite-dimensional approximation and the spectral decomposition, we can write the estimated covariance matrix $\hat{\mathbb{K}}$ as $\hat{\mathbb{K}} = \sum \hat{\lambda}_k\hat{\boldsymbol{\phi}}_k\hat{\boldsymbol{\phi}}_k^*$, where $\hat{\lambda}_k$ and

$\hat{\boldsymbol{\phi}}_k = \left\{\hat{\phi}_k^1, \hat{\phi}_k^2\right\}'$, are the k-th eigenvalue and the k-th orthonormal vector, respectively. Now note that we can rewrite our model in a more general form as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{\theta} + \mathbf{e}, \tag{3.2}$$

where $\boldsymbol{\theta} = (\xi_{11}, ..., \xi_{1N}, ..., \xi_{n1}, ..., \xi_{nN})'$ is the vector of principal component scores and $\boldsymbol{W}$ is an $nm \times nN$ block diagonal matrix, where the i-th diagonal block for $i = 1, ..., n$ is

$$\begin{bmatrix} \hat{\phi}_1(t_1)f_{s_i} & \dots & \hat{\phi}_N(t_1)f_{s_i} \\ \hat{\phi}_1(t_2)f_{s_i} & \dots & \hat{\phi}_N(t_2)f_{s_i} \\ \vdots & \ddots & \vdots \\ \hat{\phi}_1(t_m)f_{s_i} & \dots & \hat{\phi}_N(t_m)f_{s_i} \end{bmatrix}.$$

Under the above formulations related to model (3.2), we make use of the following theorem to estimate the principal component score $\boldsymbol{\theta}$.

Theorem 3.2. *Let*

$$var\begin{pmatrix} \boldsymbol{\theta} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \boldsymbol{H} & 0 \\ 0 & \boldsymbol{R} \end{pmatrix},$$

where $\boldsymbol{H} = \mathbf{I}_{n\times n} \otimes diag\{\lambda_1, ..., \lambda_N\}$ *and* $\boldsymbol{R} = \sigma^2\mathbf{I}_{nm\times nm}$ *are dispersion matrices for the random effect* $\boldsymbol{\theta}$ *and the error effect* $\mathbf{e}$*, respectively. Clearly,* $\boldsymbol{H}$ *and* $\boldsymbol{R}$ *are non-singular and positive definite matrices. Also, let* $\boldsymbol{V} = \boldsymbol{W}\boldsymbol{H}\boldsymbol{W}' + \boldsymbol{R}$*, then the best linear unbiased predictor (BLUP) for* $\boldsymbol{\theta}$ *is given by*

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \boldsymbol{H}\boldsymbol{W}'\boldsymbol{V}^{-1}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu}) \\ &= (\boldsymbol{W}'\boldsymbol{R}^{-1}\boldsymbol{W} + \boldsymbol{H}^{-1})^{-1}\boldsymbol{W}'\boldsymbol{R}^{-1}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu}). \end{aligned}$$

Proof. The proof is obtained using the results suggested by Henderson *et al.* (1959), Greven *et al.* (2010) and Yao *et al.* (2005). □

Concerning the question of the number of eigenfunctions, N, one possible method is to use an AIC-type criterion (Yao *et al.* , 2005). An alternative method is to choose the number of components with a leave-one-curve-out cross-validation technique (Rice and Silverman , 1991). Greven *et al.* (2010) and Soltani *et al.* (2017) proposed the Fraction of Variation Explained (FVE) as a method to select the random effects in linear mixed models. To be more precise, FVE states that when the eigenvalues are in decreasing

order, then the minimum number N satisfying

$$\frac{\sum_{i=1}^{N} \lambda_i + \sigma^2}{\sum_{i=1}^{\infty} \lambda_i + \sigma^2} > \alpha,$$

gives the sufficient number of significant factors, where α is a positive number less than one but close to it.

3.3 Estimation of the Fixed Effect Term

We consider here the estimation of the parameters in the fixed effects part of the model (1.1). In particular, we need to estimate each of the I functional parameters $\eta_i(t)$ with $i = 1, ..., I$. To find a smooth estimate of the mean function of this model, i.e. the total fixed effect population mean surface given by $\mu(t, s) = \boldsymbol{\eta}'(t)\mathbf{g}_s$, we can follow Krivobokova *et al.* (2007) and Lin and Carroll (2000). We can use this estimator for step 1 of the algorithm. One basic method to perform the estimation of fixed effect parameters in step 3 is the weighted least square estimation method, where the weighted matrix is the inverse of the covariance function of the random effect part shown in Section 3.1.

Now, we suggest an alternative fixed effects method to estimate the functional fixed effect parameters, η_i, $i = 1, \ldots, I$. In this method, we consider the fixed effects model as an alternative to model (1.1). We consider the bivariate random functional $\boldsymbol{U}$ as fixed, but unknown nuisance functional parameter. With this assumption, the only random variable is ε_{s_i}, which is a zero-mean Gaussian variable with variance σ^2. So, the ordinary least square estimator is equivalent to

$$\min_{\boldsymbol{\eta}} \left\| \sum_s \boldsymbol{\mathcal{Y}}_s - \boldsymbol{\eta}'(t)\mathbf{g}_s - \sum_k \hat{\xi}_{s_k} \hat{\boldsymbol{\phi}}_k \mathbf{f}_s \right\|^2,$$

where $\boldsymbol{\mathcal{Y}}_s = (\mathcal{X}_s(t_1), ..., \mathcal{X}_s(t_M))$. Also, $\hat{\boldsymbol{\phi}}_k$ and $\hat{\xi}_{s_k}$ are the estimators of $\boldsymbol{\phi_k}$ and the best linear unbiased predictor of ξ_{s_k}, respectively, which are already obtained in Sections (3.1) and (3.2).

4 Prediction

The basic idea of kriging is to predict the functional values at an unobserved site s_0 with the information given in the observed sites $s_1, ..., s_n$ and under the structure given by the linear spatial mixed effect model (1.1). Also, assume we have estimated the unknown parameters, $\eta_i(t)$, $i = 1, ..., I$, $\mathbb{K}$ and σ^2. Assume we are interested in predicting the value of the functional process $\mathcal{X}_{s_0}(t)$ at the unobserved location s_0. An unbiased predictor will be a linear weighted function of the observed functional process at other sites (following the model (1.1) and providing a minimum variance of the prediction error).

Let $\mathcal{Z}_{\mathbf{s}}(t) = (\mathcal{X}_{s_1}(t), ..., \mathcal{X}_{s_n}(t))'$ be the vector of observations, and consider the model (1.1) in the matrix form

$$\mathcal{Z}_{\mathbf{s}}(t) = X_1(\mathbf{s})\boldsymbol{\eta}(t) + X_2(\mathbf{s})\boldsymbol{U}(t) + \boldsymbol{\varepsilon}_{\mathbf{s}}(t), \quad t \in T, \tag{4.1}$$

where $X_1(\mathbf{s}) = [g_{s_k,i}]$ and $X_2(\mathbf{s}) = [f_{s_k,j}]$ are two matrices of size $n \times I$ and $n \times 2$, $1 \leq k \leq n$, $1 \leq i \leq I$, $j = 1, 2$. Also $\boldsymbol{\varepsilon}_{\mathbf{s}}(t) = (\varepsilon_{s_1}(t), ..., \varepsilon_{s_n}(t))'$ represents a vector of functional errors. Therefore $\mathbb{E}\boldsymbol{\varepsilon}_{\mathbf{s}}(t) = 0$ and $\mathbb{E}\mathcal{Z}_{\mathbf{s}}(t) = X_1(\mathbf{s})\boldsymbol{\eta}(t)$.

Clearly, at s_0, and following (1.1) and (1.2) we have

$$\mathcal{X}_{s_0}(t) = \boldsymbol{\eta}'(t)\mathbf{g}_{s_0} + \boldsymbol{U}'(t)\mathbf{f}_{s_0} + \varepsilon_{s_0}(t),$$

where, following (1.1), $\mathbf{g}_{s_0} = (g_1(s_0), \ldots, g_I(s_0))'$ and $\mathbf{f}_{s_0} = (f_1(s_0), f_2(s_0))'$. A functional universal kriging predictor (Caballero *et al.* , 2013) for $\mathcal{X}_{s_0}(t)$ is given by

$$\hat{\mathcal{X}}_{s_0}(t) = \boldsymbol{\lambda}'\mathcal{Z}_{\mathbf{s}}(t),$$

where the components of $\boldsymbol{\lambda}$ are real numbers obtained such that $\hat{\mathcal{X}}_{s_0}(t)$ is an unbiased estimator of $\mathcal{X}_{s_0}(t)$. This will be the case if $\mathbb{E}\hat{\mathcal{X}}_{s_0}(t) = \mathbb{E}\mathcal{X}_{s_0}(t)$, $t \in T$, i.e. if

$$\boldsymbol{\lambda}' X_1(\mathbf{s})\boldsymbol{\eta}(t) = \boldsymbol{\eta}'(t)\mathbf{g}_{s_0},$$

or equivalently

$$\boldsymbol{\lambda}' X_1(\mathbf{s}) = \mathbf{g}'_{s_0}.$$

Finally, $\boldsymbol{\lambda}$ is chosen by minimizing the mean square prediction error, i.e.

$$\min_{\boldsymbol{\lambda}} MSE(s_0) = \min_{\boldsymbol{\lambda}} \int Var(\hat{\mathcal{X}}_{s_0}(t) - \mathcal{X}_{s_0}(t))dt.$$

Note that we can write $Var(\hat{\mathcal{X}}_{s_0}(t) - \mathcal{X}_{s_0}(t))$ in a simpler form

$$\begin{aligned} Var(\hat{\mathcal{X}}_{s_0}(t) - \mathcal{X}_{s_0}(t)) &= Var(\boldsymbol{\lambda}'\mathcal{Z}_{\mathbf{s}}(t) - \mathcal{X}_{s_0}(t)) \\ &= Var(\boldsymbol{\lambda}'\mathcal{Z}_{\mathbf{s}}(t)) + Var(\mathcal{X}_{s_0}(t)) - 2Cov(\boldsymbol{\lambda}'\mathcal{Z}_{\mathbf{s}}(t), \mathcal{X}_{s_0}(t)) \\ &= \boldsymbol{\lambda}' Var(\mathcal{Z}_{\mathbf{s}}(t))\boldsymbol{\lambda} + Var(\mathcal{X}_{s_0}(t)) - 2\boldsymbol{\lambda}' Cov(\mathcal{Z}_{\mathbf{s}}(t), \mathcal{X}_{s_0}(t)), \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\Sigma} = Var(\mathcal{Z}_{\mathbf{s}}(t)) &= Var(X_1(\mathbf{s})\boldsymbol{\eta}(t) + X_2(\mathbf{s})\mathbf{U}(t) + \boldsymbol{\varepsilon}_{\mathbf{s}}(t)) \\ &= X_2(\mathbf{s}) Var\mathbf{U}(t) X_2'(\mathbf{s}) + Var(\boldsymbol{\varepsilon}_{\mathbf{s}}(t)) \\ &= X_2(\mathbf{s})\mathbb{K}(t,t)X_2'(\mathbf{s}) + \sigma^2\mathbf{1}, \end{aligned}$$

and

$$\begin{aligned} \sigma_0^2(t) = Var(\mathcal{X}_{s_0}(t)) &= Var(\boldsymbol{\eta}'(t)\mathbf{g}_{s_0} + \mathbf{U}'(t)\mathbf{f}_{s_0} + \varepsilon_{s_0}(t)) \\ &= \mathbf{f}_{s_0}' Var\mathbf{U}(t)\mathbf{f}_{s_0} + \sigma^2 \\ &= \mathbf{f}_{s_0}'\mathbb{K}(t,t)\mathbf{f}_{s_0} + \sigma^2. \end{aligned}$$

Note also that

$$\mathbf{c}(t) = Cov(\mathcal{Z}_{\mathbf{s}}(t), \mathcal{X}_{s_0}(t)) = [Cov(\mathcal{X}_{s_1}(t), \mathcal{X}_{s_0}(t)), \ldots, Cov(\mathcal{X}_{s_n}(t), \mathcal{X}_{s_0}(t))]',$$

where for each $i = 1, ..., n$

$$\begin{aligned} Cov(\mathcal{X}_{s_i}(t), \mathcal{X}_{s_0}(t)) &= Cov(\boldsymbol{\eta}'(t)\mathbf{g}_{s_i} + \mathbf{U}'(t)\mathbf{f}_{s_i} + \varepsilon_{s_i}(t), \boldsymbol{\eta}'(t)\mathbf{g}_{s_0} + \mathbf{U}'(t)\mathbf{f}_{s_0} + \varepsilon_{s_0}(t)) \\ &= \mathbf{f}_{s_i}'\mathbb{K}(t,t)\mathbf{f}_{s_0}. \end{aligned}$$

Now, let us consider the function

$$\phi(\boldsymbol{\lambda}, \mathbf{v}) = \int_{\Omega} (\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda})(t)dt - 2\int_{\Omega} \boldsymbol{\lambda}'\mathbf{c}(t)dt + \int_{\Omega} \sigma_0^2(t)dt - 2\left[\boldsymbol{\lambda}'\mathbf{X}_1(s) - \mathbf{g}'(s_0)\right]\mathbf{v}, \tag{4.2}$$

where $\mathbf{v}$ is the vector of Lagrange multipliers for the condition of unbiasedness. The minimum of ϕ is found in those values of $\boldsymbol{\lambda}$ and $\mathbf{v}$ for which the partial derivatives $\dfrac{\partial\phi}{\partial\boldsymbol{\lambda}}$ and $\dfrac{\partial\phi}{\partial\mathbf{v}}$ are zero. As a specific case of Caballero *et al.* (2013), λ and $\hat{\chi}_{s_0}(t)$ are achieved as

$$\boldsymbol{\lambda} = \left[\int_{\Omega} \boldsymbol{\Sigma}dt\right]^{-1}\left[\int_{\Omega} \mathbf{c}(t)dt\right] + \left[\int_{\Omega} \boldsymbol{\Sigma}dt\right]^{-1} \mathbf{X}_1(s)\mathbf{W}_1^{-1}\left\{\mathbf{g}(s_0) - \mathbf{W}_2\right\},$$

and

$$\hat{\chi}_{s_0}(t) = \left\{ \left[\int_\Omega \mathbf{c}'(t)dt \right] \left[\int_\Omega \mathbf{\Sigma} dt \right]^{-1} + (\mathbf{g}(s_0) - \mathbf{W}_2)' \mathbf{W}_1^{-1} \mathbf{X}_1'(s) \left[\int_\Omega \mathbf{\Sigma} dt \right]^{-1} \right\} \mathcal{Z}_s(t),$$

where $\mathbf{W}_1 = \mathbf{X}_1'(s) \left[\int_\Omega \mathbf{\Sigma} dt \right]^{-1} \mathbf{X}_1(s)$ and $\mathbf{W}_2 = \mathbf{X}_1'(s) \left[\int_\Omega \mathbf{\Sigma} dt \right]^{-1} \int_\Omega \mathbf{c}(t)dt$.

5 Simulation Studies

This section includes two subsections. In Subsection 5.1 a simulation study is presented to examine the estimation procedure introduced in Section 3. The second simulation study in Subsection 5.2 takes into consideration the feasibility of the prediction method mentioned in Section 4.

5.1 Parameters Estimation Procedure

In this subsection, we conduct a simulation study to analyze the estimation procedure presented in Section 3. In particular, we simulated $L = 1000$ data sets from the spatial functional mixed effects model

$$\mathcal{X}_s(t) = \sin(\pi t)x^2 + \cos(\pi t)xy + U_1(t)x + U_2(t)y + \varepsilon_s(t), t \in [0,1], \tag{5.1}$$

where the spatial location $s = (x, y)$ is a pair of integers in the rectangular window which is scaled to $[-4,4] \times [-4,4]$, the fixed effects parameters are $\beta_0(t) = \sin(\pi t)$ and $\beta_1(t) = \cos(\pi t)$. We let $var(\varepsilon_s(t)) = \sigma^2 = 0.05$, $N = 4$ and $\lambda_k = 2^{1-k}$, $k = 1, \ldots, 4$. We generated all the random functions in points $\{\frac{t-0.5}{D}, t = 1, \ldots, 40\} \subset [0,1]$. Let the PC scores have normal distributions, i.e, $\xi_{s,k} \sim \mathcal{N}(0, \lambda_k)$, for all $s \in [-4,4] \times [-4,4]$ and $k = 1, \ldots, 4$. Finally, let the eigenfunctions be

$$\begin{array}{ll} \phi_1^1(t) = \sin(2\pi t), & \phi_1^2(t) = \sqrt{2}/2, \\ \phi_2^1(t) = \cos(2\pi t), & \phi_2^2(t) = \sin(6\pi t), \\ \phi_3^1(t) = \sin(4\pi t), & \phi_3^2(t) = \cos(6\pi t), \\ \phi_4^1(t) = \cos(4\pi t), & \phi_4^2(t) = \sin(8\pi t), \quad t \in (0,1). \end{array}$$

In each simulation iteration $\hat{U}(t) = \sum_{i=1}^{4} \hat{\xi}_k \boldsymbol{\phi}_k$, where $\hat{\xi}_k$ is a number to be simulated based on $N(0, \lambda_k)$.

The mean estimation of σ^2 is $\hat{\sigma}^2 = 0.0947$. Table 1 shows the summary statistics and Figure 1 depicts the box plot of the scaled differences between the estimated and the true scores $\frac{\hat{\xi}_{ik} - \xi_{ik}}{\sqrt{\lambda_k}}$, $k = 1, 2, 3, 4$. As can be noted from Table 1 and Figure 1, the scaled differences between the estimated and the true scores have a standard normal distribution.

Table 1: The minimum, first and third quartiles, mean, median and maximum of the scaled differences between the estimated and the true scores in Subsection 5.1.

	min	1 th quartile	mean	median	3 th quartile	max
$\frac{\hat{\xi}_{s,1} - \xi_{s,1}}{\sqrt{\lambda_1}}$	−2.8508	−0.5037	−0.0052	−0.0121	0.4991	3.1330
$\frac{\hat{\xi}_{s,2} - \xi_{s,2}}{\sqrt{\lambda_2}}$	−2.9203	−0.4990	−0.0014	−0.0007	0.4866	3.5097
$\frac{\hat{\xi}_{s,3} - \xi_{s,3}}{\sqrt{\lambda_3}}$	−3.2004	−0.5417	0.0020	−0.0074	0.5407	3.7634
$\frac{\hat{\xi}_{s,4} - \xi_{s,4}}{\sqrt{\lambda_4}}$	−3.4066	−0.5723	−0.0151	−0.0161	0.5441	3.9189

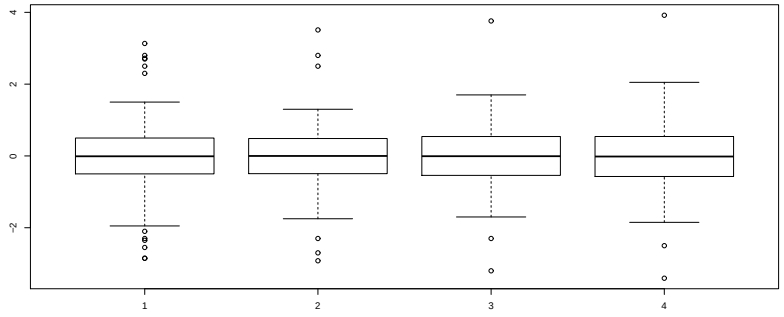

Figure 1: The boxplot of the scaled differences between the estimated and the true scores in Subsection 5.1.

Figure 2 shows the true, the estimated and a 90% confidence interval for the functional parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$. Here, the black line shows the mean of the estimated functions, the dash-dotted lines illustrates the true function, and the dashed lines depict the pointwise 5th and 95th percentiles of the estimated functions. We note that the mean and the true functions are close to each other and the true functions are located within the confidence limits. Also, the MSE is given by

$$\frac{1}{L}\sum_{l=1}^{L}\int(\hat{\beta}_{l,k}(t)-\beta_k(t))^2dt, \quad k=0,1,$$

where L is the number of iterations and $\hat{\beta}_{l,k}$ is the estimate of β_k for the l-th iteration, calculated for each estimation of β_0 and β_1 giving 0.0008 and 0.0012, respectively. The MSE indicates that both β_0 and β_1 were accurately estimated.

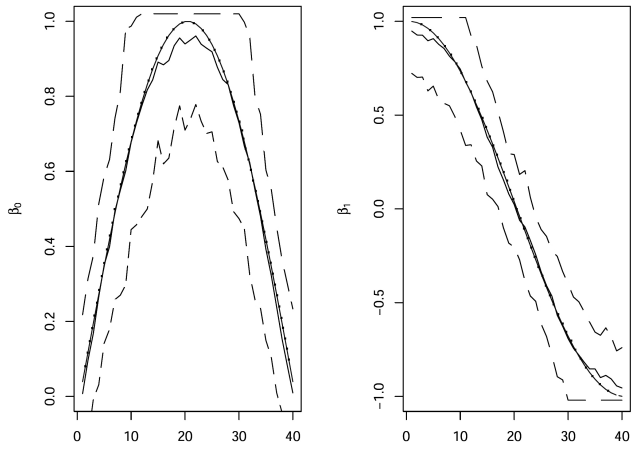

Figure 2: Parameter estimation for the fixed effects part in Subsection 5.1.

Figure 3 shows the true, the estimated and a 90% confidence interval for $\boldsymbol{\phi}_k^0$, $\boldsymbol{\phi}_k^1$, $k=1,\ldots,4$. The black, dash-dotted and dashed lines are similar to what we mentioned for the fixed effects in Figure 2. We note that the estimated functions are very close to the true functions and the confidence band covers the true functions.

Moreover, it is noteworthy to mention that $\hat{K}$ is a high dimensional matrix. Therefore, it is not usual to estimate K, instead, the estimation of the eigenvalues and

eigenfunctions were employed, noting the result according to Mercer Lemma (2.1), $\mathbb{K}(t,t') = \sum_{k=1}^{\infty} \lambda_k \boldsymbol{\phi}_k(t)\boldsymbol{\phi}_k^*(t')$.

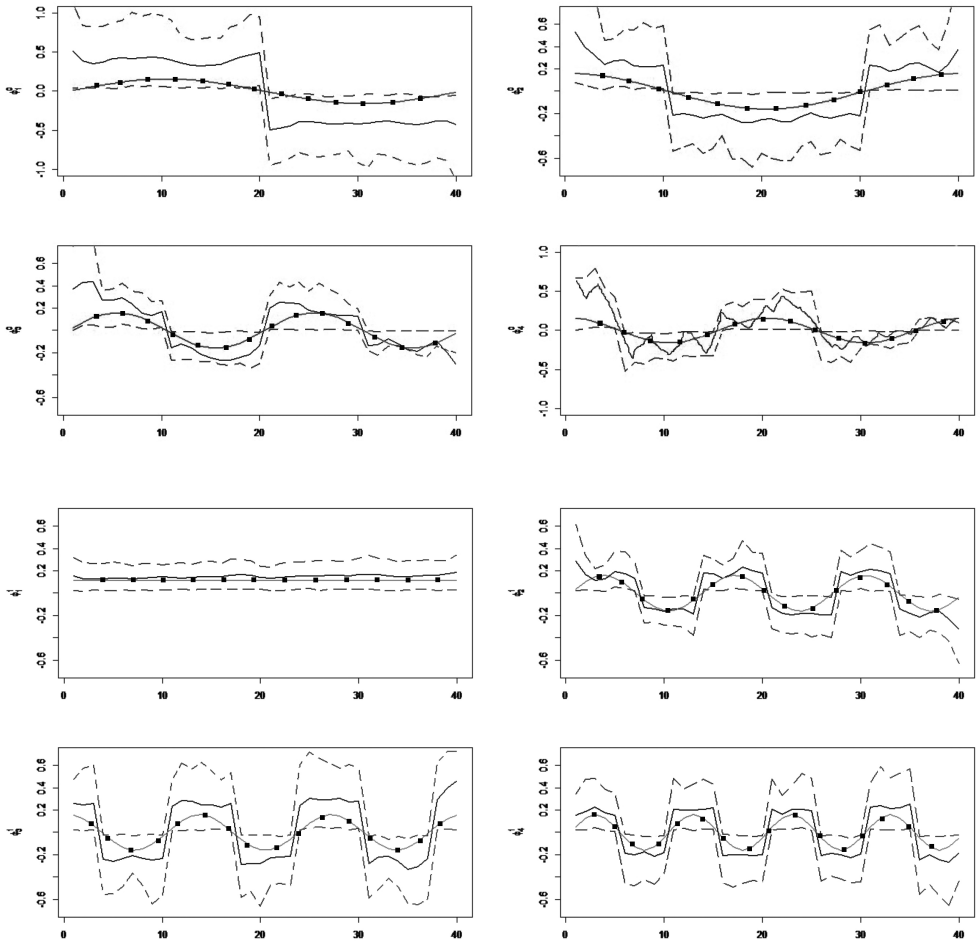

Figure 3: Estimation of the eigenvectors in Subsection 5.1.

5.2 Prediction

In this subsection, we generated a spatial functional $\hat{\mathcal{X}}_{s_0}$ at site $s_0 = (5, 1)$ using the model (5.1). We then generated the functional mixed effect (5.1) for all spatial locations $s = (x, y)$ in the rectangular window $[-4, 4] \times [-4, 4]$ for 100 times. In each iteration, $\mathcal{X}_{s_0}$ was predicted using the suggested method in Section 4. The dash-dotted line in Figure 4 shows $\hat{\mathcal{X}}_{s_0}$. The black line is the mean of the predicted functions, and the dashed lines are the pointwise 5th and 95th percentiles of the predicted functions. As can be seen, the mean and $\hat{\mathcal{X}}_{s_0}$ are very close to each other. Also, the confidence lines cover $\hat{\mathcal{X}}_{s_0}$, which justifies that our proposed prediction method works well.

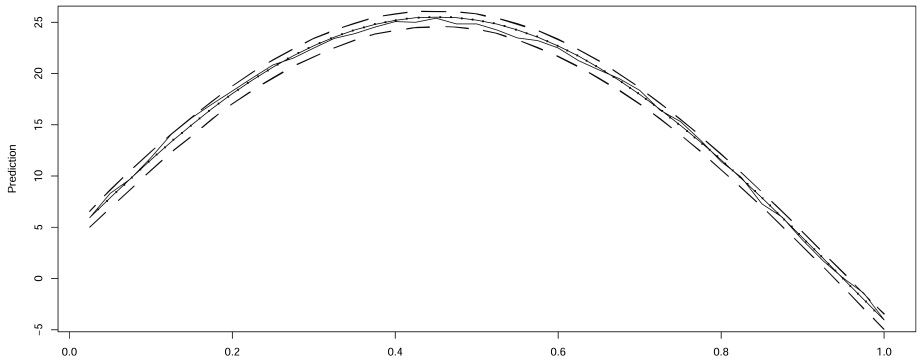

Figure 4: Prediction of the new site $s_0 = (5, 1)$ in Subsection 5.2.

6 Real Data Analysis

To analyze a real data set we use the Canadian temperature data set which is a well known data set in FDA consisting of mean daily temperatures for 35 Canadian weather stations. These temperatures of the weather stations were in fda package in R and Ramsay and Silverman's home page (http://www.functionaldata.org) and the geographical coordinates are in Canadian weather office homepage (http://www.climate.weatheroffice .ec.gc.ca/climateData). Since Canada is a very widespread country, modeling the Canadian temperature is interesting.

We consider the model

$$\begin{aligned} \mathcal{X}_s(t) &= \beta_0(t) + \beta_1(t)x + \beta_2(t)y + \beta_3(t)x^2 + \beta_4(t)xy + \beta_5(t)y^2 \\ &+ U_1(t)x + U_2(t)y + \varepsilon_s(t), \end{aligned}$$

where x and y represent the normalized geographical coordinates of the weather stations and $\beta_i, i = 0, \ldots, 5$ are the functional parameters. This data set has been analyzed by numerous authors. Both Ramsay and Silverman (2005) and Giraldo *et al.* (2010) suggested a Fourier basis with 65 functions to smooth the temperatures. We used a Fourier basis with 65 functions to estimate the $\beta_i, i = 0, \ldots, 5$, too. Figure 5 shows the estimation of these functional parameters.

Just one principal component of the covariance function shows 82.67% of explained average information. In particular, 50.62% of the variation is explained by U_1 and 32.05% is explained by U_2. These values justify that both U_1 and U_2 should remain in the model. Also, it is important to mention that $\hat{\lambda}_1 = 16.24$ and $\hat{\sigma}^2 = 3.40$.

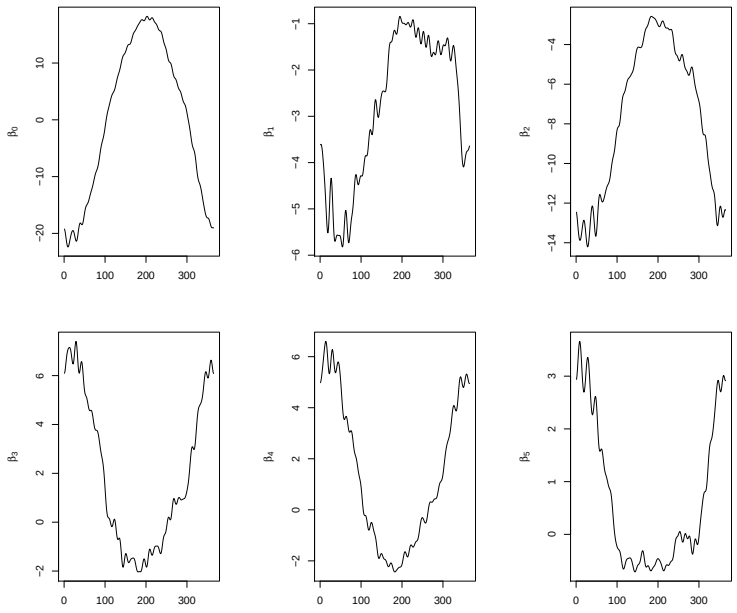

Figure 5: The estimations of the functional parameters for the Canadian data set.

In Figure 6, we show a four cell diagram. In the top-left plot we depict the average daily temperature between 1960 and 1994 for 35 weather stations in Canada. The top-right plot shows the weather prediction in all stations using our suggested Kriging method. For this, each individual site was temporarily removed and its weather was krigged using the temperature data of the other sites. The bottom-left plot shows the residual curves. We note that the residuals are around zero, which justifies that our model results in better estimation than those of Caballero *et al.* (2013). Finally, the bottom-right plot depicts the average daily temperature reported at just one station, the Fredericton weather station. The black line shows the average temperature while the grey one shows our estimation which are very close of to each other. The sum of the squared errors for the Fredericton weather station for the ordinary kriging for function-valued data is 944.737, using universal kriging (Caballero *et al.* , 2013) for functional data we get 918.7242, and for our prediction method we obtain 277.9509.

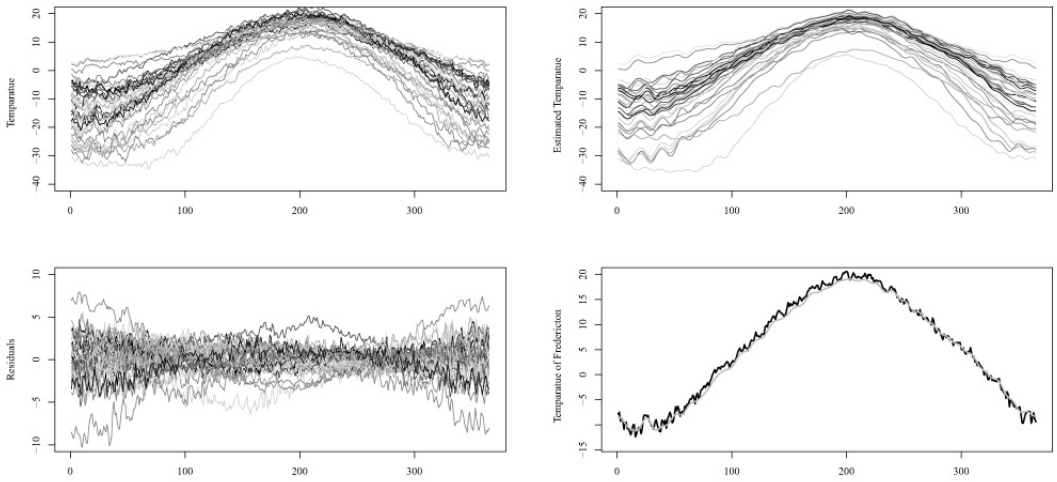

Figure 6: Canadian temperatures: The top-left plot depicts the average daily temperature between 1960 and 1994 for 35 weather stations in Canada. The top-right plot shows the weather prediction in all stations. The bottom-left plot shows the residual curves. The bottom-right plot depicts the average daily temperature reported at Fredericton weather station.

7 Conclusions

The standard functional data techniques are often based on the condition that the functional observations are independent, but this remains a problem in many applied contexts. Spatial functional data analysis arises as a new branch of functional statistics that handles correlated functional data.

In this context, we have introduced a functional mixed effects random model for modeling spatial data. Our model encompasses fixed and random effects, both being linear combinations of L^2 functions and random elements, respectively. Estimation and prediction procedures are depicted. The proposed model adapts easily to non-stationary cases, and it can be considered a step forward in the analysis of spatial functional data.

There are, however, several points for further improvements. Even if the number of terms N in the expansions is selected using an objective method, the combination of N with the number m of temporal observations can be dramatically high bringing computational problems. A more global objective method dealing with both sizes at the same time would ease this problem. Another topic is the selection of the orthonormal basis functions. The question of which functions (wavelet basis, Fourier basis, etc) should provide better fittings to a given data set is open for further research.

References

Abramovich, F. and Angelini, C. (2006), Bayesian maximum a posteriori multiple testing procedure. *Sankhya*, **68**, 436–460.

Antoniadis, A., Paparoditis, E., and Sapatinas, T. (2006), A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society B*, **68**, 837–857.

Aston, J. A. D., Chiou, J. M., and Evans, J. P. (2010), Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **59(2)**, 297–317

Bosq, D. (2000), *Linear Process in Function Spaces. Theory and Applications*. Springer-Verlag: Berlin.

Caballero, W., Giraldo, R., and Mateu J. (2013), A universal kriging approach for spatial

functional data. *Stochastic Environmental Research and Risk Assessment*, **27(7)**, 1553–1563.

Cardot, H., Crambes, C., Kneip, A., and Sarda, P. (2007), Smoothing spline estimators in functional linear regression with errors-in-variables. *Computational Statistics and Data Analysis*, **51**, 4832–4848.

Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010), Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21**, 224–239.

Demidenko , E. (2004), *Mixed Models, Theory and Applications*. Wiley.

Ferraty, F. and Vieu, P. (2006), *Non Parametric Functional Data Analysis. Theory and Practice*. New York: Springer.

Giraldo, R., Delicado, P., and Mateu, J. (2010), Geostatistics for functional data: An ordinary kriging approach. *Environmental and Ecological Statistics*, **9**, 16–21.

Giraldo, R., Delicado, P., and Mateu, J. (2011), Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, **18(3)**, 411–426.

Green, P. J. and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. Chapman & Hall Ltd.

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010), Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, **4**, 1022–1054.

Gromenko, O., Kokoszka, P., Zhu, L., and Sojka, J. (2012), Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *The Annals of Applied Statistics*, **6**, 669–696.

Guillas, S. and Lai, M. J. (2010), Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics*, **22**, 477–497.

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krosigk, C. N. (1959), Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **13**, 192–218.

Horvath, L., Huskova, M., and Kokoszka, P. (2010), Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis*, **101**, 352–367.

Krafty, R. T., Hall, M., and Guo, W. (2011), Functional mixed effects spectral analysis. *Biometrika*, **98(3)**, 583–598.

Krivobokova, T. and Kauermann, G. (2007), A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, **102**, 1328–1337.

Lin, X. and Carroll, R. J. (2000), Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, **95**, 520–534.

Morris, J. S. and Carroll. R. J. (2006), Wavelet–based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 179–199.

Nerini, D., Monestiez, P., and Mante, C. (2010), Cokriging for spatial functional data. *Journal of Multivariate Analysis*, **101(2)**, 409–418.

Ramsay, J. O. and Silverman, B. W. (2002), *Applied Functional Data Analysis*. New York: Springer.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*. New York: Springer.

Reiss, P. T. and Ogden, R. T. (2010), Functional generalized linear models with images as predictors. *Biometrics*, **66**, 61–69.

Rice, J. and Silverman, B. (1991), Estimating the mean and covariance Structure Nonparametrically when the data are curves. *Journal of the Royal Statistical Society, ser. B*, **53**, 233–243.

Scheipl. F., Staicu, A. M., and Greven, S. (2015), Functional additive mixed models. *Journal of Computational and Graphical Statistics*, **24**, 477–501.

Soltani, A. R., Nematollahi, A. R., and Nasirzadeh, R. (2017), On convergence of sample and population Hilbertian functional principal component. *Bulletin of the Iranian Mathematical Society*, **43(2)**, 467–475.

Staicu, A. M., Crainiceanu, C. M., and Carroll, R. J. (2010), Fast Methods for Spatially Correlated Multilevel Functional Data. *Biostatistics*, **11**, 177–194.

Yao, F., Muller, H. G., and Wang, J. L. (2005), Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–590.

Zipunnikov V., Caffo B. S., Yousem D. M., Davatzikos C., Schwartz B. S., and Crainiceanu, C. (2011), Functional principal component model for high–dimensional brain imaging. *NeuroImage*, **58(3)**, 772–784.