Estimation of Trace-variogram Using Legendre-Gauss Quadrature

Gilberto Sassi \textsuperscript{a}, and Chang Chian\textsuperscript{b}

\textsuperscript{a}Universidade Federal de Bahia
\textsuperscript{b}Universidade de São Paulo

Abstract. Functional Data Analysis is known for its application in several fields of science. In some cases, functional datasets are constituted by spatially indexed curves. The primary goal of this paper is to supply a straightforward and precise approach to interpolate these curves, i.e., the aim is to estimate a curve at an unmonitored location. It is proven that the best linear unbiased estimator for this unsampled curve is the solution of a linear system, where the coefficients and the constant terms of the system are formed using a function called trace-variogram. In this paper, we propose using Legendre-Gauss quadrature to estimate the trace-variogram. This estimator’s suitable numerical properties are shown in simulation studies for normal and non-normal datasets. Simulation results indicated that the proposed methodology outperforms the established estimation procedure. An \texttt{R} package was built and is available at the CRAN repository. The novel estimation methodology is illustrated with a real dataset on temperature curves from 35 weather stations in Canada.

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). It is suitable for applications where we need to analyze an observation from a family \( \{X(t_j)\}_{j=1}^{J} \), where \( t_1, \ldots, t_J \) may be equally spaced and \( t_j \in (t_{\min}, t_{\max}), \forall j = 1, \ldots, J \). When the interval between \( t_j \) and \( t_{j+1} \) gets smaller, we could consider this observation as sampled from a random continuous family \( \chi = \{X(t) \mid t \in (t_{\min}, t_{\max})\} \). Furthermore, there are cases where the underlying data are clearly a 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 functional linear regression for multivariate responses and developed a locally sparse estimation for the functional coefficients. 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 a 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) analyzed a glaucoma scleral strain dataset using a Bayesian functional mixed model capable of detecting nonparametric covariate effects and serial and nested interfunctional correlation. Zamani, Hashemi and Haghbin (2019) proposed an improvement in the Portmanteau test of Gabrys and Kokoszka (2007) 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 available curves. In Giraldo, Delicado and Mateu (2011), the weights of this linear combination are scalars; in Giraldo, Delicado and Mateu (2010) the weights are 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. 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 formulating 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 in Geostatistics. Salazar, Giraldo and Porcu (2015) generalized the method of kriging proposed by Giraldo, Delicado and Mateu (2010) and Giraldo, Delicado and Mateu (2011).

In this paper, we assume the observations to be \( n \) curves \( \chi_{s_1}(t), \ldots, \chi_{s_n}(t) \) at locations \( s_1, \ldots, s_n \) in a specified region, where \( s_i = (\theta_i, \eta_i) \), \( i = 1, \ldots, n \), \( \theta_i \) is the latitude and \( \eta_i \) is the longitude. The aim is to estimate the unobserved curve \( \chi_{s_0}(t) \) where \( s_0 \notin \{s_1, \ldots, s_n\} \). The method proposed by Giraldo, Delicado and Mateu (2011), and referred to in this paper as Functional Ordinary Kriging Predictor (FOKP), is straightforward: the curve \( \chi_{s_0}(t) \) at \( s_0 \notin \{s_1, \ldots, s_n\} \) is a linear combination of all the curves \( \chi_{s_1}(t), \ldots, \chi_{s_n}(t) \), i.e.,

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

where \( \lambda_1, \ldots, \lambda_n \) is the solution of the linear system

\[
\begin{pmatrix}
\gamma(||s_1 - s_1||) & \gamma(||s_1 - s_2||) & \cdots & \gamma(||s_1 - s_n||) \\
\gamma(||s_2 - s_1||) & \gamma(||s_2 - s_2||) & \cdots & \gamma(||s_2 - s_n||) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(||s_n - s_1||) & \gamma(||s_n - s_2||) & \cdots & \gamma(||s_n - s_n||)
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_n
\end{pmatrix} =
\begin{pmatrix}
\gamma(||s_1 - s_0||) \\
\gamma(||s_2 - s_0||) \\
\vdots \\
\gamma(||s_n - s_0||)
\end{pmatrix},
\]

where \( \mu \) is a constant from the method of Lagrange multipliers, the function

\[
\gamma(h) = \int \gamma(h;t)dt,
\]

is called trace-variogram, and \( \gamma(h,t) \) is the semivariogram as defined in Section 3. More precisely, for each \( t \), a weakly stationary and isotropic spatial process is assumed, and the semivariogram can be computed. Furthermore, the trace-variogram is an integration of \( \gamma(h,t) \) over \( t \). Usually, the integral in equation (1) is approximated using a modified version of the empirical semivariogram (see Giraldo, Delicado and Mateu, 2011, for more details).

In this paper, we propose to estimate the trace-variogram using the Legendre-Gauss quadrature. We apply the exponential model to approximate the semivariogram for each quadrature point. The main advantage of the Legendre-Gauss quadrature is that a very high-order accuracy can be obtained with just a few points, often fewer than 20. This approach produces a smaller mean squared error, as seen from the simulation results and the real dataset application on temperature from 35 weather stations in Canada. All \texttt{R} codes are available in a GitHub directory \texttt{gilberto-sassi/geoFourierFDA}. We have also organized an \texttt{R} package called \texttt{geoFourierFDA} available in the CRAN repository.
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 Functional Ordinary Kriging Predictor 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 on temperature curves from 35 stations in 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, in Functional Data Analysis observations are viewed as functions defined over a set \(B \subseteq \mathbb{R}^p, p \in \mathbb{N}\), usually \(p \in \{1, 2\}\) Before we consider the kriging methods for spatial functional data in Section 3, we formally introduce some concepts on a functional variable, functional data, and functional dataset (see Ferraty and Vieu, 2006, for more details).

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

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

An observation \(\chi\) of \(\chi\) is referred to as functional data.

**Remark.**

1. When \(p = 1\), a functional variable is called a random curve, and functional data is called a curve.
2. We denote a functional variable by \(\chi(t)\) and 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 \(\chi_1(t), \ldots, \chi_n(t)\).

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. This paper uses the Fourier polynomials to smooth these curves as explained in Section 3.

3 A Kriging Method for Functional Data

This sections presents the problem of estimating a curve \(\chi_{s_0}(t)\) at an unmonitored location \(s_0\). First, we begin explaining the Functional Ordinary Kriging Predictor (FOKP) for functional datasets, where the curves are smoothed using Fourier polynomials. Then, we present our proposal to estimate the trace-variogram used to approximated \(\chi_{s_0}(t)\) as established in Theorem 3.1.

This paper assumes that we have a pointwise isotropic and weakly stationary spatial process. More precisely, let \(s_i = (\theta_i, \eta_i) \in \mathbb{R}^2, i = 1, \ldots, n\), be locations of a compact region \(D \subseteq \mathbb{R}^2\), where \(\theta_i\) and \(\eta_i\) are the latitude and the longitude, respectively, and \(\chi_{s_i}(t), t \in T\), are square-integrable curves, then we have

1. \(E[\chi_{s}(t)] = \mu(t), \forall t \in T, \forall s \in D\),
2. \(\text{Cov}(\chi_{s_i}(t), \chi_{s_j}(t)) = C(h; t), \forall s_i, s_j \in D, \forall t \in T\) and \(h = ||s_i - s_j||\), where \(||s_i - s_j||\) is the Euclidean distance between \(s_i\) and \(s_j\).
3. \( \frac{1}{2} \text{Var}(\chi_{s_i}(t) - \chi_{s_j}(t)) = \gamma(h; t), \forall s_i, s_j \in D, \forall t \in T, h = \|s_i - s_j\|, \) 
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 the goal is to obtain an estimate for the curve \( \chi_{s_0}(t) \), where \( s_0 \not\in \{s_1, \ldots, s_n\} \). Then the estimate of \( \chi_{s_0}(t) \) is given by:

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

(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

\[
\arg \min_{\lambda_1, \ldots, \lambda_n} \mathbb{E} \left[ \|\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\|_2^2 \right] \quad \text{subject to} \quad \mathbb{E} [\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)] = 0, \forall t \in T,
\]

(3)

where \( \|f(t)\|_2^2 = \int_T |f(t)|^2 dt \).

In order to solve the optimization problem described by equation (3), Theorem 3.1, proved by Giraldo, Delicado and Mateu (2011), establishes an equivalence between solving this optimization problem and solving a linear system.

**Theorem 3.1.** Under some regularity conditions, solving the nonlinear optimization problem

\[
\arg \min_{\lambda_1, \ldots, \lambda_n} \mathbb{E} \left[ \|\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\|_2^2 \right] \quad \text{subject to} \quad \mathbb{E} [\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)] = 0, \forall t \in T,
\]

is equivalent to solving

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

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

The first thing we have to note in Theorem 3.1 is that, in practice, for each location \( s_i, i = 1, \ldots, n \), we observe a time series \( \chi_{s_i}(t_j), j = 1, \ldots, m \), where \( t_j \in [-\pi, \pi] \). If we assume \( t \in [-\pi, \pi] \) and that the curves can be approximated by a Fourier polynomial given by

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

(4)

for suitable values of \( a_0^i, a_1^i, \ldots, a_p^i, b_1^i, \ldots, b_p^i \), \( i = 1, \ldots, n \). There are several ways to choose \( p \). 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 than or equal to \( x \), and \( m \) is the length of the observed time series \( \chi_{s_i}(t_1), \ldots, \chi_{s_i}(t_m), i = 1, \ldots, n \).

Coefficients \( a_0^i, a_1^i, \ldots, a_p^i, b_1^i, b_2^i, \ldots, b_p^i \) are least squared error estimates, i.e., they minimize the following sum

\[
S(a_0^i, a_1^i, a_2^i, \ldots, a_p^i, b_1^i, b_2^i, \ldots, b_p^i) = \sum_{l=1}^{m} \left[ \chi_{s_i}(t_l) - \left( \frac{a_0^i}{2} + \sum_{k=1}^{p} \left[ a_k^i \cos(k \cdot t_l) + b_k^i \sin(k \cdot t_l) \right] \right) \right]^2.
\]
For simplicity, we use a vectorial notation:
\[ \chi_{s_i}(t) = \mathbf{c}_i^\top \varphi(t), \]
where
\[ \mathbf{c}_i = (a_{i0}, a_{i1}, a_{i2}, \ldots, a_{ip}, b_{i1}, b_{i2}, \ldots, b_{ip})^\top, \]
\[ \varphi(t) = \left( \frac{1}{2}, \cos(t), \cos(2t), \ldots, \cos(pt), \sin(t), \sin(2t), \ldots, \sin(pt) \right)^\top. \]

Another critical aspect of Theorem 3.1 is the estimation of the trace-variogram. Giraldo, Delicado and Mateu (2011) have estimated the trace-variogram using a modified version of the empirical semivariogram. In this paper, we refer to the Functional Ordinary Kriging Predictor using this estimate of trace-variogram as Ordinary Kriging for Functional Data (OKFD).

Next, we present an alternative approach to estimate the trace-variogram, which has a smaller mean square error than OKFD.

In order to estimate the trace-variogram, we have to solve the following integration
\[ \gamma(h) = \int_T \gamma(h,t) dt. \] (5)

Let \( f : [a, b] \rightarrow \mathbb{R} \) be a continuous function, then the integral over \([a, b]\) can be approximated by
\[ \int_a^b f(u) du \approx \frac{b - a}{2} \sum_{k=0}^s \omega_k f \left( \frac{(b - a) z_k + b + a}{2} \right), \]
where \( z_0, z_1, z_2, \ldots, z_s \) are the zeros of the Legendre polynomial \( p_{s+1}(x) \) (Khuri, 2003, pp. 441) of degree \( s + 1 \) and \( \omega_0, \omega_1, \omega_2, \ldots, \omega_s \) are solutions of the following linear system
\[ \int_{-1}^{1} u^j du = \sum_{k=0}^s \omega_k z_k^j, \quad j = 0, 1, \ldots, s. \]

Consequently, we can approximate the trace-variogram by
\[ \gamma(h) = \int_{-\pi}^{\pi} \gamma(h,t) dt \approx \pi \sum_{k=0}^s w_k \gamma(h, \pi z_k). \] (6)

Under the pointwise isotropic and weakly stationary spatial assumption, for each \( t \in [-\pi, \pi] \), we can estimate the semivariogram using the exponential model given by
\[ \gamma(h,t) = \sigma_t^2 \left( 1 - \exp \left( -\frac{h}{\phi_t} \right) \right), \]
where \( \sigma_t^2 > 0 \) and \( \phi_t > 0 \). In other words, for a fixed value \( t \in [-\pi, \pi] \), we can estimate \( \sigma_t^2 \) and \( \phi_t \) using the sample \( \chi_{s_i}(t) = c_1 \varphi(t); \chi_{s_2}(t) = c_2 \varphi(t); \ldots; \chi_{s_n}(t) = c_n \varphi(t) \). If we knew the semivariogram \( \gamma(\cdot, t) \) at the points \( t = \pi z_1, \pi z_2, \ldots, \pi z_s \), we could approximate the trace-variogram \( \gamma(\cdot) \) using equation (6). More precisely, for each \( t = \pi z_k \), we have the following sample
\[ X_{s_1}(\pi z_k) = c_1^\top \varphi(\pi z_k); X_{s_2}(\pi z_k) = c_2^\top \varphi(\pi z_k); \ldots; X_{s_n}(\pi z_k) = c_n^\top \varphi(\pi z_k), \quad k = 1, \ldots, s, \]
and

\[
\begin{pmatrix}
X_{s_1}(\pi z_k) \\
X_{s_2}(\pi z_k) \\
\vdots \\
X_{s_n}(\pi z_k)
\end{pmatrix} \sim N(\mu_k; \Sigma_k),
\]

where

\[
\mu_k = \\
\begin{pmatrix}
\mu_{\pi z_k} \\
\mu_{\pi z_k} \\
\vdots \\
\mu_{\pi z_k}
\end{pmatrix}; \quad \Sigma_k = \\
\begin{pmatrix}
c_{11} & c_{12} & \cdots & c_{1n} \\
c_{21} & c_{22} & \cdots & c_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
c_{n1} & c_{n2} & \cdots & c_{nn}
\end{pmatrix}
\]

and \(c_{ij}^k = \sigma_{\pi z_k}^2 \exp\left(-\frac{||s_i - s_j||}{\phi_{\pi z_k}}\right), \ i, j = 1, 2, 3, \ldots, n\). We estimate \(\sigma_{\pi z_k}, \mu_{\pi z_k}\) and \(\phi_{\pi z_k}\) using the maximum likelihood estimate where the likelihood function is given by

\[
l(\sigma_{\pi z_k}, \phi_{\pi z_k}, \mu_{\pi z_k}) = \det(2\pi \Sigma_k)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(X - \mu_k)^\top \Sigma_k^{-1}(X - \mu_k)\right), \quad (7)
\]

where \(X = (X_{s_1}(\pi z_k) \ X_{s_2}(\pi z_k) \ \cdots \ X_{s_n}(\pi z_k))^\top\). Finally, the trace-variogram is approximated by

\[
\gamma(h) \approx \pi \sum_{k=0}^{s} \omega_k \sigma_{\pi z_k}^2 \left(1 - \exp\left(-\frac{h}{\phi_{\pi z_k}}\right)\right). \quad (8)
\]

All methods explained in this section were implemented in R language and are available at a GitHub directory gilberto-sassi/geoFourierFDA. Furthermore, all codes have been organized as an R package named geoFourierFDA available at the CRAN repository.

## 4 Simulation Study

In this section, we evaluate the proposed methodology for estimating the curve \(\chi_{s_0}(t)\) proposed in Section 3. We assume normal and non-normal datasets, and we computed mean square error and bias of estimates.

First, we generate \(n\) locations \(s_1, s_2, \ldots, s_n\) on \(D = [-5, 5] \times [-5, 5]\), i.e., the latitude \(\theta\) ranges from \(-5\) to \(5\) and the longitude \(\eta\) ranges from \(-5\) to \(5\), and the aim is to estimate the curve at location \(s_0 = (\theta_0, \eta_0) = (0, 0)\). In order to simulate the curves at locations \(s_1, s_2, \ldots, s_n, s_0\), we sampled a random matrix

\[
M = \\
\begin{pmatrix}
m_{0,1} & m_{0,2} & m_{0,3} & \cdots & m_{0,n+1} \\
m_{1,1} & m_{1,2} & m_{1,3} & \cdots & m_{1,n+1} \\
m_{2,1} & m_{2,2} & m_{2,3} & \cdots & m_{2,n+1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
m_{n+1,1} & m_{n+1,2} & m_{n+1,3} & \cdots & m_{n+1,n+1}
\end{pmatrix}, \quad (9)
\]

where \(s_{n+1} = s_0\). We also simulated a vector \(\alpha = (\alpha_0 \ \alpha_1 \ \cdots \ \alpha_p \ \alpha_{p+1} \ \cdots \ \alpha_{2p})^\top\) from the multivariate normal distribution \(N(0, I_{2p+1})\).
4.1 Results from Simulation in the First Scenario – Normality

Then simulated curves are given by

\[
\chi_{s_i}(t) = \frac{a_0}{2} + \sum_{k=1}^{p} (\alpha_k \cos(k \cdot t) + \alpha_{k+p} \sin(k \cdot t)) + \\
\frac{m_{a,i}}{2} + \sum_{k=1}^{p} (m_{k,i} \cos(k \cdot t) + m_{k+p,i} \sin(k \cdot t)) + \epsilon_t, \quad \forall t \in [-\pi, \pi], \ i = 1, \ldots, n,
\]

and

\[
\chi_{s_0}(t) = \frac{a_0}{2} + \sum_{k=1}^{p} (\alpha_k \cos(k \cdot t) + \alpha_{k+p} \sin(k \cdot t)) + \\
\frac{m_{a,n+1}}{2} + \sum_{k=1}^{p} (m_{k,n+1} \cos(k \cdot t) + m_{k+p,n+1} \sin(k \cdot t)) + \epsilon_t, \quad \forall t \in [-\pi, \pi],
\]

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

Finally, we build time series \((\chi_{s_i}(t_1) \chi_{s_i}(t_2) \cdots \chi_{s_i}(t_m))^\top\) for each location \(s_i, \ i = 0, 1, \ldots, n\), and we use these time series \((\chi_{s_i}(t_1) \chi_{s_i}(t_2) \cdots \chi_{s_i}(t_m))^\top, \ i = 1, \ldots, n\), to estimate the curve \(\chi_{s_0}(t)\), and then we compute a mean square error and a bias measures given by

\[
mse = \frac{1}{m} \sum_{k=1}^{m} (\chi_{s_0}(t_k) - \hat{\chi}_{s_0}(t_k))^2,
\]

\[
bias = \frac{1}{m} \sum_{k=1}^{m} (\chi_{s_0}(t_k) - \hat{\chi}_{s_0}(t_k)).
\]

In this paper, we used two scenarios: normality and non-normality assumptions. In the first scenario, for each \(t \in [-\pi, \pi]\) and for each \(s \in D\), \(\chi_s(t)\) has a normal distribution, and in the second scenario, for each \(t \in [-\pi, \pi]\) and for each \(s \in D\), \(\chi_s(t)\) does not have a normal distribution. In the first scenario, each column \(m_{k,i} = (m_{k,1} m_{k,2} \cdots m_{k,n+1})^\top\) of the matrix \(M\) from equation (9) is sampled from \(N(0, \Sigma_k)\), \(k = 0, 1, \ldots, 2p\), where \(\Sigma_k = (\sigma_{i,j}^k)_{i,j=1}^{n+1}\) is a covariance matrix with

\[
\sigma_{i,j}^k = \exp(-2 \cdot \|s_i - s_j\|), \quad i, j = 1, \ldots, n, n+1.
\]

In the second scenario, each column \(m_{k,i} = (m_{k,1} m_{k,2} \cdots m_{k,n+1})^\top\) of the matrix \(M\) from equation (9) is sampled from the non-normal distribution proposed by Vale and Maurelli (1983) with skewness equals to \(-2\), kurtosis equals to \(2\), and with the covariance matrix given by \(\Sigma_k = (\sigma_{i,j}^k)_{i,j=1}^{n+1}\) where

\[
\sigma_{i,j}^k = \exp(-2 \cdot \|s_i - s_j\|), \quad i, j = 1, \ldots, n, n+1.
\]

In the simulation study, we used \(n = 25, 50, 75, 100\) locations and time series with length \(m = 1000\). For each \(n\), we made \(L = 1000\) replications.

4.1 Results from Simulation in the First Scenario – Normality

In Table 1, we present the mean and the standard deviation 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.
Table 1 Mean and standard deviation (SD) of $\text{mse}_l, l = 1, \ldots, 1000,$ and $\text{bias}_l, l = 1, \ldots, 1000,$ for normal simulated data.

<table>
<thead>
<tr>
<th>Number of locations</th>
<th>Mean (SD) of $\text{mse}$</th>
<th>Mean (SD) of $\text{bias}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New$^1$</td>
<td>OKFD$^2$</td>
</tr>
<tr>
<td>25</td>
<td>10.8 (3.34)</td>
<td>17.4 (11.02)</td>
</tr>
<tr>
<td>50</td>
<td>9.95 (3.21)</td>
<td>16.47 (5.08)</td>
</tr>
<tr>
<td>75</td>
<td>9.03 (3.39)</td>
<td>15.98 (5.84)</td>
</tr>
<tr>
<td>100</td>
<td>8.7 (3.27)</td>
<td>14.82 (4.31)</td>
</tr>
</tbody>
</table>

$^1$ New refers to Functional Ordinary Kriging Predictor using the trace-variogram estimation method presented in this work.

$^2$ OKFD refers to Functional Ordinary Kriging Predictor using the trace-variogram estimation method proposed by Giraldo, Delicado and Mateu (2011).

4.2 Results from Simulation in the Second Scenario – Non-normality

In Table 2, we present the mean and the standard deviation of $\text{mse}$ and $\text{bias}$ for the $L = 1000$ replications. The $\text{bias}$ is similar in both approaches, but, even in this context of non-normality, the estimation method for $\chi_{s_0}(t)$ using the trace variogram proposed in this paper has a smaller $\text{mse}$.

Table 2 Mean and standard deviation (SD) of $\text{mse}_l, l = 1, \ldots, 1000,$ and $\text{bias}_l, l = 1, \ldots, 1000,$ for non-normal simulated data.

<table>
<thead>
<tr>
<th>Number of locations</th>
<th>Mean (SD) of $\text{mse}$</th>
<th>Mean (SD) of $\text{bias}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New$^1$</td>
<td>OKFD$^2$</td>
</tr>
<tr>
<td>25</td>
<td>13.23 (5.07)</td>
<td>21.58 (17.91)</td>
</tr>
<tr>
<td>50</td>
<td>12.44 (4.86)</td>
<td>19.01 (6.66)</td>
</tr>
<tr>
<td>75</td>
<td>11.91 (5.18)</td>
<td>18.43 (7.07)</td>
</tr>
<tr>
<td>100</td>
<td>11.59 (4.99)</td>
<td>18.75 (7.12)</td>
</tr>
</tbody>
</table>

$^1$ New refers to Functional Ordinary Kriging Predictor using the trace-variogram estimation method presented in this work.

$^2$ OKFD refers to Functional Ordinary Kriging Predictor using the trace-variogram estimation method proposed by Giraldo, Delicado and Mateu (2011).
5 Application

We illustrate the ideas presented in this paper with a dataset containing daily measurements of mean temperature recorded at 35 weather stations in Canada. This dataset is included in the R packages `geofd` (Giraldo, Mateu and Delicado, 2012) and `fda` (Ramsay et al., 2018), and it consists of 35 time series where each time series is a daily mean temperature that can be downloaded at the website weather.gc.ca. This dataset is also available with the R package `geoFourierFDA`, where the proposed methods in this paper have been implemented.

In order to compare the estimation method proposed in this paper with the OKFD estimation method, we use a cross-validation measure proposed by Giraldo, Delicado and Mateu (2011), which is given by

$$\text{mse}_i = \frac{1}{365} \sum_{j=1}^{365} (\chi_{s_i}(t_j) - \hat{\chi}_{s_i}(t_j))^2, \quad i = 1, \ldots, 35,$$

where $\hat{\chi}_{s_i}(t_j), j = 1, \ldots, 365$, is the estimated curve, $\chi_{s_i}(t_j), j = 1, \ldots, 365$, is the observed time series, and $t_j = -\pi + \frac{2\pi}{365 - 1} (j - 1), j = 1, \ldots, 365$. The results presented in Table 3 show that the newly proposed estimation method is competitive compared to the OKFD estimation method.

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD) of $\text{mse}_i$, $i = 1, \ldots, 35$</th>
</tr>
</thead>
<tbody>
<tr>
<td>New 1</td>
<td>6.59 (9.57)</td>
</tr>
<tr>
<td>OKFD 2</td>
<td>189.67 (542.89)</td>
</tr>
</tbody>
</table>

1*New* refers to Functional Ordinary Kriging Predictor using the trace-variogram estimation method presented in this work.

2*OKFD* refers to Functional Ordinary Kriging Predictor using the trace-variogram estimation method proposed by Giraldo, Delicado and Mateu (2011).

6 Conclusion

This paper aims to estimate the curve $\chi_{s_0}(t)$ at a location $s_0$ out of observed sample. The model in this paper is straightforward: $\chi_{s_0}(t)$ is a linear combination of all curves where the weights are chosen to give an unbiased estimate with minimum expected square error. The curves in the functional dataset were smoothed using Fourier polynomials. It was proved by Giraldo, Delicado and Mateu (2011) that these weights are the solution of a linear system of equations $Ax = b$, where the matrix of coefficients $A$ and the column vector of solutions $b$ are computed using the trace-variogram function. Originally, a modified version of the empirical semi-variogram was used in the estimation procedure of trace-variogram. Here, we propose an alternative method using the Legendre-Gauss quadrature. We have compared the new estimation method with the established estimation procedure using simulated datasets with 25, 50, 75 and 100 curves, and we have assumed pointwise normality and non-normality of these curves. The new approach has a similar bias in all simulated scenarios, but the mean square error has a better performance. The proposed estimation method was successfully
applied to a real dataset: 35 temperature curves in Canada. We applied a cross-validation measure to compare the two estimation methods, from which we observed that the new proposal has a better performance in agreement with the simulation study. All R codes are available in a GitHub directory `gilberto-sassi/geoFourierFDA`, and we have also organized an R package called `geoFourierFDA` that is available in the CRAN repository. In the future, the method presented in this paper can be adapted for other basis functions, including B-Splines and Wavelets.

**Bibliography**


