ARTICLE TEMPLATE

How to interpolate geographically located curves: a precise and straightforward kriging approach to function-valued data

Gilberto Pereira Sassi (corresponding author)[a], Chang Chiann[b]

[a]Federal University of Bahia, Salvador, Bahia – Brazil; [b]University of São Paulo, São Paulo, São Paulo – Brazil

ARTICLE HISTORY

Compiled February 7, 2020

Abstract

Functional Data Analysis research has been highlighted for its in several fields of science and functional datasets can be spatially indexed curves, which are denominated Spatial Functional Data Analysis. The main goal of this paper is to supply a straightforward and precise approach to interpolate curves, i.e., our aim is to estimate a curve at an unmonitored location. It has been proved that best linear unbiased estimator for this unsampled curve is solution of a linear system, where the dependent and coefficient matrix use a function called trace-variogram. In this paper, we propose the use of Legendre-Gauss quadrature to estimate the trace-variogram. Simulations, assuming normality and nonnormality, showed excellent numerical properties: similar bias and smaller mean square error than the established estimation procedure were observed. The novel estimation methodology is illustrated with a real dataset of temperature curves from 35 weather stations of Canada.

KEYWORDS

Trace Variogram; Kriging; Functional Data Analysis; Fourier Series; Spatial Analysis

CONTACT Gilberto Pereira Sassi (corresponding author). Email: `gilberto.sassi@ufba.br`, Chang Chiann. Email: `chang@ime.usp.br`

1. Introduction

Functional Data Analysis is concerned with analyzing data presented in the form of curves or functions. It can be argued that the concepts of Functional Data Analysis were formalized and introduced by Ramsay and Dalzell (1991), and it is suitable for applications where we need to analyze an observation from a family $\{X(t_j)\}_{j=1,...,J}$ where $t_1, \ldots, t_J$ are equally spaced and $t_j \in (t_{min}, t_{max}), j = 1, \ldots, J$, when the interval between t_j and t_{j+1} gets smaller, we could consider this observation as sampled from a random continuous family $\boldsymbol{\chi} = \{X(t) \mid t \in (t_{min}; t_{max})\}$. Furthermore, there are cases where the underlying data are clearly a continuous function even when the sample is scattered, for example, the child growth curve and the electrical consumption curve (Ferraty and Vieu 2006). Since the seminal paper of Ramsay Ramsay and Dalzell (1991), Functional Data Analysis has been widely developed and applied in various branches of Statistics, such as geostatistics, linear model, item response theory and others. More recently, Fang et al. (2020) considered the functional linear regression for multivariate responses and developed a locally sparse estimation for the coefficient functions. Chen, Goldsmith, and Ogden (2019) proposed a method to model dynamic positron emission tomography (PET) data from multiple subjects simultaneously, where impulse response functions (IRF) are estimated using linear mixed effects functional data model. Beyaztas and Shang (2019) proposed a robust method to forecast functional time series based on the minimum density power divergence estimator of Basu et al. (1998). Lee et al. (2019), used a Bayesian functional mixed model to analyze a glaucoma scleral strain dataset, where they could pick up nonparametric covariate effects, serial and nested interfunctional correlation. Zamani, Hashemi, and Haghbin (2019) proposed a improvement in the Portmanteau test of Gabrys and Kokoszka (2007) of independecy of functional observations, which is specially suited to small samples.

In Geostatistics and Functional Data Analysis, Giraldo, Delicado, and Mateu (2010) and Giraldo, Delicado, and Mateu (2011) proposed two kriging methods for Spatial Functional Data, where the curve at an unmonitored spot is a linear combination of all avaliable curves. In Giraldo, Delicado, and Mateu (2011), the weights of this linear

2

combination is a scalar and this method is referred as Ordinary Kriging; in Giraldo, Delicado, and Mateu (2010) the weights are also curves. In both methods, the weights are estimated by an unbiased minimum square error estimator. Caballero, Giraldo, and Mateu (2013) extended these models for non-stationary spatial processes and Ignaccolo, Mateu, and Giraldo (2014) developed a kriging method with external drift where it was possible to use exogenous variables. Menafoglio, Guadagnini, and Secchi (2014) focused on the formulation of new geostatistical models and methods for functional compositional data. Reyes, Giraldo, and Mateu (2012) proposed a methodology to extend the kriging predictor for functional data to the case where the mean function was not constant by considering an approach based on the classical residual kriging method used in Geostatistics. Salazar, Giraldo, and Porcu (2015) introduced a predictor, similar to the Functional Ordinary Kriging from Giraldo Giraldo, Delicado, and Mateu (2011), where they used probability density functions with support on a finite interval. Lee, Zhu, and Toscas (2015) introduced a methodology for the analysis of spatial functional data using low-rank smoothers such as penalized splines and tensor products for space-time interpolation, prediction and forecasting using Functional Data Analysis and modeling the spatial structure explicitly. Nerini, Monestiez, and Manté (2010) generalized the method of kriging proposed by Giraldo, Delicado, and Mateu (2010) and Giraldo, Delicado, and Mateu (2011).

In this paper, we have observed n curves $\chi_{\boldsymbol{s}_1}(t), \cdots, \chi_{\boldsymbol{s}_n}(t)$ at locations $\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n$ in a specified region, where $\boldsymbol{s}_i = (\theta_i, \eta_i)$, $i = 1, \ldots, n$, θ_i is the latitute and η_i is the longitude. The aim is to estimate the unobserved curve $\chi_{\boldsymbol{s}_0}(t)$ where $\boldsymbol{s}_0 \notin \{\boldsymbol{s}_1, \cdots \boldsymbol{s}_n\}$. The ideia proposed by Giraldo Giraldo, Delicado, and Mateu (2011) was uncomplicated: the curve $\chi_{\boldsymbol{s}_0}(t)$ at $\boldsymbol{s}_0 \notin \{\boldsymbol{s}_1, \cdots \boldsymbol{s}_n\}$ is a linear combination of all the curves $\chi_{\boldsymbol{s}_1}(t), \cdots, \chi_{\boldsymbol{s}_n}(t)$, i.e., $\chi_{\boldsymbol{s}_0}(t) = \sum_{i=1}^n \lambda_i \chi_{\boldsymbol{s}_i}(t)$, where $\lambda_1, \cdots, \lambda_n$ is the solution of the

linear system

$$
\begin{pmatrix}
\gamma(\|s_1 - s_1\|) & \gamma(\|s_1 - s_2\|) & \ldots & \gamma(\|s_1 - s_n\|) & 1 \\
\gamma(\|s_2 - s_1\|) & \gamma(\|s_2 - s_2\|) & \ldots & \gamma(\|s_2 - s_n\|) & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(\|s_n - s_1\|) & \gamma(\|s_n - s_2\|) & \ldots & \gamma(\|s_n - s_n\|) & 1 \\
1 & 1 & \ldots & 1 & 0
\end{pmatrix}
\cdot
\begin{pmatrix}
\lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ -\mu
\end{pmatrix}
=
\begin{pmatrix}
\gamma(\|s_1 - s_0\|) \\ \gamma(\|s_2 - s_0\|) \\ \vdots \\ \gamma(\|s_n - s_0\|) \\ 1
\end{pmatrix},
$$

where μ is a constant and

$$
\gamma(h) = \int \gamma(h; t) dt, \tag{1}
$$

is called trace-variogram and $\gamma(h, t)$ is the semivariogram as defined in Section 3. More precisely, for each t we assume a weakly stationary and isotropic spatial process and we can compute the semivariogram for this process for $\chi_{s_1}(t), \cdots, \chi_{s_1}(t)$, and the trace-variogram is an integration of semivariogram $\gamma(h; t)$. Usually, the integral in the equation 1 is approximated using an modified version of empirical semivariogram (see Giraldo, Delicado, and Mateu 2011, for more details). In this paper, we propose an approach using Legendre-Gauss quadrature, which is intuitive since it explicitly uses the definition of trace-variogram. For illustration, the new approach is applied to a real dataset of temperature curves of 365 days from 35 weather stations of Canada.

This paper is organized as follows: in Section 2 we briefly introduce the concept of Functional Data; in Section 3 we present a novel methodology to estimate the trace-variogram and the ordinary kriging method for functional datasets; in Section 4 we compare the new estimation process for the trace-variogram with the estimation process proposed by Giraldo, Delicado, and Mateu (2011) using simulation studies; in Section 5 we illustrate the kriging method using a real dataset of temperature curves with 35 stations of Canada; and, finally, in Section 6 we make our final considerations.

2. Functional data analysis

Traditionally, Statistics deals with information from observations $x_1, \ldots, x_T$, which may be scalars, vectors or matrices. On the other hand, Functional Data Analysis, observations are viewed as functions defined over some set $B \subset \mathbb{R}^p, p \in \mathbb{N}$, usually $p \in \{1, 2\}$ [1]. Before we consider the kriging methods for spatial functional data in Section 3, we formally introduce some concepts about functional variable, functional data and functional dataset according to Ferraty and Vieu (2006).

Definition 2.1. A function $\boldsymbol{\chi} : \Omega \to L^2(B)$ is said to be a functional variable if its realizations (or values) are functions defined on $B \subset \mathbb{R}^p$ and assumed to belong to

$$L^2(B) = \left\{ \chi : B \longrightarrow \mathbb{R} \mid \int_B \sum_{k=1}^{p} |\chi(u_k)|^2 du_1 \cdots du_p < \infty \right\}.$$

An observation χ of $\boldsymbol{\chi}$ is referred to as functional data.

Remark 1.

(1) When $p = 1$, a functional variable is called a random curve and a functional data is called a curve.

(2) We denote a functional variable by $\boldsymbol{\chi}(t)$ and a functional data by $\chi(t)$.

Definition 2.2. A functional dataset $\chi_1(t), \ldots, \chi_n(t)$ is a set of observations of n functional variables $\boldsymbol{\chi}_1(t), \ldots, \boldsymbol{\chi}_n(t)$.

Generally, due to finite resolution, the curves in a functional dataset are available only at a finite grid of points and a smoothing technique is required. In this paper, we have used the Fourier Series to smooth these curves as explained in Section 3.

3. A Kriging method for spatial functional data

Throughout this paper, we assume a pointwise isotropic and weakly stationary spatial process. More precisely, let $\boldsymbol{s}_i = (\theta_i, \eta_i) \in \mathbb{R}^2$, $i = 1, \ldots, n$, be locations of a compact

---

[1]For more details about Functional Data Analysis, the website `http://www.psych.mcgill.ca/misc/fda/` organized and maintained by Ramsay and the textbooks of Ramsay (2006) and Horváth and Kokoszka (2012) are excellent references.

region $D \in R^2$, where θ_i and η_i are the latitude and the longitude, respectively, and $\chi_{s_i}(t), t \in T$, square-integrable curves, then we have

(1) $\mathrm{E}(\chi_s(t)) = m(t), \forall t \in T, \forall s \in D$,

(2) $\mathrm{Cov}(\chi_{s_i}(t), \chi_{s_j}(t)) = C(h; t), \forall s_i, s_j \in D, \forall t \in T$ and $h = \|s_i - s_j\|$, where $\|s_i - s_j\|$ is the Euclidean distance between s_i and s_j,

(3) $\dfrac{1}{2} \mathrm{Var}(\chi_{s_i}(t) - \chi_{s_j}(t)) = \gamma(h; t), \forall s_i, s_j \in D, \forall t \in T, h = \|s_i - s_j\|$,

where D is a subset of $\mathbb{R}^2$ containing a set with positive area.

Suppose we have a functional dataset $\chi_{s_1}(t), \ldots, \chi_{s_n}(t)$ at locations $s_1, \ldots, s_n$ and our goal is to obtain an estimate for the curve $\chi_{s_0}(t)$ where $s_0 \notin \{s_1, \ldots, s_n\}$. Giraldo, Delicado, and Mateu (2011) proposed the following estimator:

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^{n} \hat{\lambda}_i \chi_{s_i}(t), \quad \hat{\lambda}_1, \ldots, \hat{\lambda}_n \in \mathbb{R}, \tag{2}$$

where $\hat{\chi}_{s_0}(t)$ is called functional ordinary kriging predictor, and $\hat{\lambda}_1 \ldots, \hat{\lambda}_n$ are the solution of the following nonlinear optimization problem

$$\underset{\hat{\lambda}_1, \ldots, \hat{\lambda}_n}{\arg \min} \mathrm{E}\left[\|\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\|_{L^2}^2\right] \text{ subject to } \mathrm{E}(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) = 0, \forall t \in [0, 1], \tag{3}$$

where $\|(\chi(t))\|_{L^2}^2 = \int_T |\chi(t)|^2 dt$.

In order to solve the optimization problem described by equation (3), the theorem 3.1 establishes an equivalence between solving this optimization problem and solving a linear system.

Theorem 3.1 (Giraldo, Delicado, and Mateu (2011)). *Under some regularity conditions, solving the nonlinear optimization problem*

$$\underset{\hat{\lambda}_1, \ldots, \hat{\lambda}_n}{\arg \min} \mathrm{E}\left[\|\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\|_{L^2}^2\right] \text{ subject to } \mathrm{E}(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) = 0, \forall t \in T,$$

is equivalent to solve

$$
\begin{pmatrix}
\gamma(\|\boldsymbol{s}_1 - \boldsymbol{s}_1\|) & \gamma(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|) & \ldots & \gamma(\|\boldsymbol{s}_1 - \boldsymbol{s}_n\|) & 1 \\
\gamma(\|\boldsymbol{s}_2 - \boldsymbol{s}_1\|) & \gamma(\|\boldsymbol{s}_2 - \boldsymbol{s}_2\|) & \ldots & \gamma(\|\boldsymbol{s}_2 - \boldsymbol{s}_n\|) & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(\|\boldsymbol{s}_n - \boldsymbol{s}_1\|) & \gamma(\|\boldsymbol{s}_n - \boldsymbol{s}_2\|) & \ldots & \gamma(\|\boldsymbol{s}_n - \boldsymbol{s}_n\|) & 1 \\
1 & 1 & \ldots & 1 & 0
\end{pmatrix}
\cdot
\begin{pmatrix}
\lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ -\mu
\end{pmatrix}
=
\begin{pmatrix}
\gamma(\|\boldsymbol{s}_1 - \boldsymbol{s}_0\|) \\
\gamma(\|\boldsymbol{s}_2 - \boldsymbol{s}_0\|) \\
\vdots \\
\gamma(\|\boldsymbol{s}_n - \boldsymbol{s}_0\|) \\
1
\end{pmatrix},
$$

where μ is a constant and $\gamma(h) = \int_T \gamma(h;t)dt$ is called trace-variogram.

At each location $\boldsymbol{s}_i, i = 1, \ldots, n$, we have a time series $\chi_{\boldsymbol{s}_i}(t_j), j = 1, \ldots, m$, where $t_j \in [-\pi, \pi]$. If we assume the curves $\chi_{\boldsymbol{s}_i}(t), t \in [-\pi, \pi]$ are continuous, then these curves can be approximated by

$$
\chi_{\boldsymbol{s}_i}(t) \approx \frac{a_0^i}{2} + \sum_{k=1}^{p} \left[a_k^i \cos(k \cdot t) + b_k^i \sin(k \cdot t) \right],
$$

for suitable values of $a_0^i, a_1^i, \cdots, a_p^i, b_1^i, \cdots, b_p^i, \ i = 1, \ldots, n$. There are several ways to establish the value of p, in this paper, we have used Sturge's rule (see Scott 2009, for more details). More precisely,

$$
p = \lceil 1 + \log_2(m) \rceil,
$$

where $\lceil x \rceil$ is the smallest integer greater or equal to x, and m is the length of the observe time series $\chi_{\boldsymbol{s}_i}(t_1), \ldots, \chi_{\boldsymbol{s}_i}(t_m), i = 1, \ldots, n$.

The coefficients $a_0^i, a_1^i, a_2^i, \cdots, a_p^i, b_1^i, b_2^i, \cdots, b_p^i$ minimize the following sum

$$
S(a_0^i, a_1^i, a_2^i, \cdots, a_p^i, b_1^i, b_2^i, \cdots, b_p^i) = \sum_{l=1}^{m} \left(\chi_{\boldsymbol{s}_i}(t_l) - \left(\frac{a_0^i}{2} + \sum_{k=1}^{p} \left[a_k^i \cos(k \cdot t_l) + b_k^i \sin(k \cdot t_l) \right] \right) \right)^2.
$$

For simplicity, we have used a vectorial notation:

$$
\chi_{\boldsymbol{s}_i}(t) = \boldsymbol{c}_i^\top \boldsymbol{\varphi}(t),
$$

where

$$\mathbf{c}_i = \left(a_0^i, a_1^i, a_2^i, \cdots, a_p^i, b_1^i, b_2^i, \cdots, b_p^i \right)^\top,$$

$$\boldsymbol{\varphi}(t) = \left(\frac{1}{2}, \cos(t), \cos(2t), \cdots, \cos(pt), \sin(t), \sin(2t), \cdots, \sin(pt) \right)^\top.$$

Under the pointwise isotropic and weakly stationary spatial assumption, for each $t \in [-\pi, \pi]$ we can estimate the semivariogram using a exponential model such as

$$\gamma(h, t) = \sigma_t^2 \left(1 - \exp\left(-\frac{h}{\phi_t} \right) \right),$$

where $\sigma_t^2 > 0$ and $\phi_t > 0$. Note that for a fixed value $t \in [-\pi, \pi]$, we can estimate σ_t^2 and ϕ_t using the sample $\chi_{\mathbf{s}_1}(t) = \mathbf{c}_1 \varphi(t); \chi_{\mathbf{s}_2}(t) = \mathbf{c}_2 \varphi(t); \cdots; \chi_{\mathbf{s}_n}(t) = \mathbf{c}_n \varphi(t)$.

In order to estimate the trace-variogram we have to solve the following integration

$$\gamma(h) = \int_{-\pi}^{\pi} \gamma(h, t) dt. \tag{4}$$

Let $f : [a, b] \longrightarrow \mathbb{R}$ a continuous function, then the integral over $[a, b]$ can be approximated by

$$\int_a^b f(u) du \approx \frac{b-a}{2} \sum_{k=0}^{m} \omega_k f\left[\frac{(b-a)z_k + b + a}{2} \right],$$

where $z_0, z_1, z_2, \ldots, z_m$ are the zeros of the Legendre polynomial $p_{m+1}(x)$ (Khuri 2003, pp. 441) of degree $m + 1$ and $\omega_0, \omega_1, \omega_2, \cdots, \omega_m$ are solutions of the following linear system

$$\int_{-1}^{1} u^j du = \sum_{k=0}^{m} \omega_k z_k^j, \qquad j = 0, 1, \cdots, m.$$

Consequently, we can approximate the trace-variogram by

$$\gamma(h) = \int_{-\pi}^{\pi} \gamma(h, t)dt$$

$$\approx \pi \sum_{k=0}^{m} w_k \gamma(h, \pi z_k). \tag{5}$$

If we knew the semivariogram $\gamma(\cdot, t)$ at the points $t = \pi z_1, \pi z_2, \cdots, \pi z_m$, we could approximate the trace-variogram $\gamma(\cdot)$ using equation (5). More precisely, for each $t = \pi z_k$, we have the following sample

$$X_{\boldsymbol{s}_1} = \boldsymbol{c}_1^\top \boldsymbol{\varphi}(\pi z_k); X_{\boldsymbol{s}_2} = \boldsymbol{c}_2^\top \boldsymbol{\varphi}(\pi z_k); \cdots ; X_{\boldsymbol{s}_n} = \boldsymbol{c}_n^\top \boldsymbol{\varphi}(\pi z_k)$$

and

$$\begin{pmatrix} X_{\boldsymbol{s}_1} \\ X_{\boldsymbol{s}_2} \\ \vdots \\ X_{\boldsymbol{s}_n} \end{pmatrix} \sim N\left(\boldsymbol{\mu_k}; \Sigma_k\right).$$

where

$$\boldsymbol{\mu_k} = \begin{pmatrix} \mu_k \\ \mu_k \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} m(\pi z_k) \\ m(\pi z_k) \\ \vdots \\ m(\pi z_k) \end{pmatrix} ; \qquad \Sigma_k = \begin{pmatrix} c_{11}^k & c_{12}^k & \cdots & c_{1n}^k \\ c_{21}^k & c_{22}^k & \cdots & c_{2n}^k \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1}^k & c_{n2}^k & \cdots & c_{nn}^k \end{pmatrix}$$

and $c_{ij}^k = \sigma_k^2 \exp\left(-\dfrac{\|\boldsymbol{s}_i - \boldsymbol{s}_j\|}{\phi_k}\right)$, $i, j = 1, 2, 3, \ldots, n$. We estimate σ_k and ϕ_k using the maximum likelihood estimate where the likelihood function is given by

$$l(\sigma_k, \phi_k, \mu_k) = \det(2\pi \Sigma_k)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_k)\right),$$

where $\boldsymbol{X} = (X_{\boldsymbol{s}_1}, X_{\boldsymbol{s}_2}, \cdots, X_{\boldsymbol{s}_n})$. Finally, the trace-variogram is approximated by

$$\gamma(h) \approx \pi \sum_{k=0}^{n} \omega_k \sigma_k^2 \left(1 - \exp\left(\frac{h}{\phi_k}\right)\right).$$

4. Simulation study

In this section, our aim is to evaluate the estimation method of the curve $\chi_{\boldsymbol{s}_0}(t)$ proposed in Section 3 using mean square error and bias, when we can assume normality of data and when the normality presumption is not suitable.

First, we generate n locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n$ on $D = [-5,5] \times [-5,5]$, i.e., the latitude θ ranges from -5 to 5 and the longitude η ranges from -5 to 5, and the aim is to estimate the curve at location $\boldsymbol{s}_0 = (\theta_0, \eta_0) = (0, 0)$. In order to simulate the curves at locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n, \boldsymbol{s}_0$, we sampled a random matrix given by

$$M = \begin{pmatrix} m_{0,1} & m_{0,2} & m_{0,3} & \cdots & m_{0,n+1} \\ m_{1,1} & m_{1,2} & m_{1,3} & \cdots & m_{1,n+1} \\ m_{2,1} & m_{2,2} & m_{2,3} & \cdots & m_{2,n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{2p,1} & m_{2p,2} & m_{2p,3} & \cdots & m_{2p,n+1} \end{pmatrix} \tag{6}$$

where $\boldsymbol{s}_{n+1} = \boldsymbol{s}_0$. We also simulated a vector $\boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_p & \alpha_{p+1} & \cdots & \alpha_{2p} \end{pmatrix}^\top$ from the multivariate normal distribution $N(\boldsymbol{0}, I_{2p+1})$.

Then simulated curves are given by

$$\chi_{\boldsymbol{s}_i}(t) = \frac{\alpha_0}{2} + \sum_{k=1}^{p} (\alpha_k \cos(k \cdot t) + \alpha_{k+p} \sin(k \cdot t)) +$$

$$+ m_{o,i} + \sum_{k=1}^{p} (m_{k,i} \cos(k \cdot t) + m_{k+p,i} \sin(k \cdot t)) + \epsilon_t, \quad \forall t \in [-\pi, \pi], \ i = 1, \ldots, n,$$

and

$$\chi_{\boldsymbol{s}_0}(t) = \frac{\alpha_0}{2} + \sum_{k=1}^{p}\left(\alpha_k\cos(k\cdot t) + \alpha_{k+p}\sin(k\cdot t)\right) +$$

$$+ m_{o,n+1} + \sum_{k=1}^{p}(m_{k,n+1}\cos(k\cdot t) + m_{k+p,n+1}\sin(k\cdot t)) + \epsilon_t, \quad \forall t \in [-\pi,\pi]$$

where $\epsilon_t \sim N(0,1)$.

Finally, we build time series $\left(\chi_{\boldsymbol{s}_i}(t_1) \quad \chi_{\boldsymbol{s}_i}(t_2) \quad \cdots \quad \chi_{\boldsymbol{s}_i}(t_m)\right)^{\top}$ for each location $\boldsymbol{s}_i$, $i = 0, 1, \ldots, n$, and we use the time series $\left(\chi_{\boldsymbol{s}_i}(t_1) \quad \chi_{\boldsymbol{s}_i}(t_2) \quad \cdots \quad \chi_{\boldsymbol{s}_i}(t_m)\right)^{\top}$, $i = 1, \ldots, n$ to estimate the curve $\chi_{\boldsymbol{s}_0}(t)$, and then we compute a mean square error and a bias given by

$$mse = \frac{1}{m}\sum_{k=1}^{m}(\chi_{\boldsymbol{s}_0}(t_k) - \hat{\chi}_{\boldsymbol{s}_0}(t_k))^2,$$

$$bias = \frac{1}{m}\sum_{k=1}^{m}(\chi_{\boldsymbol{s}_0}(t_k) - \hat{\chi}_{\boldsymbol{s}_0}(t_k)).$$

In this paper, we have used two scenarios: normality and nonnormality assumptions. In the first scenario, for each $t \in [-\pi,\pi]$ and for each $\boldsymbol{s} \in D$, $\chi_{\boldsymbol{s}}(t)$ has normal distribution, and in the second scenario, for each $t \in [-\pi,\pi]$ and for each $\boldsymbol{s} \in D$ does not have normal distribution.

In the first scenario, each column $m_{k,\cdot} = \left(m_{k,1} \quad m_{k,2} \quad \cdots \quad m_{k,n+1}\right)^{\top}$ of the matrix M from equation (6) is sampled from $N(\boldsymbol{0}, \Sigma_k)$, $k = 0, 1, \ldots, 2p$, where $\Sigma_k = (\sigma_{i,j}^k)_{i,j=1}^{n+1}$ is a covariance matrix with

$$\sigma_{i,j}^k = \exp\left(-2\cdot\|\boldsymbol{s}_i - \boldsymbol{s}_j\|\right), \quad i,j = 1, \ldots, n, n+1.$$

In the second scenario, each column $m_{k,\cdot} = \left(m_{k,1} \quad m_{k,2} \quad \cdots \quad m_{k,n+1}\right)^{\top}$ of the matrix M from equation (6) is sampled from the nonnormal distribution proposed by Vale and Maurelli (1983) with the covariance matrix given by $\Sigma_k = (\sigma_{i,j}^k)_{i,j=1}^{n+1}$ where

$$\sigma_{i,j}^k = \exp\left(-2\cdot\|\boldsymbol{s}_i - \boldsymbol{s}_j\|\right), \quad i,j = 1, \ldots, n, n+1,$$

with skeness equals to -2 and kurtosis equals to 2.

In our simulation study, we have used $n = 25, 50, 75, 100$ locations and time series with length $m = 1000$. For each n, we made $L = 1000$ replications.

Results from simulation in the first scenario

In the Table 1, we present the mean of mse and $bias$ for the $L = 1000$ replications. The $bias$ in both methods are similar, but the mean square error (mse) is smaller for the method proposed in this paper. This conclusion can be visualized in the Figures 1, 2, 3 and 4 where we present the boxplot for mse for $L = 1000$ replications, and in the Figures 6, 7, 8 and 9 where we present the boxplot for $bias$ for $L = 1000$ replications.

Table 1.: Mean of mse and $bias$ for normal simulated data. **New** refers to the new methodology proposed in this paper, and **OKFD** refers to methodology proposed by Giraldo Giraldo, Delicado, and Mateu (2011).

Number of locations	Mean of mse		Mean of $bias$	
	New	OKFD	New	OKFD
25	10.88	21.08	-0.02	-0.03
50	9.47	16.34	-0.01	-0.01
75	9.03	15.98	0.01	0.00
100	8.70	14.82	-0.04	-0.03

Results from simulation in the second scenario

In the Table 2, we present the mean of mse and $bias$ for the $L = 1000$ replications. The $bias$ is similar in both approaches, but, even in this context of nonnormality, the estimation of trace variogram proposed in this paper has equal ou smaller mse. We note that for datasets with few locations the two estimation methods give similar results, but when datasets have many locations our proposal has smaller mse. This analysis can be visualized in the figures 11, 12, 13, 14, where we displayed the boxplot for the mse for $L = 1000$ replications, and in the figures 16, 17, 18, 19, where we displayed the boxplot for the $bias$ for $L = 1000$ replications.
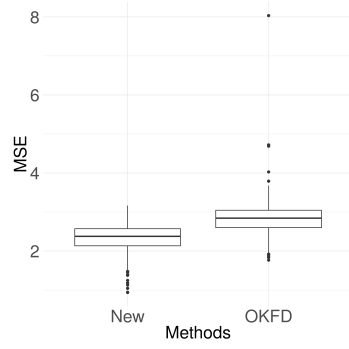
Figure 1.: $n = 25$
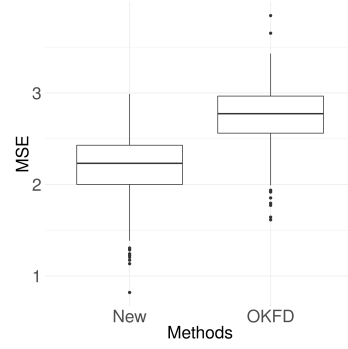
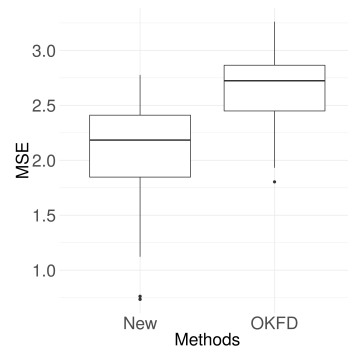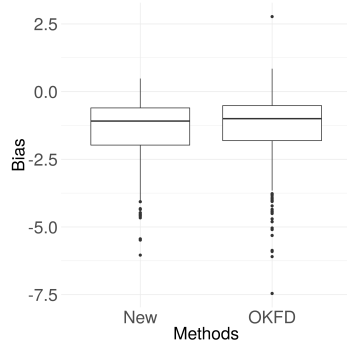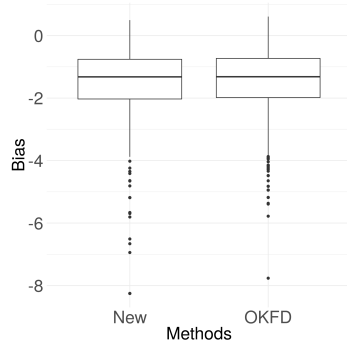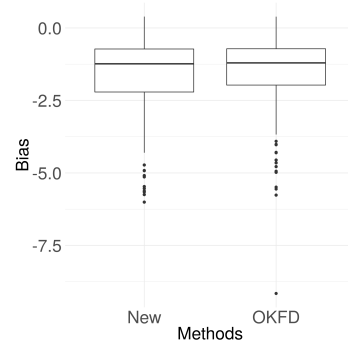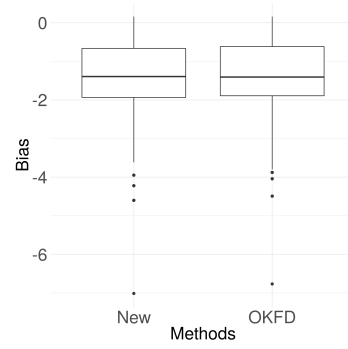Figure 2.: $n = 50$
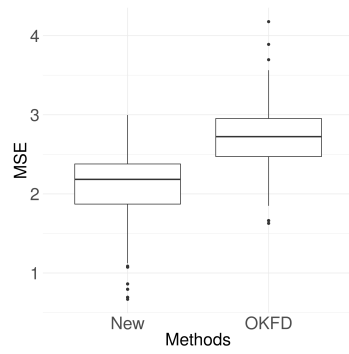
Figure 3.: $n = 75$

Figure 4.: $n = 100$

Figure 5.: *mse* for normal data for the estimation method in this paper (New) for $L = 1000$ replications and for the estimation method proposed by Giraldo, Delicado, and Mateu (2011) (OKFD).

Table 2.: Mean of *mse* and *bias* for nonnormal simulated data. **New** refers to the new methodology proposed in this paper, and **OKFD** refers to methodology proposed by Giraldo Giraldo, Delicado, and Mateu (2011).

Number of locations	Mean of *mse*		Mean of *bias*	
	New	OKFD	New	OKFD
25	13.23	21.58	0.00	0.02
50	12.44	19.01	0.05	0.04
75	11.91	18.43	0.02	0.03
100	11.59	18.75	-0.05	-0.04

Figure 6.: $n = 25$

Figure 7.: $n = 50$

Figure 8.: $n = 75$

Figure 9.: $n = 100$

Figure 10.: *bias* for normal data for the estimation method in this paper (New) for $L = 1000$ replications and for the estimation method proposed by Giraldo, Delicado, and Mateu (2011) (OKFD).
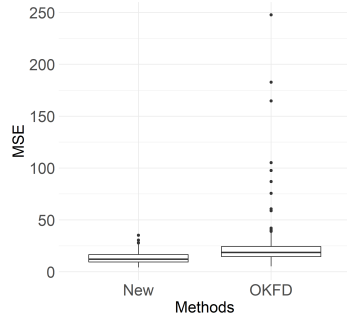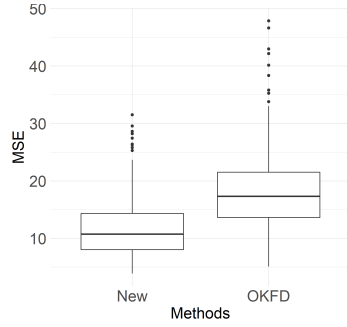
Figure 11.: $n = 25$
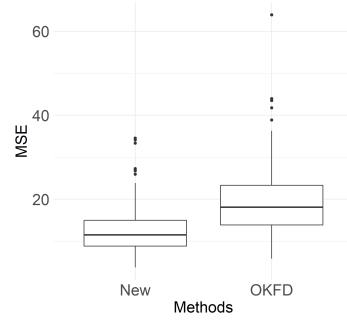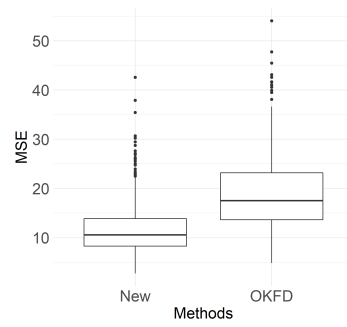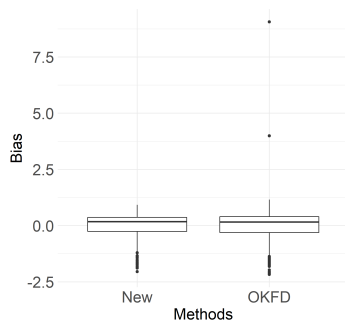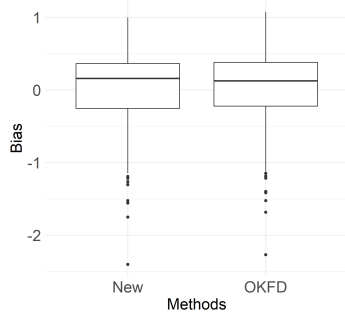

Figure 12.: $n = 50$

Figure 13.: $n = 75$

Figure 14.: $n = 100$

Figure 15.: mse for nonnormal data for the estimation method in this paper (New) for $L = 1000$ replications and for the estimation method proposed by Giraldo, Delicado, and Mateu (2011) (OKFD).
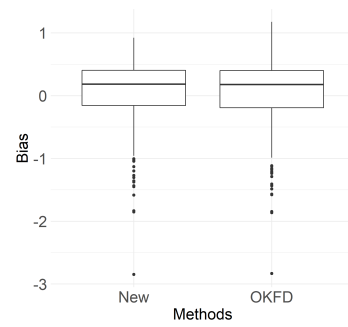
Figure 16.: $n = 25$

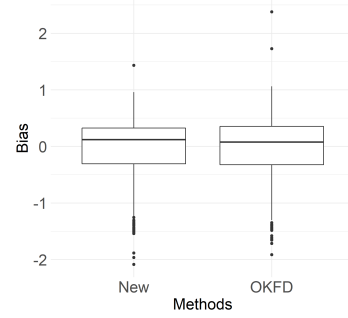Figure 17.: $n = 50$

Figure 18.: $n = 75$

Figure 19.: $n = 100$

Figure 20.: *bias* for the estimation method in this paper (New) for $L = 1000$ replications and for the estimation method proposed by Giraldo, Delicado, and Mateu (2011) (OKFD).

5. Application

We illustrate the ideas presented above with a dataset containing daily mean temperature measurements recorded at 35 weather stations of Canada. This dataset is included in the R packages `geofd` and `fda` (Giraldo, Mateu, and Delicado 2012, Ramsay et al. (2018)), it consists of 35 time series where each time series is a daily mean temperature and can be downloaded at site weather.gc.ca. In the Figure 21, we show all the weather stations considered in this paper.
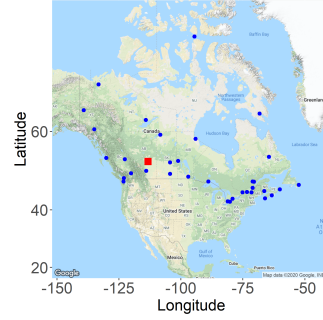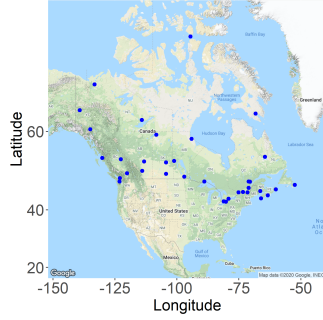

Figure 21.: 35 weather stations of Canada.   Figure 22.: Edmonton's station ($\boldsymbol{s}_0$).

In order to compare the estimation method proposed in this paper with the method presented by Giraldo, Delicado, and Mateu (2011), we estimate the curve at each location in the dataset using all remaining curves in the dataset and for each station we have computed the following mean square error measure

$$mse_i = \frac{1}{365} \sum_{j=1}^{365} (\chi_{\boldsymbol{s}_i}(t_j) - \hat{\chi}_{\boldsymbol{s}_i}(t_j))^2, \quad i = 1, \dots, 35,$$

where $\hat{\chi}_{\boldsymbol{s}_i}(t_j), j = 1, \dots, 365$ is the estimated curve and $\chi_{\boldsymbol{s}_i}(t_j), j = 1, \dots, 365$ is the observed time series, where $t_j = -\pi + \frac{2\pi}{365-1}(j-1)$, $j = 1, \dots, 365$. We present the results in the Table 3, and we conclude that our estimation method is a competitive estimation procedure for this dataset when compared to established estimation procedure of Giraldo, Delicado, and Mateu (2011). The Figure 22 shows the location of Edmonton's station in Canada. We show the box plot for $mse_i, i = 1, \dots, 35$ in the Figure 23 and we also show the box plot for $\log(mse_i), i = 1, \dots, 35$ in the Figure 24 in order to give a better insight of results.

Table 3.: Summary of *mse* of each station. **New** refers to the new methodology proposed in this work, and **OKFD** refers to metodology proposed by Giraldo, Delicado, and Mateu (2011).

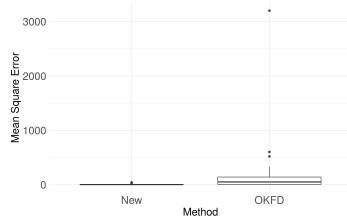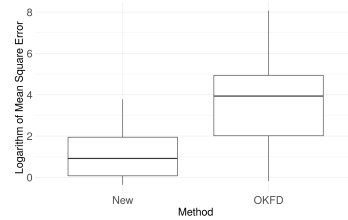|	New	OKFD
Mean of $mse_i, i = 1, \ldots, 35$	6.59	189.67

Figure 23.: Box plot for mse_i.

Figure 24.: Box plot for $\log(mse_i)$.

As example, we considered the station located at Edmonton as s_0 and use all the remaining stations in the dataset to estimate the daily mean temperature curve in Edmonton's station. The Figure 25 shows the estimated curve using our proposal and the observed time series.
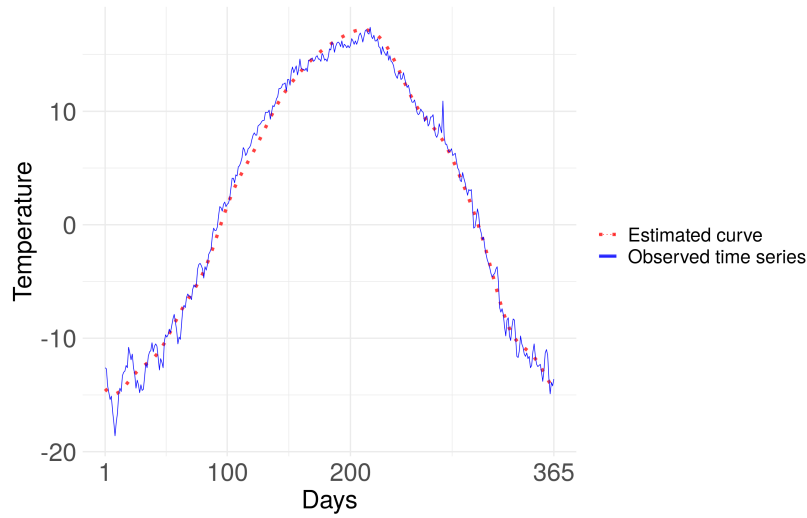

Figure 25.: Estimated curve at Edmonton station.

18

6. Conclusion

In this paper, our aim is to estimate the curve $\chi_{\boldsymbol{s}_0}(t)$ at a location $\boldsymbol{s}_0$ out of observed sample. We have proposed a simple model, where the $\chi_{\boldsymbol{s}_0}(t)$ is a linear combination of all curves where the weights are chosen to give an unbiased estimate with minimum expected square error. It was proved by Giraldo, Delicado, and Mateu (2011) that these weights are solution of a linear system where the independent and coefficient matrix use the trace-variogram. Originally, an empirical version of trace-variogram was used in the estimation procedure of trace-variogram. Here, we propose an alternative method using the Legendre-Gauss quadrature. In our proposal, we use the structure of trace-variogram, whose definition involves an integration, to reduce the problem to estimate the trace-variogram in a problem to estimate a semivariogram in Geostatistics in some quadrature points. We compare our estimation method with established estimation procedure using simulated datasets with $25, 50, 75$ e 100 curves and assuming pointwise normality and nonnormality of theses curves. In all simulated scenaries, our approach has similar bias, but the mean square error has a better performance. We finish illustrating the results with a real dataset: 35 curves of temperature from the maritime provinces. We use a out-of-sample measure to compare the two estimation method, where we observed that the our proposal has a better performance in agreement with our simulation study. Additionally, future works will adapt the new estimation procedure of an unmonitered curve presented in this paper for other basis functions including B-splines and wavelets.

References

Basu, Ayanendranath, Ian R Harris, Nils L Hjort, and MC Jones. 1998. "Robust and efficient estimation by minimising a density power divergence." *Biometrika* 85 (3): 549–559.

Beyaztas, Ufuk, and Han Lin Shang. 2019. "Forecasting functional time series using weighted likelihood methodology." *Journal of Statistical Computation and Simulation* 89 (16): 3046–3060.

Caballero, William, Ramón Giraldo, and Jorge Mateu. 2013. "A Universal Kriging Approach for Spatial Functional Data." *Stochastic Environmental Research and Risk Assessment* 27

(7): 1553–1563.

Chen, Yakuan, Jeff Goldsmith, and R. Todd Ogden. 2019. "Functional Data Analysis of Dynamic PET Data." *Journal of the American Statistical Association* 114 (526): 595–609.

Fang, Kuangnan, Xiaochen Zhang, Shuangge Ma, and Qingzhao Zhang. 2020. "Smooth and locally sparse estimation for multiple-output functional linear regression." *Journal of Statistical Computation and Simulation* 90 (2): 341–354.

Ferraty, Frédéric, and Philippe Vieu. 2006. *Nonparametric Functional Data Analysis: Theory and Practice.* Springer-Verlag.

Gabrys, Robertas, and Piotr Kokoszka. 2007. "Portmanteau test of independence for functional observations." *Journal of the American Statistical Association* 102 (480): 1338–1348.

Giraldo, R, P Delicado, and J Mateu. 2011. "Ordinary Kriging for Function-Valued Spatial Data." *Environmental and Ecological Statistics* 18 (3): 411–426.

Giraldo, Ramón, Pedro Delicado, and Jorge Mateu. 2010. "Continuous Time-Varying Kriging for Spatial Prediction of Functional Data: An Environmental Application." *Journal of agricultural, biological, and environmental statistics* 15 (1): 66–82.

Giraldo, Ramón, Jorge Mateu, and Pedro Delicado. 2012. "`geofd`: an R Package for Function-Valued Geostatistical Prediction." *Revista Colombiana de Estadística* 35 (3): 385–407.

Horváth, Lajos, and Piotr Kokoszka. 2012. *Inference for Functional Data with Applications.* Vol. 200. Springer-Verlag.

Ignaccolo, Rosaria, Jorge Mateu, and Ramon Giraldo. 2014. "Kriging with External Drift for Functional Data for Air Quality Monitoring." *Stochastic Environmental Research and Risk Assessment* 28 (5): 1171–1186.

Khuri, André I. 2003. *Advanced calculus with applications in statistics.* Vol. 486. John Wiley & Sons.

Lee, D-J, Z Zhu, and P Toscas. 2015. "Spatio-Temporal Functional Data Analysis for Wireless Sensor Networks Data." *Environmetrics* 26 (5): 354–362.

Lee, Wonyul, Michelle F Miranda, Philip Rausch, Veerabhadran Baladandayuthapani, Massimo Fazio, J Crawford Downs, and Jeffrey S Morris. 2019. "Bayesian semiparametric functional mixed models for serially correlated functional data, with application to glaucoma data." *Journal of the American Statistical Association* 114 (526): 495–513.

Menafoglio, Alessandra, Alberto Guadagnini, and Piercesare Secchi. 2014. "A Kriging Approach Based on Aitchison Geometry for the Characterization of Particle-Size Curves in Heterogeneous Aquifers." *Stochastic Environmental Research and Risk Assessment* 28 (7):

1835–1851.

Nerini, David, Pascal Monestiez, and Claude Manté. 2010. "Cokriging for Spatial Functional Data." *Journal of Multivariate Analysis* 101 (2): 409–418.

Ramsay, J. O., Hadley Wickham, Spencer Graves, and Giles Hooker. 2018. *fda: Functional Data Analysis*. R package version 2.4.8.

Ramsay, James O. 2006. *Functional Data Analysis*. John Wiley & Sons.

Ramsay, James O, and CJ Dalzell. 1991. "Some Tools for Functional Data Analysis." *Journal of the Royal Statistical Society B* 53 (3): 539–572.

Reyes, Adriana, Ramón Giraldo, and Jorge Mateu. 2012. "Residual Kriging for Functional Data. Application to the Spatial Prediction of Salinity Curves." *Revista Colombiana de Estadística* 35 (3): 385–407.

Salazar, Elías, Ramón Giraldo, and Emilio Porcu. 2015. "Spatial Prediction for Infinite-Dimensional Compositional Data." *Stochastic Environmental Research and Risk Assessment* 29 (7): 1737–1749.

Scott, David W. 2009. "Sturges' rule." *Wiley Interdisciplinary Reviews: Computational Statistics* 1 (3): 303–306.

Vale, C. David, and Vincent A. Maurelli. 1983. "Simulating multivariate nonnormal distributions." *Psychometrika* 48 (3): 465–471.

Zamani, Atefeh, Maryam Hashemi, and Hossein Haghbin. 2019. "Improved functional portmanteau tests." *Journal of Statistical Computation and Simulation* 89 (8): 1423–1436.