

# A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space

Alessandra Menafoglio, Piercesare Secchi

*MOX – Department of Mathematics  
Politecnico di Milano  
Piazza Leonardo da Vinci 32, 20133 Milano, Italy  
e-mail: [alessandra.menafoglio@polimi.it](mailto:alessandra.menafoglio@polimi.it)  
e-mail: [piercesare.secchi@polimi.it](mailto:piercesare.secchi@polimi.it)*

and

Matilde Dalla Rosa

*Eni S.p.A, Exploration & Production Division  
Via Emilia 1, San Donato (MI), Italy  
e-mail: [matilde.dalla.rosa@eni.com](mailto:matilde.dalla.rosa@eni.com)*

**Abstract:** We address the problem of predicting spatially dependent functional data belonging to a Hilbert space, with a Functional Data Analysis approach. Having defined new global measures of spatial variability for functional random processes, we derive a Universal Kriging predictor for functional data. Consistently with the new established theoretical results, we develop a two-step procedure for predicting georeferenced functional data: first model selection and estimation of the spatial mean (drift), then Universal Kriging prediction on the basis of the identified model. The proposed methodology is applied to daily mean temperatures curves recorded in the Maritimes Provinces of Canada.

**AMS 2000 subject classifications:** Primary 62H11; secondary 62M20, 62M30.

**Keywords and phrases:** Functional data analysis, Sobolev metrics, spatial prediction, variogram.

Received October 2012.

## 1. Introduction

Functional Data Analysis (FDA, Ramsay and Silverman (2005)) has recently received a great deal of attention in the literature because of the increasing need to analyze infinite-dimensional data, such as curves, surfaces and images. Whenever functional data are spatially dependent, FDA methods relying on the assumption of independence among observations could fail because consistency problems may arise (Hörmann and Kokoszka, 2011; Horváth and Kokoszka, 2012).

In the presence of spatial dependence, not only *ad hoc* estimation and regression techniques need to be developed (e.g., Gromenko et al. (2012) and Yamamoto and Tanaka (2003)), but also other topics need to be addressed. Among them, spatial prediction assumes a key role: the extension of kriging techniques (Cressie, 1993) to the functional setting meets the need of interpolating complex data collected in a limited number of spatial locations and thus could find application in different areas of industrial and environmental research. Nevertheless, a relatively small body of literature has been produced on this topic: indeed, theoretical results in this direction have been recently derived (Giraldo et al. (2008b, 2011); Delicado et al. (2010); Giraldo et al. (2010a); Monestiez and Nerini (2008); Nerini et al. (2010)) moving from the pioneering work by Goulard and Voltz (1993), but this theory is still limited to stationary functional stochastic processes valued in  $L^2$ .

However, in geophysical and environmental applications, natural phenomena are typically very complex and they rarely show a uniform behavior over the spatial domain: in these cases, non-stationary methods are needed. To this end, two techniques for kriging non-stationary functional data belonging to  $L^2$  have been developed concurrently with the present work, proposing a Residual Kriging approach (Reyes et al., 2012) and a Universal Kriging approach (Caballero et al., 2013). These methods, however, are worked out specifically for functional data belonging to  $L^2$  and do not allow to threat functional data valued in general Hilbert spaces, for which non-stationary kriging techniques are yet to be developed. In this work, we tackle this problem both from a theoretical point of view and from a computational one.

The methodological effort is here devoted to establish a general and coherent theoretical framework for Universal Kriging prediction in any separable Hilbert space, not just  $L^2$ . For instance, in our setting both pointwise and differential properties characterizing the functional data can be explicitly incorporated in the measures of spatial dependence – namely trace-variogram and trace-covariogram – if data are assumed to belong to a proper Sobolev space (see Remark 7 and Section 3).

Together with the theoretical results – presented in Section 2, new algorithms to perform spatial prediction are developed in Section 4 and their performance is tested through a simulation study illustrated as supplementary material in (Menaoglio et al., 2013). Two main goals move this part of the work: first to select an optimal linear model for the spatial mean – i.e. the drift – in the absence of a priori information, second to estimate the structure of spatial dependence of the associated residuals, which is that involved in the kriging prediction.

Finally, the case study that first motivated this work is presented in Section 5. It originates from a meteorological application and concerns the analysis of daily mean temperature curves recorded in the Maritimes Provinces of Canada. The aim of the study is to predict the whole space-time field of temperatures on the basis of the available data, deriving furthermore estimates for the temperature spatial trend. The problem of spatial prediction of temperatures is of interest in microclimate prediction as well as in hydrological and forest ecosystem modeling. It has been already faced in the literature about kriging for functional

data by means of stationary techniques (e.g., (Giraldo et al., 2010a)); here a non-Euclidean distance is adopted for the spatial domain and a drift term is modeled. We will show that the introduction of a drift term has a strong influence on the analysis in terms of cross-validation performance and prediction accuracy, besides allowing a climatical interpretation of the results.

## 2. Universal Kriging for functional random fields

### 2.1. Preliminaries and definitions

Let  $(\Omega, \mathfrak{F}, \mathbb{P})$  a probability space and  $H$  a separable Hilbert space endowed with the inner product  $\langle \cdot, \cdot \rangle$  and the induced norm  $\|\cdot\|$ , whose points are functions  $\mathcal{X} : \mathcal{T} \rightarrow \mathbb{R}$ , where  $\mathcal{T}$  is a compact subset of  $\mathbb{R}$ . Call *functional random variable* a measurable function  $\mathbf{X} : \Omega \rightarrow H$ , whose realization  $\mathcal{X}$ , called *functional datum*, is an element of  $H$  (Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012).

Consider on  $H$  a (functional) random field:

$$\{\chi_s, s \in D \subseteq \mathbb{R}^d\}, \quad (1)$$

that is a set of functional random variables  $\chi_s$  of  $H$ , indexed by a continuous spatial vector  $s$  varying in  $D \subseteq \mathbb{R}^d$  (usually  $d = 2$ ).

In this framework, a functional dataset  $\chi_{s_1}, \dots, \chi_{s_n}$  is the collection of  $n$  observations of the random field (1) relative to  $n$  locations  $s_1, \dots, s_n \in D$ ; in nontrivial situations a vector of observations  $\chi_{\bar{s}} = (\chi_{s_1}, \dots, \chi_{s_n})^T$  is characterized by a structure of spatial dependence reflecting the covariance structure of the generating random process (1). The aim of this work is the prediction of the realization  $\chi_{s_0}$  in an unsampled site  $s_0 \in D$ , through a geostatistical approach, based on global definitions of covariogram and variogram.

For  $1 \leq p < \infty$  denote with  $L^p(\Omega; H)$  the vector space of (equivalence classes of) measurable functions  $\mathbf{X} : \Omega \rightarrow H$  with  $\|\mathbf{X}\| \in L^p(\Omega)$  – i.e.  $\int_{\Omega} \|\mathbf{X}(\omega)\|^p \mathbb{P}(d\omega) = \mathbb{E}[\|\mathbf{X}\|^p] < \infty$  where  $\mathbb{E}$  indicates the expected value, that is a Banach space with respect to the norm  $\|\mathbf{X}\|_{L^p(\Omega; H)} = (\mathbb{E}[\|\mathbf{X}\|^p])^{1/p}$ .

In this work, we assume that the following condition holds.

**Assumption 1** (Square-integrability). Each element  $\chi_s, s \in D$ , of the random field (1) belongs to  $L^2(\Omega; H)$ .

When Assumption 1 is true, the expected value  $m_s$  of the random field (1) can be defined by Bochner integral as:

$$m_s = \int_{\Omega} \chi_s(\omega) \mathbb{P}(d\omega), \quad s \in D. \quad (2)$$

The expected value (2) coincides, for almost all  $t \in \mathcal{T}$ , with its pointwise definition  $m_s(t) = \mathbb{E}[\chi_s(t)]$  (Dunford and Schwartz, 1958). Moreover, for any  $x \in H$ ,  $\langle x, m_s \rangle = \mathbb{E}[\langle x, \chi_s \rangle]$ .

A global measure of spatial dependence is provided by the following:

**Definition 2.** The (*global*) covariance function of the process (1) is the function  $C : D \times D \rightarrow \mathbb{R}$ :

$$C(\mathbf{s}_i, \mathbf{s}_j) = \text{Cov}(\chi_{\mathbf{s}_i}, \chi_{\mathbf{s}_j}) := \mathbb{E}[\langle \chi_{\mathbf{s}_i} - m_{\mathbf{s}_i}, \chi_{\mathbf{s}_j} - m_{\mathbf{s}_j} \rangle], \quad (3)$$

When Assumption 1 holds true,  $C$  is a positive definite function:

$$\sum_i \sum_j \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) \geq 0, \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D, \forall \lambda_i, \lambda_j \in \mathbb{R},$$

and defines a scalar product on  $L^2(\Omega; H)$ . The function  $C$  will also be called *trace-covariogram* because of its relation – for every fixed couple  $\mathbf{s}_i, \mathbf{s}_j \in D$  – with the cross-covariance operator  $C_{\mathbf{s}_i, \mathbf{s}_j} : H \rightarrow H$  defined, for  $x \in H$ , by:

$$C_{\mathbf{s}_i, \mathbf{s}_j} x = \mathbb{E}[\langle \chi_{\mathbf{s}_i} - m_{\mathbf{s}_i}, x \rangle (\chi_{\mathbf{s}_j} - m_{\mathbf{s}_j})]. \quad (4)$$

Indeed, by applying Parsival Identity and following the arguments presented in (Hörmann and Kokoszka (2011), Section 3), one can easily prove the following:

**Proposition 3.** For every couple of locations  $\mathbf{s}_i, \mathbf{s}_j$  in  $D$ ,  $C(\mathbf{s}_i, \mathbf{s}_j)$  is the trace of the corresponding cross-covariance operator  $C_{\mathbf{s}_i, \mathbf{s}_j}$ :

$$C(\mathbf{s}_i, \mathbf{s}_j) = \sum_{k=1}^{\infty} \langle C_{\mathbf{s}_i, \mathbf{s}_j} e_k, e_k \rangle, \quad (5)$$

where  $\{e_k, k \in \mathbb{N}\}$  is any orthonormal basis of  $H$ . In particular:

$$|C(\mathbf{s}_i, \mathbf{s}_j)| \leq \sum_{k=1}^{\infty} |\lambda_k^{(\mathbf{s}_i, \mathbf{s}_j)}|,$$

being  $\lambda_k^{(\mathbf{s}_i, \mathbf{s}_j)}$ ,  $k = 1, 2, \dots$ , the singular values of the cross-covariance operator  $C_{\mathbf{s}_i, \mathbf{s}_j}$ .

Notice that the trace of  $C_{\mathbf{s}_i, \mathbf{s}_j}$  is well defined by  $\sum_{k=1}^{\infty} \langle C_{\mathbf{s}_i, \mathbf{s}_j} e_k, e_k \rangle$ , since  $C_{\mathbf{s}_i, \mathbf{s}_j}$  is a trace-class Hilbert-Schmidt operator (Bosq, 2000) and thus the series converges absolutely for any orthonormal basis  $\{e_k, k \geq 1\}$  of  $H$  and the sum does not depend on the choice of the basis (Zhu (2007), Theorem 1.24).

Expression (3) induces a notion of global variance and of variogram, as well as new global definitions of second-order and intrinsic stationarity.

**Definition 4.** The (*global*) variance of the process (1) is the function  $\sigma^2 : D \rightarrow [0, +\infty]$ :

$$\sigma^2(\mathbf{s}) = \text{Var}(\chi_{\mathbf{s}}) = \mathbb{E}[\|\chi_{\mathbf{s}} - m_{\mathbf{s}}\|^2], \quad \mathbf{s} \in D. \quad (6)$$

The (*global*) semivariogram of the process (1) is the function  $\gamma : D \times D \rightarrow [0, +\infty]$ :

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2} \text{Var}(\chi_{\mathbf{s}_i} - \chi_{\mathbf{s}_j}), \quad \mathbf{s}_i, \mathbf{s}_j \in D. \quad (7)$$

The (*global*) variogram of the process (1) is defined as  $2\gamma$ .

Function (7) has the same properties as its finite-dimensional analogue (Chilès and Delfiner, 1999); in particular it is a conditionally negative definite function:

$$\sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{s}_i, \mathbf{s}_j) \leq 0, \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D, \forall \lambda_i, \lambda_j \in \mathbb{R} \quad \text{s.t.} \quad \sum_i \lambda_i = 0.$$

Relations with covariance operators can be established for the global variance and the global semivariogram. Indeed, let  $C_{\mathbf{s}, \mathbf{s}} : H \rightarrow H$  be the covariance operator of  $\chi_{\mathbf{s}}$  and  $\Gamma_{\mathbf{s}_i - \mathbf{s}_j, \mathbf{s}_i - \mathbf{s}_j} : H \rightarrow H$  half the covariance operator of the increment  $\chi_{\mathbf{s}_i} - \chi_{\mathbf{s}_j}$ , that is, for  $x \in H$ :

$$\begin{aligned} C_{\mathbf{s}, \mathbf{s}} x &= \mathbb{E}[\langle \chi_{\mathbf{s}} - m_{\mathbf{s}}, x \rangle (\chi_{\mathbf{s}} - m_{\mathbf{s}})]; \\ \Gamma_{\mathbf{s}_i - \mathbf{s}_j, \mathbf{s}_i - \mathbf{s}_j} x &= \frac{1}{2} \mathbb{E}[\langle \chi_{\mathbf{s}_i} - \chi_{\mathbf{s}_j} - (m_{\mathbf{s}_i} - m_{\mathbf{s}_j}), x \rangle (\chi_{\mathbf{s}_i} - \chi_{\mathbf{s}_j} - (m_{\mathbf{s}_i} - m_{\mathbf{s}_j}))]. \end{aligned}$$

Both operators are trace-class, self-adjoint, positive and Hilbert-Schmidt (e.g., Bosq (2000)), therefore, from Proposition 3, it is straightforward to prove that for any fixed  $\mathbf{s}, \mathbf{s}_i, \mathbf{s}_j \in D$  the variance  $\sigma^2(\mathbf{s})$  and the semivariogram  $\gamma(\mathbf{s}_i, \mathbf{s}_j)$  coincide with the trace of  $C_{\mathbf{s}, \mathbf{s}}$  and  $\Gamma_{\mathbf{s}_i - \mathbf{s}_j, \mathbf{s}_i - \mathbf{s}_j}$  respectively:

$$\sigma^2(\mathbf{s}) = \sum_{k=1}^{\infty} \langle C_{\mathbf{s}, \mathbf{s}} e_k, e_k \rangle = \sum_{k=1}^{\infty} \lambda_k^{(\mathbf{s}, \mathbf{s})} \quad (8)$$

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \sum_{k=1}^{\infty} \langle \Gamma_{\mathbf{s}_i - \mathbf{s}_j, \mathbf{s}_i - \mathbf{s}_j} e_k, e_k \rangle = \sum_{k=1}^{\infty} \zeta_k^{(\mathbf{s}_i - \mathbf{s}_j, \mathbf{s}_i - \mathbf{s}_j)}, \quad (9)$$

being  $\{\lambda_k^{(\mathbf{s}, \mathbf{s})}\}$  and  $\{\zeta_k^{(\mathbf{s}_i - \mathbf{s}_j, \mathbf{s}_i - \mathbf{s}_j)}\}$  the (non negative) eigenvalues of  $C_{\mathbf{s}, \mathbf{s}}$  and  $\Gamma_{\mathbf{s}_i - \mathbf{s}_j, \mathbf{s}_i - \mathbf{s}_j}$  respectively. For this reason, the global semivariogram will also be called *trace-semivariogram*.

Concerning the notion of stationarity, new global definitions of second-order and intrinsic stationarity can be stated as follow.

**Definition 5.** A process  $\{\chi_{\mathbf{s}}, \mathbf{s} \in D \subseteq \mathbb{R}^d\}$  is said to be (*globally*) *second-order stationary* if the following conditions hold:

- (i)  $\mathbb{E}[\chi_{\mathbf{s}}] = m, \quad \forall \mathbf{s} \in D \subseteq \mathbb{R}^d;$
- (ii)  $\text{Cov}(\chi_{\mathbf{s}_i}, \chi_{\mathbf{s}_j}) = \mathbb{E}[(\chi_{\mathbf{s}_i} - m, \chi_{\mathbf{s}_j} - m)] = C(\mathbf{h}), \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D \subseteq \mathbb{R}^d, \mathbf{h} = \mathbf{s}_i - \mathbf{s}_j.$

A process  $\{\chi_{\mathbf{s}}, \mathbf{s} \in D \subseteq \mathbb{R}^d\}$  is said to be (*globally*) *intrinsically stationary* if:

- (i)  $\mathbb{E}[\chi_{\mathbf{s}}] = m, \quad \forall \mathbf{s} \in D \subseteq \mathbb{R}^d;$
- (ii')  $\text{Var}(\chi_{\mathbf{s}_i} - \chi_{\mathbf{s}_j}) = \mathbb{E}[\|\chi_{\mathbf{s}_i} - \chi_{\mathbf{s}_j}\|^2] = 2\gamma(\mathbf{h}), \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D \subseteq \mathbb{R}^d, \mathbf{h} = \mathbf{s}_i - \mathbf{s}_j.$

Finally, the condition of isotropy can be established as follow.

**Definition 6.** A second-order stationary random process is said to be *isotropic* if:

$$\text{Cov}(\chi_{\mathbf{s}_i}, \chi_{\mathbf{s}_j}) = C(\|\mathbf{h}\|), \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D \subseteq \mathbb{R}^d, \mathbf{h} = \mathbf{s}_i - \mathbf{s}_j,$$

where  $\|\cdot\|$  is a norm on  $D$ .

**Remark 7.** When  $H = L^2$  and global second-order stationarity and isotropy for the process (1) are in force, the trace-semivariogram (7) corresponds to the integrated version of pointwise semivariograms  $\gamma(h_{i,j}; t) = \frac{1}{2} \text{Var}(\chi_{\mathbf{s}_i}(t) - \chi_{\mathbf{s}_j}(t))$  (assumed to exist a.e.):

$$\gamma(h_{i,j}) = \int_{\mathcal{T}} \gamma(h_{i,j}; t) dt, \quad (10)$$

being  $h_{i,j} = \|\mathbf{s}_i - \mathbf{s}_j\|$ . This has been introduced in (Giraldo et al., 2008a) with the name of *trace-semivariogram*. However our definition is more general and permits the analysis of functional data in more complex situations. For instance, we might want to take explicitly into account the regularity of the elements of  $H$  – which captures the dependence along the coordinate  $t \in \mathcal{T}$  – by assuming that  $H$  is an appropriate Sobolev space and working with the inner product consistent with this assumption.

In particular, let  $\mathcal{H}^k, k \geq 1$ , be the subset of  $L^2$  consisting of the equivalence classes of functions with weak derivatives  $D^\alpha \mathcal{X}, \alpha \leq k$ , in  $L^2$ :

$$\mathcal{H}^k(\mathcal{T}) = \{\mathcal{X} : \mathcal{T} \rightarrow \mathbb{R}, \text{ s.t. } D^\alpha \mathcal{X} \in L^2, \forall \alpha \leq k, \alpha \in \mathbb{N}\}.$$

By considering on  $\mathcal{H}^k$  the usual inner product and norm, the resulting trace-variogram is ( $D^0 \chi_{\mathbf{s}} = \chi_{\mathbf{s}}$ ):

$$2\gamma(\mathbf{s}_i, \mathbf{s}_j) = \text{Var}(\chi_{\mathbf{s}_i} - \chi_{\mathbf{s}_j})_{\mathcal{H}^k} = \sum_{\alpha=0}^k \text{Var}(D^\alpha \chi_{\mathbf{s}_i} - D^\alpha \chi_{\mathbf{s}_j})_{L^2},$$

where  $\text{Var}(D^\alpha \chi_{\mathbf{s}_i} - D^\alpha \chi_{\mathbf{s}_j})_{L^2}$  are the trace-variograms in  $L^2$  relative to the weak derivative random fields  $\{D^\alpha \chi_{\mathbf{s}}, \mathbf{s} \in D\}$ ,  $0 \leq \alpha \leq k$  (assumed to exist, for every  $\alpha = 0, 1, \dots, k$ ), which might significantly influence the overall trace-variogram.

The choice of the most proper Sobolev space may allow to distinguish among functional random fields which might appear similar from a spatial dependence point of view, but indeed very different in the structure of dependence along the coordinate  $t \in \mathcal{T}$  (see Section 3).

Moreover, suppose the random field to be the random path of a stochastic dynamical system,  $\{\chi_\tau, \tau \in \mathcal{D} \subset \mathbb{R}\}$ , whose state  $\chi_\tau$  is a functional random variable belonging to a Sobolev space  $H$  – determined by the equations which govern the dynamics of the system – (Arnold, 2003). In dynamical system theory, the Sobolev norm of the state coincides with (twice) the energy of the system. Therefore, the choice of the most proper Sobolev space for geostatistical analysis implies a precise physical meaning for the measure of stochastic variability: indeed, the global variance represents (twice) the mean energy of the system, while the trace-variogram (twice) the mean energy of the increments between two states.

In the light of the Proposition 3, existence of strong, (globally) second-order and (globally) intrinsic stationary functional processes can be established by

direct construction as in (Hörmann and Kokoszka, 2011). Considering an orthonormal basis  $\{e_j, j \geq 1\}$  of  $H$ , every functional random process (1) with constant mean  $m$  admits the following expansion:

$$\chi_s = m + \sum_{j \geq 1} \xi_j(s) e_j. \quad (11)$$

The scalar fields  $\xi_j(s) = \langle \chi_s - m, e_j \rangle$ ,  $j = 1, 2, \dots$ , determine the stationarity of the functional process. In fact, as proved in (Hörmann and Kokoszka, 2011), process (1) is strong stationary if and only if the scalar fields  $\xi_j(s)$  are strictly stationary for all  $j \geq 1$ ; moreover the random element  $\chi_s$ ,  $s \in D$ , belongs to  $L^2(\Omega; H)$  if and only if the sequence  $\{\xi_j(s)\}_{j \geq 1}$  belongs to  $\ell^2(\Omega; \mathbb{R})$  (i.e.  $\sum_{j \geq 1} \mathbb{E}[\xi_j(s)^2] < \infty$ ). Furthermore, second-order stationarity of each scalar field  $\xi_j(s)$ ,  $j = 1, 2, \dots$ , ensures that the cross-covariance operator  $C_{s,s+h}$  depends only on the increment vector  $h \in D$ , for every  $s \in D$ , which is in fact a sufficient condition for the functional process to be globally second-order stationary. This condition can be weakened in order to obtain the following necessary and sufficient condition for global second-order stationarity:

$$\sum_{j \geq 1} \mathbb{E}[\xi_j(s)\xi_j(s+h)] = C(h), \quad (12)$$

for each  $s, h \in D$  and for some real-valued function  $C$ . As a consequence, a necessary condition for global second-order stationarity is the independence of the  $\ell^2$ -norm of the sequence  $\{\xi_j(s)\}_{j \geq 1}$  from the location  $s \in D$ .

As in finite-dimensional theory, intrinsic stationarity is a weaker condition than second-order stationarity. Indeed, a  $d$ -dimensional isotropic Brownian motion  $\{W_s, s \in D \subseteq \mathbb{R}^d\}$  can be seen as a functional random field on  $H = L^2([0, 1])$ , such that each element  $\mathcal{W}_s : [0, 1] \rightarrow \mathbb{R}$ ,  $s \in D$ , is a functional random variable whose realization  $\mathcal{W}_s(\omega, \cdot)$ ,  $\omega \in \Omega$ , is constant over the domain  $[0, 1]$ :  $\mathcal{W}_s(\omega, t) = W_s(\omega)$ , for all  $t \in [0, 1]$ . Obviously, each  $\mathcal{W}_s$  belongs to  $L^2(\Omega, H)$  and:

$$\text{Var}(\mathcal{W}_{s_i} - \mathcal{W}_{s_j}) = \mathbb{E}[(W_{s_i} - W_{s_j})^2] = \|s_i - s_j\|,$$

while

$$\text{Cov}(\mathcal{W}_{s_i}, \mathcal{W}_{s_j}) = \mathbb{E}[W_{s_i} W_{s_j}] = (\|s_i\| + \|s_j\| - \|s_i - s_j\|),$$

which is not a function of  $(s_i - s_j)$ .

## 2.2. Universal Kriging predictor

Consider a non-stationary random process  $\{\chi_s, s \in D\}$ , whose elements are representable as:

$$\chi_s = m_s + \delta_s, \quad (13)$$

where  $m_s$ , the drift, describes the non-constant spatial mean variation, while the residual term  $\delta_s$  is supposed to be a zero-mean, second-order stationary and

isotropic random field, i.e.:

$$\begin{cases} \mathbb{E}[\chi_s] = m_s, & s \in D \subseteq \mathbb{R}^d; \\ \mathbb{E}[\delta_s] = 0, & s \in D \subseteq \mathbb{R}^d; \\ \text{Cov}(\delta_{s_i}, \delta_{s_j}) = \mathbb{E}[\langle \delta_{s_i}, \delta_{s_j} \rangle] = C(\|h\|), & \forall s_i, s_j \in D \subseteq \mathbb{R}^d, h = s_i - s_j. \end{cases}$$

As in classical geostatistics (Cressie, 1993), assume the following linear model for the drift  $m_s$ :

$$m_s(t) = \sum_{l=0}^L a_l(t) f_l(s), \quad s \in D, t \in \mathcal{T}, \quad (14)$$

where  $f_0(s) = 1$  for all  $s \in D$ ,  $f_l(\cdot)$ ,  $l = 1, \dots, L$ , are known functions of the spatial variable  $s \in D$  and  $a_l(\cdot) \in H$ ,  $l = 0, \dots, L$ , are functional coefficients independent from the spatial location. Hence it is supposed that the dependence of the mean  $m_s$  on the spatial variable  $s \in D$  is explained by the family of functions  $\{f_l(\cdot)\}_{l=1, \dots, L}$ , that are constant with respect to the variable  $t \in \mathcal{T}$ ; in the meantime, the functional nature of the drift  $m_s$  is preserved thanks to the introduction of the functional coefficients  $a_l(\cdot)$ ,  $l = 0, \dots, L$ .

For most applications, these assumptions are not too restrictive: in fact this model is able to describe precisely the drift term whenever it is a separable function or in the presence of a scalar external drift.

Given  $n$  observations  $\chi_{s_1}, \dots, \chi_{s_n}$  sampled from a realization of  $\{\chi_s, s \in D\}$ , our next goal is the formulation of the Universal Kriging predictor of the variable  $\chi_{s_0}$  located in  $s_0 \in D$ , which is the best linear unbiased predictor (BLUP):

$$\chi_{s_0}^* = \sum_{i=1}^n \lambda_i^* \chi_{s_i},$$

whose weights  $\lambda_1^*, \dots, \lambda_n^* \in \mathbb{R}$  minimize the global variance of the prediction error under the unbiasedness constraint:

$$(\lambda_1^*, \dots, \lambda_n^*) = \underset{\substack{\lambda_1, \dots, \lambda_n \in \mathbb{R}: \\ \chi_{s_0}^\lambda = \sum_{i=1}^n \lambda_i \chi_{s_i}}}{\operatorname{argmin}} \text{Var}(\chi_{s_0}^\lambda - \chi_{s_0}) \quad \text{s.t.} \quad \mathbb{E}[\chi_{s_0}^\lambda] = m_{s_0}. \quad (15)$$

In (15) both the variance to be minimized and the unbiasedness constraint are well defined since the linear predictor  $\chi_{s_0}^\lambda$  (and thus  $\chi_{s_0}^\lambda - \chi_{s_0}$ ) belongs to the same space  $H$  as the variables  $\chi_{s_1}, \dots, \chi_{s_n}$ , because  $H$  is closed with respect to linear combinations of its elements.

From the unbiasedness constraint, the following set of restrictions on the weights can be easily derived:

$$\sum_{i=1}^n \lambda_i f_l(s_i) = f_l(s_0), \quad \forall l = 0, \dots, L. \quad (16)$$

By including (16) in the minimization problem through  $L + 1$  Lagrange multipliers,  $\mu_0, \dots, \mu_L$ , problem (15) can be solved by minimizing the functional  $\Phi$ :

$$\Phi = \text{Var}(\boldsymbol{\chi}_{\mathbf{s}_0}^{\lambda} - \boldsymbol{\chi}_{\mathbf{s}_0}) + 2 \sum_{l=0}^L \mu_l \left( \sum_{i=1}^n \lambda_i f_l(\mathbf{s}_i) - f_l(\mathbf{s}_0) \right),$$

easily reduced to:

$$\begin{aligned} \Phi = & \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) + C(0) - 2 \sum_{i=1}^n \lambda_i C(\mathbf{s}_i, \mathbf{s}_0) \\ & + 2 \sum_{l=0}^L \mu_l \left( \sum_{i=1}^n \lambda_i f_l(\mathbf{s}_i) - f_l(\mathbf{s}_0) \right). \end{aligned} \quad (17)$$

Under suitable assumptions on the sampling design – namely  $\Sigma = (C(h_{i,j})) \in \mathbb{R}^{n \times n}$  positive definite and  $\mathbb{F}_{\mathbf{s}} = (f_l(\mathbf{s}_i)) \in \mathbb{R}^{n \times (L+1)}$  of full rank, the functional (17) admits a unique global minimum that can be found solving the following linear system:

$$\begin{pmatrix} C(0) & \cdots & C(h_{1,n}) & 1 & f_1(\mathbf{s}_1) & \cdots & f_L(\mathbf{s}_1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C(h_{n,1}) & \cdots & C(0) & 1 & f_1(\mathbf{s}_n) & \cdots & f_L(\mathbf{s}_n) \\ 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ f_1(\mathbf{s}_1) & \cdots & f_1(\mathbf{s}_n) & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f_L(\mathbf{s}_1) & \cdots & f_L(\mathbf{s}_n) & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu_0 \\ \mu_1 \\ \vdots \\ \mu_L \end{pmatrix} = \begin{pmatrix} C(h_{0,1}) \\ \vdots \\ C(h_{0,n}) \\ 1 \\ f_1(\mathbf{s}_0) \\ \vdots \\ f_L(\mathbf{s}_0) \end{pmatrix}, \quad (18)$$

where  $C(h_{i,j})$  denotes the trace-covariogram function of the residual process  $\{\delta_{\mathbf{s}}, \mathbf{s} \in D\}$ , evaluated in  $h_{i,j} = \|\mathbf{s}_i - \mathbf{s}_j\|$ .

Notice that, by combining (3) and (7), one can easily extend to the functional case the well-known relation between trace-covariogram and trace-semivariogram:

$$\gamma(h_{i,j}) = C(0) - C(h_{i,j}),$$

that allows to express equivalently the linear system (18) in terms of the semivariogram function  $\gamma$ .

In addition, we can associate to the pointwise prediction  $\boldsymbol{\chi}_{\mathbf{s}_0}^*$  in  $\mathbf{s}_0$  a measure of its global variability through the *Universal Kriging variance*:

$$\begin{aligned} \sigma_{UK}^2(\mathbf{s}_0) &= C(0) - \sum_{i=1}^n \lambda_i C(h_{i,0}) - \sum_{l=0}^L \mu_l f_l(\mathbf{s}_0) \\ &= \sum_{i=1}^n \lambda_i \gamma(h_{i,0}) + \sum_{l=0}^L \mu_l f_l(\mathbf{s}_0), \quad \mathbf{s}_0 \in D; f_0(\mathbf{s}) = 1, \forall \mathbf{s} \in D. \end{aligned} \quad (19)$$

Observe that both kriging system (18) and the kriging variance (19) have exactly the same form of the finite-dimensional corresponding expressions, indicating the consistency of our extensions with the real-valued random field case (if  $H = \mathbb{R}$ , the trace-covariogram reduces to the usual covariogram).

Moreover, when considering the very specific case treated at the beginning of Remark 7, the Universal Kriging system (18) reduces to the Ordinary Kriging system already presented in (Giraldo et al., 2008a). Furthermore, with the same arguments of Remark 7, one can easily see that if  $H = L^2$  the Universal Kriging system (18) coincides with that derived in (Caballero et al., 2013).

Finally, global second-order stationarity of the residual process has been assumed for the construction of the optimal predictor: however, as in classical theory, the Ordinary Kriging predictor is also well defined under the hypothesis of intrinsic stationarity. In fact, second-order stationarity for the residuals has to be required whenever the mean  $m_s$  is not constant and the residual trace-variogram is unknown: in such a case, the trace-covariogram is needed for the generalized least squares estimate of the drift (see Subsection 2.4).

### 2.3. Variogram estimation

In order to determine the Universal Kriging predictor in  $s_0$  by solving (18), an estimation of the trace-covariogram or, as usually preferred, of the trace-semivariogram is needed. As in classical geostatistics, semivariogram estimation can be performed by first determining an empirical estimator and then fitting a valid model. The latter step is necessary in order to fulfil the requirements on the trace-semivariogram function, e.g. conditional negative definiteness.

Suppose to know the realization  $\delta_{s_1}, \dots, \delta_{s_n}$  of the residual process  $\{\delta_s, s \in D\}$ , in the  $n$  sampling locations  $s_1, \dots, s_n$  of the domain  $D$  in which we observe the functional dataset  $\chi_{s_1}, \dots, \chi_{s_n}$ . Recall that the residual process is zero-mean second-order stationary and isotropic, so that:

$$\gamma(h) = \frac{1}{2} \mathbb{E}[\|\delta_{s_i} - \delta_{s_j}\|^2], \quad \forall s_i, s_j \in D \subseteq \mathbb{R}^d, h = \|s_i - s_j\|.$$

Following the approach adopted in (Giraldo et al., 2008a) and by analogy with the finite-dimensional case, a method-of-moments estimator can be used:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} \|\delta_{s_i} - \delta_{s_j}\|^2, \quad (20)$$

where  $N(h)$  indicates the set of all couples of sites separated by a distance  $h$  and  $|N(h)|$  is its cardinality. In applications, since it is hardly possible to calculate an estimate  $\hat{\gamma}(h)$  for every value of  $h$ , a discretized version  $\hat{\gamma}(\mathbf{h}) = (\hat{\gamma}(h_1), \dots, \hat{\gamma}(h_K))$  of  $\hat{\gamma}(h)$  can be used instead, for  $K$  classes of distance centered in  $h_1, \dots, h_K$ .

For the fitting step, a least squares criterion can be used, minimizing the distance between the empirical estimate  $\hat{\gamma}(\mathbf{h})$  and a parametric valid model  $\gamma(\mathbf{h}; \boldsymbol{\vartheta})$ , properly chosen among the classical families of valid (semi)variogram

models (Cressie, 1993). Indeed, in classical geostatistics there exists a number of parametric families of valid models that can be used in the functional case as well, since the trace-semivariogram is a real valued function which has to fulfil the same set of requirements as its finite-dimensional analogue. As an alternative, *ad hoc* constructed valid models can be tested for conditional negative definiteness by means of spectral methods (Armstrong and Diamond, 1984).

#### 2.4. Drift estimation

Although the drift coefficients are not directly included in the Universal Kriging system (18), their estimation is necessary in order to assess the trace-semivariogram of the residual process  $\{\delta_s, s \in D\}$ , since, in general, this is unobserved.

Assuming the dichotomy (13) and the linear model (14), the original process can be expressed as:

$$\chi_s = \sum_{l=0}^L a_l f_l(s) + \delta_s, \quad s \in D. \quad (21)$$

Hence, the compact matrix form for model (21) for the random vector  $\chi_{\vec{s}} = (\chi_{s_1}, \dots, \chi_{s_n})^T$  – whose realization  $\chi_{\vec{s}}$  belongs to the product space  $H^n = H \times H \times \dots \times H$  – is:

$$\chi_{\vec{s}} = \mathbb{F}_{\vec{s}} a_{\vec{l}} + \delta_{\vec{s}}, \quad (22)$$

where  $a_{\vec{l}} = (a_0, \dots, a_L)^T$  is the vector of coefficients,  $\delta_{\vec{s}} = (\delta_{s_1}, \dots, \delta_{s_n})^T$  is the random vector of spatially-correlated residuals and  $\mathbb{F}_{\vec{s}} = (f_l(s_i)) \in \mathbb{R}^{n \times (L+1)}$  is the design matrix.

The theory of linear models in functional data analysis (FDA, Ramsay and Silverman (2005)) has been developed under the founding hypothesis of independent and identically distributed residuals, so that the ordinary least squares approach developed in that framework inevitably turns out to be somewhat inadequate in the presence of correlated residuals. In order to properly take into account the structure of spatial dependence existing among observations, we propose a generalized least squares criterion (GLS) with weighting matrix  $\Sigma^{-1}$ , the inverse of the  $n \times n$  covariance matrix  $\Sigma$  of  $\chi_{\vec{s}}$ .

Indeed, a measure of the statistical distance among functional random variables  $\mathcal{X}, \mathcal{Y}$  in  $H^n$  can be provided through the following extension of the notion of Mahalanobis distance (Mahalanobis, 1936):

$$d_{\Sigma^{-1}}(\mathcal{X}, \mathcal{Y}) = \|\mathcal{X} - \mathcal{Y}\|_{\Sigma^{-1} - H^n} = \|\Sigma^{-1/2}(\mathcal{X} - \mathcal{Y})\|_{H^n},$$

where  $\Sigma^{-1/2}$  indicates the square root matrix of  $\Sigma^{-1}$  (or, equivalently, the inverse of the square root matrix of  $\Sigma$ ) and  $\|\cdot\|_{H^n}$  denotes the norm in  $H^n$ , defined as  $\|\mathcal{X}\|_{H^n}^2 = \sum_{i=1}^n \|\mathcal{X}_i\|_H^2$ , for  $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_n) \in H^n$ .

The GLS estimator  $\hat{a}_{\vec{l}}^{GLS} = (\hat{a}_0^{GLS}, \dots, \hat{a}_L^{GLS})^T$  can be determined by solving the following optimal problem:

$$\min_{\hat{a}_{\vec{l}} \in H^{L+1}} \Phi^{GLS}(\hat{a}_{\vec{l}}) \quad (23)$$

where the functional  $\Phi^{GLS}$  to be minimized corresponds to the functional Mahalanobis distance between fitted values  $\widehat{\mathbf{m}}_{\vec{s}} = \mathbb{F}_{\vec{s}} \widehat{\mathbf{a}}_{\vec{l}}$  and observed data:

$$\Phi^{GLS}(\widehat{\mathbf{a}}_{\vec{l}}) = \|\chi_{\vec{s}} - \mathbb{F}_{\vec{s}} \widehat{\mathbf{a}}_{\vec{l}}\|_{\Sigma^{-1} - H^n}^2 = \|\chi_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}}\|_{\Sigma^{-1} - H^n}^2. \quad (24)$$

**Proposition 8.** *If  $\text{rank}(\mathbb{F}_{\vec{s}}) = L+1 \leq n$  and  $\text{rank}(\Sigma) = n$ , there exists a unique vector  $\widehat{\mathbf{a}}_{\vec{l}}^{GLS}$  solving the estimation problem (23), which admits the following explicit representation:*

$$\widehat{\mathbf{a}}_{\vec{l}}^{GLS} = (\mathbb{F}_{\vec{s}}^T \Sigma^{-1} \mathbb{F}_{\vec{s}})^{-1} \mathbb{F}_{\vec{s}}^T \Sigma^{-1} \chi_{\vec{s}}. \quad (25)$$

Moreover, the (unique) GLS drift estimator  $\widehat{\mathbf{m}}_{\vec{s}}$  is:

$$\widehat{\mathbf{m}}_{\vec{s}} = \mathbb{F}_{\vec{s}} (\mathbb{F}_{\vec{s}}^T \Sigma^{-1} \mathbb{F}_{\vec{s}})^{-1} \mathbb{F}_{\vec{s}}^T \Sigma^{-1} \chi_{\vec{s}}. \quad (26)$$

Since estimators (25) and (26) are linear, their mean and variance-covariance matrix can be easily derived:

$$\mathbb{E}[\widehat{\mathbf{a}}_{\vec{l}}^{GLS}] = a_{\vec{l}}; \quad \text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}^{GLS}) = (\mathbb{F}_{\vec{s}}^T \Sigma^{-1} \mathbb{F}_{\vec{s}})^{-1}; \quad (27)$$

$$\mathbb{E}[\widehat{\chi}_{\vec{s}}] = m_{\vec{s}}; \quad \text{Cov}(\widehat{\mathbf{m}}_{\vec{s}}) = \mathbb{F}_{\vec{s}} (\mathbb{F}_{\vec{s}}^T \Sigma^{-1} \mathbb{F}_{\vec{s}})^{-1} \mathbb{F}_{\vec{s}}^T. \quad (28)$$

Besides being unbiased, the following result holds.

**Proposition 9.** *The estimator  $\widehat{\mathbf{a}}_{\vec{l}}^{GLS}$  is the BLUE (Best Linear Unbiased Estimator) for the coefficients  $a_{\vec{l}}$ , i.e. for any other linear unbiased estimator  $\widehat{\mathbf{a}}_{\vec{l}} = \mathbb{A} \chi_{\vec{s}} + \mathbf{b}$  of  $a_{\vec{l}}$ , the matrix:*

$$\text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}) - \text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}^{BLUE})$$

is positive semi-definite. As a consequence,  $\widehat{\mathbf{m}}_{\vec{s}}^{GLS}$  is the BLUE for the mean vector  $m_{\vec{s}}$ .

Let  $\Sigma^{GLS}$  be the  $n \times n$  covariance matrix of the estimator  $\widehat{\delta}_{\vec{s}} = \chi_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}}^{GLS}$ ,  $\Sigma^{GLS} = \mathbb{E}[\widehat{\delta}_{\vec{s}} \widehat{\delta}_{\vec{s}}^T]$ , then the identity:

$$\Sigma = \text{Cov}(\widehat{\mathbf{m}}_{\vec{s}}^{GLS}) + \Sigma^{GLS}, \quad (29)$$

can be verified through orthogonality arguments (for details see the Appendix). Expression (29) provides a decomposition of the covariance matrix  $\Sigma$  in a part depending on the variability of the drift estimator  $\widehat{\mathbf{m}}_{\vec{s}}^{GLS}$  and a component representing the dependence structure of the estimated residual process.

Although an unbiased estimator of the covariance matrix  $\Sigma^{GLS}$  represents a natural estimator of  $\Sigma$ , it provides a biased estimation of the spatial-dependence structure, underestimating it for a quantity:  $\mathbb{B} = \text{Cov}(\widehat{\mathbf{m}}_{\vec{s}}^{GLS}) = \mathbb{F}_{\vec{s}} (\mathbb{F}_{\vec{s}}^T \Sigma^{-1} \mathbb{F}_{\vec{s}})^{-1} \mathbb{F}_{\vec{s}}^T$ . However, the bias of such an estimator proved to be negligible in all the performed simulations, as shown in (Menaoglio et al. (2013)).

Finally, observe that when  $\Sigma = \sigma^2 \mathbb{I}$ , that is precisely the case of (globally) uncorrelated residuals, OLS and GLS criteria coincide. From the point of view

of the residual variogram, the (global) uncorrelated case corresponds to a pure nugget structure, because the mean squared norm of the discrepancy among uncorrelated observations is equal to the variance  $\sigma^2$  of the process, being thus independent from their separating distance. Therefore, the estimation of the residuals variogram (or semivariogram), besides allowing the analysis of the spatial dependence structure, will be the leading tool in the determination of the most proper procedures for the statistical treatment of the observations, as will be clarified in the next sections.

### 3. Trace-variograms in Sobolev spaces: An example

The purpose of this first example is to show how the choice of the space  $H$  data are assumed to belong to might heavily influence the way in which the spatial dependence is modeled (see also Remark 7 in Section 2).

To see this, we now consider two globally second-order stationary and isotropic functional random fields,  $\{\chi_s^{(m)}, s \in D\}$ ,  $m = 1, 2$ , built by direct construction as in (11) by combining 7 independent, zero mean, second-order stationary and isotropic scalar random fields  $\{\xi_j(s), s \in D\}$ ,  $j = 1, \dots, 7$ , as:

$$\chi_s^{(1)} = \sum_{k=1}^7 \xi_k^{(1)}(s) e_k = \sum_{k=1}^7 \xi_k(s) e_k \quad (30)$$

$$\chi_s^{(2)} = \sum_{k=1}^{25} \xi_k^{(2)}(s) e_k = \sum_{k=19}^{25} \xi_{k-18}(s) e_k, \quad (31)$$

where  $\{e_k, k \geq 1\}$  denotes the Fourier basis.

An example of functional data generated by processes (30) and (31) is shown in the left panels of Figure 1a and 1b respectively, where, for each dataset, 5 of the 100 simulated curves are represented. Here, the functional datasets  $\chi_{s_1}^{(m)}, \dots, \chi_{s_{100}}^{(m)}$ ,  $m = 1, 2$ , are obtained by combining according to (30) and (31) a set of realizations of the scalar fields  $\xi_j$ ,  $j = 1, \dots, 7$ , simulated following the scheme described in Section S.1 of the supplementary material (Menafoglio et al., 2013). The different behavior of the curves is evident: the first dataset has a less fluctuating pattern along the coordinate  $t \in [0, 1]$ , since only the first 3 frequencies are excited; conversely, the second dataset is characterized by a very fluctuating pattern, due to the higher order basis truncation involving only the 10<sup>th</sup> to 12<sup>th</sup> frequencies. First order derivatives are represented in the right panels of Figure 1a and 1b.

Notice that, by construction, each realization of both processes belongs not only to  $L^2$ , but also to  $\mathcal{H}^1$ ; moreover, both processes are globally second-order stationary either in  $L^2$  or in  $\mathcal{H}^1$ . Indeed,  $L^2$  trace-variograms, as well as  $\mathcal{H}^1$  trace-variograms, can be explicitly computed.

Assume first  $H = L^2$ . For  $s_i, s_j \in D$ ,  $m = 1, 2$ , ( $N_1 = 7$ ,  $N_2 = 25$ ):

$$2\gamma^{(m)}(s_i, s_j)_{L^2} = \mathbb{E}[\|\chi_{s_i}^{(m)} - \chi_{s_j}^{(m)}\|_{L^2}^2] = \sum_{k=1}^{N_m} \mathbb{E}[|\xi_k^{(m)}(s_i) - \xi_k^{(m)}(s_j)|^2].$$

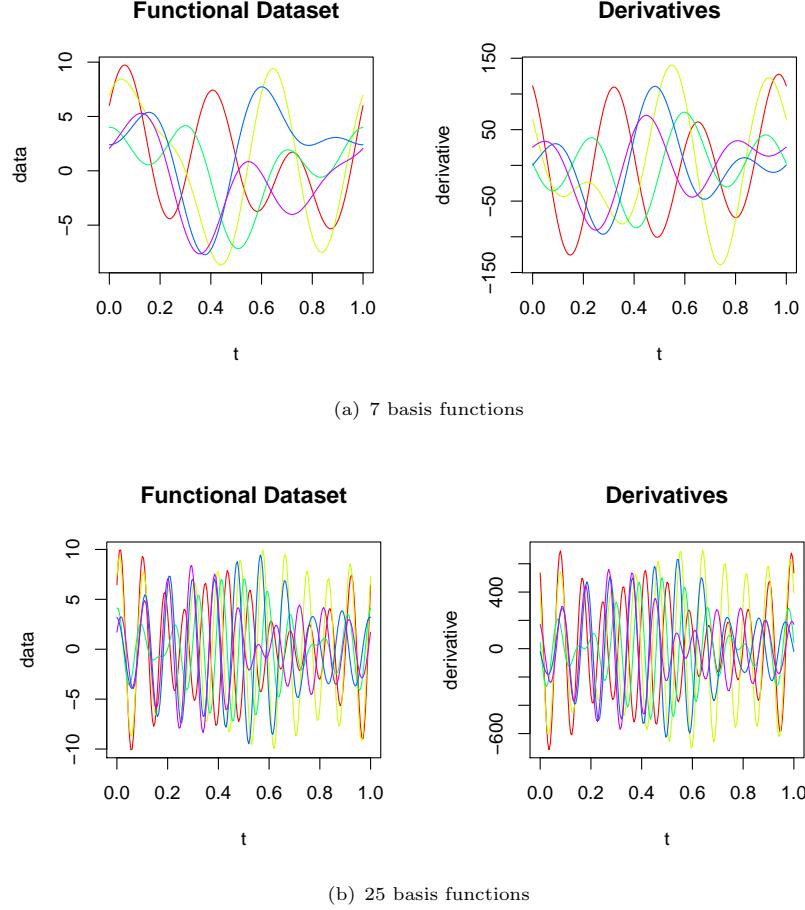


FIG 1. First 5 data of the functional datasets and corresponding derivatives. On the left: 5 data of the first dataset, built on a 7 Fourier functions basis. On the right: 5 data of the second dataset, built on a 25 Fourier functions basis, assuming non-zero only the last 7 coefficients.

Therefore, for both functional random fields, the  $L^2$  trace-variograms coincide:

$$2\gamma_{L^2}^{(1)} = 2\gamma_{L^2}^{(2)} = \sum_{k=1}^7 2\gamma_{\xi_k},$$

where  $2\gamma_{\xi_k}$  indicate the variogram of the field  $\xi_k$ ,  $k = 1, \dots, 7$ . Obviously, also their empirical estimates coincide (Figure 2a). Notice that the different behavior of the two datasets along the coordinate  $t$  is lost when inspecting  $L^2$  trace-variograms: they are able to capture only the structure of spatial dependence determined by the fields  $\xi_j$ ,  $j = 1, \dots, 7$ , ignoring the possibly different associated frequencies.

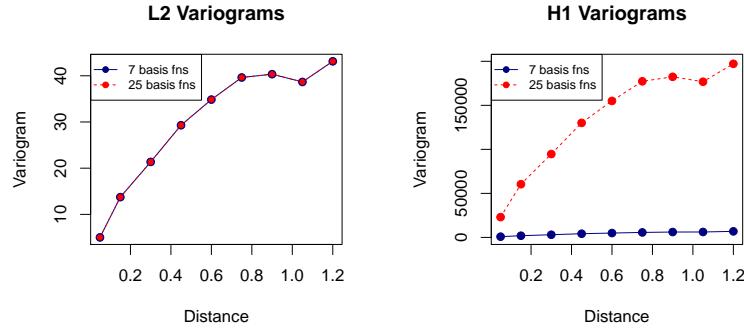


FIG 2. Empirical trace-variograms in  $L^2$  (on the left) and  $\mathcal{H}^1$  (on the right).

Information regarding curve fluctuation can be modeled through first derivative: we thus assume  $H = \mathcal{H}^1$ . Trace-variograms in  $\mathcal{H}^1$  can be computed by:

$$2\gamma^{(m)}(\mathbf{s}_i, \mathbf{s}_j)_{\mathcal{H}^1} = 2\gamma^{(m)}(\mathbf{s}_i, \mathbf{s}_j)_{L^2} + \text{Var}(D\chi_{\mathbf{s}_i}^{(m)} - D\chi_{\mathbf{s}_j}^{(m)})_{L^2}.$$

However, since:

$$\begin{aligned} \text{Var}(D\chi_{\mathbf{s}_i}^{(m)} - D\chi_{\mathbf{s}_j}^{(m)})_{L^2} &= \mathbb{E}[\|D\chi_{\mathbf{s}_i}^{(m)} - D\chi_{\mathbf{s}_j}^{(m)}\|_{L^2}^2] - \|\mathbb{E}[D\chi_{\mathbf{s}_i}^{(m)} - D\chi_{\mathbf{s}_j}^{(m)}]\|_{L^2}^2 \\ &= \sum_{k=1}^{N_m} \left\lfloor \frac{k}{2} \right\rfloor^2 \pi^2 \mathbb{E}[|\xi_k(\mathbf{s}_i) - \xi_k(\mathbf{s}_j)|^2], \end{aligned}$$

$2\gamma_{\mathcal{H}^1}^{(1)}$  does not coincide with  $2\gamma_{\mathcal{H}^1}^{(2)}$ . Indeed:

$$\begin{aligned} 2\gamma_{\mathcal{H}^1}^{(1)} &= 2\gamma_{L^2}^{(1)} + \sum_{k=2}^7 \left\lfloor \frac{k}{2} \right\rfloor^2 \pi^2 2\gamma_{\xi_k} = \sum_{k=1}^7 \left( 1 + \left\lfloor \frac{k}{2} \right\rfloor^2 \pi^2 \right) 2\gamma_{\xi_k}; \\ 2\gamma_{\mathcal{H}^1}^{(2)} &= 2\gamma_{L^2}^{(2)} + \sum_{k=19}^{25} \left\lfloor \frac{k}{2} \right\rfloor^2 \pi^2 2\gamma_{\xi_{k-18}} = \sum_{k=19}^{25} \left( 1 + \left\lfloor \frac{k}{2} \right\rfloor^2 \pi^2 \right) 2\gamma_{\xi_{k-18}}. \end{aligned}$$

Notice that, for  $k = 1, \dots, 7$ , the weights associated to the variogram  $2\gamma_{\xi_k}$  depends on the frequency associated to  $\xi_k$ , a greater weight being assigned to a higher frequency.

Figure 2b shows the empirical  $\mathcal{H}^1$  trace-variograms estimated from the two datasets. Even if the shapes of the estimates are not completely appreciable from the Figure, they are very similar, without showing notable differences with respect to  $L^2$  estimates; however the orders of magnitude of the horizontal asymptotes – twice the sills – are significantly different. Indeed, the different variances characterizing the two fields appear clearly: the curve corresponding to  $2\gamma_{\mathcal{H}^1}^{(2)}$  (in red) is much higher than the other (in blue), since the energy of the

random field  $\{\chi_s^{(2)}, s \in D\}$  is much higher than that of the other. In conclusion, the example clearly evidences that the choice of the space for the analysis has to be carefully taken, according to the dataset structure and, above all, to the purposes of the analysis. Indeed, if the aim of the analysis is purely spatial, then  $L^2$  space is rich enough for exploratory analysis and for kriging prediction. In other situations, the choice of a Sobolev space might be needed instead. This is the case of the dynamical system example in Remark 7 of Section 2 or of the example presented here.

#### 4. Algorithms

**Drift estimation** In order to compute the Universal Kriging prediction coherently with the established theoretical results, an iterative algorithm is necessary. Indeed, both the GLS drift estimator  $\hat{m}_{\vec{s}}^{GLS}$  and the system (18) depend substantially on the residual covariance structure, that can be assessed only once an estimation of the residual process – obtainable by difference from the estimate  $\hat{m}_{\vec{s}}^{GLS}$  – is available.

Therefore we propose to initialize the procedure to the ordinary least squares (OLS) estimate, computing at each step the residual estimate and the related trace-(semi)variogram structure, as well as the update of the drift estimate on the basis of the structure of spatial dependence currently available.

Having reached convergence, proved to be within five iterations by simulations, the final estimate of the (semi)variogram model can be used to solve the Universal Kriging system (18) – by exploiting (2.2) – deriving the desired prediction. The described algorithm is summarized in Algorithm 10.

**Algorithm 10.** Given a realization  $\chi_{\vec{s}} = (\chi_{s_1}, \dots, \chi_{s_n})$  of the nonstationary random field  $\{\chi_s, s \in D\}$ ,  $D \subset \mathbb{R}^d$ , representable as in (13):

1. Estimate the drift vector  $m_{\vec{s}}$  through the OLS method ( $\hat{m}_{\vec{s}}^{OLS} = \mathbb{F}_{\vec{s}}(\mathbb{F}_{\vec{s}}^T \mathbb{F}_{\vec{s}})^{-1} \mathbb{F}_{\vec{s}}^T \chi_{\vec{s}}$ ) and set  $\hat{m}_{\vec{s}} := \hat{m}_{\vec{s}}^{OLS}$ .
2. Compute the residual estimate  $\hat{\delta}_{\vec{s}} = (\hat{\delta}_{s_1}, \dots, \hat{\delta}_{s_n})$  by difference  $\hat{\delta}_{\vec{s}} = \chi_{\vec{s}} - \hat{m}_{\vec{s}}$ .
3. Estimate the trace-semivariogram  $\gamma(\cdot)$  of the residual process  $\{\delta_s, s \in D\}$  from  $\hat{\delta}_{\vec{s}}$  first with the empirical estimator (20), then fitting a valid model  $\gamma(\cdot; \hat{\vartheta})$ . Derive from  $\gamma(\cdot; \hat{\vartheta})$  the estimate  $\hat{\Sigma}$  of  $\Sigma$ .
4. Estimate the drift vector  $m_{\vec{s}}$  with  $\hat{m}_{\vec{s}}^{GLS}$ , obtained from  $\chi_{\vec{s}}$  using (26).
5. Repeat 2.–4. until convergence has been reached.

For computational efficiency reasons, the step 4. can be performed through the auxiliary uncorrelated vector  $\tilde{\chi}_{\vec{s}} = L^{-1}\chi_{\vec{s}}$ , where  $L$  appears in the Cholesky decomposition  $\hat{\Sigma} = LL^T$ . Indeed,  $\tilde{\chi}_{\vec{s}}$  is an estimate of  $\tilde{\chi}_{\vec{s}} = \Sigma^{-1/2}\chi_{\vec{s}}$  since the inverse Cholesky factor  $L^{-1}$  provides an estimate of  $\Sigma^{-1/2}$  (for details see the proof of Proposition 8 in the Appendix).

**Drift model selection** Although knowledge of the functions  $f_l$ , for  $l = 1, \dots, L$  ( $f_0(s) = 1$  for all  $s \in D$ ), is one of the underlying assumptions for

the procedure detailed in Algorithm 10, in most applications no ‘a priori’ information is available about the family  $\{f_l\}_{l=0,\dots,L} = \{f_0, \dots, f_L\}$  (e.g., no scalar external drift for the observed phenomenon is known). Therefore a model selection step before the application of Algorithm 10 is needed. In order to handle the model selection problem, we propose first to choose a number of candidate regressors families – e.g. the  $2^5$  polynomials of order lower than 2 – then to select the optimal set of regressors with a predictive criterion.

Formally, consider  $N_f$  collections of functions  $f_{\vec{l}}^k = \{f_0^k, \dots, f_L^k\}$ , corresponding to  $N_f$  possible drifts  $m_s^k = \sum_{l=0}^L a_l f_l^k(s)$ ,  $s \in D$ ,  $k = 1, \dots, N_f$ . The aim of the proposed method is the determination of a permutation  $\{(1), \dots, (N_f)\}$  of the set of indexes  $\{1, \dots, N_f\}$  according to the mean squared error of prediction:

$$MSE_k = \mathbb{E}[\|\chi_s - \chi_s^{*k}\|^2], \quad k = 1, \dots, N_f.$$

that can be assessed by a cross-validation (leave-on-out) technique combined with a Universal Kriging prediction, based on a proper drift estimate. The proposed procedure is summed up in the following Algorithm.

**Algorithm 11.** Given a realization  $\chi_{s_1}, \dots, \chi_{s_n}$  of the nonstationary random field  $\{\chi_s, s \in D\}$  and  $N_f$  collections of functions  $f_{\vec{l}}^k = \{f_0^k, \dots, f_L^k\}$  (candidate forms for the drift):

1. Fix a collection  $f_{\vec{l}}^k$ ,  $k = 1, \dots, N_f$ ;
2. Compute the GLS drift estimate  $\hat{m}_{\vec{s}}^{GLS,k}$ , the residual estimate  $\hat{\delta}_{\vec{s}}^k$  and the corresponding trace-semivariogram model  $\gamma_k(\cdot)$  applying  $M$  iterations of Algorithm 10 ( $M = 1$  for OLS estimate);
3. For each fixed  $i = 1, \dots, n$ , predict  $\chi_{s_i}$  from  $\chi_{\vec{s}-i} = (\chi_{s_j})_{j \neq i}$  through the Universal Kriging predictor  $\chi_{s_i}^{*k}$  solving (18) with  $\gamma = \gamma_k$  and  $f_{\vec{l}} = f_{\vec{l}}^k$ ;
4. Compute the sample mean squared error:  $MSE_k = \frac{1}{n} \sum_{i=1}^n \|\chi_{s_i} - \chi_{s_i}^{*k}\|^2$ ;
5. Repeat 1.-4. for every collection  $f_{\vec{l}}^k$ ,  $k = 1, \dots, N_f$ ;
6. Sort  $\{MSE_1, \dots, MSE_{N_f}\}$  in increasing order, determining the optimal permutation  $\{(1), \dots, (N_f)\}$  of  $\{1, \dots, N_f\}$ ; order the collections  $\{f_{\vec{l}}^k\}_{k=1,\dots,N_f}$  according to  $\{(1), \dots, (N_f)\}$ ,  $\{f_{\vec{l}}^{(k)}\}_{(k)=1,\dots,N_f}$ ;
7. For  $k = 1, \dots, N_f$ :
  - a. Check the second-order stationarity of the residual semivariogram model  $\gamma_{(k)}(\cdot)$  relative to the  $(k)$ -th model;
  - b. If  $\gamma_{(k)}(\cdot)$  proves to be second-order stationary, select the optimal drift model as:

$$m_s^{opt} = \sum_{l=0}^L a_l f_l^{(k)}(s), \quad s \in D,$$

and stop the procedure.

Note that step 7. of Algorithm 11 guarantees the stationarity of the residuals and thus ensures that the Universal Kriging hypotheses are fulfilled by the selected drift model. The residual second-order stationarity can be checked through the analysis of the residual empirical (semi)variogram with the same

criteria used in finite-dimensional geostatistics (e.g., presence of a sill close to the estimated variance and sub-quadratic growth for increasing distances). Moreover the adoption of a predictive criterion contributes to avoid over-fitting: very complex models are unable to filter the noise in the observed data, therefore the selection of a too complex drift structure would also catch part of their stochastic variability, reducing considerably the predictive power of the model.

In order to obtain the final Universal Kriging prediction, Algorithms 10 and 11 need to be combined. Two main choices can be made, according to computational efficiency or estimation accuracy criteria.

The first possibility is to consider for step 2. of Algorithm 11 the OLS estimation method, which actually corresponds to the very first iteration of the Algorithm 10. By making this choice a three-step procedure is finally obtained: first drift model selection by Algorithm 11, second GLS estimation by Algorithm 10, finally Universal Kriging prediction. This choice aims mainly in controlling the computational costs, ignoring the possible influences of the drift estimation method on the prediction (not always negligible).

The second possible choice is the integration of Algorithms 10 and 11, by considering GLS estimation method during step 2. of Algorithm 11 and then using the drift estimate of the selected model, available at the end of Algorithm 11, for Universal Kriging prediction. This choice does not preserve the computational costs from becoming high in the presence of many candidates families, but permits to perform a more precise drift model selection, which contributes to make the kriging prediction more accurate. Moreover, the fairly high speed of convergence of Algorithms 10 – within 5 iterations in all the simulations – and the consideration of a moderate number of drift candidates contribute to control the computational efficiency of the procedure. For these reasons, we make the latter choice for the case study illustrated in the next section.

We explore the performance of the Algorithms 10 and 11 with an extensive simulation study presented as supplementary material in (Menafoglio et al., 2013).

## 5. A case study: Analysis of Canada's Maritime Provinces Temperatures

**Analysis of averaged temperatures data** The proposed methodology will be now applied to the Canada's Maritime Provinces Temperatures dataset (available in R package `geofd` (Giraldo et al., 2010b)), that collects daily mean temperatures data, observed in 35 meteorological stations located in Canada's Maritimes Provinces (Figure 3). This region consists of three provinces, Nova Scotia, New Brunswick and Prince Edward Island, located in the south-eastern part of Canada (Figure 3, first panel), whose very distinctive feature is the exposition toward the sea: indeed, especially because of the Gulf Stream coming from the Ocean, the Provinces climate is temperate, characterized by mild winters and cool summers (Stanley, 2002).

For each sampled site (Figure 3, second panel), identified by geographical coordinates (longitude, latitude), the original raw data consist of 365 measure-

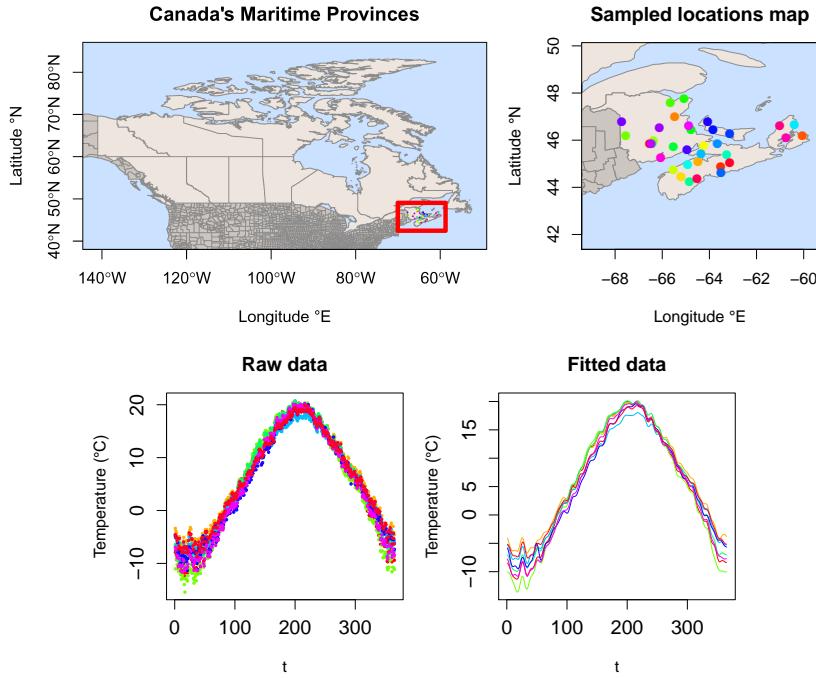


FIG 3. *Canada's Maritime Provinces Temperatures dataset, averaged over 1960–1994. From left to right: map of Canada highlighting the Maritimes region; zoom of Maritime Provinces and sampled locations; 8 raw data; 8 fitted data.*

ments (one per day), obtained by averaging, over the years 1960 to 1994, the daily mean temperatures recorded by the Meteorological Service of Canada. This dataset, besides being very similar to the Canadian Weather dataset handled in (Ramsay and Silverman, 2005), has been analyzed in the literature concerning geostatistical theory for stationary and isotropic functional processes (e.g., (Giraldo, 2009; Giraldo et al., 2010a)).

Coherently with previous analyses, the Hilbert space  $H$  has been set to be  $L^2$  and raw data (Figure 3, third panel) have been projected on a basis of 65 Fourier function, selected in (Giraldo, 2009) through a non parametric functional cross-validation procedure (Figure 3, last right panel).

Denote with  $\{\chi_s, s \in D \subset \mathbb{R}^d\}$  the random field of temperature functions and call  $D$  the spatial domain, endowed with the non-Euclidean metric induced by the geodesic distance that, assuming a spherical approximation for the Earth, can be explicitly computed as:

$$d_g(s_1, s_2) = 2R_m \arcsin \left( \sqrt{\sin^2 \left( \frac{\zeta_1 - \zeta_2}{2} \right) + \cos(\zeta_1) \cos(\zeta_2) \sin^2 \left( \frac{\varphi_1 - \varphi_2}{2} \right)} \right), \quad (32)$$

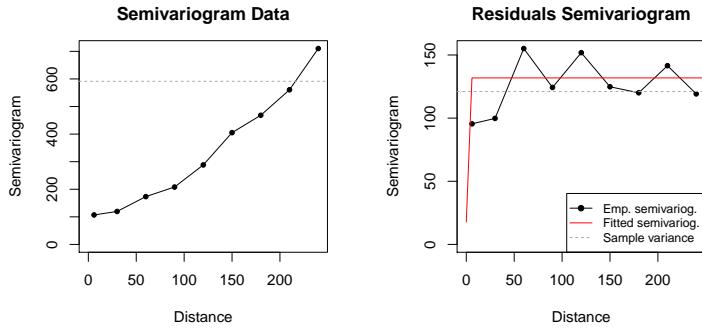


FIG 4. Estimated trace-semivariograms from data (on the left) and from residuals (on the right).

where  $\mathbf{s}_i = (\zeta_i, \varphi_i)$ , (longitude and latitude)  $i = 1, 2$  e  $R_m \simeq 6371$  km indicates the Earth's mean radius. Although the validity of usual parametric variogram models is not guaranteed in non-Euclidean spaces (Curriero, 2006), both the spherical and the exponential models are valid in the spherical geometry (Huang et al., 2011) and thus can be used in this case.

By a first stationary analysis of the data through the trace-semivariogram empirical estimate, represented in the left panel of Figure 4, the non-stationarity of the field is apparent (super-quadratic growth for increasing distances, no evidence of a sill close to the sample variance of the data). Therefore, we analyze the data by means of Algorithms 10 and 11, searching the optimal drift model among polynomials of degree lower than 2.

The linear model singled out by the Algorithm 11 is model 23:

$$m(\mathbf{s}, t) = a_0(t) + a_1(t)y + a_2(t)x^2 + a_3(t)xy, \quad \mathbf{s} = (x, y), \quad t \in \mathcal{T} = [0, 365], \quad (33)$$

where the coordinates are identified with latitude and longitude,  $(x, y) = (\zeta, \varphi)$ . Concerning the residuals structure of spatial dependence, the right panel of Figure 4 shows that the parametric model that best fits the empirical trace-semivariogram estimate is a pure-nugget model, meaning that the estimated residuals are uncorrelated. Therefore, the spatial variability characterizing the data is mostly explained by the deterministic drift term, while the residuals do not seem to contribute to the spatial correlation of the stochastic process.

In such a case, Universal Kriging predictor reduces to the drift estimate – i.e. the prediction which would have been obtained via FDA linear models, which in fact provides the best predictive performance among the functional forms tested by the Algorithm 11. In particular, it is better performing – in terms of cross-validation errors – than the Ordinary Kriging predictor (i.e. drift model 1) computed by using geodesic distance (Table 1, second column), as well as by using Euclidean distance – adopted in previous works (e.g., (Giraldo et al., 2010a)) – (Table 1, third column). Indeed, cross-validation statistics obtained

TABLE 1

*Comparison of cross-validation squared error statistics computed by Universal Kriging (UKFD), Ordinary Kriging (OKFD, Giraldo et al. (2010a)) – using geodesic and Euclidean distance. The per cent reduction of UKFD error with respect to OKFD is reported between brackets*

	UKFD (Pure nugget; Geod. dist.)	OKFD (Geod. dist.)	OKFD (Eucl. dist.)
Median	99.1 ( $\downarrow$ 31%, 32%)	144.3	144.6
Mean (MSE)	155.4 ( $\downarrow$ 8%, 13%)	168.8	179.2

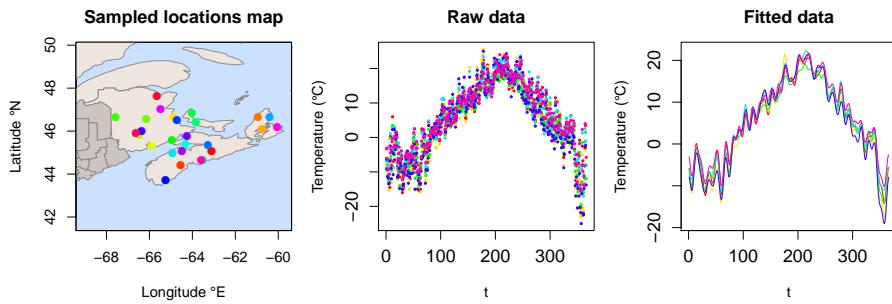


FIG 5. *Canada's Maritime Provinces Temperatures dataset, year 1980. From left to right: map of Maritime Provinces and sampled locations; 6 raw data; 6 fitted temperature curves.*

with the proposed methodology are improved at least of 8% with respect to the stationary methods and at least of 13% with respect to the analyses already presented in the literature<sup>1</sup>.

The fact that the residuals do not show a non-trivial structure of spatial dependence might be due to the average over 34 years made on the original data, which may have masked the small scale variability. For this reason, we will now apply our methodology to a one-year dataset, collecting the measurements recorded in the same area, during the year 1980.

**Analysis of 1980 temperatures data** The dataset analyzed in this second part of the case study collects daily mean temperatures recorded, along the (leap) year 1980, in 27 meteorological stations located in the same region considered before (Figure 5, left panel). The raw data (Figure 5, central panel), available on Natural Resources of Canada website (2012), have been projected as before on a basis of 65 Fourier functions, obtaining the functional dataset represented in the right panel of Figure 5. Choices for the functional and spatial metrics have been taken coherently with the previous analysis ( $H = L^2$  and geodesic distance on the spatial domain).

<sup>1</sup>The codes for computing the stationary predictors are available in `geofd` R package. Cross-validation statistics are here computed with respect to fitted data and are thus different from statistics reported in previous works (e.g., (Giraldo et al., 2010a)), that refer to raw data instead.

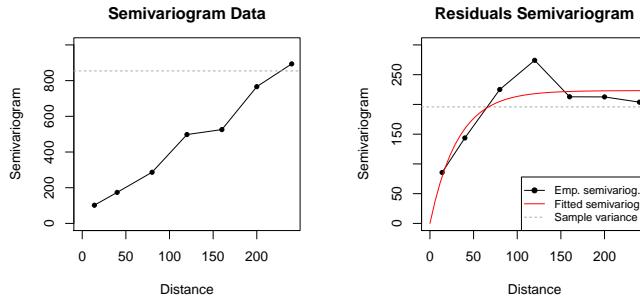


FIG 6. Estimated trace-semivariograms from data (on the left) and from residuals (on the right).

Although the empirical trace-semivariogram estimated from the data (Figure 6, left panel) seems not so far from stationarity, we proceed in applying our procedure, since among polynomials of degree lower than two also the stationary model is tested by Algorithm 11. The selected model is model 31:

$$m(\mathbf{s}, t) = a_0(t) + a_1(t)y + a_2(t)x^2 + a_3(t)y^2 + a_4(t)xy, \quad \mathbf{s} = (x, y), \quad t \in \mathcal{T} = [0, 366],$$

which provided the best cross-validation results.

By observing the residuals trace-semivariogram, a strong correlation among residuals can be recognized; in particular, the exponential structure appears suitable for fitting the empirical semivariogram. Therefore, in this case, GLS method is the most appropriate for estimating the drift, while Universal Kriging seems the most appropriate method to perform optimal spatial prediction.

Figure 7 shows the contour plots of the GLS drift estimate (upper panels) and of the Universal Kriging prediction (lower panels), obtained by fixing the time coordinate  $t$  to the Spring Equinox (21<sup>st</sup> March, first panels), the Summer Solstice (21<sup>st</sup> June, second panels), the Autumn Equinox (23<sup>rd</sup> September, third panels) and the Winter Solstice (21<sup>st</sup> December, fourth panels).

The first interesting result to be noticed is the climatical interpretation emerging from the obtained maps. The exposition of the Maritimes region towards the sea plays a key role indeed, due to the alternation of Atlantic warm-humid currents with freezing streams coming from the internal Canadian regions. These currents circulations significantly influences the temperatures and clearly reflects on drift contour lines (Figure 7, upper panels) with a clear rotation, which begins during the springtime and continues until September under the influence of Gulf Stream from South (third panel). Indeed, the early spring drift map (first panel) presents colder temperatures in the internal part of New Brunswick, while the warmer temperatures are recorded in the South; early summer panel (second panel) presents the opposite spatial behavior instead, featured by a warmer zone in the continental region and a cooler area along the sea. Moreover, notice the different rotation speed during the year – much faster in the transition from spring to summer and from summer to autumn than during the other months –

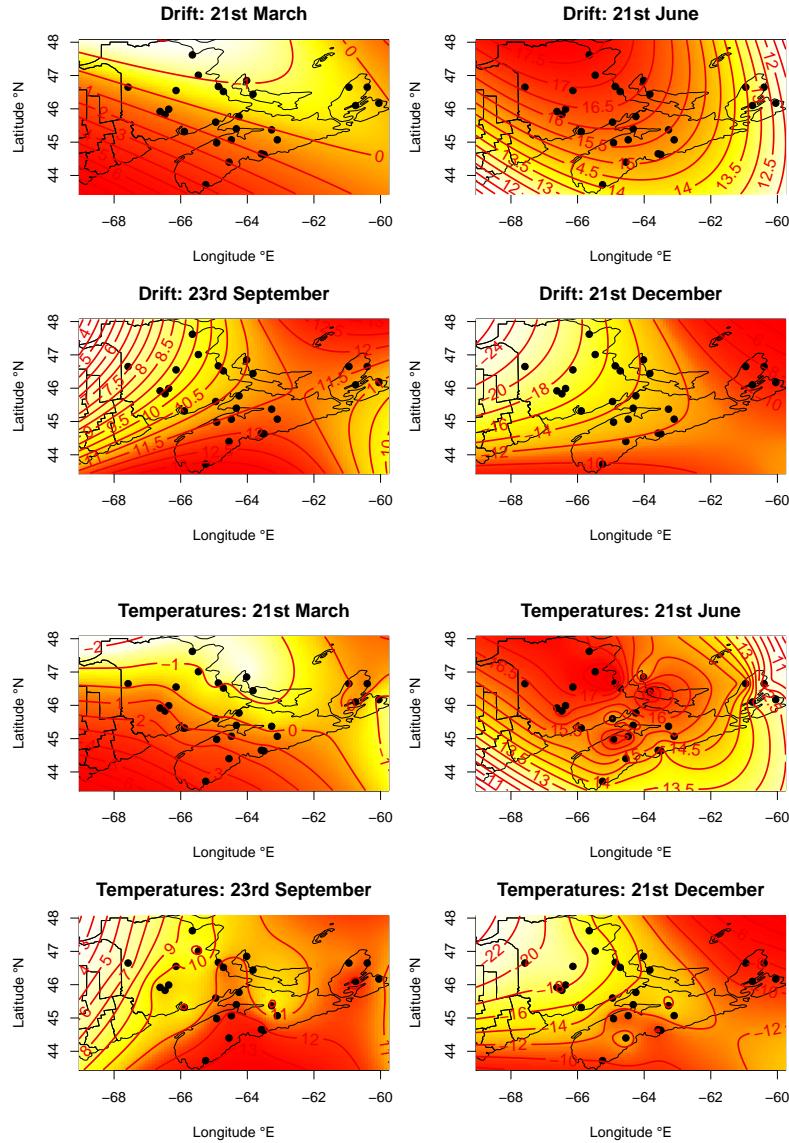


FIG 7. Drift estimation and Universal Kriging prediction contour plots for the Spring Equinox (21<sup>st</sup> March), to the Summer Solstice (21<sup>st</sup> June), to the Autumn Equinox (23<sup>rd</sup> September) and to the Winter Solstice (21<sup>st</sup> December).

that reflects the climatical trend in the region, featured by long lasting cold seasons and shorter warm periods.

Together with the drift rotation speed, the complexity in the spatial behavior (Figure 7, lower panels) seems to change along the temporal coordinate. Indeed,

TABLE 2  
*Comparison of cross-validation squared error statistics computed by Universal Kriging (UKFD), and Ordinary Kriging (OKFD) using in both cases geodesic distance. The reduction of UKFD error with respect to OKFD is reported between brackets*

	UKFD (Geod. dist.)	OKFD (Geod. dist.)
Median	190.2 ( $\downarrow$ 8%)	205.9
Mean (MSE)	263.5 ( $\downarrow$ 14%)	306.8

Universal Kriging maps relative to colder seasons (first, third and fourth panels) point out a much stronger influence of the drift component on the prediction with respect to the summer season (second panel); the latter is featured by very local structures instead, which seem to be strongly related to the geographical configuration of the area – notice in particular the low temperature zones marked off by the Bay of Fundy and by the Atlantic Ocean. The interpretability of our results supports the conclusions drawn from the simulation study presented in the supplementary material (Menafoglio et al., 2013): our methodology applied to real data provides fairly accurate results also locally, although curves are handled as points of an infinite-dimensional space, under global assumptions.

Besides being climatically interpretable, the obtained results are consistent with the seasonal reference maps published by Natural Resources Canada, providing a further validation of the model.

Finally, cross-validation analysis has been performed, comparing non-stationary results with those obtained by applying a stationary model. Table 2 reports cross-validation statistics relative to Universal and Ordinary Kriging, using in both cases the geodesic distance (first and second column respectively). A significant reduction in the prediction error is due to the introduction of the drift term: indeed, moving from Ordinary to Universal Kriging the error decreases at least of 8% – if we consider the median value, presenting a 14% reduction in mean.

Concerning the local errors along the temporal coordinate  $t$ , the modeling of a non-constant spatial mean makes the prediction unbiased and thus prevents the systematic overestimation or underestimation of the data, which occurs instead in OKFD prediction, in Bon Accord and Truro respectively (Figure 8a).

Moreover, by considering the spatial distribution of cross-validation errors (Figure 8b), it clearly appears that the most significant increase of predictive power is obtained in peripheral zones, in particular Bon Accord (NS) and Truro (NB) (western part of the maps). This kind of improvement is explained by the increased flexibility reached through the introduction of a drift term. This drives the prediction in peripheral areas and allows to reach more extreme predicted values, above all during the winter season where the drift is more influent on the prediction. For instance, observe the NW corner of Universal Kriging maps computed for the 1<sup>st</sup> January (Figure 9a): with OKFD the most extreme predicted temperatures are around -9C, while UKFD prediction reaches values below -16° C.

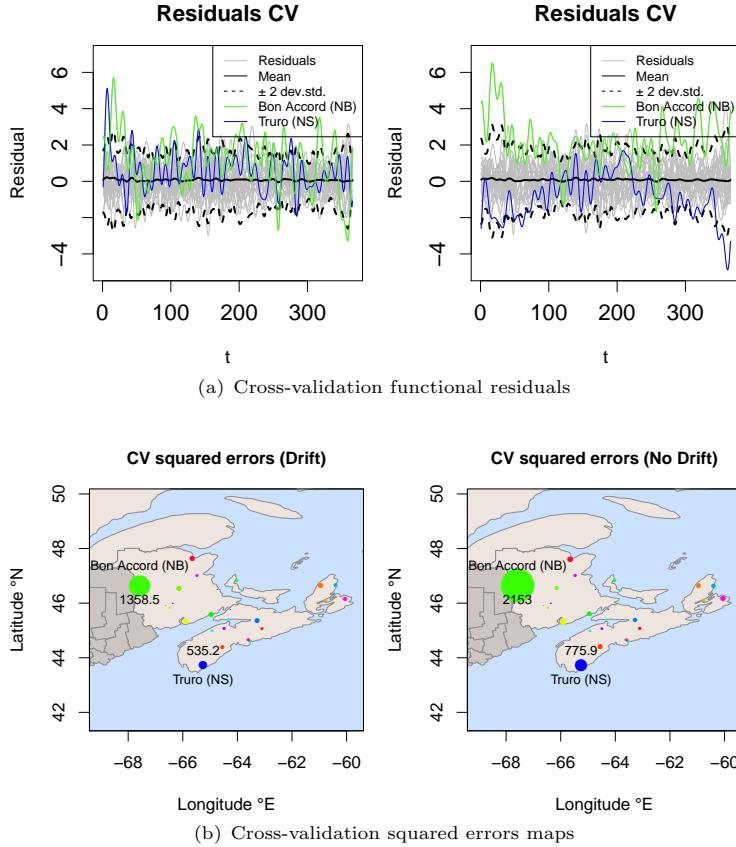


FIG 8. Comparison of cross-validation results for Universal Kriging and Ordinary Kriging, using in both cases the geodesic distance. In Subfigure (a): functional residuals for the 27 locations (grey lines) highlighting Bon Accord (NS) (green line) and Truro (NB) (blue line). In Subfigure (b): squared errors map; the dimension of the points is proportional to the cross-validation squared error; Bon Accord (NS) (green points) and Truro (NB) (blue points) are marked in the western part of the maps.

On the other hand, the additional flexibility obtained by introducing the drift term contributes to mitigate the smoothing effect of kriging; this reflects on a very accurate local prediction that reproduces the local structures much better than the Ordinary Kriging interpolation. For example, look at the local structures that arise during the summer period between the Bay of Fundy and the Atlantic Ocean in Figure 9b: they are very well reproduced by UKFD prediction, while they are severely smoothed in the OKFD interpolation.

Therefore, the non-stationary prediction, obtained by applying our procedure, proves to be much more satisfactory than the stationary interpolations in terms both of global prediction error as well as of local behavior: Universal Kriging prediction is precise and flexible, besides being simple and easy to compute.

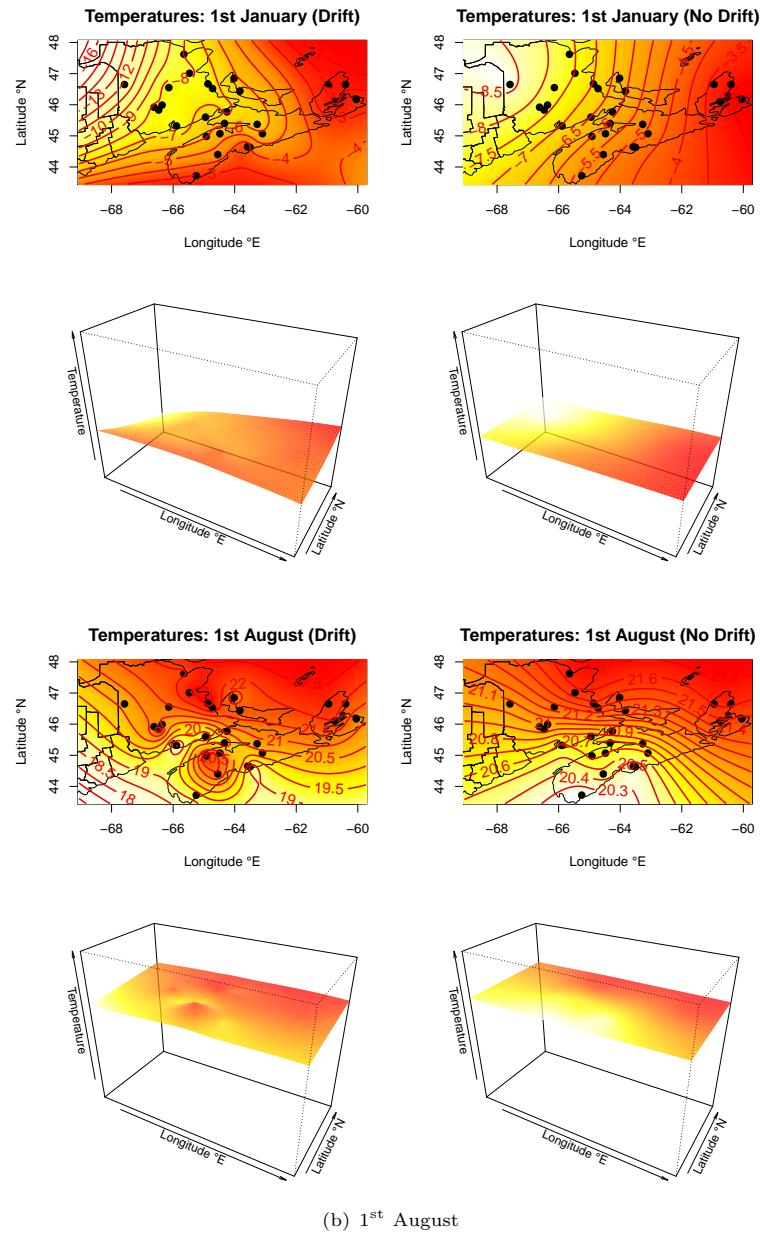


FIG 9. Comparison of the results obtained with Universal Kriging and Ordinary Kriging, using in both cases the geodesic distance. Upper panels show contours maps, lower panels represent the associated 3D plots.

## 6. Conclusions and further research

In this work, a new kriging methodology for non-stationary spatially dependent functional data has been developed. On one hand, the theoretical effort has been spent for the formulation of a coherent framework, based on minimal assumptions. On the other hand, the developed algorithms aimed at making our theoretical results applicable on real data, through reliable and efficient procedures.

The development of inferential tools for spatially dependent functional data is still one of the most challenging topic to be addressed: the significance of regressors coefficients should be tested during drift model selection and kriging confidence bands should be provided together with point-wise prediction. To this end, a possible immediate perspective is given by the extension to the georeferenced functional case of non-parametric resampling methods like the bootstrap – e.g., (Efron and Tibshirani, 1993) and more recently, in the field of FDA, (Ferraty et al., 2010), which would allow to avoid distributional assumptions by means of a computer-intensive technique.

Developing statistical models and inferential procedures for general Hilbert spaces, instead of working out *ad hoc* techniques for the  $L^2$  space, opens broad perspectives of research: indeed, it may allow the integration of the kriging methodology, which is in fact an interpolation technique, with the physical model underlying the observed phenomenon. In this direction, more complex linear models (e.g., FDA Total Model (Ramsay and Silverman, 2005)) would be worth investigating in order to model more precisely the drift term, possibly including more complex regressors which might influence or drive the physical system.

## 7. Appendix: Proofs

*Proof of Proposition 8.* Consider the auxiliary optimal problem:

$$\min_{\hat{\mathbf{a}}_{\vec{l}} \in H^{L+1}} \tilde{\Phi}^{OLS}(\hat{\mathbf{a}}_{\vec{l}}) \quad (34)$$

where:

$$\tilde{\Phi}^{OLS}(\hat{\mathbf{a}}_{\vec{l}}) = \|\tilde{\chi}_{\vec{s}} - \tilde{\mathbb{F}}_{\vec{s}} \hat{\mathbf{a}}_{\vec{l}}\|_{H^n}^2, \quad (35)$$

with  $\tilde{\chi}_{\vec{s}} = \Sigma^{-1/2} \chi_{\vec{s}}$ , whose components are uncorrelated, and  $\tilde{\mathbb{F}}_{\vec{s}} = \Sigma^{-1/2} \mathbb{F}_{\vec{s}}$ .

It is easily seen the equivalence of the estimation problems (23) and (34), as  $\Phi^{GLS}(\hat{\mathbf{a}}_{\vec{l}}) = \tilde{\Phi}^{OLS}(\hat{\mathbf{a}}_{\vec{l}})$ .

Assume that  $\Sigma$  is known and denote with  $\tilde{V}$  the closed subspace of  $H^n$  generated by linear combination of  $\tilde{\mathbb{F}}_{\vec{s}}$  columns with coefficients in  $H$  and let  $\tilde{V}^\perp$  be its orthogonal complement:

$$\tilde{V} = \{\tilde{\mathbf{v}} \in H^n : \tilde{\mathbf{v}} = \tilde{\mathbb{F}}_{\vec{s}} \hat{\mathbf{a}}_{\vec{l}}, \quad \hat{\mathbf{a}}_{\vec{l}} \in H^L\}, \quad (36)$$

$$\tilde{V}^\perp = \{\tilde{\mathbf{w}} \in H^n : \langle \tilde{\mathbf{w}}, \tilde{\mathbf{v}} \rangle_{H^n} = 0, \quad \forall \tilde{\mathbf{v}} \in \tilde{V}\}. \quad (37)$$

The estimator  $\widehat{\tilde{\mathbf{m}}}_{\vec{s}} = \widetilde{\mathbb{F}}_{\vec{s}} \widehat{\mathbf{a}}_{\vec{l}}$  of  $\tilde{\mathbf{m}}_{\vec{s}}$  – mean vector of  $\tilde{\chi}_{\vec{s}}$  – minimizing  $\tilde{\Phi}^{OLS}$  is the projection of  $\tilde{\chi}_{\vec{s}}$  on  $\widetilde{V}$ , while the residual vector  $\widehat{\tilde{\delta}}_{\vec{s}} = \tilde{\chi}_{\vec{s}} - \widehat{\tilde{\mathbf{m}}}_{\vec{s}}$  is the projection of  $\tilde{\chi}_{\vec{s}}$  on  $\widetilde{V}^\perp$ :

$$\widehat{\tilde{\mathbf{m}}}_{\vec{s}} = P_{\widetilde{V}} \tilde{\chi}_{\vec{s}} \quad (38)$$

$$\widehat{\tilde{\delta}}_{\vec{s}} = P_{\widetilde{V}^\perp} \tilde{\chi}_{\vec{s}} \quad (39)$$

If  $\text{rank}(\mathbb{F}_{\vec{s}}) = L + 1$  and  $\text{rank}(\Sigma) = n$ , then  $\text{rank}(\widetilde{\mathbb{F}}_{\vec{s}}) = L + 1$ , which ensures the existence and uniqueness of the projections (38) and (39).

Moreover, the projection (38) can be explicitly computed pre-multiplying  $\tilde{\chi}_{\vec{s}}$  by the orthogonal projection matrix  $\widetilde{\mathbb{H}} = \widetilde{\mathbb{F}}_{\vec{s}} (\widetilde{\mathbb{F}}_{\vec{s}}^T \widetilde{\mathbb{F}}_{\vec{s}})^{-1} \widetilde{\mathbb{F}}_{\vec{s}}^T$ , deriving directly the following linear expressions:

$$\begin{aligned} \widehat{\mathbf{a}}_{\vec{l}}^{GLS} &= (\widetilde{\mathbb{F}}_{\vec{s}}^T \widetilde{\mathbb{F}}_{\vec{s}})^{-1} \widetilde{\mathbb{F}}_{\vec{s}}^T \tilde{\chi}_{\vec{s}}; \\ \widehat{\tilde{\mathbf{m}}}_{\vec{s}} &= \widetilde{\mathbb{H}} \tilde{\chi}_{\vec{s}} = \widetilde{\mathbb{F}}_{\vec{s}} (\widetilde{\mathbb{F}}_{\vec{s}}^T \widetilde{\mathbb{F}}_{\vec{s}})^{-1} \widetilde{\mathbb{F}}_{\vec{s}}^T \tilde{\chi}_{\vec{s}}. \end{aligned} \quad (40)$$

Applying the inverse transformation, expressions (25) and (26) can be finally obtained.  $\square$

*Proof of Proposition 9.* Let  $\widehat{\mathbf{a}}_{\vec{l}}$  be a generic linear estimator of the coefficients  $a_{\vec{l}}$ :

$$\widehat{\mathbf{a}}_{\vec{l}} = \mathbb{A} \chi_{\vec{s}} + \mathbf{b}, \quad (41)$$

with  $\mathbb{A} \in \mathbb{R}^{L+1,n}$ ,  $\mathbf{b} \in H^{L+1}$ . The unbiasedness condition translates into the constraints:

$$\mathbb{A} \mathbb{F}_{\vec{s}} = \mathbb{I}_n \quad (42)$$

$$\mathbf{b} = \mathbf{0}, \quad a.e. \quad (43)$$

where  $\mathbb{I}_n$  is the identity matrix in  $\mathbb{R}^n$ ,  $\mathbf{0} \in H^{L+1}$  is the vector of  $L+1$  identically zero functions.

By definition of optimality of  $\widehat{\mathbf{a}}_{\vec{l}}^{BLUE}$ , for every other linear unbiased estimator  $\widehat{\mathbf{a}}_{\vec{l}}$ , the matrix:

$$\text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}) - \text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}^{BLUE}),$$

is positive semi-definite, or equivalently:

$$\mathbf{x}^T (\text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}) - \text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}^{BLUE})) \mathbf{x} \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

For a generic linear estimator (41), under the unbiasedness constraints (43), the variance-covariance matrix is:

$$\text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}) = \mathbb{A} \Sigma \mathbb{A}^T;$$

and, by the inequality (Shumway and Dean, 1968):

$$\boldsymbol{\alpha} \mathbb{C}^{-1} \boldsymbol{\alpha}^T \geq \boldsymbol{\alpha} \mathbb{D} (\mathbb{D}^T \mathbb{C} \mathbb{D})^{-1} \mathbb{D}^T \boldsymbol{\alpha}^T,$$

that holds for  $\mathbb{C} \in \mathbb{R}^{n,n}$  semidefinite positive,  $\mathbb{D} \in \mathbb{R}^{n,L+1}$  and  $\boldsymbol{\alpha} \in \mathbb{R}^n$ , a lower bound for  $\mathbf{x}^T \text{Cov}(\widehat{\mathbf{a}}_{\vec{l}}) \mathbf{x}$  can be obtained by setting  $\mathbb{C} = \Sigma^{-1}$ ,  $\boldsymbol{\alpha} = \mathbf{x}^T \mathbb{A}$  and  $\mathbb{D} = \mathbb{F}_{\vec{s}}$ :

$$\mathbf{x}^T \mathbb{A} \Sigma \mathbb{A}^T \mathbf{x} \geq \mathbf{x}^T \mathbb{A} \mathbb{F}_{\vec{s}} (\mathbb{F}_{\vec{s}}^T \Sigma^{-1} \mathbb{F}_{\vec{s}})^{-1} \mathbb{F}_{\vec{s}}^T \mathbb{A}^T \mathbf{x}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

The lower bound is reached for:

$$\mathbb{A}^{BLUE} = (\mathbb{F}_{\vec{s}}^T \Sigma^{-1} \mathbb{F}_{\vec{s}})^{-1} \mathbb{F}_{\vec{s}}^T \Sigma^{-1}.$$

Hence, the optimal linear estimator is:

$$\widehat{\mathbf{a}}_{\vec{l}}^{BLUE} = (\mathbb{F}_{\vec{s}}^T \Sigma^{-1} \mathbb{F}_{\vec{s}})^{-1} \mathbb{F}_{\vec{s}}^T \Sigma^{-1} \boldsymbol{\chi}_{\vec{s}} \equiv \widehat{\mathbf{a}}_{\vec{l}}^{GLS},$$

that in particular minimizes the mean square errors  $MSE_l$  simultaneously for every  $l = 0, \dots, L$ :

$$MSE_l = \mathbb{E}[\|\widehat{\mathbf{a}}_l - \mathbf{a}_l\|^2] = (\mathbb{A} \Sigma \mathbb{A}^T)_{ll},$$

subject to the unbiasedness constraints (42) and (43).

As a consequence, by linearity,  $\widehat{\mathbf{m}}_{\vec{s}}$  is the BLUE for the drift.  $\square$

*Proof of decomposition of variance (29).* Let  $\Sigma^{GLS}$  be the  $n \times n$  covariance matrix of the estimator  $\widehat{\boldsymbol{\delta}}_{\vec{s}}$ ,  $\Sigma^{GLS} = \mathbb{E}[\widehat{\boldsymbol{\delta}}_{\vec{s}} \widehat{\boldsymbol{\delta}}_{\vec{s}}^T]$  and consider the following matrix notations:

$$\begin{aligned} gg^T &= (\langle g_i, g_j \rangle), \quad g = (g_1, \dots, g_n) \in H^n \\ \mathbb{E}[\mathbb{A}] &= (\mathbb{E}[\mathbb{A}_{ij}]), \quad \mathbb{A} = (\mathbb{A}_{ij}) \in \mathbb{R}^n, \end{aligned}$$

Then:

$$\begin{aligned} \Sigma &:= \text{Cov}(\boldsymbol{\chi}_{\vec{s}}) = \mathbb{E}[(\boldsymbol{\chi}_{\vec{s}} - m_{\vec{s}})(\boldsymbol{\chi}_{\vec{s}} - m_{\vec{s}})^T] \\ &= \mathbb{E}[\Sigma^{1/2}(\tilde{\boldsymbol{\chi}}_{\vec{s}} - \tilde{m}_{\vec{s}})(\tilde{\boldsymbol{\chi}}_{\vec{s}} - \tilde{m}_{\vec{s}})^T (\Sigma^{1/2})^T] \\ &= \mathbb{E}[\Sigma^{1/2}(\tilde{\boldsymbol{\chi}}_{\vec{s}} \pm \widehat{\mathbf{m}}_{\vec{s}} - \tilde{m}_{\vec{s}})(\tilde{\boldsymbol{\chi}}_{\vec{s}} \pm \widehat{\mathbf{m}}_{\vec{s}} - \tilde{m}_{\vec{s}})^T (\Sigma^{1/2})^T] \\ &= \Sigma^{1/2} \mathbb{E}[(\tilde{\boldsymbol{\chi}}_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}})(\tilde{\boldsymbol{\chi}}_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}})^T + (\widehat{\mathbf{m}}_{\vec{s}} - \tilde{m}_{\vec{s}})(\widehat{\mathbf{m}}_{\vec{s}} - \tilde{m}_{\vec{s}})^T] (\Sigma^{1/2})^T \\ &= \mathbb{E}[\Sigma^{1/2}(\tilde{\boldsymbol{\chi}}_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}})(\tilde{\boldsymbol{\chi}}_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}})^T (\Sigma^{1/2})^T] \\ &\quad + \mathbb{E}[\Sigma^{1/2}(\widehat{\mathbf{m}}_{\vec{s}} - \tilde{m}_{\vec{s}})(\widehat{\mathbf{m}}_{\vec{s}} - \tilde{m}_{\vec{s}})^T (\Sigma^{1/2})^T] \\ &= \mathbb{E}[(\boldsymbol{\chi}_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}})(\boldsymbol{\chi}_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}})^T] + \mathbb{E}[(\widehat{\mathbf{m}}_{\vec{s}} - m_{\vec{s}})(\widehat{\mathbf{m}}_{\vec{s}} - m_{\vec{s}})^T] \\ &= \mathbb{E}[\widehat{\boldsymbol{\delta}}_{\vec{s}} \widehat{\boldsymbol{\delta}}_{\vec{s}}^T] + \mathbb{E}[(\widehat{\mathbf{m}}_{\vec{s}} - m_{\vec{s}})(\widehat{\mathbf{m}}_{\vec{s}} - m_{\vec{s}})^T] \\ &= \Sigma^{GLS} + \text{Cov}(\widehat{\mathbf{m}}_{\vec{s}}), \end{aligned}$$

where the third equality holds because:

$$\begin{aligned} &\mathbb{E}[(\tilde{\boldsymbol{\chi}}_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}})(\widehat{\mathbf{m}}_{\vec{s}} - \tilde{m}_{\vec{s}})^T] \\ &= \mathbb{E}[(\mathbb{I} - \widetilde{\mathbb{F}}_{\vec{s}} (\widetilde{\mathbb{F}}_{\vec{s}}^T \widetilde{\mathbb{F}}_{\vec{s}})^{-1} \widetilde{\mathbb{F}}_{\vec{s}}^T) \tilde{\boldsymbol{\chi}}_{\vec{s}} \tilde{\boldsymbol{\chi}}_{\vec{s}}^T \widetilde{\mathbb{F}}_{\vec{s}} (\widetilde{\mathbb{F}}_{\vec{s}}^T \widetilde{\mathbb{F}}_{\vec{s}})^{-1} \widetilde{\mathbb{F}}_{\vec{s}}^T] + \mathbb{E}[(\tilde{\boldsymbol{\chi}}_{\vec{s}} - \widehat{\mathbf{m}}_{\vec{s}})(\tilde{m}_{\vec{s}})^T] \\ &= (\mathbb{I} - \widetilde{\mathbb{F}}_{\vec{s}} (\widetilde{\mathbb{F}}_{\vec{s}}^T \widetilde{\mathbb{F}}_{\vec{s}})^{-1} \widetilde{\mathbb{F}}_{\vec{s}}^T) \widetilde{\mathbb{F}}_{\vec{s}} (\widetilde{\mathbb{F}}_{\vec{s}}^T \widetilde{\mathbb{F}}_{\vec{s}})^{-1} \widetilde{\mathbb{F}}_{\vec{s}}^T - (\tilde{m}_{\vec{s}} - \tilde{m}_{\vec{s}})(\tilde{m}_{\vec{s}})^T = (0) \quad \square \end{aligned}$$

## Supplementary Material

### Simulation Study

(doi: [10.1214/13-EJS843SUPP](https://doi.org/10.1214/13-EJS843SUPP); .pdf). In this supplement, the performance of Algorithms 10 and 11 is tested through an extended simulation study.

## References

- ARMSTRONG, M. and DIAMOND, P. (1984). Testing variogram for positive-definiteness. *Mathematical geology* 16(4), 407–421. [MR0774542](#)
- ARNOLD, L. (2003). *Random Dynamical Systems* (Second ed.). Springer. [MR1374107](#)
- BOSQ, D. (2000). *Linear Processes in Function Spaces*. Springer, New York. [MR1783138](#)
- CABALLERO, W., GIRALDO, R., and MATEU, J. (2013). A universal kriging approach for spatial functional data. *Stochastic Environmental Research and Risk Assessment* 27, 1553–1563.
- CHILÈS, J. P. and DELFINER, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York. [MR1679557](#)
- CRESSIE, N. (1993). *Statistics for Spatial data*. John Wiley & Sons, New York. [MR1239641](#)
- CURRIERO, F. (2006). On the use of non-euclidean distance measures in geostatistics. *Mathematical Geology* 38, 907–926. [MR2325912](#)
- DELICADO, P., GIRALDO, R., COMAS, C., and MATEU, J. (2010). Statistics for spatial functional data. *Environmetrics* 21(3–4), 224–239. [MR2842240](#)
- DUNFORD, N. and SCHWARTZ, J. (1958). *Linear Operators I*. Interscience, New York.
- EFRON, B. and TIBSHIRANI, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC. [MR1270903](#)
- FERRATY, F., KEILEGOM, I. V., and VIEU, F. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics* 37(2), 286–306. [MR2682301](#)
- FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York. [MR2229687](#)
- GIRALDO, R. (2009). *Geostatistical analysis of functional data*. Ph. D. thesis, Universitat Politècnica da Catalunya, Barcellona.
- GIRALDO, R., DELICADO, P., and MATEU, J. (2008a). Geostatistics for functional data: An ordinary kriging approach. Technical report. Universitat Politècnica de Catalunya, <http://hdl.handle.net/2117/1099>.
- GIRALDO, R., DELICADO, P., and MATEU, J. (2008b). *Point-wise kriging for spatial prediction of functional data*, Chapter 22, pp. 135–142. Functional and Operatorial Statistics. Proceedings of the First International Workshop on Functional and Operatorial Statistics. Springer, Toulouse, France. [MR2490341](#)
- GIRALDO, R., DELICADO, P., and MATEU, J. (2010a). Continuous time-varying kriging for spatial prediction of functional data: An environmental

- application. *Journal of Agricultural, Biological, and Environmental Statistics* 15(1), 66–82. [MR2755385](#)
- GIRALDO, R., DELICADO, P., and MATEU, J. (2010b). *Geofd: A package for prediction for functional data*. Ph. D. thesis, Universitat Politècnica de Catalunya.
- GIRALDO, R., DELICADO, P., and MATEU, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*. 18(3), 411–426. [MR2832903](#)
- GOULARD, M. and VOLTZ, M. (1993). Geostatistical interpolation of curves: A case study in soil science. In A. Soares (Ed.), *Geostatistics Tróia '92*, Volume 2, pp. 805–816. Kluwer Academic, Dordrecht.
- GROMENKO, O., KOKOSZKA, P., ZHU, L., and SOJKA, J. (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Annals of Applied Statistics* 6(2), 669–696. [MR2976487](#)
- HÖRMANN, S. and KOKOSZKA, P. (2011). Consistency of the mean and the principal components of spatially distributed functional data. In F. Ferraty (Ed.), *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, pp. 169–175. Physica-Verlag HD. [MR2867568](#)
- HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer. [MR2920735](#)
- HUANG, C., ZHANG, H., and ROBESON, S. M. (2011). On the validity of commonly used covariance and variogram functions on the sphere. *Mathematical Geosciences* 43(6), 721–733. [MR2824128](#)
- MAHALANOBIS, P. C. (1936). On the generalised distance in statistics. *Proceedings National Institute of Science, India* 2(1), 49–55.
- MENAFOGLIO, A., DALLA ROSA, M., and SECCHI, P. (2013). Supplement to “A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space”. DOI: [10.1214/00-EJS843SUPP](https://doi.org/10.1214/00-EJS843SUPP).
- MONESTIEZ, P. and NERINI, D. (2008). A cokriging method for spatial functional data with applications in oceanology. In S. Dabo-Niang and F. Ferraty (Eds.), *Functional and Operatorial Statistics*, Chapter 36, pp. 237–242. Springer. [MR2490355](#)
- Natural Resources of Canada website (2012).
- NERINI, D., MONESTIEZ, P., and MANTÉ, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis* 101(2), 409–418. [MR2564350](#)
- RAMSAY, J. and SILVERMAN, B. (2005). *Functional data analysis* (Second ed.). Springer, New York. [MR2168993](#)
- REYES, A., GIRALDO, R., and MATEU, J. (2012). Residual kriging for functional data. Application to the spatial prediction of salinity curves. Reporte interno de investigación no. 20, Universidad Nacional de Colombia.
- SHUMWAY, R. H. and DEAN, W. (1968). Best linear unbiased estimation for multivariate stationary processes. *Technometrics* 10(3), 523–534.
- STANLEY, D. (2002). *Canada's Maritime provinces* (First ed.). Lonely Planet Publications, Marybirnong.

- YAMANISHI, Y. and TANAKA, Y. (2003). Geographically weighted functional multiple regression analysis: A numerical investigation. *Journal of Japanese Society of Computational Statistics* (15), 307–317. [MR2027947](#)
- ZHU, K. (2007). *Operator Theory in Function Spaces* (Second ed.). American Mathematical Society. [MR2311536](#)