Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 204–229
https://doi.org/10.1214/20-BJPS466

Recent developments in complex and spatially correlated functional data

Israel Martínez-Hernández and Marc G. Genton
King Abdullah University of Science and Technology

Abstract. As high-dimensional and high-frequency data are being collected on a large scale, the development of new statistical models is being pushed forward. Functional data analysis provides the required statistical methods to deal with large-scale and complex data by assuming that data are continuous functions, for example, realizations of a continuous process (curves) or continuous random field (surfaces), and that each curve or surface is considered as a single observation. Here, we provide an overview of functional data analysis when data are complex and spatially correlated. We provide definitions and estimators of the first and second moments of the corresponding functional random variable. We present two main approaches: The first assumes that data are realizations of a functional random field, that is, each observation is a curve with a spatial component. We call them *spatial functional data*. The second approach assumes that data are continuous deterministic fields observed over time. In this case, one observation is a surface or manifold, and we call them *surface time series*. For these two approaches, we describe software available for the statistical analysis. We also present a data illustration, using a high-resolution wind speed simulated dataset, as an example of the two approaches. The functional data approach offers a new paradigm of data analysis, where the continuous processes or random fields are considered as a single entity. We consider this approach to be very valuable in the context of big data.

1 Introduction

The statistical analysis of large, complex, and high-dimensional data has become a significant challenging problem. Due to the rapid development of complex, performant technologies, data can now be collected on a large scale, resulting in high-dimensional and high-frequency data, sometimes necessitating high-performance computing, which is often a limitation for practitioners; see Galeano and Peña (2019) for a general view of data science and big data. Among various approaches, functional data analysis (FDA) provides statistical methods to handle large-scale and complex data (Chen et al., 2017, Giraldo, Dabo-Niang and Martínez, 2018). For a general introduction to FDA, the reader is referred to Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), and Kokoszka and Reimherr (2017). FDA assumes that observations (called functional data) have characteristics that vary along a continuum, for example, curves or surfaces. Thus, FDA deals with data that are defined on a space that is intrinsically infinite-dimensional.

The approach of FDA has several advantages as a realization of a continuous process or a continuous random field can be considered as functional data. In this case, the stationarity of the process (random field) is not needed since FDA treats the whole curve (surface) as a single entity. Thus, FDA is part of object data analysis (Menafoglio and Secchi, 2017). FDA is useful when the number of variables, p, is bigger than the sample size, n: $p \gg n$. In particular,

Key words and phrases. Functional data, functional random field, manifold data, spatial functional data, spatial statistics, spatio-temporal statistics, surface data.
Received January 2020; accepted January 2020.

FDA can analyze longitudinal data. Smoothness is an important property of functional data, in contrast with multivariate data analysis, where smoothness has no meaning. Thus, FDA extracts additional information contained in a continuous function or in its derivative. Although in practice, each continuous functions, say $y_i(v)$, is observed on a finite set of points, the continuity is obtained with smoothing techniques. In the process of smoothing, FDA does not require observed $y_i(v)$ on a regular grid, that is, a sample of $y_i(v)$ and $y_j(v)$ can be collected on a different set of points; $\mathbf{v}_i = \{v_{1_i}, \ldots, v_{n_i}\}$ and $\mathbf{v}_j = \{v_{1_j}, \ldots, v_{n_j}\}$, respectively. In general, the methods of FDA are essentially nonparametric and can model complex and spatially correlated data.

Functional data can also have a spatial component, because data are collected somewhere at some time (Haining, 2003, p. 15). If the functional data are curves with a spatial component, we call them *spatial functional data* (SFD). Thus, a dataset of SFD has the form $y(\mathbf{s}_1; v), \ldots, y(\mathbf{s}_n; v)$, where $\mathbf{s}_i \in D$ represents the locations in a given region D, and v represents the continuous parameter of the functional data. For example, $y(\mathbf{s}_i; v)$ can be the daily wind speed observed at location $\mathbf{s}_i$, with v indicating the time within a day, or $y(\mathbf{s}_i; v)$ can be the spectrum of brain activity at location $\mathbf{s}_i$, with v representing the frequency. If the continuous parameter v represents time, as for the example of wind speed, then SFD can be related to spatio-temporal data where the temporal dependence is captured through the continuity of the curve.

The combination of FDA and spatial statistics provides a powerful tool to deal with complex and large spatial data. This combination is attracting interest, and much research is focused on this topic. Nerini, Monestiez and Manté (2010) proposed a spatial functional linear model, and they analyzed data in Oceanography. Zhou et al. (2010) proposed mixed effects models for spatially correlated hierarchical functional data. Ruiz-Medina (2011) extended the spatial autoregressive processes and the spatial moving average processes to the Hilbert space. Giraldo, Delicado and Mateu (2012) proposed a methodology for clustering spatially correlated functional data; see also, Jiang and Serban (2012) and Romano, Balzanella and Verde (2017). Staicu, Crainiceanu and Carroll (2010) proposed a methodology for functional models with a hierarchical structure where the functions at the lowest hierarchy level are spatially correlated. Delicado et al. (2010), Ruiz-Medina (2012), and Mateu and Romano (2017) provided surveys of SFD. Menafoglio and Secchi (2017) presented a review of complex and spatially dependent data, such as curves and surfaces. Some references to SFD with a Bayesian perspective are Baladandayuthapani et al. (2008), Zhang et al. (2016), Song and Mallick (2019), and Rekabdarkolaee et al. (2019).

Functional data can also have a complex domain, for example, a two-dimensional Euclidean domain. Spatial data can be considered as functional data with the same domain as the corresponding random field, for instance if observations are dense over the region or if the domain is not a subset of the Euclidean space; see, for example, Alfeld, Neamtu and Schumaker (1996), Wahba (1981), and Gneiting (2013) for spatial data over complex domains. With the FDA approach, a realization of a random field is considered as a single point observation of the functional data. Similarly to the case when data are curves, the continuous surface (or the manifold) needs to be estimated. For that estimation, one can use a tensor product of univariate B-splines (Eilers and Marx, 1996, Wood, 2006, Qingguo and Longsheng, 2010, Xiao, Li and Ruppert, 2013). Another way is to approximate the continuous data using finite elements analysis; see Ramsay (2002), Duchamp and Stuetzle (2003), and Sangalli, Ramsay and Ramsay (2013). These types of functional data can also be observed over time; they are then called a functional time series. Functional time series can be related to spatio-temporal data when observations are dense over the domain. Thus, the approach of functional time series can be used in complex and large spatio-temporal data.

In this paper, our goal is to provide a review of complex and spatially correlated functional data, with two approaches using either spatial functional data or surface (manifold) data. In

both approaches, we focus on covariance functions and modeling. The remainder of our paper is organized as follows: In Section 2, we present basic concepts of functional random fields that include the mean, the covariance, and other important concepts related to the covariance. In this section, we also present different estimators of the various objects defined. In Section 3, we describe how to model small-scale and large-scale variations of spatial functional data, as well as corresponding methods to estimate the parameters involved. Section 4 (first approach) presents a brief description of the concept of functional kriging. In Section 5, we describe the second approach based on surface data, which can be considered as an alternative to the analysis of spatio-temporal data, and describe how to model and estimate the continuous surface. In Section 6, we present some software available for the analysis of spatial functional data and surface (manifold) data. In Section 7, we provide an example of the two approaches using a high-resolution wind speed simulated dataset in Dumat Al Jandal, Saudi Arabia. Section 8 ends the paper with some discussions.

2 Functional random fields

2.1 Basic concepts

In this section, we introduce the basic concepts of SFD. Let (Ω, F, P) be a probability space. Without loss of generality, we assume that the domain $\mathcal{T}$ of the curves is $\mathcal{T} = [0, 1]$, and let $\mathcal{H} = L^2([0, 1])$ be the Hilbert space of square integrable functions defined on $[0, 1]$, equipped with the inner product $\langle f, g\rangle = \int_0^1 f(v)g(v)\,\mathrm{d}v$. We denote by $\|\cdot\|_{\mathcal{H}}$ the norm in $\mathcal{H}$ induced by the inner product. A random variable $X : \Omega \to \mathcal{H}$ taking values in the Hilbert space $\mathcal{H}$ is called a *functional random variable* (Ramsay and Silverman, 2005, Ferraty and Vieu, 2006, Horváth and Kokoszka, 2012). Let $D \subset \mathbb{R}^2$ be a fixed study area (e.g., a country). A random field $\{X(\mathbf{s}) : \mathbf{s} \in D\}$ taking values in $\mathcal{H}$ is called a *functional random field*, that is, for each location $\mathbf{s} \in D$, $X(\mathbf{s}) : \Omega \to \mathcal{H}$ is a functional random variable.

We denote by $X(\mathbf{s}_0; v)$ the functional random variable at a fixed location $\mathbf{s}_0$ and $v \in [0, 1]$, and we denote by $X(\mathbf{s}_0; v_0)$ the scalar random variable obtained by evaluating $X(\mathbf{s}; v)$ at $\mathbf{s} = \mathbf{s}_0$ and $v = v_0$. Lastly, we use f to denote a function in $\mathcal{H}$.

Let $X(\mathbf{s}; v)$ be a functional random field with $\mathbb{E}(\|X(\mathbf{s}; v)\|_{\mathcal{H}}) < \infty$, for all $\mathbf{s} \in D$. The mean $\mu(\mathbf{s}; v) := \mathbb{E}\{X(\mathbf{s}; v)\}$ of $X(\mathbf{s}; v)$ is defined as an element of $\mathcal{H}$ such that

$$\langle \mu(\mathbf{s}; \cdot), f\rangle = \mathbb{E}\langle X(\mathbf{s}; \cdot), f\rangle, \quad \forall f \in \mathcal{H},$$

where the dot in $(\mathbf{s}; \cdot)$ indicates the integrated variable in $[0, 1]$. This implies that $\mathbb{E}\{X(\mathbf{s}; v_0)\} = \mu(\mathbf{s}; v_0)$ for almost all $v_0 \in [0, 1]$.

The covariance is one of the most studied objects in spatial statistics, due to its relevance for prediction. If $\mathbb{E}\{\|X(\mathbf{s}; v)\|_{\mathcal{H}}^2\} < \infty$, then the covariance operator at locations $\mathbf{s}_1$ and $\mathbf{s}_2$ is defined as an operator $C(\mathbf{s}_1, \mathbf{s}_2; \bullet) : \mathcal{H} \to \mathcal{H}$ such that

$$\begin{aligned} C(\mathbf{s}_1, \mathbf{s}_2; f)(\cdot) &= \mathbb{E}\big[\langle X(\mathbf{s}_1; \cdot) - \mu(\mathbf{s}_1; \cdot), f\rangle\{X(\mathbf{s}_2; \cdot) - \mu(\mathbf{s}_2; \cdot)\}\big] \\ &= \int_0^1 \sigma(\mathbf{s}_1, \mathbf{s}_2; u, \cdot) f(u)\,\mathrm{d}u, \quad f \in \mathcal{H}, \end{aligned} \tag{2.1}$$

where $\sigma(\mathbf{s}_1, \mathbf{s}_2; u_0, v_0) := \mathbb{E}[\{X(\mathbf{s}_1; u_0) - \mu(\mathbf{s}_1; u_0)\}\{X(\mathbf{s}_2; v_0) - \mu(\mathbf{s}_2; v_0)\}]$ is the point-wise covariance, called the *kernel* of $C(\mathbf{s}_1, \mathbf{s}_2; \bullet)$. This definition can be written in terms of a tensor operation as follows: $C(\mathbf{s}_1, \mathbf{s}_2; f) = \mathbb{E}[\{X(\mathbf{s}_1; \cdot) - \mu(\mathbf{s}_1; \cdot)\} \otimes \{X(\mathbf{s}_2; \cdot) - \mu(\mathbf{s}_2; \cdot)\}(f)]$.

A common assumption, in practice, is the stationarity condition of a process, which is defined as follows.

Definition 1 (Weak stationarity). A functional random field $X(\mathbf{s}; v)$ is said to be (weakly) stationary if

1. $\mathbb{E}(\|X(\mathbf{s}; v)\|_{\mathcal{H}}^2) < \infty$,
2. $\mu(\mathbf{s}; v) = \mu(v)$, that is, the mean does not depend on the location $\mathbf{s}$, and
3. $C(\mathbf{s}_1 + \mathbf{h}, \mathbf{s}_2 + \mathbf{h}; \cdot) = C(\mathbf{s}_1, \mathbf{s}_2; \cdot)$, for all $\mathbf{s}_1, \mathbf{s}_2, \mathbf{h} \in D$.

The last condition is equivalent to the property that the covariance operator depends only on the increments $\mathbf{s}_1 - \mathbf{s}_2$. This means, there exists a covariance operator $\tilde{C}(\mathbf{s}; \cdot) : \mathcal{H} \to \mathcal{H}$ such that

$$C(\mathbf{s}_1, \mathbf{s}_2; f) = \tilde{C}(\mathbf{s}_1 - \mathbf{s}_2; f), \quad f \in \mathcal{H},$$

and so the variance operator can be written as $\mathrm{Var}\{X(\mathbf{s}; v)\} = C(\mathbf{s}, \mathbf{s}; \cdot) = \tilde{C}(\mathbf{0}; \cdot)$. Thus, for convenience, we write $C(\mathbf{0}; \cdot)$ to denote the variance operator $C(\mathbf{s}, \mathbf{s}; f)$ of the stationary functional random field $X(\mathbf{s}; v)$.

Now, we define the concept of isotropy for SFD.

Definition 2 (Isotropy). A stationary functional random field $X(\mathbf{s}; v)$ is said to be isotropic if there exists a covariance operator $\tilde{C}_0(h; \cdot) : \mathcal{H} \to \mathcal{H}$ such that

$$C(\mathbf{s}_1, \mathbf{s}_2; f) = \tilde{C}_0(h; f), \quad f \in \mathcal{H},$$

where $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$, for all $\mathbf{s}_1, \mathbf{s}_2 \in D$.

In spatial statistics, the variogram plays an important role to make inference. The extension of this concept to functional random fields is as follows: The variogram operator Γ is defined as the variance operator of the difference between the functional random field at two locations $\mathbf{s}_1$, $\mathbf{s}_2$, that is, $\Gamma(\mathbf{s}_1, \mathbf{s}_2; \cdot) := \frac{1}{2}\mathrm{Var}\{X(\mathbf{s}_1; v) - X(\mathbf{s}_2; v)\}$. If the functional random field $X(\mathbf{s}; v)$ has a finite second moment, then we have $2\Gamma(\mathbf{s}_1, \mathbf{s}_2; f) = C(\mathbf{s}_1, \mathbf{s}_1; f) + C(\mathbf{s}_2, \mathbf{s}_2; f) - C(\mathbf{s}_1, \mathbf{s}_2; f) - C(\mathbf{s}_2, \mathbf{s}_1; f)$, for $f \in \mathcal{H}$. Thus, if the functional random field is stationary, then there exists a variogram operator $\tilde{\Gamma}(\mathbf{s}; \cdot) : \mathcal{H} \to \mathcal{H}$ such that it satisfies

$$\Gamma(\mathbf{s}_1, \mathbf{s}_2; f) = \tilde{\Gamma}(\mathbf{s}_1 - \mathbf{s}_2; f) = C(\mathbf{0}; f) - C(\mathbf{s}_1 - \mathbf{s}_2; f), \quad f \in \mathcal{H}, \tag{2.2}$$

where the corresponding kernel is $\gamma(\mathbf{s}_1 - \mathbf{s}_2; u, v) = \sigma(\mathbf{0}; u, v) - \sigma(\mathbf{s}_1 - \mathbf{s}_2; u, v)$. Furthermore, if $X(\mathbf{s}; v)$ is isotropic, then there exists $\tilde{\Gamma}_0(h; \cdot) : \mathcal{H} \to \mathcal{H}$ such that

$$\Gamma(\mathbf{s}_1, \mathbf{s}_2; f) = \tilde{\Gamma}_0(h; f), \quad f \in \mathcal{H},$$

where $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$, for all $\mathbf{s}_1, \mathbf{s}_2 \in D$.

Remark 1. Unlike the finite dimensional case (multivariate random field) where the covariance is composed of matrices, here, the covariance is composed of operators, since the space of functional data is intrinsically infinite-dimensional.

Another way to describe the second-order spatial dependence of the functional random field is by using a "global" measure. This global measure is the trace-covariogram $\sigma_{\mathrm{tr}} : D \times D \to \mathbb{R}$ (Giraldo, Delicado and Mateu, 2011, Menafoglio, Secchi and Dalla Rosa, 2013) defined as

$$\sigma_{\mathrm{tr}}(\mathbf{s}_1, \mathbf{s}_2) = \mathbb{E}\big\{\big\langle X(\mathbf{s}_1; \cdot) - \mu(\mathbf{s}_1; \cdot), X(\mathbf{s}_2; \cdot) - \mu(\mathbf{s}_2; \cdot)\big\rangle\big\} = \int_0^1 \sigma(\mathbf{s}_1, \mathbf{s}_2; v, v)\,\mathrm{d}v. \tag{2.3}$$

The trace-covariogram computes the covariance of the inner product of the functional random field at two locations. Thus, it summarizes the covariance on the diagonal, and so, in general,

it depends only on the locations. If the functional random field is stationary, then there exists $\tilde{\sigma}_{\text{tr}} : D \to \mathbb{R}$ such that it depends only on the separation vector $\mathbf{s}_1 - \mathbf{s}_2$, that is,

$$\sigma_{\text{tr}}(\mathbf{s}_1, \mathbf{s}_2) = \tilde{\sigma}_{\text{tr}}(\mathbf{s}_1 - \mathbf{s}_2).$$

In addition, if the functional random field is isotropic, then there exists $\tilde{\sigma}_{\text{tr},0} : \mathbb{R} \to \mathbb{R}$ such that

$$\sigma_{\text{tr}}(\mathbf{s}_1, \mathbf{s}_2) = \tilde{\sigma}_{\text{tr},0}(h),$$

where $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$.

Similarly, the trace-variogram is defined in terms of the inner product of the difference, that is,

$$\gamma_{\text{tr}}(\mathbf{s}_1 - \mathbf{s}_2) = \frac{1}{2}\mathbb{E}\{\langle X(\mathbf{s}_1; \cdot) - X(\mathbf{s}_2; \cdot), X(\mathbf{s}_1; \cdot) - X(\mathbf{s}_2; \cdot)\rangle\} - \frac{1}{2}\|\mu(\mathbf{s}_1; \cdot) - \mu(\mathbf{s}_2; \cdot)\|^2_{\mathcal{H}}.$$

We observe that, if $X(\mathbf{s}; v)$ is stationary, then $\sigma(\mathbf{s}; u, v) = \sigma(\mathbf{0}; u, v) - \gamma(\mathbf{s}; u, v)$. Thus, the trace-variogram satisfies

$$\gamma_{\text{tr}}(\mathbf{s}_1 - \mathbf{s}_2) = \sigma_{\text{tr}}(\mathbf{0}) - \sigma_{\text{tr}}(\mathbf{s}_1 - \mathbf{s}_2).$$

Trace-covariogram and trace-variogram are also important for optimization problems. Especially if we want to use the criterion of minimizing equations of the form $\mathbb{E}(\langle X, Y\rangle)$, as in (2.7) below.

2.2 Estimation

Now, we describe estimators of the mean μ, the covariance C, and the variogram Γ. For this purpose, we assume that $X(\mathbf{s}; v)$ is an isotropic (stationary) functional random field with mean $\mu(v)$ and covariance operator C. Let $x(\mathbf{s}_1; v), \ldots, x(\mathbf{s}_n; v)$ be observations of the functional random field $X(\mathbf{s}; v)$. We assume that the observations $x(\mathbf{s}_i; v)$ are given in the functional form. Although in real data, $x(\mathbf{s}_i; v)$ are observed on a finite set of points $v_{i1}, \ldots, v_{im}$, the continuous curves should be estimated (see Ramsay and Silverman, 2005).

The main feature of spatial data is that "nearby" data look similar, and an estimator must take into account such spatial dependence. Otherwise, it will not have desirable properties, such as consistency.

2.2.1 *Mean estimation.* We describe two different approaches to obtain an estimator of the mean $\mu(v)$ (Gromenko et al., 2012). A model of the mean can be written as

$$X(\mathbf{s}; v) = \mu(v) + \varepsilon(\mathbf{s}; v), \tag{2.4}$$

where $\varepsilon(\mathbf{s}; v)$ is an isotropic functional random field with zero mean and covariance operator C.

The first approach is similar to the kriging method. Specifically, this is defined as a weighting of the observed curves:

$$\hat{\mu}(v) = \sum_{i=1}^{n} w_i x(\mathbf{s}_i; v), \tag{2.5}$$

where the weights w_i are estimated by solving the optimization problem

$$\min_{w_1,\ldots,w_n} \mathbb{E}\{\langle \hat{\mu} - \mu, \hat{\mu} - \mu\rangle\} = \min_{w_1,\ldots,w_n} \mathbb{E}\left\{\left\|\sum_{i=1}^{n} w_i x(\mathbf{s}_i; \cdot) - \mu\right\|^2_{\mathcal{H}}\right\}, \tag{2.6}$$

subject to the condition $\sum_{i=1}^{n} w_i = 1$. Using the Lagrange multiplier method, this leads to solve

$$\sum_{i=1}^{n} w_i = 1, \qquad \sum_{i=1}^{n} w_i \, \sigma_{\mathrm{tr},\varepsilon}(\mathbf{s}_i, \mathbf{s}_j) - \lambda = 0, \quad j = 1, \ldots, n, \tag{2.7}$$

where $\sigma_{\mathrm{tr},\varepsilon}(\mathbf{s}_1, \mathbf{s}_2)$ is the trace-covariogram of $\varepsilon(\mathbf{s}; v)$. Thus, the estimation problem (2.6) becomes estimating the matrix $\{\sigma_{\mathrm{tr},\varepsilon}(\mathbf{s}_i, \mathbf{s}_j)\}_{i,j=1}^{n}$. Since the functional random field $\varepsilon(\mathbf{s}; v)$ is unobserved, a common approach is to use an iterative procedure. At the first iteration, an initial estimator of $\mu(v)$ is obtained by assuming that $\varepsilon(\mathbf{s}; v)$ is spatially uncorrelated, i.e., $\hat{\mu}_0(v) = \frac{1}{n}\sum_{i=1}^{n} x(\mathbf{s}_i; v)$. Next, $\hat{\mu}_0(v)$ is subtracted from the data $x(\mathbf{s}_i; v)$, then an initial estimator of $\sigma_{\mathrm{tr},\varepsilon}(\mathbf{s}_i, \mathbf{s}_j)$ is obtained as described below in (2.11). At the second iteration, the mean is re-estimated by solving (2.7), with the initial information of $\sigma_{\mathrm{tr},\varepsilon}(\mathbf{s}_i, \mathbf{s}_j)$. This process is repeated until convergence.

The second approach uses finite basis functions similarly to the cokriging method on the coefficients of the basis functions, see Goulard and Voltz (1993), Nerini, Monestiez and Manté (2010), and Giraldo, Delicado and Mateu (2011). Let $\{\eta_1(v), \ldots, \eta_K(v)\}$ be basis functions, for example, Fourier basis functions or B-spline basis functions. Then, the observed curves are approximated as

$$x(\mathbf{s}_i; v) \approx \sum_{k=1}^{K} z_k(\mathbf{s}_i)\eta_k(v), \quad i = 1, \ldots, n, \tag{2.8}$$

where $z_k(\mathbf{s}_i) = \langle x(\mathbf{s}_i; \cdot), \eta_k \rangle$, which is a scalar for each i and k. Let $Z_k(\mathbf{s})$ be the corresponding scalar random field with realization $\{z_k(\mathbf{s}_i)\}_{i=1}^{n}$. Then, by using (2.4), the mean can be approximated as

$$\mu(v) \approx \sum_{k=1}^{K} \mathbb{E}\{Z_k(\mathbf{s})\}\eta_k(v),$$

and $\mathbb{E}\{Z_k(\mathbf{s})\}$ should be estimated. For this, from (2.4) we have that

$$Z_k(\mathbf{s}_i) = \mu_k^z + \varepsilon^k(\mathbf{s}_i), \quad i = 1, \ldots, n,$$

where $\mu_k^z = \langle \mu, \eta_k \rangle = \mathbb{E}\{Z_k(\mathbf{s})\}$, and $\varepsilon^k(\mathbf{s}_i) = \langle \varepsilon(\mathbf{s}_i; \cdot), \eta_k \rangle$, for each $k = 1, \ldots, K$. Thus, $Z_k(\mathbf{s}_i)$ is an isotropic (stationary) scalar random field, for $k = 1, \ldots, K$. Once μ_k^z is estimated, we obtain the estimator of the mean

$$\hat{\mu}(v) = \sum_{k=1}^{K} \hat{\mu}_k^z \eta_k(v).$$

To estimate μ_k^z, we can use the ordinary least squares (OLS) estimator, by minimizing

$$\sum_{i=1}^{n} \{z_k(\mathbf{s}_i) - w_i \mu_k^z\}^2,$$

with $w_i = 1$ for all $i = 1, \ldots, n$, or a weighted least squares (WLS) estimator, that is, w_i are estimated with the information of the covariance matrix of $\varepsilon^k(\mathbf{s})$. Also, we can use a generalized least squares (GLS) estimator, which can perform better than OLS or WLS; see Schabenberger and Gotway (2005) and Cressie (2015) for the development of these estimators.

Another way to estimate μ_k^z is by using Maximum Likelihood Estimators (MLEs). For this, assume that, for each i, $X(\mathbf{s}_i; v)$ is a Gaussian process in $[0, 1]$. Then, for each k,

$\mathbf{z}_k = \{z_k(\mathbf{s}_1), \ldots, z_k(\mathbf{s}_n)\}^\top$ is a realization of $\mathbf{Z}_k \sim \mathcal{N}_n(\mu_k^z \mathbf{1}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the covariance of $\varepsilon^k(\mathbf{s})$ described by some valid parametric covariance model, for example, Matérn, exponential, or spherical. Then, the log likelihood is

$$l(\boldsymbol{\theta}, \mu_k^z) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\det\{\boldsymbol{\Sigma}(\boldsymbol{\theta})\} - \frac{1}{2}(\mathbf{z}_k - \mu_k^z\mathbf{1})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\mathbf{z}_k - \mu_k^z\mathbf{1}). \tag{2.9}$$

To read more details of this estimator, see Stein (1999).

The coefficients in estimator (2.5) can be also considered as operators $\omega_i : \mathcal{H} \to \mathcal{H}$. In this case the estimator of the mean takes the form $\hat{\mu}(v) = \sum_{i=1}^n \omega_i\{x(\mathbf{s}_i; \cdot)\}(v)$, where the coefficients, which are operators, need to be estimated under a certain constraint on $\sum_{i=1}^n \omega_i$, see Nerini, Monestiez and Manté (2010). This approach takes into account the information of the whole curve to define a specific weight for each point $v \in [0, 1]$. Thus, it is expected to obtain better results than estimator (2.5), but the estimation procedure can be complicated.

In the context of a nonstationary functional random field, the mean can depend on the location $\mathbf{s}$, in this case, $\mu(\mathbf{s}; v)$ can be represented by a linear model $\mu(\mathbf{s}; v) = \sum_{l=0}^L a_l(\mathbf{s}) f_l(v)$, where $f_l(v)$, $l = 0, \ldots, L$, are elements of $\mathcal{H}$ independent of the spatial location $\mathbf{s}$, and $a_0(\mathbf{s}) = 1$ (Menafoglio, Secchi and Dalla Rosa, 2013, Caballero, Giraldo and Mateu, 2013). The latter approach is described with more details in Section 3.1.

2.2.2 *Covariance estimation.* Now, we describe how to estimate the covariance. We assume that $\mathbb{E}\{X(\mathbf{s}; v)\} = 0$. Let $N(h)$ be a set of pairs of indexes defined as $N(h) = \{(i, j) : \|\mathbf{s}_i - \mathbf{s}_j\| = h\}$, and let $|N(h)|$ be the cardinality of $N(h)$. The empirical covariance operator of the functional random field is defined as

$$\hat{C}(h; f) = \frac{1}{|N(h)|} \sum_{(i,j)\in N(h)} x(\mathbf{s}_i) \otimes x(\mathbf{s}_j)(f).$$

In practice, the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$ is not considered to be exactly h, instead $\|\mathbf{s}_i - \mathbf{s}_j\| \in (h - \delta, h + \delta)$, with $\delta > 0$. Also, since it is almost impossible to obtain $\hat{C}(h; f)$ for all h, discretized values of h, $h_1, \ldots, h_m$, are computed.

The usual approach in the scalar (multivariate) random field case is to fit a valid parametric model to the empirical covariance by least squares methods (Cressie, 2015). However, in spatial functional data, for each h, $\hat{C}(h; \cdot)$ is an operator. Thus, it requires new mathematical developments to define a "covariance model" in this context. Finite basis functions have been used to overcome the modeling of the covariance function in $\mathcal{H}$ (Nerini, Monestiez and Manté, 2010). Recall that $C(h; \cdot)$ is represented by its kernel $\sigma(h; u, v)$ (see (2.1)). Now, we assume that $x(\mathbf{s}_i; v)$ is approximated by a finite set of basis functions, as in (2.8). Then, we obtain

$$\sigma(h; u, v) = \mathbb{E}\{X(\mathbf{s}+\mathbf{h})X(\mathbf{s})\} = \boldsymbol{\eta}^\top(u)\mathbb{E}\{\mathbf{Z}(\mathbf{s}+\mathbf{h})\mathbf{Z}^\top(\mathbf{s})\}\boldsymbol{\eta}(v),$$

where $\boldsymbol{\eta}(v) = \{\eta_1(v), \ldots, \eta_K(v)\}^\top$, $\mathbf{Z}(\mathbf{s}) = \{Z_1(\mathbf{s}), \ldots, Z_K(\mathbf{s})\}^\top$, and $Z_k(\mathbf{s}) = \langle X(\mathbf{s}; \cdot), \eta_k\rangle$ is the scalar random field with mean zero, and so $\mathbf{Z}(\mathbf{s})$ is a multivariate random field with covariance

$$\boldsymbol{\Sigma}(h) = \mathrm{Cov}\{\mathbf{Z}(\mathbf{s}+\mathbf{h}), \mathbf{Z}(\mathbf{s})\} = \{\boldsymbol{\Sigma}_{kl}(h)\}_{k,l=1}^K,$$

where $\boldsymbol{\Sigma}_{kl}(h) = \mathrm{Cov}\{Z_k(\mathbf{s}+\mathbf{h}), Z_l(\mathbf{s})\}$ for $k, l = 1, \ldots, K$, $h = \|\mathbf{h}\|$. Thus, estimating the covariance operator C is equivalent to estimating the covariance of a multivariate random field $\mathbf{Z}(\mathbf{s})$, which can be modeled with a valid covariance function $\boldsymbol{\Sigma}(h; \boldsymbol{\theta})$ (Genton and Kleiber, 2015).

If $\boldsymbol{\eta}(v)$ is a set of orthogonal basis functions, then we can estimate each marginal-covariance functions $\boldsymbol{\Sigma}_{kk}(h)$, for $k = 1, \ldots, K$, separately. In addition, if $X(\mathbf{s}; v)$ is assumed

to be a Gaussian process, we can use the log likelihood (2.9) with $\mu_k^z = 0$, that is, maximizing $l(\boldsymbol{\theta}, 0)$ as function of $\boldsymbol{\theta}$. Also, the restricted maximum likelihood (REML) estimation can be used in this case, see Stein (1999).

Once the estimator $\hat{\boldsymbol{\Sigma}}(h; \boldsymbol{\theta})$ is obtained, then we define an estimator of $\sigma(h; u, v)$ as follows,

$$\hat{\sigma}(h; u, v) = \boldsymbol{\eta}^{\mathsf{T}}(u)\hat{\boldsymbol{\Sigma}}(h; \boldsymbol{\theta})\boldsymbol{\eta}(v) \tag{2.10}$$

and then, $\hat{C}(h; f)(\cdot) = \int_0^1 \hat{\sigma}(h; u, \cdot) f(u)\, \mathrm{d}u$.

In the estimation of the covariance of Z_k, we can also use a nonparametric approach (Hall, Fisher and Hoffmann, 1994, Hall and Patil, 1994). Bayesian approaches can be found in, for example, Banerjee, Carlin and Gelfand (2015) and Diggle and Ribeiro (2007).

As mentioned before, the trace-covariogram is a measure that describes dependence globally, in the sense that it integrates the kernel $\sigma(\mathbf{s}; u, v)$ on the diagonal. Also, the trace-covariogram appears in several optimization problems (Delicado et al., 2010, Giraldo, Delicado and Mateu, 2011, Menafoglio, Secchi and Dalla Rosa, 2013), that make its estimation important. The empirical trace-covariogram is defined as

$$\hat{\sigma}_{\mathrm{tr}}(h) = \frac{1}{|N(h)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(h)} \int_0^1 x(\mathbf{s}_i; v) x(\mathbf{s}_j; v)\, \mathrm{d}v. \tag{2.11}$$

This empirical estimator is computed on discrete values $h_1, \ldots, h_m$ of h. Then, we can fit any covariance model by least squares methods to the empirical estimates $\hat{\sigma}_{\mathrm{tr}}(h_1), \ldots, \hat{\sigma}_{\mathrm{tr}}(h_m)$.

We observe that estimating the trace-covariogram is much easier than estimating the covariance C, where the inner products $\langle x(\mathbf{s}_i; \cdot), x(\mathbf{s}_j; \cdot)\rangle = \int_0^1 x(\mathbf{s}_i; v) x(\mathbf{s}_j; v)\, \mathrm{d}v$ can be computed using the `R` (R Core Team, 2019) package *fda* (Ramsay et al., 2018).

2.2.3 *Variogram estimation*. One of the advantages of using the variogram is the robustness under misspecification of the mean. Also, the variogram can be used to estimate the covariance. The empirical variogram operator of the functional random field is defined as

$$\hat{\Gamma}(h; f) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} \{x(\mathbf{s}_i) - x(\mathbf{s}_j)\} \otimes \{x(\mathbf{s}_i) - x(\mathbf{s}_j)\}(f), \quad f \in \mathcal{H}.$$

Similarly to the estimator of the covariance operator, if a basis function $\boldsymbol{\eta}(v)$ is assumed, then the corresponding kernel γ is obtained as

$$\begin{aligned} 2\gamma(h; u, v) &= \mathbb{E}\big[\{X(\mathbf{s}+\mathbf{h}) - X(\mathbf{s})\}^2\big] \\ &= \boldsymbol{\eta}^{\mathsf{T}}(u)\mathbb{E}\big[\{\mathbf{Z}(\mathbf{s}+\mathbf{h}) - \mathbf{Z}(\mathbf{s})\}\{\mathbf{Z}(\mathbf{s}+\mathbf{h}) - \mathbf{Z}(\mathbf{s})\}^{\mathsf{T}}\big]\boldsymbol{\eta}(v), \end{aligned}$$

where $h = \|\mathbf{h}\|$. Thus, we need to estimate the variogram of the multivariate random field $\mathbf{Z}(\mathbf{s}) = \{Z_1(\mathbf{s}), \ldots, Z_K(\mathbf{s})\}^{\mathsf{T}}$, to obtain the estimator of the kernel $\gamma(h; u, v)$.

The corresponding empirical trace-variogram is defined as

$$\hat{\gamma}_{\mathrm{tr}}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} \int_0^1 \{x(\mathbf{s}_i; v) - x(\mathbf{s}_j; v)\}^2\, \mathrm{d}v. \tag{2.12}$$

Then, following the common method used in spatial statistics, a variogram model is fitted to $\hat{\gamma}_{\mathrm{tr}}(h_1), \ldots, \hat{\gamma}_{\mathrm{tr}}(h_m)$.

The principal component analysis is generally important in statistics because of its applicability to dimensional reduction techniques. In the context of functional data, a function $\zeta \in \mathcal{H}$ is called *eigenfunction* of the operator $C(\mathbf{s}, \mathbf{s}; \cdot)$, if $C(\mathbf{s}, \mathbf{s}; \zeta) = \lambda\zeta$ with λ a positive real number. The estimation of eigenfunctions can be obtained from the estimated covariance

operator $\hat{C}(\mathbf{s}, \mathbf{s}; \cdot)$. If we are only interested in the eigenfunctions, we can obtain the estimators without estimating the covariance operator, which allows us to reduce computational costs (Zhou et al., 2010, Gromenko et al., 2012). Also see Liu, Ray and Hooker (2017) for functional principal component analysis of spatial functional data.

3 Modeling functional random fields

Spatial statistics studies the variations among the observed data at different locations. The spatial variation is generally described through the mean and the covariance (Haining, 2003). The mean represents the large-scale variations, and the covariance represents the small-scale variations.

In this section, we describe the statistical models for spatial functional data. We denote the observed functional data as $y(\mathbf{s}_1; v), \ldots, y(\mathbf{s}_n; v)$, and we denote by $Y(\mathbf{s}; v)$ the corresponding functional random field with realization $y(\mathbf{s}_i; v)$, $i = 1, \ldots, n$. The functional random field $X(\mathbf{s}; v)$ denotes a stationary functional random field with covariance function C_X, and $\varepsilon(\mathbf{s}; v)$ denotes a functional white noise, that is, $\mathbb{E}\{\varepsilon(\mathbf{s}, v)\} = 0$, with a covariance such that $C_\varepsilon(\mathbf{s}_1, \mathbf{s}_2; \cdot) = 0$, if $\mathbf{s}_1 \neq \mathbf{s}_2$.

3.1 Large-scale variation

Regression models with covariates in terms of the locations can be used to model large-scale spatial variations. Such regression models must account for spatial dependence. In general, estimators obtained from regression models are smooth functions defined in D. In the context of spatial functional data, functional regression models in $\mathcal{H}$ extend models of finite-dimensional data to model large-scale variations. In Caballero, Giraldo and Mateu (2013), Menafoglio, Secchi and Dalla Rosa (2013), and Reyes, Giraldo and Mateu (2015), covariates are assumed to be separable in the spatial component and the continuity of the data. In this case, the model for large-scale variations is

$$Y(\mathbf{s}; v) = \sum_{l=0}^{L} a_l(\mathbf{s}) f_l(v) + \varepsilon(\mathbf{s}; v), \tag{3.1}$$

where $f_l \in \mathcal{H}$ are independent of $\mathbf{s}$, $a_0(\mathbf{s}) := 1$, $\{a_l(\mathbf{s})\}_{l=1}^{L}$ are known scalar regressors, and $\varepsilon(\mathbf{s}; v)$ is the functional white noise. For example, with $L = 5$, we could specify the scalar regressors $a_l(\mathbf{s})$ as

$$a_1(\mathbf{s}) = s_1, \qquad a_2(\mathbf{s}) = s_2, \qquad a_3(\mathbf{s}) = s_1 s_2, \qquad a_4(\mathbf{s}) = s_1^2, \quad \text{and} \quad a_5(\mathbf{s}) = s_2^2,$$

where $\mathbf{s} = (s_1, s_2)$ denotes the coordinates of a spatial location. In this example, $f_0(v), \ldots, f_5(v)$ need to be estimated. Thus, model (3.1) is a functional regression model with scalar covariates $a_l(\mathbf{s})$, $l = 0, 1, \ldots, L$.

The mean of the functional random field $Y(\mathbf{s}; v)$ in (3.1) is

$$\mathbb{E}\big\{Y(\mathbf{s}; v)\big\} = \mu(\mathbf{s}; v) = \sum_{l=0}^{L} a_l(\mathbf{s}) f_l(v),$$

and the covariance C_Y of $Y(\mathbf{s}; v)$ is $C_Y = C_\varepsilon$, which is zero at $(\mathbf{s}_i, \mathbf{s}_j)$ if $i \neq j$. Then, the spatial variation is described through the mean of the functional random field $Y(\mathbf{s}; v)$. The covariates $a_l(\mathbf{s})$ capture the spatial dependence, and the set of functions $\{f_l(v)\}_{l=0}^{L}$ carries the continuity of the functional data.

The estimators of $f_l(v)$ in model (3.1) can be obtained using the OLS method. The matrix form of the model can then be written as

$$\mathbf{y}(v) = \mathbf{A}\mathbf{f}(v) + \boldsymbol{\varepsilon}(v),$$

where $\mathbf{y}(v) = \{y(\mathbf{s}_1; v), \ldots, y(\mathbf{s}_n; v)\}^\mathsf{T}$, $\mathbf{A} = \{a_l(\mathbf{s}_i)\}_{i,l}$, $i = 1, \ldots, n$, $l = 0, 1, \ldots, L$, is the design matrix, $\mathbf{f}(v) = \{f_0(v), \ldots, f_L(v)\}^\mathsf{T}$, and $\boldsymbol{\varepsilon}(v) = \{\varepsilon(\mathbf{s}_1; v), \ldots, \varepsilon(\mathbf{s}_n; v)\}^\mathsf{T}$. Then, the OLS estimator is obtained by solving the optimization problem

$$\min_{f_0, \ldots, f_L \in \mathcal{H}} \sum_{i=1}^{n} \left\| y(\mathbf{s}_i; \cdot) - \sum_{l=0}^{L} a_l(\mathbf{s}_i) f_l \right\|_{\mathcal{H}}^2. \tag{3.2}$$

Under some conditions (Menafoglio, Secchi and Dalla Rosa, 2013), (3.2) admits a unique solution

$$\hat{\mathbf{f}}(v) = (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}\mathbf{y}(v).$$

Thus, the drift estimator is obtained as

$$\hat{\boldsymbol{\mu}}(v) = \mathbf{A}(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}\mathbf{y}(v). \tag{3.3}$$

Since the estimator $\hat{\boldsymbol{\mu}}(v)$ in (3.3) is a linear combination of the observed curves, $\hat{\boldsymbol{\mu}}(v)$ inherits the continuity property of $\mathbf{y}(v)$.

The basis functions approach offers another alternative to obtain estimators of $\{f_l(v)\}$ (Reyes, Giraldo and Mateu, 2015). Each component of the model (3.1) can be assumed to be in the space generated by finite basis functions, i.e.,

$$y(\mathbf{s}_i; v) = \sum_{k=1}^{K} z_{ik}\eta_k(v), \qquad f_l(v) = \sum_{k=1}^{K} b_{lk}\eta_k(v), \quad \text{and} \quad \varepsilon(\mathbf{s}_i; v) = \sum_{k=1}^{K} e_{ik}\eta_k(v),$$

for $i = 1, \ldots, n$, and $l = 0, 1, \ldots, L$, where $\boldsymbol{\eta}(v) = \{\eta_1(v), \ldots, \eta_K(v)\}^\mathsf{T}$ is the basis function. In this case, the matrix form of the model (3.1) is

$$\mathbf{Z}\boldsymbol{\eta}(v) = \mathbf{AB}\boldsymbol{\eta}(v) + \mathbf{E}\boldsymbol{\eta}(v),$$

where $\mathbf{Z} = \{z_{ik}\}$, $\mathbf{B} = \{b_{lk}\}$, and $\mathbf{E} = \{e_{ik}\}$, $i = 1, \ldots, n$, $l = 0, 1, \ldots, L$, and $k = 1, \ldots, K$. The corresponding normal equation is

$$\mathbf{A}^\mathsf{T}\mathbf{Z}\mathbf{J}_{\boldsymbol{\eta}} = \mathbf{A}^\mathsf{T}\mathbf{AB}\mathbf{J}_{\boldsymbol{\eta}},$$

where $\mathbf{J}_{\boldsymbol{\eta}} = \int \boldsymbol{\eta}(v)\boldsymbol{\eta}^\mathsf{T}(v)\,\mathrm{d}v$. The solution for $\mathbf{B}$ is found by vectorizing the normal equation, that is

$$\mathrm{vec}(\hat{\mathbf{B}}) = (\mathbf{J}_{\boldsymbol{\eta}}^\mathsf{T} \otimes \mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathrm{vec}(\mathbf{A}^\mathsf{T}\mathbf{Z}\mathbf{J}_{\boldsymbol{\eta}}). \tag{3.4}$$

Consequently, $\hat{\boldsymbol{\mu}}(v) = \mathbf{A}\hat{\mathbf{B}}\boldsymbol{\eta}(v)$.

In the context of explicative modeling, we can use other spatial functional data as covariates to describe the mean $\mu(\mathbf{s}; v)$ (Ignaccolo, Mateu and Giraldo, 2014). For example, the mean can be modeled as $\mu(\mathbf{s}; v) = \beta_0(v) + \sum_{p=1}^{P} \beta_p(v)U_p(\mathbf{s}; v)$, where $\{U_p(\mathbf{s}; v)\}_{p=1}^{P}$ are the functional covariates, and $\beta_0(v), \beta_1(v), \ldots, \beta_P(v)$ are the functional parameters to be estimated.

3.2 Small-scale variation

Small-scale variations are usually represented through the covariance structure. The modeling of covariance is one of the most studied subjects in spatial statistics (Stein, 1999, Cressie, 2015, Genton and Kleiber, 2015). For finite-dimensional data, the exponential, the Gaussian, and the Matérn are examples of parametric covariance function models.

In this section, we assume that the mean is constant over locations, and without loss of generality set to be zero. A model to describe the small-scale variations is

$$Y(\mathbf{s}; v) = X(\mathbf{s}; v) + \varepsilon(\mathbf{s}; v), \tag{3.5}$$

where $\varepsilon(\mathbf{s}; v)$ represents the functional white noise, and is assumed to be uncorrelated with $X(\mathbf{s}; v)$.

The mean of $Y(\mathbf{s}; v)$ in model (3.5) is zero, and the covariance is such that

$$C_Y(\mathbf{s}_i, \mathbf{s}_j; f) = C_X(\mathbf{s}_i, \mathbf{s}_j; f) + \mathbb{1}(i = j)C_\varepsilon(\mathbf{s}_i, \mathbf{s}_j; f), \quad f \in \mathcal{H}, \tag{3.6}$$

where $\mathbb{1}(\cdot)$ is the indicator function. Thus, we need to consider the additional term $C_\varepsilon(\mathbf{s}, \mathbf{s}; \cdot)$ in the estimation. Similarly, as in Section 2.2, a basis function approach can be used to obtain the estimators.

Let $\mathbf{Z}(\mathbf{s}) = \{Z_1(\mathbf{s}), \ldots, Z_K(\mathbf{s})\}^\mathsf{T}$ be the multivariate random field obtained from the projection of $Y(\mathbf{s}; v)$ onto the basis functions $\{\eta_k\}_{k=1}^K$, that is, each component of $\mathbf{Z}(\mathbf{s})$ is defined as $Z_k(\mathbf{s}) = \langle Y(\mathbf{s}; \cdot), \eta_k \rangle$, which are scalar random fields with mean zero. From (3.5), for each $k = 1, \ldots, K$, we have that the process $Z_k(\mathbf{s})$ is such that

$$Z_k(\mathbf{s}) = \langle X(\mathbf{s}), \eta_k \rangle + \langle \varepsilon(\mathbf{s}), \eta_k \rangle.$$

The variance of $Z_k(\mathbf{s})$ is

$$\begin{aligned}\mathbb{E}\{\langle Y(\mathbf{s}; \cdot), \eta_k \rangle \langle Y(\mathbf{s}; \cdot), \eta_k \rangle\} &= \langle C_Y(\mathbf{s}, \mathbf{s}; \eta_k), \eta_k \rangle \\ &= \langle C_X(\mathbf{s}, \mathbf{s}; \eta_k), \eta_k \rangle + \langle C_\varepsilon(\mathbf{s}, \mathbf{s}; \eta_k), \eta_k \rangle,\end{aligned}$$

where the last equality is obtained using (3.6). This implies that the random field $Z_k(\mathbf{s})$ has a nugget effect $\langle C_\varepsilon(\mathbf{s}, \mathbf{s}; \eta_k), \eta_k \rangle$ that should be considered when fitting a covariance (variogram) model.

The covariance estimator of $Y(\mathbf{s}; v)$ is obtained from (2.10), after estimating the covariance $\mathbf{\Sigma}(h) = \text{Cov}\{\mathbf{Z}(\mathbf{s}_1), \mathbf{Z}(\mathbf{s}_2)\}$, $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$, from the data $\{\langle y(\mathbf{s}_i; \cdot), \eta_k \rangle\}_{i=1}^n$, $k = 1, \ldots, K$. These ideas are extended to estimating the variogram of $Y(\mathbf{s}; v)$ by estimating the matrix-variogram of $\mathbf{Z}(\mathbf{s})$.

In the case of the trace-variogram of $Y(\mathbf{s}; v)$, from (3.5) we have that the mean of the inner product satisfies $\mathbb{E}\{\langle Y(\mathbf{s}_i; \cdot), Y(\mathbf{s}_j; \cdot) \rangle\} = \mathbb{E}\{\langle X(\mathbf{s}_i; \cdot), X(\mathbf{s}_j; \cdot) \rangle\} + \mathbb{E}\{\langle \varepsilon(\mathbf{s}_i; \cdot), \varepsilon(\mathbf{s}_j; \cdot) \rangle\}$. That is, the trace-covariogram of $Y(\mathbf{s}; v)$ is such that

$$\sigma_{\text{tr,Y}}(\mathbf{s}_1, \mathbf{s}_2) = \sigma_{\text{tr,X}}(\mathbf{s}_1, \mathbf{s}_2) + \mathbb{1}(\mathbf{s}_1 = \mathbf{s}_2)\sigma_{\text{tr},\varepsilon}(\mathbf{s}_1, \mathbf{s}_2).$$

Then, when estimating $\sigma_{\text{tr,Y}}(\mathbf{s}_1, \mathbf{s}_2)$ using $\{\langle y(\mathbf{s}_i; \cdot), y(\mathbf{s}_j; \cdot) \rangle\}_{i,j}$, as described in Section 2.2.2, one should consider the nugget effect $\sigma_{\text{tr},\varepsilon}(\mathbf{s}, \mathbf{s})$. The same is true in the case of the trace-variogram.

3.3 Large-scale and small-scale variations

Datasets often have both a trend component (large-scale variation) and a spatial variability (small-scale variation), for example, temperature data show an increasing tendency and a small variability around this tendency. In the context of functional data correlated only in time, Martínez-Hernández and Genton (2020) proposed a method to estimate trend using tensor product surfaces.

A model for spatial functional data can be written as

$$Y(\mathbf{s}; v) = \mu(\mathbf{s}; v) + X(\mathbf{s}; v) + \varepsilon(\mathbf{s}; v).$$

As before, the mean can be expressed as $\mu(\mathbf{s}; v) = \sum_{l=0}^L a_l(\mathbf{s}) f_l(v)$. Similarly to (3.1), the model to estimate the parameters of the mean can be written as

$$Y(\mathbf{s}; v) = \sum_{l=0}^{L} a_l(\mathbf{s}) f_l(v) + \epsilon(\mathbf{s}; v), \tag{3.7}$$

where the residual $\epsilon(\mathbf{s}; v) := X(\mathbf{s}; v) + \varepsilon(\mathbf{s}; v)$ is now a functional random field with mean zero and covariance $C_\epsilon(\mathbf{s}_1, \mathbf{s}_2; \cdot)$. Thus, unlike model (3.1), in which the residuals are not correlated, model (3.7) has spatially correlated residuals. Because of this correlation, we can use the GLS method instead of the OLS method (Menafoglio, Secchi and Dalla Rosa, 2013). Let $\boldsymbol{\Sigma}$ be the trace-variogram matrix of $\epsilon(\mathbf{s}_i; v)$ at different distances of the locations $\mathbf{s}_i$. Then, the GLS estimator of $\mu(\mathbf{s}_i; v)$ is

$$\hat{\boldsymbol{\mu}}(v) = \mathbf{A}\big(\mathbf{A}^\top \boldsymbol{\Sigma}^{-1} \mathbf{A}\big)^{-1} \mathbf{A}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}(v), \tag{3.8}$$

where $\mathbf{A}$ is the design matrix, and $\mathbf{y}(v)$ the evaluation of $y(\mathbf{s}; v)$ at n locations, both defined in Section 3.1.

The trace-variogram matrix $\boldsymbol{\Sigma}$ cannot be estimated directly from $\epsilon(\mathbf{s}_i; v)$, because we do not observe them. The approach commonly used, to solve this problem, is to use the following iterative procedure:

1. Compute the initial estimator $\hat{\mu}_0(\mathbf{s}_i; v)$ as in (3.3), assuming that $\{\epsilon(\mathbf{s}_i; v)\}_{i=1}^n$ are not spatially correlated.
2. Compute the residuals $\hat{\epsilon}(\mathbf{s}_i; v) = y(\mathbf{s}_i; v) - \hat{\mu}_0(\mathbf{s}_i; v)$, for $i = 1, \ldots, n$.
3. Estimate the initial empirical trace-variogram $\hat{\gamma}_{\mathrm{tr},0}(h)$ as in (2.12), using $\{\hat{\epsilon}(\mathbf{s}_i; v)\}_{i=1}^n$, and then, obtain an estimator $\hat{\boldsymbol{\Sigma}}_0$ of $\boldsymbol{\Sigma}$ by fitting a parametric model with a nugget effect.
4. Re-estimate the mean to get $\hat{\mu}_1(\mathbf{s}_i; v)$ by using (3.8) with $\hat{\boldsymbol{\Sigma}}_0$.
5. Repeat steps 2–4 until convergence.

Once the mean $\mu(\mathbf{s}; v)$ is estimated, it is removed from the data. Then, the covariance is estimated as in Section 3.2, using the spatial functional data $\{y(\mathbf{s}_i; v) - \hat{\mu}(\mathbf{s}_i; v)\}_{i=1}^n$.

The iterative procedure can also be performed with the finite basis function $\boldsymbol{\eta}(v)$. In Step 1, we can use (3.4), and in Step 4, use the estimator $\operatorname{vec}(\hat{\mathbf{B}}) = (\mathbf{J}_{\boldsymbol{\eta}}^\top \otimes \mathbf{A}^\top \hat{\boldsymbol{\Sigma}}_0 \mathbf{A})^{-1} \operatorname{vec}(\mathbf{A}^\top \hat{\boldsymbol{\Sigma}}_0 \mathbf{Z} \mathbf{J}_{\boldsymbol{\eta}})$ to obtain $\hat{\mu}_1$.

4 Kriging for functional random fields

In spatial statistics, the concept of kriging (co-kriging for the multivariate setting) is a synonym of optimal interpolation. The main goal is to be able to predict at locations where data are not observed. This predictor is a linear combination of the observed data, such that it is the best linear unbiased predictor under squared loss. Here, we briefly mention the concept of kriging and we redirect readers to the references provided.

Let $\mathbf{s}_0 \in D$ be the location at which the curve will be predicted. Let $\{\Psi_1, \ldots, \Psi_n\}$ be linear operators from $\mathcal{H}$ to $\mathcal{H}$. In general, kriging can be defined as

$$\hat{x}(\mathbf{s}_0; v) = \sum_{i=1}^{n} \Psi_i\big\{x(\mathbf{s}_i; v)\big\}, \tag{4.1}$$

where the coefficients Ψ_i are obtained by minimizing the square norm of the error prediction, $\hat{x}(\mathbf{s}_0; v) - X(\mathbf{s}_0; v)$, with an additional constraint of unbiasedness. That is,

$$\min_{\Psi_1, \ldots, \Psi_n} \mathbb{E}\big\{\big\langle \hat{x}(\mathbf{s}_0; \cdot) - X(\mathbf{s}_0; \cdot), \hat{x}(\mathbf{s}_0; \cdot) - X(\mathbf{s}_0; \cdot)\big\rangle\big\}$$

$$\text{s.t. } \mathbb{E}\big\{\hat{x}(\mathbf{s}_0; v)\big\} = \mathbb{E}\big\{X(\mathbf{s}_0; v)\big\}.$$

A particular case of the coefficients operators Ψ_i in (4.1) is the so-called *kernel operators*, which are defined as $\Psi_i(f)(v) = \int_0^1 \lambda_i(v, u) f(u)\, \mathrm{d}u$, $f \in \mathcal{H}$. In this case, the estimation is through functions $\lambda_i(v, u)$. Other cases are $\Psi_i(f)(v) = \lambda_i(v) f(v)$ and $\Psi_i(f)(v) = w_i f(v)$, with w_i scalars. The latter corresponds to a simple ponderation of the observed curves, that is,

$\hat{x}(\mathbf{s}_0; v) = \sum_{i=1}^{n} w_i x(\mathbf{s}_i; v)$. All these cases can be fitted into ordinary or universal kriging. For ordinary kriging, see Goulard and Voltz (1993), Nerini, Monestiez and Manté (2010), Giraldo, Delicado and Mateu (2010), and Giraldo, Delicado and Mateu (2011), as well as the review paper by Delicado et al. (2010). For universal kriging, see Menafoglio, Secchi and Dalla Rosa (2013), Caballero, Giraldo and Mateu (2013), Reyes, Giraldo and Mateu (2015), and Menafoglio, Grujic and Caers (2016). For co-kriging (multivariate functional random fields), see Bohorquez, Giraldo and Mateu (2017) and Grujic et al. (2018). An alternative approach to kriging, which is based on a tensor function space, can be found in Aguilera-Morillo, Durbán and Aguilera (2017).

5 Surface time series

In many phenomena, data can be collected in the form of a surface, called *surface data* in our case, or *manifold data* for more complex structures. For example, we can have data arising from neuroimaging (Lila, Aston and Sangalli, 2016), from two-dimensional time-frequency domains (Aston, Pigoli and Tavakoli, 2017), from satellite images (Zhang, Clayton and Townsend, 2011), and functional data with two-dimensional domain (Crainiceanu et al., 2011, Morris et al., 2011). Here, we consider surface data as functional, that is, the atoms of the functional random variable are continuous surfaces. Surface data provide an alternative approach to analyzing spatial data, where the continuous realization of a random field is considered as a unit. This approach can have computational advantages, especially if the locations, where data are observed, are dense in space. Particularly, spatio-temporal data (Cressie and Wikle, 2011) can be considered as surface data that are observed over time (surface time series). Surface data capture the spatial dependence through the continuity of the surface (see, e.g., Bernardi et al., 2017). Moreover, the approach of surface data can be applied to a nonplanar spatial domain, such as a sphere or a general two-dimentional manifold (Dassi et al., 2015, Wilhelm et al., 2016, Ettinger, Perotto and Sangalli, 2016, Menafoglio and Secchi, 2017, Greco, Ventrucci and Castelli, 2018). Kriging can be applied to these complex domains by using an appropriate distance, but the covariance models do not necessarily guarantee a positive definite covariance, for example, the Matérn covariance (Gneiting, 2013). Here, we focus on spatio-temporal data, where at each time point we observe a surface.

The functional time series approach to spatial statistics has been studied by Ruiz-Medina, Salmerón and Angulo (2007). In Aston, Pigoli and Tavakoli (2017), a tensor product Hilbert space was considered to propose a separability test for the covariance operators of random surfaces.

5.1 Basic concepts

Let (Ω, F, P) be a probability space, and let $\mathcal{H}$ be the Hilbert space defined as the set of functions with domain $D \subset \mathbb{R}^2$, $\mathcal{H} = \{f : D \to \mathbb{R} : \int_D |f(\mathbf{s})|^2 \, \mathrm{d}\mathbf{s} < \infty\}$, and with the inner product $\langle f, g \rangle = \int_D f(\mathbf{s}) g(\mathbf{s}) \, \mathrm{d}\mathbf{s}$. The norm induced by the inner product is denoted by $\|\cdot\|_{\mathcal{H}}$. Thus, a random variable $X : \Omega \to \mathcal{H}$ is a functional random variable with a surface as atom. We denote by $X(\mathbf{s})$ this functional random variable with $\mathbf{s} = (s_1, s_2) \in D$.

A functional time series is a sequence of functional random variables $\{X_t(\mathbf{s}); t \in \mathbb{Z}\}$ in $\mathcal{H}$. Bosq (2000) is a monograph on linear processes in function spaces, including functional time series in a Hilbert space. Hörmann and Kokoszka (2012) reviewed functional time series. Also, see Ramsay and Silverman (2005), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012) for a general introduction to functional data analysis. In this paper, we assume that $\mathbb{E}\{\|X_t(\mathbf{s})\|_{\mathcal{H}}^2\} < \infty$ for all $t \in \mathbb{Z}$.

The mean of the surface time series $\{X_t(\mathbf{s})\}$ is defined as $\mu_t(\mathbf{s}) = \mathbb{E}\{X_t(\mathbf{s})\}$, where $\mu_t(\mathbf{s})$ is such that $\mathbb{E}(\langle X_t, f\rangle) = \langle \mu_t, f\rangle$ for all $f \in \mathcal{H}$. The covariance function at lag $h \in \mathbb{N}$ is defined as

$$C_{X_{t-h},X_t}(f) = \mathbb{E}\big\{\langle X_{t-h} - \mu_{t-h}, f\rangle (X_t - \mu_t)\big\}, \quad f \in \mathcal{H}.$$

This covariance function can be expressed as

$$C_{X_{t-h},X_t}(f)(\cdot) = \int \sigma_{t-h,t}(\cdot, \mathbf{s}) f(\mathbf{s})\, d\mathbf{s},$$

where $\sigma_{t-h,t}(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{Cov}\{X_{t-h}(\mathbf{s}_1), X_t(\mathbf{s}_2)\}$. The stationarity condition is important for statistical inference.

Definition 3 (Weak stationarity). A surface time series $\{X_t(\mathbf{s}); t \in \mathbb{Z}\}$ is said to be (weakly) stationary if

1. $\mathbb{E}\{X_t(\mathbf{s})\} = \mu(\mathbf{s})$ for all $t \in \mathbb{Z}$, and
2. $C_{X_{t_1+h},X_{t_2+h}}(f) = C_{X_{t_1},X_{t_2}}(f)$ for all $t_1, t_2 \in \mathbb{Z}$, $h \in \mathbb{N}$, and $f \in \mathcal{H}$.

If the surface time series is stationary, we write C_h for the covariance functions instead of $C_{X_t,X_{t+h}}$. The definition of stationarity does not require stationarity over the space D, for example, for each t, data can be a realization of a nonstationary random field. In general, the covariance function C_0 describes only the spatial dependence, whereas C_h, for $h \neq 0$, describes the dependency over time of the surface time series.

A *surface white noise* is a stationary surface time series with a zero mean and a covariance function $C_h = 0$, if $h \neq 0$. Thus, the surface white noise can have a spatial correlation at each time point, but not across time.

Similarly as before, the eigenfunctions are defined as functions $\zeta \in \mathcal{H}$ such that

$$C_0(\zeta)(\mathbf{s}) = \lambda\zeta(\mathbf{s}),$$

where λ is positive and is the corresponding eigenvalue. Moreover, the covariance operator C_0 can be decomposed in terms of the eigenfunctions, that is

$$C_0(f)(\mathbf{s}) = \sum_{j=1}^{\infty} \lambda_j \langle \zeta_j, f\rangle \zeta_j(\mathbf{s}),$$

where ζ_j, $j = 1, 2, \ldots$, are the eigenfunctions of C_0 with eigenvalues λ_j. The eigenvalues are such that $\sum_{j=1}^{\infty} \lambda_j = \mathbb{E}\{\|X_0(\mathbf{s})\|_{\mathcal{H}}^2\} < \infty$. This operator C_0 is nuclear and therefore Hilbert-Schmidt. Similarly as in the finite dimensional case, eigenfunctions are important to reduce dimensionality.

5.2 Estimation

Now, we describe estimators of the mean $\mu(\mathbf{s})$ and the covariance operator C_h. Let $\{x_t(\mathbf{s})\}_{t=1}^T$ be a realization of a stationary surface time series $X_t(\mathbf{s})$ with mean $\mu(\mathbf{s})$. The sample mean is defined as

$$\hat{\mu}(\mathbf{s}) = \frac{1}{T}\sum_{t=1}^{T} x_t(\mathbf{s}). \tag{5.1}$$

The sample mean $\hat{\mu}(\mathbf{s})$ is an unbiased estimator of $\mu(\mathbf{s})$.

The empirical covariance at lag h of $X_t(\mathbf{s})$ is defined as

$$\hat{C}_h(f)(\mathbf{s}) = \frac{1}{T-h}\sum_{t=1}^{T-h} \langle x_t - \hat{\mu}, f\rangle \big\{x_{t+h}(\mathbf{s}) - \hat{\mu}(\mathbf{s})\big\}, \quad f \in \mathcal{H}, \tag{5.2}$$

and the corresponding empirical kernel is defined as

$$\hat{\sigma}_h(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{T-h} \sum_{t=1}^{T-h} \{x_{t+h}(\mathbf{s}_1) - \hat{\mu}(\mathbf{s}_1)\}\{x_t(\mathbf{s}_2) - \hat{\mu}(\mathbf{s}_2)\}.$$

The empirical covariance operator $\hat{C}_h$ is an unbiased estimator of C_h, see Bosq (2000). For papers related to the mean and the covariance functions, refer to previous studies by authors Hörmann and Kokoszka (2010), Horváth, Kokoszka and Reeder (2013), and Horváth, Kokoszka and Rice (2014).

In time series analysis, there is no direct modeling of the covariance function. Instead, a model for the process is proposed, and the covariance function is derived from the model. We adopt this idea in the next section (Section 5.3).

5.3 Modeling

This section discusses two topics: the continuous estimation of the surface and the modeling of the continuous surface series.

5.3.1 *Estimating the continuous surface.* In practice, data are observed on a finite set of points, that is, for each $t = 1, \ldots, T$, we observe n_t points $\{y_{t,i} = y_t(\mathbf{s}_i)\}_{i=1}^{n_t}$ of the functional data $Y_t(\mathbf{s})$ on a set of points $\{\mathbf{s}_1, \ldots, \mathbf{s}_{n_t}\} \subset D$, and possibly with measurement errors. Thus, for each t, the unknown surface (deterministic field) $y_t(\mathbf{s})$ needs to be estimated. This continuous surface estimate can be associated with kriging in classical spatial data analysis.

Because the procedure to estimate $y_t(\mathbf{s})$ is independent of t, we drop the subindex t in the sequel, and we consider n as the sample size. Thus, a model of $y(\mathbf{s})$ can be written as

$$y_i = y(\mathbf{s}_i) + \varepsilon_i,$$

where $\{\varepsilon_i\}_{i=1}^{n}$ represent the measurement errors that are spatially uncorrelated. The function $y(\mathbf{s})$ describes the spatial structure of the phenomenon being studied.

To estimate $y(\mathbf{s})$, one can extend the smoothness techniques of the curves described in Ramsay and Silverman (2005) to surfaces (manifolds). In particular, one can extend the spline smoothing. The extension of spline smoothing to surfaces is an important research area. Extensions have been done over Euclidean domains and non-Euclidean domains, including the spherical domain. One extension is to use the tensor product of univariate B-splines (Eilers and Marx, 1996, Wood, 2006, Qingguo and Longsheng, 2010, Xiao, Li and Ruppert, 2013). In this case, an estimator of $y(\mathbf{s})$ has the form

$$\hat{y}(\mathbf{s}) = \sum_{k=1}^{K_1} \sum_{l=1}^{K_2} \theta_{kl} \eta_k(s_1) \nu_l(s_2),$$

where $\{\eta_k\}_{k=1}^{K_1}$ and $\{\nu_l\}_{l=1}^{K_2}$ are B-splines basis functions for s_1 and s_2 coordinates, respectively, with $\mathbf{s} = (s_1, s_2)$. In the estimation procedure, smoothness properties are imposed through a penalization term. To read about some approaches of bivariate smoothing, see Ruppert, Wand and Carroll (2003), Lai and Schumaker (2007), and Wood (2017).

In general, the estimation of $y(\mathbf{s})$ can be formulated as the minimization of the sum of squared errors with a penalization term. The penalization term measures the roughness of the fitted surface and can carry partial information of $y(\mathbf{s})$. That is, the optimization problem is written as

$$\sum_{i=1}^{n} \{y_i - y(\mathbf{s}_i)\}^2 + \lambda P(y), \tag{5.3}$$

where $P(y)$ is the penalization term, and λ the smoothness parameter which controls the smoothness of the estimated surface. A popular penalization is the thin-plate energy, which is defined as $P(y) = \int \{(\frac{\partial^2 y}{\partial s_1^2})^2 + 2(\frac{\partial^2 y}{\partial s_1 \partial s_2})^2 + (\frac{\partial^2 y}{\partial s_2^2})^2\} \mathrm{d}\mathbf{s}$. The resulting estimator is called the thin plate splines (Duchon, 1977). A Bayesian adaptive thin plate spline was proposed in Yue and Speckman (2010). Another example of penalization involves the Laplacian, that is $P(y) = \int (\frac{\partial^2 y}{\partial s_1^2} + \frac{\partial^2 y}{\partial s_2^2})^2 \mathrm{d}\mathbf{s}$ (Wood, Bravington and Hedley, 2008, Sangalli, Ramsay and Ramsay, 2013). The definition of the penalty term depends on each specific problem. For example, in the Laplacian case, the unique penalty parameter λ that controls both directions s_1 and s_2 implies an isotropic smoothing. In contrast, if $P(y) = \lambda_1 P_1(y) + \lambda_2 P_2(y)$, where $P_i(y)$ is a penalty term in the ith coordinate, then it results in anisotropic smoothing.

In general, the penalty term can be defined in terms of a partial differential equation (PDE). For example, $P(y) = \int (Ly - u)^2 \mathrm{d}\mathbf{s}$, where L is a differential operator and $Ly = u$ is a PDE (see, e.g., Azzimonti et al., 2015, Sangalli, 2020). The advantage of the penalty term with PDE is that it can handle complex domains with boundary conditions or interior holes, and varies depending on the phenomena being studied. The PDE is such that it contains information about the phenomena, and it regularizes the estimation with values of λ. The solution of (5.3) may not have a closed form, but it can be approximated by using finite elements analysis, see Ramsay (2002), Duchamp and Stuetzle (2003), and Sangalli, Ramsay and Ramsay (2013).

We describe the solution of (5.3) using the finite elements analysis technique. Let $\mathcal{M}$ be a mesh of D. Let $\{\phi_1(\mathbf{s}), \ldots, \phi_K(\mathbf{s})\}$ be basis functions that are piece-wise polynomials associated with the mesh $\mathcal{M}$. Then, the estimator of $y(\mathbf{s})$ is assumed to have the form

$$\hat{y}(\mathbf{s}) = \sum_{k=1}^{K} \beta_k \phi_k(\mathbf{s}),$$

where the coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^\mathsf{T}$ need to be estimated. Let $\mathbf{y} = (y_1, \ldots, y_n)^\mathsf{T}$ be the observed values over D, and let $\mathbf{P}$ be the discretization of the penalty. Then, the estimator of $\boldsymbol{\beta}$ has the form

$$\hat{\boldsymbol{\beta}} = (\Phi^\mathsf{T}\Phi + \lambda \mathbf{P})^{-1} \Phi^\mathsf{T} \mathbf{y},$$

where $\Phi = \{\phi_k(\mathbf{s}_i)\}_{i,k=1}^{n,K}$ is the $n \times K$ matrix which represents the evaluation of each basis function at the locations at which data are observed.

The time component can also be considered in the PDE. Arnone et al. (2019) proposed general forms of time-dependent PDEs in the context of time dependent surface data.

Additional studies of data over complex domains have been published. Wang and Ranalli (2007) proposed a modified thin plate spline over a complex domain; Lindgren, Rue and Lindström (2011) linked Gaussian fields via stochastic partial differential equations, where the solution is found using a finite elements analysis; Scott-Hayward et al. (2014) proposed a complex region spatial smoother using the geodesic distance; and Menafoglio, Gaetani and Secchi (2018) proposed a methodology for spatial fields of object data over complex domains.

5.3.2 *Functional autoregressive models.* Here, we assume that the functional time series consist of continuous surfaces (deterministic fields) that can be estimated with the methods described in Section 5.3.1. In the context of functional time series, the most popular model is the functional autoregressive model of order P, FAR(P). A surface time series $\{X_t(\mathbf{s})\}$ follows the FAR(P) model if $X_t(\mathbf{s}) = \sum_{p=1}^{P} \Psi_p(X_{t-p})(\mathbf{s}) + W_t(\mathbf{s})$, where each coefficient $\Psi_p : \mathcal{H} \to \mathcal{H}$ is an operator, and $\{W_t(\mathbf{s})\}$ is a surface white noise. In practice, the order P needs to be estimated (Kokoszka and Reimherr, 2013). Here, we assume $P = 1$ to illustrate the ideas and to simplify the notations.

Let $\{y_t(\mathbf{s})\}_{t=1}^{T}$ be the surface data observed over time $t = 1, \ldots, T$, and assume that it is a stationary surface time series. Then, the dependency over time can be modeled by using the FAR(1) process, that is

$$y_t(\mathbf{s}) = \mu(\mathbf{s}) + x_t(\mathbf{s}), \tag{5.4}$$

$$x_t(\mathbf{s}) = \Psi(x_{t-1})(\mathbf{s}) + W_t(\mathbf{s}), \tag{5.5}$$

where $\mu(\mathbf{s})$ represents the surface mean, i.e., the large-scale variation for all t. The unobserved $x_t(\mathbf{s})$ follows a stationary FAR(1) process with mean zero, and $\{W_t(\mathbf{s})\}$ is a surface white noise. The surface white noise $\{W_t(\mathbf{s})\}$ can be interpreted as the surface data components that describe the small-scale variation for each time t, which are not correlated over time. The dependency over time is driven by the operator Ψ.

Since $\{y_t(\mathbf{s})\}$ is assumed to be stationary, the estimation of $\mu(\mathbf{s})$ can be obtained as in (5.1). After removing the mean, the rest of the analysis is performed on the process $x_t(\mathbf{s}) = y_t(\mathbf{s}) - \hat{\mu}_t(\mathbf{s})$.

Now, we focus on the estimation of the coefficient operator Ψ. Let C_h be the covariance operator of the FAR(1) process $X_t(\mathbf{s})$ with realization $\{x_t(\mathbf{s})\}$. Then, the covariance operator of $X_t(\mathbf{s})$ satisfies

$$C_1(f) = \Psi\{C_0(f)\}, \quad f \in \mathcal{H}. \tag{5.6}$$

Moreover, for $h \in \mathbb{N}$, we have that $C_h(f) = \Psi^h\{C_0(f)\}$. So if the coefficient Ψ and C_0 are known, then we can compute the covariance function at any lag $h \in \mathbb{N}$. Thus, the estimation of Ψ is crucial.

To obtain an estimator of Ψ, we can use the estimation of C_0 and C_1 in (5.6), defining $\Psi(f) = C_1\{C_0^{-1}(f)\}$, if C_0 is invertible. In principle, we can always use the estimator (5.2) for $h = 0, 1$, and compute the inverse $\hat{C}_0^{-1}$. However, when the sample size T tends to infinity, $\hat{C}_0^{-1}$ becomes unbounded (Cardot, Ferraty and Sarda, 1999). This is because C_0 is a compact operator (Bosq, 2000). Thus, it is necessary to use some regularization methods to obtain C_0^{-1}. That is, $(C_0 + \alpha_T)^{-1}$ is computed instead of C_0^{-1} where $\alpha_T > 0$ and $\alpha_T \downarrow 0$. Alternatively, C_0^{-1} can be approximated by using only the first k eigenfunctions corresponding to the largest eigenvalues, that is, $C_0^{-1} = \sum_{j=1}^{k} \lambda_j^{-1} \zeta_j \otimes \zeta_j$, see Bosq (2000) and Kokoszka and Reimherr (2017). Let $\hat{C}_0^{-1}$ denote an estimator of the inverse operator, either using some regularization method, finite eigenfunctions or other methods (Martínez-Hernández, Genton and González-Farías, 2019). Then, the estimator of Ψ is defined as

$$\hat{\Psi}(f) = \hat{C}_1\{\hat{C}_0^{-1}(f)\}, \quad f \in \mathcal{H}.$$

Once $\mu(\mathbf{s})$ and Ψ are estimated, then the one-step-ahead prediction is obtained as

$$\hat{y}_{T+1}(\mathbf{s}) = \hat{\mu}(\mathbf{s}) + \hat{x}_{T+1}(\mathbf{s}),$$

where $\hat{x}_{T+1}(\mathbf{s}) = \hat{\Psi}(x_T)(\mathbf{s})$. The estimators of μ, Ψ, and C_h can be explicitly expressed in terms of the basis functions $\{\phi_1(\mathbf{s}), \ldots, \phi_K(\mathbf{s})\}$ from the finite elements technique. Alternative approaches can be used to predict data, for example one can extend the ideas described in Hyndman and Ullah (2007) and Aue, Norinho and Hörmann (2015) to surfaces.

The approach of the surface time series provides alternatives to analyze spatio-temporal data. The advantage of this approach is that it can handle data collected on a large scale for each time point, over a general domain, Euclidean, or non-Euclidean. With this approach, the process is modeled instead of the covariance, which is convenient when the classical covariance models are not guaranteed to be positive definite.

6 Statistical software

Here, we mention some packages available in R for SFD and surface data. The package *fda* (Ramsay et al., 2018) provides several commands to analyze and construct continuous functions. It contains several options of basis functions, such as Fourier basis functions and spline basis functions. The basis functions can be used to estimate the continuous functions as in (2.8). Once the coefficients are obtained in (2.8), we can use packages for classical spatial data, for example, *spatial* (Venables and Ripley, 2002), *gstat* (Pebesma, 2004), *RandomFields* (Schlather et al., 2019), *fields* (Nychka et al., 2017), *geoR* (Ribeiro and Diggle, 2018), *ExaGeoStatR* (Abdulah et al., 2019) for large datasets, and *spBayes* (Finley, Banerjee and Gelfand, 2015) for Bayesian analysis of hierarchical multivariate models. Thus, when data are expressed in terms of basis functions as in (2.8), we can combine the *fda* package and the packages for spatial data to obtain estimators of the mean and the covariance functions described in Section 2.2.

The package *geofd* by Giraldo, Delicado and Mateu (2015) implements kriging of functional data described in Section 4. The curves observed are pre-processed by fitting Fourier or B-splines basis functions. Also, this package provides a command to compute the trace-variogram defined in (2.12). Another package related to Section 4 is *fdagstat* (Grujic and Menafoglio, 2017). This package implements kriging, cokriging, and universal kriging, and includes the large-scale variation described in Sections 3.1 and 3.3.

For the surface data described in Section 5, one can use the package *mgcv* by Wood (2017), which allows us to smooth surfaces. The package *fdaPDE* (Lila et al., 2019) implements smoothing with PDE penalization described in Sangalli, Ramsay and Ramsay (2013) and Azzimonti et al. (2015). *INLA* (Rue, Martino and Chopin, 2009) can be used to estimate continuous surfaces. The package *Manifoldgstat* by Sartori and Torriani (2019) implements kriging for manifold-valued random fields.

Some visualization tools for functional data and functional time series are the functional boxplots (Sun and Genton, 2011, 2012) implemented in the *fda* package; for functional images and surfaces, surface boxplots are used (Genton et al., 2014). These tools are based on an order induced by a depth notion for functional data.

7 Data analysis

In this section, we illustrate general ideas of modeling spatial functional data and surface data, without a deep statistical analysis of the data. Our goal is to provide a general example, using only available packages.

We use wind data simulated from the Weather Research and Forecasting (WRF) model by Yip (2018). Each measurement corresponds to hourly wind speed from 2009 to 2014, in a 115×115 km region centered in Dumat Al Jandal, Saudi Arabia. That is where the first wind farm of the country is being built. These wind speeds are simulated on a regular grid of points in space, namely at 5-km resolution.

7.1 Spatial functional data approach

Our data are the daily wind speed, where $y(\mathbf{s}_i; v_j)$ is the wind speed at location $\mathbf{s}_i$ for hour v_j, for $v_j = 1, 2, \ldots, 24$. With the functional approach, we consider $y(\mathbf{s}_i; \cdot)$ as a single object, assuming continuity over time. To illustrate the continuous estimation of the curves, we focus on data observed on June 22, 2010. For that specific day, we fit 23 cubic B-spline basis functions for each location, that is, $\hat{y}(\mathbf{s}_i; v) = \sum_{k=1}^{23} z_k(\mathbf{s}_i)\eta_k(v)$, where $\{\eta_k\}_{k=1}^{23}$ are B-spline basis functions. The continuous curves are required to be smooth; thus, we consider a penalization

term on the second derivative of $\hat{y}(\mathbf{s}_i; v)$ with respect to v. Specifically, the estimator is such that

$$y(\mathbf{s}_i; v_j) = \hat{y}(\mathbf{s}_i; v_j) + e_{ij},$$

where the coefficients $z_k(\mathbf{s}_i)$ of $\hat{y}(\mathbf{s}_i; v)$ are obtained as the solution that minimizes

$$\sum_{j=1}^{24} \{y(\mathbf{s}_i; v_j) - \hat{y}(\mathbf{s}_i; v_j)\}^2 + \lambda \int \hat{y}''(\mathbf{s}_i; v)\, \mathrm{d}v,$$

with the smoothing parameter λ fixed. We select the optimal smoothing parameter by generalized cross-validation through the $n = 529$ locations. Figure 1 shows an example of the curve estimation at two locations. These two locations are separated by 10 km at the same latitude.

Once continuous curves are estimated with basis functions, we can compute the mean (large-scale variation). Since data are measured on a regular grid of points, and if each corresponding coefficient random field $\{Z_k(\mathbf{s}_i)\}_{i=1}^n$, with realization $\{z_k(\mathbf{s}_i)\}_{i=1}^n$, is stationary, then we can use the empirical mean to estimate each $\mathbb{E}\{Z_k(\mathbf{s})\}$. Under this scenario, Figure 2 (left) shows the result of the estimated mean curve.

On the other hand, if the coefficient random fields $\{Z_k(\mathbf{s}_i)\}_{i=1}^n$ are nonstationary, then $\mathbb{E}\{Z_k(\mathbf{s})\}$ can depend on the location $\mathbf{s}$. Under this framework, we use the `surf.gls` command defined in the *spatial* package. Figure 2 (right) shows the estimator of $\mathbb{E}\{Z_1(\mathbf{s})\}$ for the first random field coefficient $Z_1(\mathbf{s})$. The `surf.gls` command uses the GLS method to

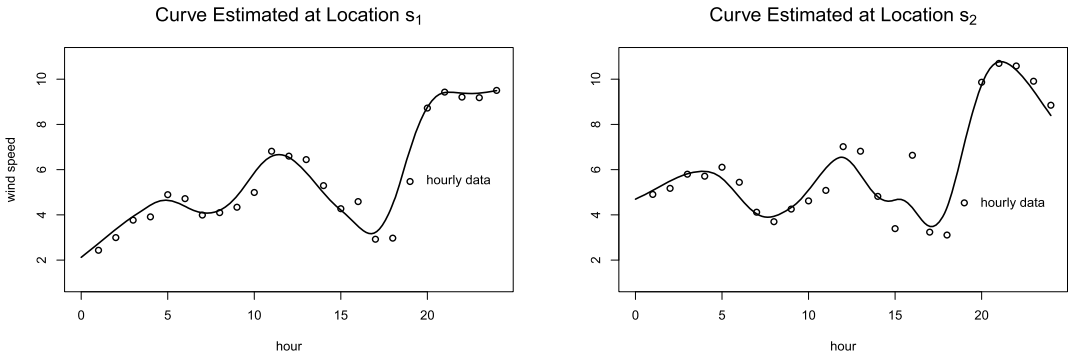

Figure 1 *Example of two estimations of the continuous curve at two different locations* $\mathbf{s}_1$ *and* $\mathbf{s}_2$ *separated by* 10 *km with latitude fixed. We use* 23 *cubic B-spline basis functions.*

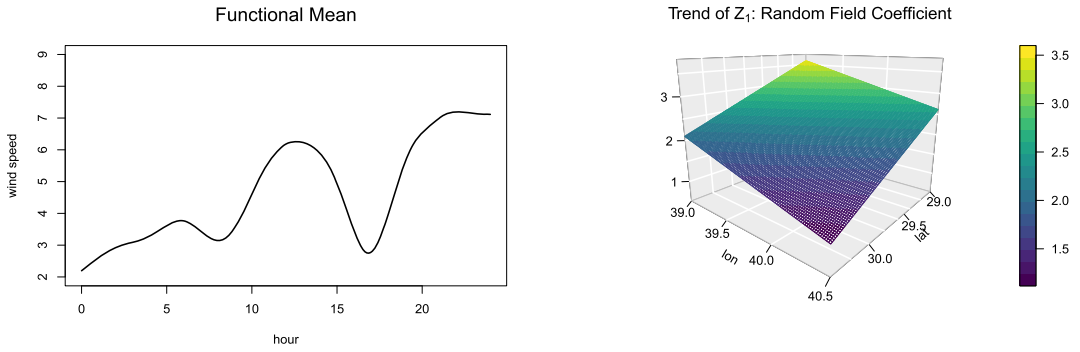

Figure 2 *Left*: *mean curve if the coefficient random fields* $\{Z_k(\mathbf{s})\}$ *are stationary. Right*: *trend estimation of the first coefficient random field* $Z_1(\mathbf{s})$ *if the coefficient random fields* $\{Z_k(\mathbf{s})\}$ *are nonstationary.*

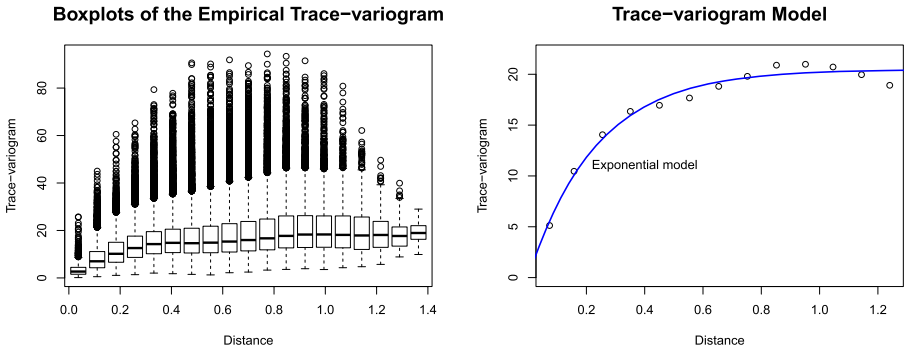

Figure 3 *Trace-variogram of the daily wind speed. Left*: *empirical. Right*: *fitted model* (*blue*).

obtain the estimator with an exponential model as the covariance function. From Figure 2 (right), we observe some evidence that the mean depends on the location $\mathbf{s}$.

For the modeling of small-scale variations, we use the trace-variogram defined in (2.12). With the assumption of isotropy, Figure 3 (left) shows boxplots of the corresponding empirical trace-variogram for different distances. The `fit.tracevariog` command in the *geofd* package allows us to fit four covariance models: spherical, exponential, Gaussian, and Matérn. In our data, the exponential model is the best model in terms of minimizing the sum of squares errors. Figure 3 (right) shows the fitted model. Finally, ordinary kriging predictors can be obtained with the command `okfd` in the *geofd* package.

We have used daily wind speed to illustrate general results. Similarly, we can use monthly or yearly wind speed with $y(\mathbf{s}_i; v)$ representing the entire year, and repeat the procedure.

7.2 Surface data approach

For illustration purposes, we consider daily average data of the hourly wind speed, that is, $y_t(\mathbf{s}_i)$ represents the daily average wind speed on day t, for $t = 1, \ldots, 365$ in the year 2010.

To estimate the continuous surface $y_t(\mathbf{s})$ described in Section 5.3.1, we use the `smooth.FEM.basis` command in the *fdaPDE* package. It uses the finite elements analysis and penalizes with the Laplacian. Figure 4 shows one surface data for a specific day, $t = 1$. In the left panel, we plot the surface data observed on the 115×115 km region centered in Dumah Al Jandal, Saudi Arabia. In the right panel, we plot the estimated continuous surface. The estimated continuous functions $\{\hat{y}_1(\mathbf{s}), \ldots, \hat{y}_{365}(\mathbf{s})\}$ is the surface time series, and can be modeled as in (5.4) and (5.5). This approach allows us to successfully forecast the next day of surface wind data, $\hat{x}_{T+1}(\mathbf{s})$ and $\hat{y}_{T+1}(\mathbf{s})$, as well as understanding its temporal dependence.

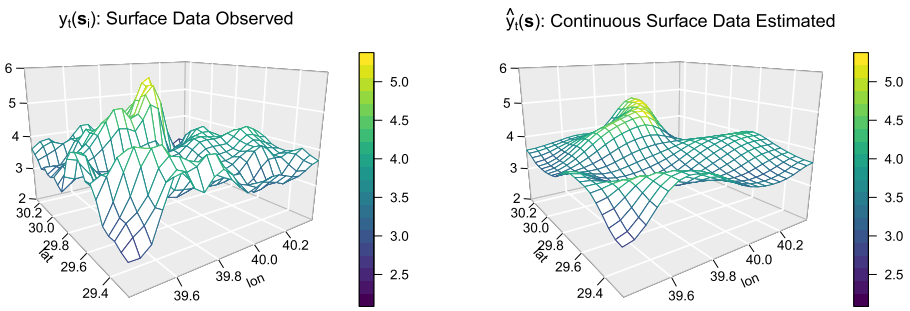

Figure 4 *Estimation of surface data for* $t = 1$. *Left*: *daily average of wind speed on a regular grid of points. Right*: *the continuous surface estimated with finite elements analysis.*

Here, we consider a rectangular region domain centered at Dumah Al Jandal, Saudi Arabia, but could also, similarly, consider the domain as being the entire kingdom of Saudi Arabia or the entire world, and study the statistical properties of the surface time series. In general, the approach of functional data can be used in more complex and large datasets.

8 Discussion

In this paper, we have provided an overview of functional data analysis in the case of spatially correlated data. We presented two main approaches, one when data are curves observed over space (spatial functional data), and the other when spatio-temporal random fields are considered as a surface (or manifold) time series. These two approaches present a new paradigm of data analysis, in which the continuous processes or random fields are considered as a single entity. Although software packages are still limited for statistical analysis, we believe that this mode of thinking can be valuable in the context of big data.

It is a welcomed fact that users can have different tools for data analysis, either with the surface time series approach or with classical spatio-temporal techniques. The choice will depend on the data, on the phenomena being studied, and on the scientific questions of interest. We appreciate all the effort and work done in both directions. We have attempted to collect all significant references, and we apologize for any unmentioned works.

Similarly, as functional data with temporal dependence arise, we can also consider spatio-temporal functional random fields (Bel et al., 2011, Lee, Zhu and Toscas, 2015) meaning that a functional time series is observed at each location. For example, $y_t(\mathbf{s}_i; v)$ can represent the functional time series of daily wind speed at location $\mathbf{s}_i$, where t is the day index, and v is the time within a day. Multivariate spatial functional data can also be considered, that is, at each location, we can observe different functional data, such as temperature, precipitation, and humidity.

We conclude that the functional approach opens new areas of research to develop methodologies and theories for the analysis of complex and large spatio-temporal datasets.

Acknowledgments

We thank Carolina Euan (KAUST), Joydeep Chowdhury (ISI), Laura Sangalli (PoliMi), Soumya Das (KAUST), and Zhuo Qu (KAUST) for comments on this paper. We are grateful to Professor Georgiy Stenchikov's group, the Atmospheric and Climate Modeling group at KAUST, for producing and providing the high-resolution WRF dataset.

This publication is based on research supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-2018-CRG7-3742.

References

Abdulah, S., Li, Y., Cao, J., Ltaief, H., Keyes, D. E., Genton, M. G. and Sun, Y. (2019). ExaGeoStatR: A package for large-scale geostatistics in R. Available at arXiv:1908.06936.

Aguilera-Morillo, M. C., Durbán, M. and Aguilera, A. M. (2017). Prediction of functional data with spatial dependence: A penalized approach. *Stochastic Environmental Research and Risk Assessment* **31**, 7–22.

Alfeld, P., Neamtu, M. and Schumaker, L. L. (1996). Fitting scattered data on sphere-like surfaces using spherical splines. *Journal of Computational and Applied Mathematics* **73**, 5–43. MR1424867 https://doi.org/10.1016/0377-0427(96)00034-9

Arnone, E., Azzimonti, L., Nobile, F. and Sangalli, L. M. (2019). Modeling spatially dependent functional data via regression with differential regularization. *Journal of Multivariate Analysis* **170**, 275–295. Special Issue on Functional Data Analysis and Related Topics. MR3913041 https://doi.org/10.1016/j.jmva.2018.09.006

Aston, J. A. D., Pigoli, D. and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics* **45**, 1431–1461. MR3670184 https://doi.org/10.1214/16-AOS1495

Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association* **110**, 378–392. MR3338510 https://doi.org/10.1080/01621459.2014.909317

Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M. and Nobile, F. (2015). Blood flow velocity field estimation via spatial regression with PDE penalization. *Journal of the American Statistical Association* **110**, 1057–1071. MR3420684 https://doi.org/10.1080/01621459.2014.946036

Baladandayuthapani, V., Mallick, B. K., Hong, M. Y., Lupton, J. R., Turner, N. D. and Carroll, R. J. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics* **64**, 64–73. MR2422820 https://doi.org/10.1111/j.1541-0420.2007.00846.x

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*. *Monographs on Statistics and Applied Probability.* **135**. Boca Raton, FL: CRC Press. MR3362184

Bel, L., Bar-Hen, A., Petit, R. and Cheddadi, R. (2011). Spatio-temporal functional regression on paleoecological data. *Journal of Applied Statistics* **38**, 695–704. MR2773575 https://doi.org/10.1080/02664760903563650

Bernardi, M. S., Sangalli, L. M., Mazza, G. and Ramsay, J. O. (2017). A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. *Stochastic Environmental Research and Risk Assessment* **31**, 23–38.

Bohorquez, M., Giraldo, R. and Mateu, J. (2017). Multivariate functional random fields: Prediction and optimal sampling. *Stochastic Environmental Research and Risk Assessment* **31**, 53–70.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. *Lecture Notes in Statistics* **149**. New York: Springer. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

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.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45**, 11–22. MR1718346 https://doi.org/10.1016/S0167-7152(99)00036-X

Chen, K., Zhang, X., Petersen, A. and Müller, H.-G. (2017). Quantifying infinite-dimensional data: Functional data analysis in action. *Statistics in Biosciences* **9**, 582–604.

Crainiceanu, C. M., Caffo, B. S., Luo, S., Zipunnikov, V. M. and Punjabi, N. M. (2011). Population value decomposition, a framework for the analysis of image populations. *Journal of the American Statistical Association* **106**, 775–790. MR2894733 https://doi.org/10.1198/jasa.2011.ap10089

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. *Wiley Series in Probability and Statistics*. Hoboken, NJ: John Wiley & Sons, Inc. MR2848400

Cressie, N. A. C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. New York: John Wiley & Sons, Inc. Paperback edition of the 1993 edition. MR3559472

Dassi, F., Ettinger, B., Perotto, S. and Sangalli, L. M. (2015). A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Applied Numerical Mathematics* **90**, 111–131. MR3300898 https://doi.org/10.1016/j.apnum.2014.10.007

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *EnvironMetrics* **21**, 224–239. MR2842240 https://doi.org/10.1002/env.1003

Diggle, P. J. and Ribeiro, P. J. Jr. (2007). *Model-Based Geostatistics*. *Springer Series in Statistics*. New York: Springer. MR2293378

Duchamp, T. and Stuetzle, W. (2003). Spline smoothing on surfaces. *Journal of Computational and Graphical Statistics* **12**, 354–381. MR1983159 https://doi.org/10.1198/1061860031743

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables* (W. Schempp and K. Zeller, eds.) 85–100. Berlin, Heidelberg: Springer. MR0493110

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121. MR1435485 https://doi.org/10.1214/ss/1038425655

Ettinger, B., Perotto, S. and Sangalli, L. M. (2016). Spatial regression models over two-dimensional manifolds. *Biometrika* **103**, 71–88. MR3465822 https://doi.org/10.1093/biomet/asv069

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. *Springer Series in Statistics*. New York: Springer. MR2229687

Finley, A. O., Banerjee, S. and Gelfand, A. E. (2015). spBayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software* **63**, 1–28.

Galeano, P. and Peña, D. (2019). Data science, big data and statistics. *TEST* **28**, 289–329. MR3962048 https://doi.org/10.1007/s11749-019-00651-9

Genton, M. G., Johnson, C., Potter, K., Stenchikov, G. and Sun, Y. (2014). Surface boxplots. *Stat* **3**, 1–11. MR4027321 https://doi.org/10.1002/sta4.39

Genton, M. G. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science* **30**, 147–163. MR3353096 https://doi.org/10.1214/14-STS487

Giraldo, R., Dabo-Niang, S. and Martínez, S. (2018). Statistical modeling of spatial big data: An approach from a functional data analysis perspective. *Statistics & Probability Letters* **136**, 126–129. MR3806852 https://doi.org/10.1016/j.spl.2018.02.025

Giraldo, R., Delicado, P. and Mateu, J. (2010). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 66–82. MR2755385 https://doi.org/10.1007/s13253-009-0012-z

Giraldo, R., Delicado, P. and Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics* **18**, 411–426. MR2832903 https://doi.org/10.1007/s10651-010-0143-y

Giraldo, R., Delicado, P. and Mateu, J. (2012). Hierarchical clustering of spatially correlated functional data. *Statistica Neerlandica* **66**, 403–421. MR2983302 https://doi.org/10.1111/j.1467-9574.2012.00522.x

Giraldo, R., Delicado, P. and Mateu, J. (2015). geofd: Spatial prediction for function value data. R package version 1.0.

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19**, 1327–1349. MR3102554 https://doi.org/10.3150/12-BEJSP06

Goulard, M. and Voltz, M. (1993) *Geostatistical Interpolation of Curves: A Case Study in Soil Science*, 805–816. Dordrecht: Springer.

Greco, F., Ventrucci, M. and Castelli, E. (2018). P-spline smoothing for spatial data collected worldwide. *Spatial Statistics* **27**, 1–17. MR3868196 https://doi.org/10.1016/j.spasta.2018.08.008

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**, 669–696. MR2976487 https://doi.org/10.1214/11-AOAS524

Grujic, O. and Menafoglio, A. (2017). fdagstat, an R package. R package version 1.0.

Grujic, O., Menafoglio, A., Yang, G. and Caers, J. (2018). Cokriging for multivariate Hilbert space valued random fields: Application to multi-fidelity computer code emulation. *Stochastic Environmental Research and Risk Assessment* **32**, 1955–1971.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Hall, P., Fisher, N. I. and Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *The Annals of Statistics* **22**, 2115–2134. MR1329185 https://doi.org/10.1214/aos/1176325774

Hall, P. and Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields* **99**, 399–424. MR1283119 https://doi.org/10.1007/BF01199899

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics* **38**, 1845–1884. MR2662361 https://doi.org/10.1214/09-AOS768

Hörmann, S. and Kokoszka, P. (2012). Chapter 7—functional time series. In *Time Series Analysis: Methods and Applications* (T. S. Rao, S. S. Rao and C. Rao, eds.), *Handbook of Statistics* **30**, 157–186. Amsterdam: Elsevier. MR3295420

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. New York: Springer. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **75**, 103–122. MR3008273 https://doi.org/10.1111/j.1467-9868.2012.01032.x

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics* **179**, 66–82. MR3153649 https://doi.org/10.1016/j.jeconom.2013.11.002

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* **51**, 4942–4956. MR2364551 https://doi.org/10.1016/j.csda.2006.07.028

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

Jiang, H. and Serban, N. (2012). Clustering random curves under spatial interdependence with application to service accessibility. *Technometrics* **54**, 108–119. MR2929427 https://doi.org/10.1080/00401706.2012.657106

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis* **34**, 116–129. MR3008019 https://doi.org/10.1111/j.1467-9892.2012.00816.x

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. *Texts in Statistical Science Series*. Boca Raton, FL: CRC Press. MR3793167

Lai, M.-J. and Schumaker, L. L. (2007). *Spline Functions on Triangulations*. *Encyclopedia of Mathematics and Its Applications.* Cambridge: Cambridge University Press. MR2355272 https://doi.org/10.1017/CBO9780511721588

Lee, D.-J., Zhu, Z. and Toscas, P. (2015). Spatio-temporal functional data analysis for wireless sensor networks data. *EnvironMetrics* **26**, 354–362. MR3366969 https://doi.org/10.1002/env.2344

Lila, E., Aston, J. A. D. and Sangalli, L. M. (2016). Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *Annals of Applied Statistics* **10**, 1854–1879. MR3592040 https://doi.org/10.1214/16-AOAS975

Lila, E., Sangalli, L. M., Ramsay, J. and Formaggia, L. (2019). fdaPDE: Functional data analysis and partial differential equations; statistical analysis of functional and spatial data, based on regression with partial differential regularizations. R package version 0.1-6.

Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **73**, 423–498. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

Liu, C., Ray, S. and Hooker, G. (2017). Functional principal component analysis of spatially correlated data. *Statistics and Computing* **27**, 1639–1654. MR3687330 https://doi.org/10.1007/s11222-016-9708-4

Martínez-Hernández, I. and Genton, M. G. (2020). Nonparametric trend estimation in functional time series with application to annual mortality rates. Available at arXiv:2001.04660.

Martínez-Hernández, I., Genton, M. G. and González-Farías, G. (2019). Robust depth-based estimation of the functional autoregressive model. *Computational Statistics & Data Analysis* **131**, 66–79. MR3906795 https://doi.org/10.1016/j.csda.2018.06.003

Mateu, J. and Romano, E. (2017). Advances in spatial functional statistics. *Stochastic Environmental Research and Risk Assessment* **31**, 1–6.

Menafoglio, A., Gaetani, G. and Secchi, P. (2018). Random domain decompositions for object-oriented kriging over complex domains. *Stochastic Environmental Research and Risk Assessment* **32**, 3421–3437.

Menafoglio, A., Grujic, O. and Caers, J. (2016). Universal kriging of functional data: Trace-variography vs cross-variography? Application to gas forecasting in unconventional shales. *Spatial Statistics* **15**, 39–55. MR3457667 https://doi.org/10.1016/j.spasta.2015.12.003

Menafoglio, A. and Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics. *European Journal of Operational Research* **258**, 401–410. MR3582352 https://doi.org/10.1016/j.ejor.2016.09.061

Menafoglio, A., Secchi, P. and Dalla Rosa, M. (2013). A universal kriging predictor for spatially dependent functional data of a Hilbert space. *Electronic Journal of Statistics* **7**, 2209–2240. MR3108813 https://doi.org/10.1214/13-EJS843

Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P. and Gutstein, H. (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *Annals of Applied Statistics* **5**, 894–923. MR2840180 https://doi.org/10.1214/10-AOAS407

Nerini, D., Monestiez, P. and Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis* **101**, 409–418. MR2564350 https://doi.org/10.1016/j.jmva.2009.03.005

Nychka, D., Furrer, R., Paige, J. and Sain, S. (2017). fields: Tools for spatial data. R package version 9.9.

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences* **30**, 683–691.

Qingguo, T. and Longsheng, C. (2010). B-spline estimation for spatial data. *Journal of Nonparametric Statistics* **22**, 197–217. MR2598962 https://doi.org/10.1080/10485250903272569

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2168993

Ramsay, J. O., Wickham, H., Graves, S. and Hooker, G. (2018). fda: Functional data analysis. R package version 2.4.8. MR3645102 https://doi.org/10.1007/978-1-4939-7190-9

Ramsay, T. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society, Series B* **64**, 307–319. MR1904707 https://doi.org/10.1111/1467-9868.00339

Rekabdarkolaee, H. M., Krut, C., Fuentes, M. and Reich, B. J. (2019). A Bayesian multivariate functional model with spatially varying coefficient approach for modeling hurricane track data. *Spatial Statistics* **29**, 351–365. MR3907916 https://doi.org/10.1016/j.spasta.2018.12.006

Reyes, A., Giraldo, R. and Mateu, J. (2015). Residual kriging for functional spatial prediction of salinity curves. *Communications in Statistics Theory and Methods* **44**, 798–809. MR3315788 https://doi.org/10.1080/03610926.2012.753087

Ribeiro, P. J. Jr. and Diggle, P. J. (2018). geoR: Analysis of geostatistical data. R package version 1.7-5.2.1.

Romano, E., Balzanella, A. and Verde, R. (2017). Spatial variability clustering for spatially dependent functional data. *Statistics and Computing* **27**, 645–658. MR3613590 https://doi.org/10.1007/s11222-016-9645-2

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* **71**, 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

Ruiz-Medina, M. (2011). Spatial autoregressive and moving average Hilbertian processes. *Journal of Multivariate Analysis* **102**, 292–305. MR2739116 https://doi.org/10.1016/j.jmva.2010.09.005

Ruiz-Medina, M. (2012). New challenges in spatial and spatiotemporal functional statistics for high-dimensional data. *Spatial Statistics* **1**, 82–91.

Ruiz-Medina, M. D., Salmerón, R. and Angulo, J. M. (2007). Kalman filtering from POP-based diagonalization of ARH(1). *Computational Statistics & Data Analysis* **51**, 4994–5008. MR2361366 https://doi.org/10.1016/j.csda.2006.07.013

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge: Cambridge University Press. MR1998720 https://doi.org/10.1017/CBO9780511755453

Sangalli, L. M. (2020). A novel approach to the analysis of spatial and functional data over complex domains. *Quality Engineering* **32**, 181–190. https://doi.org/10.1080/08982112.2019.1659357

Sangalli, L. M., Ramsay, J. O. and Ramsay, T. O. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society, Series B* **75**, 681–703. MR3091654 https://doi.org/10.1111/rssb.12009

Sartori, I. and Torriani, L. (2019). Manifoldgstat, an R package. R package.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. *Texts in Statistical Science Series*. London: Chapman & Hall/CRC. MR2134116

Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Auel, J., Menck, P. J., Gross, S., Ober, U., Ribeiro, P., Ripley, B. D., Singleton, R. and Pfaff, B. (R Core Team) (2019). RandomFields: Simulation and analysis of random fields. R package version 3.3.6.

Scott-Hayward, L. A. S., Mackenzie, M. L., Donovan, C. R., Walker, C. G. and Ashe, E. (2014). Complex region spatial smoother (CReSS). *Journal of Computational and Graphical Statistics* **23**, 340–360. MR3215814 https://doi.org/10.1080/10618600.2012.762920

Song, J. J. and Mallick, B. (2019). Hierarchical Bayesian models for predicting spatially correlated curves. *Statistics* **53**, 196–209. MR3900086 https://doi.org/10.1080/02331888.2018.1547905

Staicu, A.-M., Crainiceanu, C. M. and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics* **11**, 177–194.

Stein, M. L. (1999). *Interpolation of Spatial Data. Some Theory for Kriging*. *Springer Series in Statistics*. New York: Springer. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics* **20**, 316–334. MR2847798 https://doi.org/10.1198/jcgs.2011.09224

Sun, Y. and Genton, M. G. (2012). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *EnvironMetrics* **23**, 54–64. MR2873783 https://doi.org/10.1002/env.1136

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer. ISBN 0-387-95457-0. MR1337030 https://doi.org/10.1007/978-1-4899-2819-1

Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing* **2**, 5–16. MR0618629 https://doi.org/10.1137/0902002

Wang, H. and Ranalli, M. G. (2007). Low-rank smoothing splines on complicated domains. *Biometrics* **63**, 209–217. MR2345591 https://doi.org/10.1111/j.1541-0420.2006.00674.x

Wilhelm, M., Dedè, L., Sangalli, L. M. and Wilhelm, P. (2016). IGS: An IsoGeometric approach for smoothing on surfaces. *Computer Methods in Applied Mechanics and Engineering* **302**, 70–89. MR3461105 https://doi.org/10.1016/j.cma.2015.12.028

Cressie, N. A. C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. New York: John Wiley & Sons, Inc. Paperback edition of the 1993 edition. MR3559472

Wood, S. (2017). *Generalized Additive Models: An Introduction with R*, 2nd ed. London: Chapman & Hall. MR2206355

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* **62**, 1025–1036. MR2297673 https://doi.org/10.1111/j.1541-0420.2006.00574.x

Wood, S. N., Bravington, M. V. and Hedley, S. L. (2008). Soap film smoothing. *Journal of the Royal Statistical Society, Series B* **70**, 931–955. MR2530324 https://doi.org/10.1111/j.1467-9868.2008.00665.x

Xiao, L., Li, Y. and Ruppert, D. (2013). Fast bivariate P-splines: The sandwich smoother. *Journal of the Royal Statistical Society, Series B* **75**, 577–599. MR3065480 https://doi.org/10.1111/rssb.12007

Yip, C. M. A. (2018). Statistical characteristics and mapping of near-surface and elevated wind resources in the Middle East. Ph.D. Thesis, King Abdullah University of Science and Technology.

Yue, Y. and Speckman, P. L. (2010). Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics* **19**, 96–116. MR2654402 https://doi.org/10.1198/jcgs.2009.08124

Zhang, J., Clayton, M. K. and Townsend, P. A. (2011). Functional concurrent linear regression model for spatial images. *Journal of Agricultural, Biological, and Environmental Statistics* **16**, 105–130. MR2767762 https://doi.org/10.1007/s13253-010-0047-1

Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K. A., Majewski, T., Czerniak, B. A. and Morris, J. S. (2016). Functional CAR models for large spatially correlated functional datasets. *Journal of the American Statistical Association* **111**, 772–786. MR3538704 https://doi.org/10.1080/01621459.2015.1042581

Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V. and Carroll, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association* **105**, 390–400. MR2757206 https://doi.org/10.1198/jasa.2010.tm08737

Statistics Program
King Abdullah University of Science and Technology
Thuwal 23955-6900
Saudi Arabia
E-mail: israel.martinezhernandez@kaust.edu.sa
marc.genton@kaust.edu.sa