ARTICLE TEMPLATE

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

Gilberto Pereira Sassi[a] and Chang Chiann[b]

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

ARTICLE HISTORY
Compiled March 6, 2020

Abstract
Functional Data Analysis 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., the aim is to estimate a curve at an unmonitored location. It has been proved 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 the use of Legendre-Gauss quadrature to estimate the trace-variogram. Excellent numerical properties of this estimator were showed in simulation studies for normality dataset and non-normality dataset: smaller mean square error compared with the established estimation procedure. 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

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 [19], and it is suitable for applications where we need to analyze an observation from a family $\{X(t_j)\}_{j=1,\ldots,J}$ 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 [6]. Since the seminal paper of Ramsay Ramsay and Dalzell [19], 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. [5] considered the functional linear regression for multivariate responses and developed a locally sparse estimation for the functional coefficients. Chen et al. [4] 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 [2] proposed a robust method

CONTACT Gilberto Pereira Sassi. Email: sassi.pereira.gilberto@gmail.com

to forecast functional time series based on the minimum density power divergence estimator of Basu et al. [1]. Lee et al. [15], 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 et al. [25] proposed a improvement in the Portmanteau test of independency of Gabrys and Kokoszka [7] of functional observations, which is specially suited to small samples.

In Geostatistics and Functional Data Analysis, Giraldo et al. [8] and Giraldo et al. [9] 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 et al. [9], the weights of this linear combination is a scalar and this method is referred as Ordinary Kriging; in Giraldo et al. [8] the weights are curves. In both methods, the weights are estimated by an unbiased minimum square error estimator. Caballero et al. [3] extended these models for non-stationary spatial processes and Ignaccolo et al. [12] developed a kriging method with external drift where it was possible to use exogenous variables. Menafoglio et al. [16] focused on the formulation of new geostatistical models and methods for functional compositional data. Reyes et al. [21] 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 et al. [22] introduced a predictor, similar to the Functional Ordinary Kriging from Giraldo et al. [9], where they used probability density functions with support on a finite interval. Lee et al. [14] 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. Nerini et al. [17] generalized the method of kriging proposed by Giraldo et al. [8] and Giraldo et al. [9].

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 [9] 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

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

where μ is a constant and the function

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

is called trace-variogram where $\gamma(h, t)$ 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 the equation (1) is approximated using an modified version of empirical semivariogram [see 9, 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 et al. [9] 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 [6].

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

(1) $\mathrm{E}(\chi_{\boldsymbol{s}}(t)) = \mu(t), \forall t \in T, \forall s \in D$,
(2) $\mathrm{Cov}(\chi_{\boldsymbol{s}_i}(t), \chi_{\boldsymbol{s}_j}(t)) = C(h; t), \forall \boldsymbol{s}_i, \boldsymbol{s}_j \in D, \forall t \in T$ and $h = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$, where $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ is the Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$,

[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 [18] and Horváth and Kokoszka [11] are excellent references.

(3) $\frac{1}{2}\operatorname{Var}(\chi_{\boldsymbol{s}_i}(t) - \chi_{\boldsymbol{s}_j}(t)) = \gamma(h;t), \forall \boldsymbol{s}_i, \boldsymbol{s}_j \in D, \forall t \in T, h = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|,$

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

Suppose we have a functional dataset $\chi_{\boldsymbol{s}_1}(t), \ldots, \chi_{\boldsymbol{s}_n}(t)$ at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ and the goal is to obtain an estimate for the curve $\chi_{\boldsymbol{s}_0}(t)$ where $\boldsymbol{s}_0 \notin \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$. Giraldo et al. [9] proposed the following estimator:

$$\hat{\chi}_{\boldsymbol{s}_0}(t) = \sum_{i=1}^{n} \hat{\lambda}_i \chi_{\boldsymbol{s}_i}(t), \quad \hat{\lambda}_1, \ldots, \hat{\lambda}_n \in \mathbb{R}, \tag{2}$$

where $\hat{\chi}_{\boldsymbol{s}_0}(t)$ is called functional ordinary kriging predictor, and $\hat{\lambda}_1 \ldots, \hat{\lambda}_n$ are the solution of the following nonlinear optimization problem

$$\underset{\hat{\lambda}_1,\ldots,\hat{\lambda}_n}{\arg\min}\, \mathrm{E}\left[\|\hat{\chi}_{\boldsymbol{s}_0}(t) - \chi_{\boldsymbol{s}_0}(t)\|_{L^2}^2\right] \text{ subject to } \mathrm{E}(\hat{\chi}_{\boldsymbol{s}_0}(t) - \chi_{\boldsymbol{s}_0}(t)) = 0, \forall t \in [0,1], \tag{3}$$

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 et al. [9])**.** *Under some regularity conditions, solving the nonlinear optimization problem*

$$\underset{\hat{\lambda}_1,\ldots,\hat{\lambda}_n}{\arg\min}\, \mathrm{E}\left[\|\hat{\chi}_{\boldsymbol{s}_0}(t) - \chi_{\boldsymbol{s}_0}(t)\|_{L^2}^2\right] \textit{ subject to } \mathrm{E}(\hat{\chi}_{\boldsymbol{s}_0}(t) - \chi_{\boldsymbol{s}_0}(t)) = 0, \forall t \in T,$$

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 23, 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

$$\begin{aligned} \boldsymbol{c}_i &= \left(a_0^i, a_1^i, a_2^i, \cdots, a_p^i, b_1^i, b_2^i, \cdots, b_p^i\right)^\top, \\ \boldsymbol{\varphi}(t) &= \left(\frac{1}{2}, \cos(t), \cos(2t), \cdots, \cos(pt), \sin(t), \sin(2t), \cdots, \sin(pt)\right)^\top. \end{aligned}$$

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 $\chi_{\boldsymbol{s}_1}(t) = \boldsymbol{c}_1\varphi(t); \chi_{\boldsymbol{s}_2}(t) = \boldsymbol{c}_2\varphi(t); \cdots; \chi_{\boldsymbol{s}_n}(t) = \boldsymbol{c}_n\varphi(t)$.

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

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

where $z_0, z_1, z_2, \ldots, z_s$ are the zeros of the Legendre polynomial $p_{s+1}(x)$ [see 13, pp. 441] of degree $s + 1$ and $\omega_0, \omega_1, \omega_2, \cdots, \omega_s$ are solutions of the following linear system

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

Consequently, we can approximate the trace-variogram by

$$\begin{aligned}\gamma(h) &= \int_{-\pi}^{\pi} \gamma(h,t)dt \\ &\approx \pi \sum_{k=0}^{s} w_k \gamma\left(h, \pi z_k\right). \end{aligned} \tag{5}$$

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

$$X_{\boldsymbol{s}_1} = \boldsymbol{c}_1^\top \boldsymbol{\varphi}(\pi z_k); X_{\boldsymbol{s}_2} = \boldsymbol{c}_2^\top \boldsymbol{\varphi}(\pi z_k); \cdots; X_{\boldsymbol{s}_n} = \boldsymbol{c}_n^\top \boldsymbol{\varphi}(\pi z_k), \quad k = 1, \ldots, s,$$

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

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

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

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

4. Simulation study

In this section, the 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 assume normal dataset and non-normal dataset.

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

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

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 $N(\mathbf{0}, I_{2p+1})$.

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

$$\chi_{\boldsymbol{s}_0}(t) = \frac{\alpha_0}{2} + \sum_{k=1}^{p} (\alpha_k \cos(k \cdot t) + \alpha_{k+p} \sin(k \cdot t)) + \\ + m_{o,n+1} + \sum_{k=1}^{p} (m_{k,n+1} \cos(k \cdot t) + m_{k+p,n+1} \sin(k \cdot t)) + \epsilon_t, \quad \forall t \in [-\pi, \pi],$$

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

Finally, we build time series $\begin{pmatrix} \chi_{\boldsymbol{s}_i}(t_1) & \chi_{\boldsymbol{s}_i}(t_2) & \cdots & \chi_{\boldsymbol{s}_i}(t_m) \end{pmatrix}^\top$ for each location $\boldsymbol{s}_i,\ i = 0, 1, \ldots, n$, and we use the time series $\begin{pmatrix} \chi_{\boldsymbol{s}_i}(t_1) & \chi_{\boldsymbol{s}_i}(t_2) & \cdots & \chi_{\boldsymbol{s}_i}(t_m) \end{pmatrix}^\top,\ i = 1, \ldots, n$ 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 non-normality 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 $\boldsymbol{s} \in D$, $\chi_{\boldsymbol{s}}(t)$ does not have normal distribution. In the first scenario, each column $m_{k,\cdot} = \begin{pmatrix} m_{k,1} & m_{k,2} & \cdots & m_{k,n+1} \end{pmatrix}^\top$ of the matrix M from equation (6) is sampled from $N(\mathbf{0}, \Sigma_k),\ k = 0, 1, \ldots, 2p$, 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 $m_{k,\cdot} = \begin{pmatrix} m_{k,1} & m_{k,2} & \cdots & m_{k,n+1} \end{pmatrix}^\top$ of the matrix M from equation (6) is sampled from the non-normal distribution proposed by Vale and Maurelli [24] with skeness equals to -2, kurtosis equals to 2 and with the covariance matrix given by $\Sigma_k = (\sigma^k_{i,j})^{n+1}_{i,j=1}$ where

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

In the 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 1a, 1b, 1c and 1d where we present the boxplot for mse for $L = 1000$ replications, and in the Figures 2a, 2b, 2c and 2d where we present the boxplot for $bias$ for $L = 1000$ replications.

[Table 1 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

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 non-normality, the estimation of trace-variogram proposed in this paper has equal ou smaller mse. This analysis can be visualized in the figures 3a, 3b, 3c, 3d, where we displayed the boxplot for the mse for $L = 1000$ replications, and in the figures 4a, 4b, 4c, 4d, where we displayed the boxplot for the $bias$ for $L = 1000$ replications.

[Table 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

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` [20] and `fda` [10], 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 5, we show all the weather stations considered in this paper.

[Figure 5 about here.]

In order to compare the estimation method proposed in this paper with the method presented by Giraldo et al. [9], 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

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

where $\hat{\chi}_{\boldsymbol{s}_i}(t_j), j = 1, \ldots, 365$ is the estimated curve and $\chi_{\boldsymbol{s}_i}(t_j), j = 1, \ldots, 365$ is the observed time series, where $t_j = -\pi + \frac{2\pi}{365-1}(j-1),\ j = 1, \ldots, 365$. We present the results in the Table 3, and we conclude that the new proposed estimation method is a competitive estimation procedure for this dataset when compared to established estimation procedure of Giraldo et al. [9]. We show the box plot for $mse_i, i = 1, \ldots, 35$ in the Figure 6a and we also show the box plot for $\log(mse_i), i = 1, \ldots, 35$ in the Figure 6b in order to give a better insight of results.

[Table 3 about here.]

[Figure 6 about here.]

As example, we considered the station located at Edmonton as $\boldsymbol{s}_0$ and use all the remaining stations in the dataset to estimate the daily mean temperature curve in Edmonton's station. The Figure 7a shows the location of Edmonton's station in Canada and the Figure 7b shows the estimated curve using the new proposal and the observed time series.

[Figure 7 about here.]

6. Conclusion

In this paper, the 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 et al. [9] that these weights are the solution of a linear system where the coefficients and the constant terms of this system 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 the new proposal, we use the structure of trace-variogram, whose definition involves an integration, in order to transform the problem of estimating the trace-variogram into a problem of estimating a semivariogram in Geostatistics in some quadrature points. We compare the new estimation method with established estimation procedure using simulated datasets with $25, 50, 75$ e 100 curves and assuming pointwise normality and non-normality of theses curves. In all simulated scenaries, the new 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 weather stations of Canada. We use a out-of-sample measure to compare the two estimation methods, where we observed that the new proposal has a better performance in agreement with the 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

[1] A. Basu, I. R. Harris, N. L. Hjort, and M. Jones. Robust and efficient estimation by minimising a density power divergence. Biometrika, 85(3):549–559, 1998.

[2] U. Beyaztas and H. L. Shang. Forecasting functional time series using weighted likelihood methodology. Journal of Statistical Computation and Simulation, 89(16):3046–3060, 2019.

[3] W. Caballero, R. Giraldo, and J. Mateu. A universal kriging approach for spatial functional data. Stochastic Environmental Research and Risk Assessment, 27(7):1553–1563, 2013.

[4] Y. Chen, J. Goldsmith, and R. T. Ogden. Functional data analysis of dynamic pet data. Journal of the American Statistical Association, 114(526):595–609, 2019.

[5] K. Fang, X. Zhang, S. Ma, and Q. Zhang. Smooth and locally sparse estimation for multiple-output functional linear regression. Journal of Statistical Computation and Simulation, 90(2): 341–354, 2020.

[6] F. Ferraty and P. Vieu. Nonparametric Functional Data Analysis: Theory and Practice. Springer-Verlag, 2006.

[7] R. Gabrys and P. Kokoszka. Portmanteau test of independence for functional observations. Journal of the American Statistical Association, 102(480):1338–1348, 2007.

[8] R. Giraldo, P. Delicado, and J. Mateu. Continuous time-varying kriging for spatial prediction of functional data: An environmental application. Journal of agricultural, biological, and environmental statistics, 15(1):66–82, 2010.

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

[10] R. Giraldo, J. Mateu, and P. Delicado. `geofd`: an `R` package for function-valued geostatistical prediction. Revista Colombiana de Estadística, 35(3):385–407, 2012.

[11] L. Horváth and P. Kokoszka. Inference for Functional Data with Applications, volume 200. Springer-Verlag, 2012.

[12] R. Ignaccolo, J. Mateu, and R. Giraldo. Kriging with external drift for functional data for air quality monitoring. Stochastic Environmental Research and Risk Assessment, 28(5):1171–1186, 2014.

[13] A. I. Khuri. Advanced calculus with applications in statistics, volume 486. John Wiley & Sons, 2003.

[14] D.-J. Lee, Z. Zhu, and P. Toscas. Spatio-temporal functional data analysis for wireless sensor networks data. Environmetrics, 26(5):354–362, August 2015.

[15] W. Lee, M. F. Miranda, P. Rausch, V. Baladandayuthapani, M. Fazio, J. C. Downs, and J. S. Morris. 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, 2019.

[16] A. Menafoglio, A. Guadagnini, and P. Secchi. 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, 2014.

[17] D. Nerini, P. Monestiez, and C. Manté. Cokriging for spatial functional data. Journal of Multivariate Analysis, 101(2):409–418, 2010.

[18] J. O. Ramsay. Functional Data Analysis. John Wiley & Sons, 2006.

[19] J. O. Ramsay and C. Dalzell. Some tools for functional data analysis. Journal of the Royal Statistical Society B, 53(3):539–572, 1991.

[20] J. O. Ramsay, H. Wickham, S. Graves, and G. Hooker. fda: Functional Data Analysis, 2018. R package version 2.4.8.

[21] A. Reyes, R. Giraldo, and J. Mateu. Residual kriging for functional data. application to the spatial prediction of salinity curves. Revista Colombiana de Estadística, 35(3):385–407, 2012.

[22] E. Salazar, R. Giraldo, and E. Porcu. Spatial prediction for infinite-dimensional compositional data. Stochastic Environmental Research and Risk Assessment, 29(7):1737–1749, October 2015.

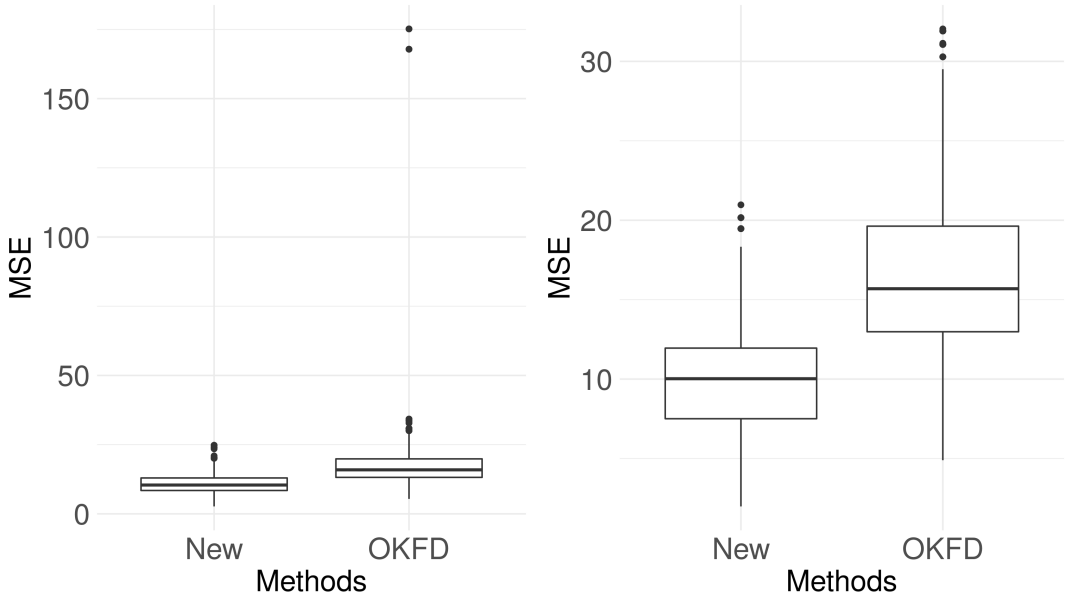
[23] D. W. Scott. Sturges' rule. Wiley Interdisciplinary Reviews: Computational Statistics, 1(3): 303–306, 2009.

[24] C. D. Vale and V. A. Maurelli. Simulating multivariate nonnormal distributions. Psychometrika,

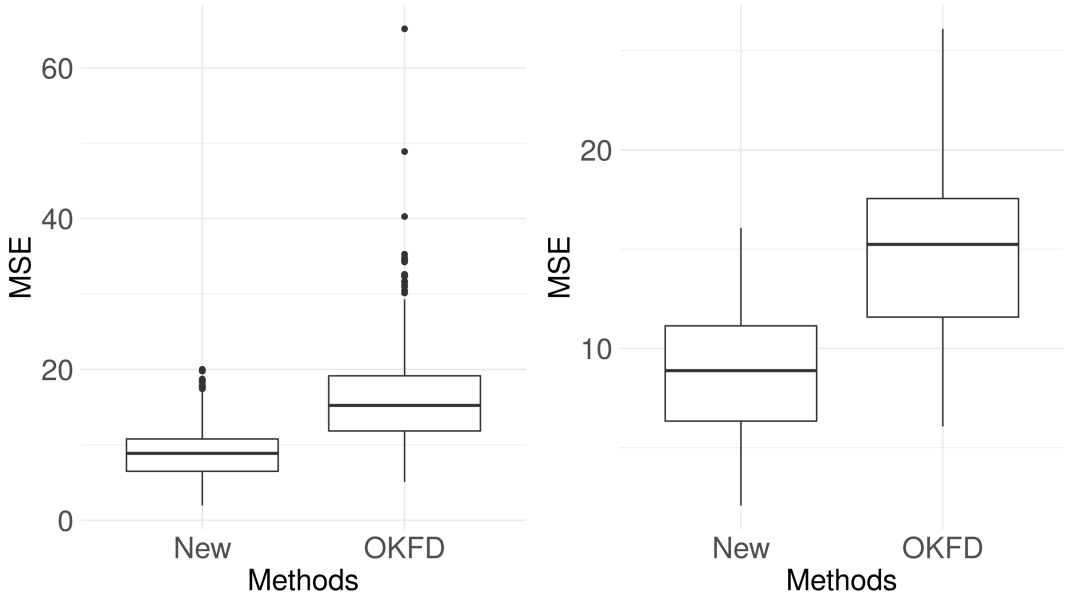
48(3):465–471, Sep 1983. ISSN 1860-0980.

[25] A. Zamani, M. Hashemi, and H. Haghbin. Improved functional portmanteau tests. Journal of Statistical Computation and Simulation, 89(8):1423–1436, 2019.

List of Figures

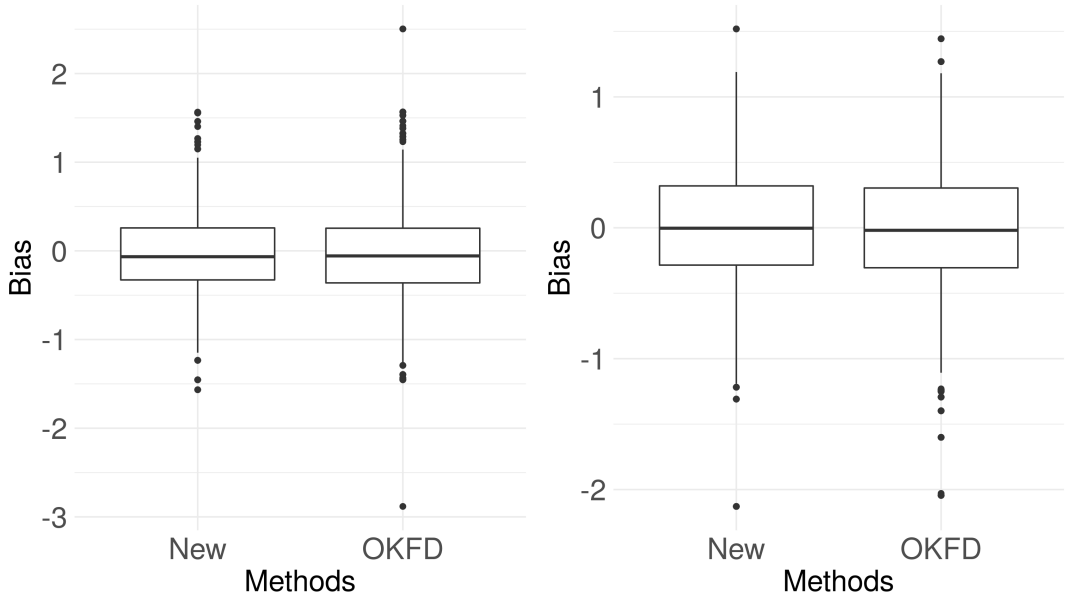


(a) $n = 25$ (b) $n = 50$



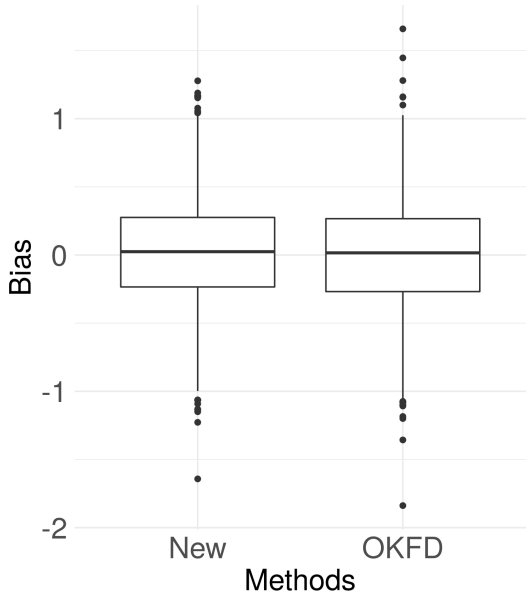
(c) $n = 75$ (d) $n = 100$

Figure 1.: mse of the estimation method proposed (New) and the estimation method proposed by Giraldo et al. [9] (OKFD) for normal dataset with L = 1000 replications.

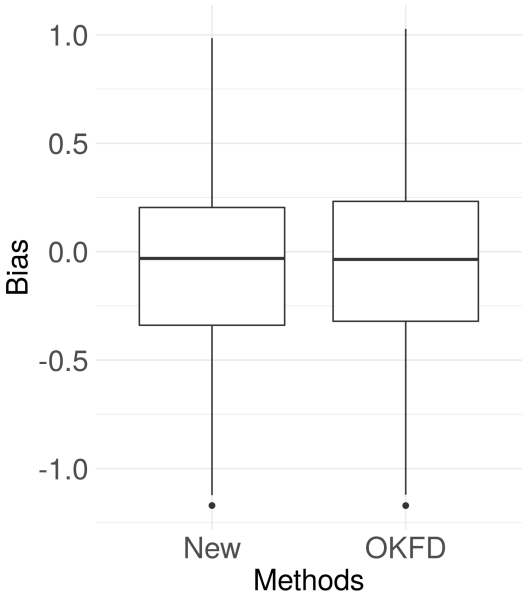


(a) $n = 25$

(b) $n = 50$

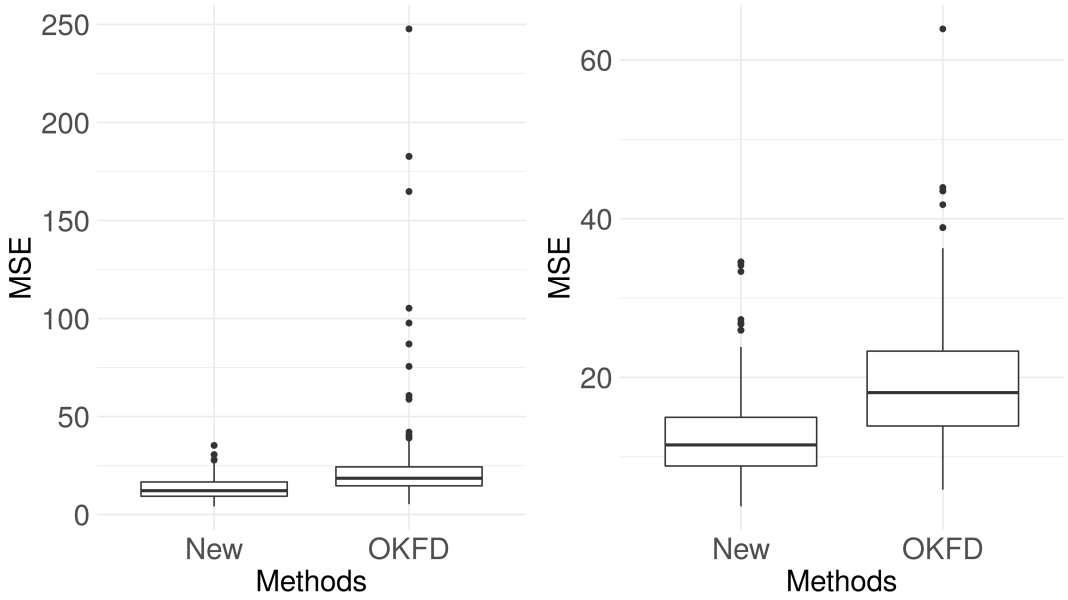


(c) $n = 75$



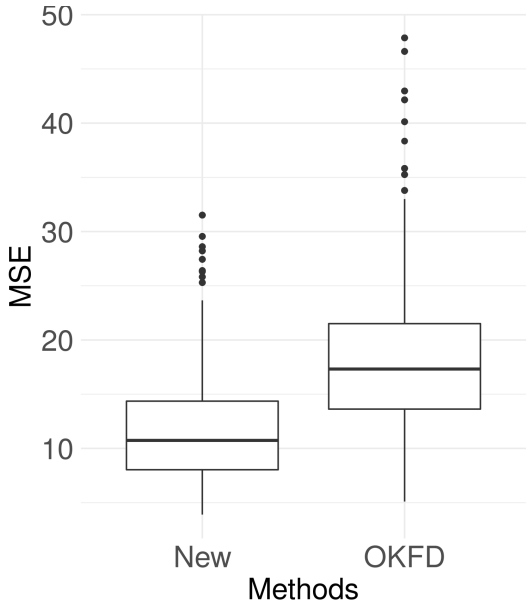
(d) $n = 100$

Figure 2.: $bias$ of the estimation method proposed (New) and the estimation method proposed by Giraldo et al. [9] (OKFD) for normal dataset with L = 1000 replications.

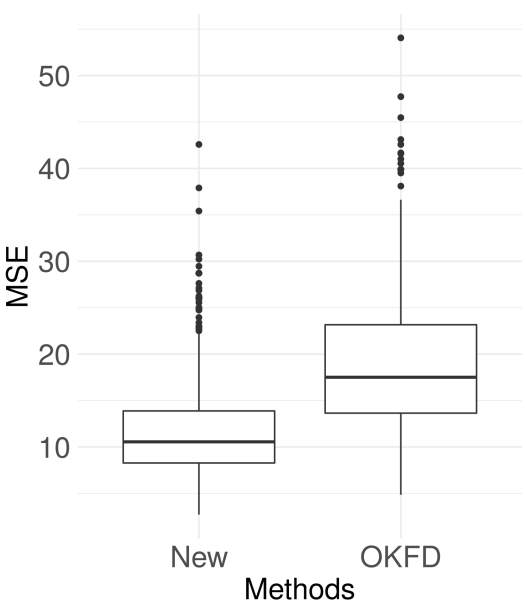


(a) $n = 25$

(b) $n = 50$

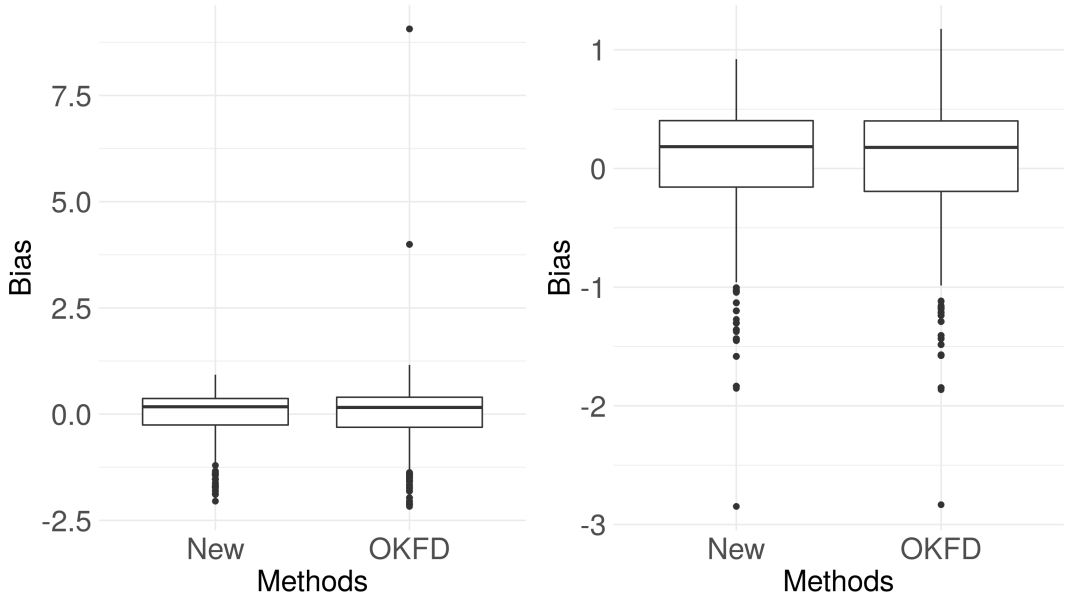


(c) $n = 75$

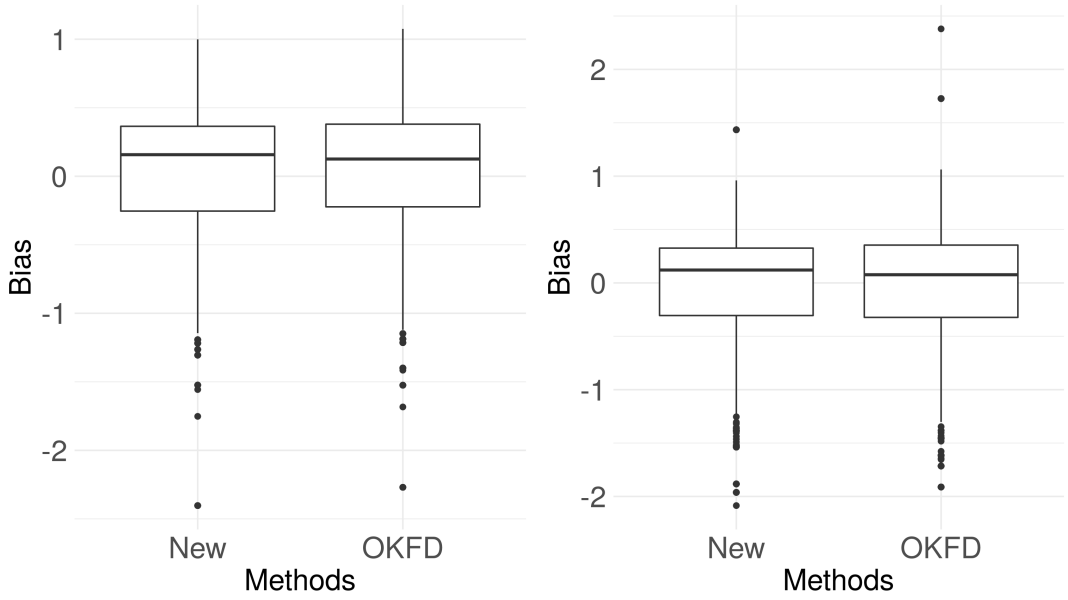


(d) $n = 100$

Figure 3.: mse of the estimation method proposed (New) and the estimation method proposed by [9] (OKFD) for non-normal dataset with L = 1000 replications.



(a) $n = 25$ (b) $n = 50$



(c) $n = 75$ (d) $n = 100$

Figure 4.: $bias$ of the estimation method proposed (New) and the estimation method proposed by [9] (OKFD) for non-normal dataset with L = 1000 replications.

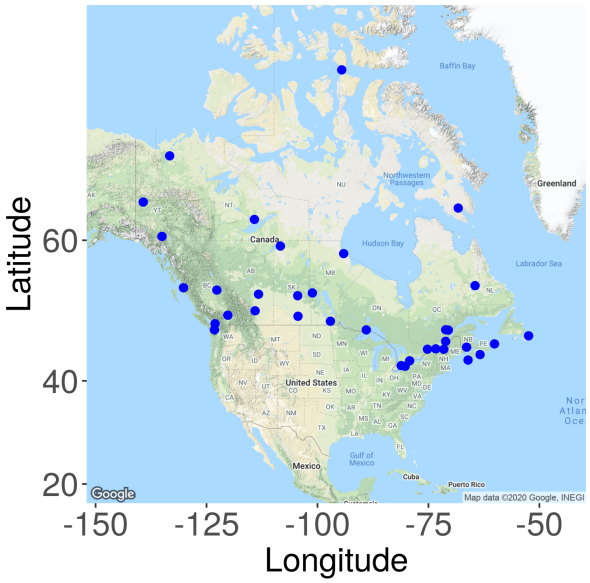
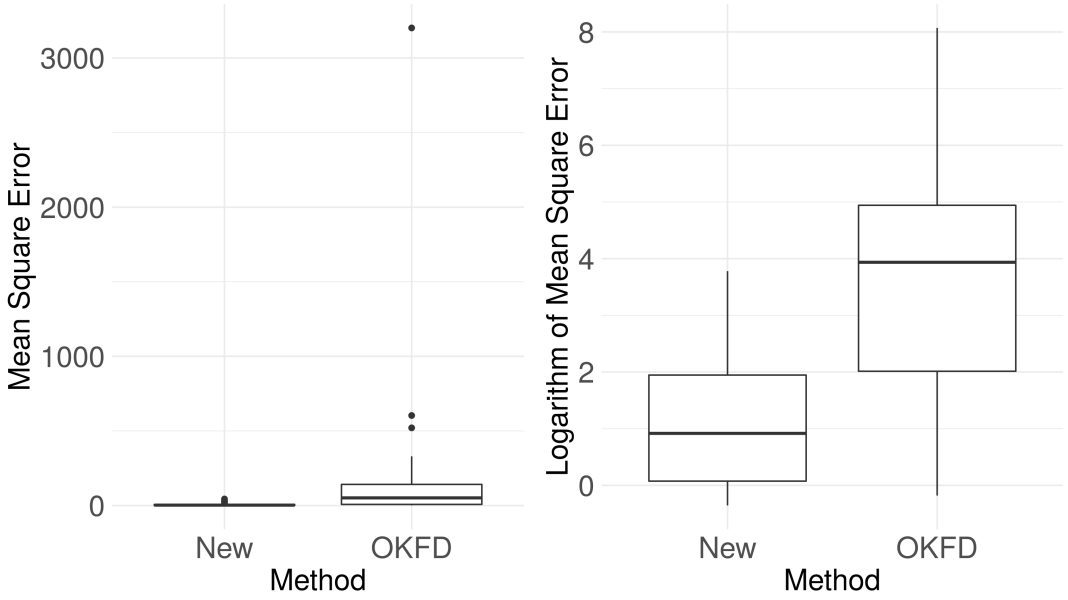


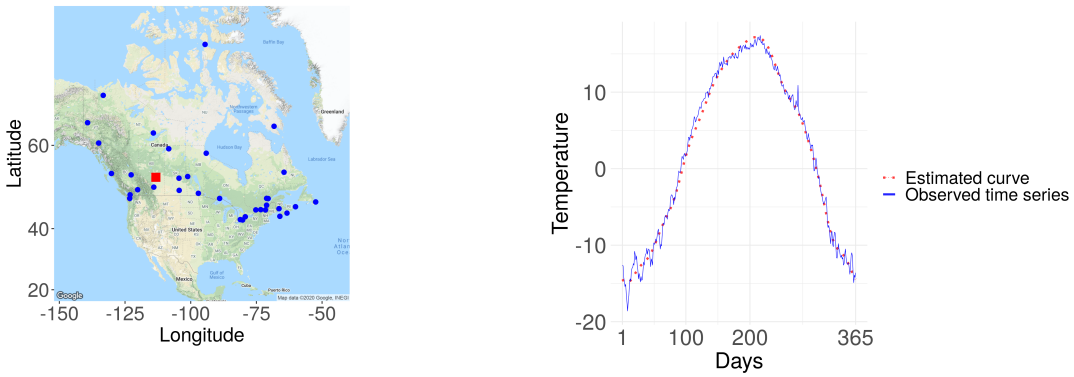
Figure 5.: 35 weather stations of Canada.



(a) Box plot for mse_i.

(b) Box plot for $\log(mse_i)$.

Figure 6.: Boxplot for mse_i and $\log(mse_i)$, $i = 1, \ldots, n$.



(a) Station s_0.

(b) Observed and estimated curve.

Figure 7.: Weather station located at Edmonton.

List of Tables

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 [9].

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

Table 2.: Mean of mse and $bias$ for non-normal simulated data. **New** refers to the new methodology proposed in this paper, and **OKFD** refers to methodology proposed by Giraldo [9].

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

Table 3.: Mean of mse of each station. **New** refers to the new methodology proposed in this work, and **OKFD** refers to metodology proposed by [9].

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