# Kriging for Hilbert-space valued random fields: the Operatorial point of view

Alessandra Menafoglio[a,*], Giovanni Petris[b]

[a] *MOX, Department of Mathematics, Politecnico di Milano, Italy*
[b] *Department of Mathematical Sciences, University of Arkansas, USA*

**Abstract**

We develop a comprehensive framework for linear spatial prediction in Hilbert spaces. We explore the problem of Best Linear Unbiased (BLU) prediction in Hilbert spaces through an original point of view, based on a new Operatorial definition of Kriging. We ground our developments on the theory of Gaussian processes in function spaces and on the associated notion of measurable linear transformation. We prove that our new setting allows (a) to derive an explicit solution to the problem of Operatorial Ordinary Kriging, and (b) to establish the relation of our novel predictor with the key concept of conditional expectation of a Gaussian measure. Our new theory is posed as a unifying theory for Kriging, which is shown to include the Kriging predictors proposed in the literature on Functional Data through the notion of finite-dimensional approximations. Our original viewpoint to Kriging offers new relevant insights for the geostatistical analysis of either finite- or infinite-dimensional georeferenced dataset.

*Keywords:* Geostatistics, Gaussian Processes, conditional expectations, measurable linear transformations

## 1. Introduction

In recent years, the increasing availability of complex and high-dimensional data has motivated a fast and extensive growth of Functional Data Analysis (FDA, e.g., Ramsay and Silverman, 2005) and Object Oriented Data Analysis (OODA, e.g., Marron and Alonso, 2014, and references therein). These new branches of statistics share the same abstract approach in interpreting each datum as a realization of a random element in a finite- or infinite-dimensional space. Properties of the space to which data are assumed to belong directly reflect on the methodologies that one can employ for the statistical analysis. For instance, the geometry of a Hilbert space allows for a class of methods based on the notions of inner product and norm (e.g., Bosq, 2000, and references therein),

*Corresponding author
*Email addresses:* `alessandra.menafoglio@polimi.it` (Alessandra Menafoglio), `GPetris@uark.edu` (Giovanni Petris)

whereas methods suitable for data in general metric spaces need to rely on the notion of distance only.

A rich body of literature has been devoted to the statistical analysis of functional data. Most works in this field rely upon the embedding of the data into a Hilbert space, particularly $L^2$, to develop parametric or non-parametric methods for their treatment (e.g., Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012). The Hilbert space embedding allows for the generalization to the functional setting of several well-known multivariate methods, such as principal component analysis (e.g., Ramsay and Silverman, 2005), K-mean clustering (e.g., Tarpey and Kinateder, 2003; Sangalli et al., 2010), or hypothesis testing (e.g., via $T^2$-Hotelling statistics, Pini et al., 2015). However, new issues emerged with the advent of FDA, such as the problem of data smoothing (e.g., Ramsay and Silverman, 2005) or curve alignment (i.e., registration, e.g., Vantini, 2012; Srivastava et al., 2011, and references therein). For an overview on FDA and its most recent advances we refer to (Cuevas, 2014) and (Bongiorno et al., 2014).

In this framework, a relatively large body of literature addresses the problem of the geostatistical characterization and prediction of spatially dependent functional data. Early works in this field focused on $L^2$ data to develop linear spatial predictors (i.e., Kriging predictors) in the form of optimal linear combinations of the data (e.g., Delicado et al., 2010; Giraldo et al., 2011; Caballero et al., 2013). Even though the $L^2$ embedding is commonly employed in FDA, several environmental applications deal with constrained or manifold data, for which the $L^2$ geometry may be inappropriate. For instance, Menafoglio et al. (2014a,b) deal with a set of constrained functional data in the form of particle-size densities, i.e., probability density functions describing the distribution of grains sizes within a given soil sample. In this case, the usual $L^2$ geometry is not appropriate, as it completely neglects the data constraints (see, e.g., Delicado, 2007, 2011).

These elements motivate the adoption of an abstract viewpoint, along the line of OODA. In this setting, Menafoglio et al. (2013) establish a Kriging theory for random fields valued in any separable Hilbert space, allowing for the analysis of a broad range of object data, such as curves, surfaces or images. The present work stands in continuity with the approach of Menafoglio et al. (2013), with whom we share the geometric viewpoint to the treatment of either finite- or infinite-dimensional data as *atoms* of the geostatistical analysis. However, we here explore the problem of linear spatial prediction in Hilbert spaces through an original point of view, based on a new operatorial definition of Kriging. In this setting, the theory of Operatorial Kriging is posed as a unifying framework for Kriging, with the scope of including either the formulations of Kriging for curves in $L^2$ (e.g., Delicado et al., 2010; Nerini et al., 2010) or that for Hilbert data (Menafoglio et al., 2013).

The remaining part of this work is organized as follow. Section 2 introduces the problem and highlights the main contributions of this work. Section 3 recalls the theory of Gaussian measures on Hilbert spaces, upon which we ground the developments of Section 4 and 5. Section 6 investigates discretizations of the Operatorial Kriging predictor, and the relation of our new theory with the existing literature works of Nerini et al. (2010); Menafoglio et al. (2013). Section 7 provides a discussion on the impact of our results from the application viewpoint and Section 8 concludes the work.

## 2. Kriging for Hilbert data: state of the art and main contributions

We denote by $D$ a $d$-dimensional spatial domain, and by $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ the locations of the available data $x_{\boldsymbol{s}_1}, ..., x_{\boldsymbol{s}_n}$. As in classical geostatistics, we assume that the latter are a partial observation of a random field $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$. Throughout this work, we assume that $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ is valued in a separable Hilbert-space $\mathcal{H}$, and that it is Gaussian and stationary (in the sense that will be clarified in Sections 3-5). Our aim is the prediction of the element $\mathcal{X}_{\boldsymbol{s}_0}$ at an unobserved location $\boldsymbol{s}_0$ in $D$.

In this setting, if the Hilbert space $\mathcal{H}$ was the one-dimensional Euclidean space $\mathbb{R}$, classical geostatistics literature would advocate the use of a Kriging predictor, that is the Best Linear Unbiased Predictor (BLUP) $\mathcal{X}^*_{\boldsymbol{s}_0} = \sum_{i=1}^n \lambda_i \mathcal{X}_{\boldsymbol{s}_i}$, whose weights minimize the variance of prediction error under the unbiasedness constraint (e.g., Cressie, 1993). This can also be interpreted — in the Gaussian setting — in terms of the conditional expectation of $\mathcal{X}_{\boldsymbol{s}_0}$ given $\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n}$.

We note that, in the scalar case, the notion of linear predictor is equivalently understood either as a linear combination of the observations or as a linear transformation of the vector of observations, i.e., any linear transformation applied to the vector of observations $(\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n})^T \in \mathbb{R}^n$ and valued in $\mathbb{R}$ acts as a linear combination of $\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n}$. Instead, when $\mathcal{H}$ is an infinite-dimensional Hilbert space, an ambiguity exists in the definition of a Kriging predictor. For instance, Giraldo et al. (2011) and Menafoglio et al. (2013) interpret the Kriging problem in terms of finding the BLUP among the predictors of the form

$$\mathcal{X}^{\boldsymbol{\lambda}}_{\boldsymbol{s}_0} = \sum_{i=1}^n \lambda_i \mathcal{X}_{\boldsymbol{s}_i}, \tag{1}$$

with $\lambda_i$ scalar weights in $\mathbb{R}$, for $i = 1, ..., n$.

Despite its simplicity, predictor (1) does not provide, in general, the best linear unbiased transformation of the vector of observations, that is the *Operatorial* Kriging predictor $\mathcal{X}^{\Lambda}_{\boldsymbol{s}_0} = \Lambda(\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n})$, for some linear operator $\Lambda : \mathcal{H} \times ... \times \mathcal{H} \to \mathcal{H}$. The operatorial viewpoint has been first considered by Nerini et al. (2010) in Reproducing Kernel Hilbert Spaces (RKHSs). These authors address the problem of finding the best predictor over the class of linear unbiased Hilbert-Schmidt transformations of the observations, i.e., of the form

$$\mathcal{X}^{\boldsymbol{B}}_{\boldsymbol{s}_0} = \sum_{i=1}^n B_i \mathcal{X}_{\boldsymbol{s}_i}, \tag{2}$$

where $B_i : \mathcal{H} \to \mathcal{H}$ are Hilbert-Schmidt linear operators and $\mathcal{X}_{\boldsymbol{s}_i}$ observations in a RKHS. Even though this class of predictors is more general than that of (1), the RKHS-embedding — which is key to the well-posedness of the problem — still appears a too restrictive setting, as, for instance, the Hilbert space $L^2$ is not a RKHS, even though it is commonly employed in FDA.

In this work we establish an Operatorial Kriging theory valid for any separable Hilbert-space, which relies upon the key notion of measurable linear transformation associated with a Gaussian measure (Mandelbaum, 1984; Luschgy, 1996) (Section 3). This broad class of operators includes linear Hilbert-Schmidt operators, and is here shown to allow for the Operatorial Kriging prediction in any finite- or infinite-dimensional separable Hilbert space.

As a first key result, Theorem 4 states that the Operatorial Kriging problem is well-posed in our setting, and provides an explicit expression for the Operatorial Kriging predictor, i.e., for

$$\mathcal{X}^*_{\boldsymbol{s}_0} = \Lambda^*(\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n}), \tag{3}$$

where $\Lambda^* : \mathcal{H} \times ... \times \mathcal{H} \to \mathcal{H}$ is the optimal measurable linear transformation associated with the Gaussian law of the vector of observations (see Section 5). Unlike previous settings, our new framework allows to interpret the Operatorial Kriging theory in terms of conditional expectation of $\mathcal{X}_{\boldsymbol{s}_0}$ given $\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n}$ (see Sections 4 and 5), similarly as in the scalar setting.

As a further key result, we will show that the existing formulations of Kriging are included in our new framework, when considering discretizations of the Operatorial Kriging problem. This will provide new relevant insights on the theory of Kriging, at the intersection between the fields of FDA, multivariate and high-dimensional statistics. In particular, we will prove that no ambiguity exists in the Kriging formulation as long as the dimension of the employed discretization is non-greater than the number $n$ of observations, with important implications in the field of Kriging large datasets.

## 3. Gaussian measures on Hilbert spaces

In this Section, we recall some preliminaries on Gaussian measures in Hilbert spaces and set the notation that will be used hereafter. We refer the reader to Bogachev (1998); Da Prato and Zabczyk (2014) for a deep dissertation on the topic.

We denote with the symbol $\mathcal{H}$ (or $\mathcal{H}_1$, $\mathcal{H}_2$) a real separable Hilbert space with norm $\|\cdot\|_{\mathcal{H}}$ and inner product $\langle \cdot, \cdot, \rangle_{\mathcal{H}}$, equipped with its Borel $\sigma$-algebra $\mathfrak{B}(\mathcal{H})$. We call $\mathcal{L}(\mathcal{H}, \mathcal{H}_1)$ the Banach space of continuous linear operators on $\mathcal{H}$ in $\mathcal{H}_1$. Further, we denote with $\mathcal{H}^\star$ the dual of $\mathcal{H}$, i.e., the space $\mathcal{L}(\mathcal{H}, \mathbb{R})$ of linear and continuous functional on $\mathcal{H}$, which is identified with $\mathcal{H}^\star$ via Riesz representation theorem. Given an operator $A$ in $\mathcal{L}(\mathcal{H}, \mathcal{H}_1)$, we denote by $A' \in \mathcal{L}(\mathcal{H}_1, \mathcal{H})$ its adjoint.

Given a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, a $\mathcal{H}$-valued random variable $\mathcal{X}$ is a measurable function on $(\Omega, \mathfrak{F}, \mathbb{P})$ in $(\mathcal{H}, \mathfrak{B}(\mathcal{H}))$, $\mathcal{X} : (\Omega, \mathfrak{F}) \to (\mathcal{H}, \mathfrak{B}(\mathcal{H}))$. We denote by $\mu_{\mathcal{X}}$ the law of $\mathcal{X}$, i.e., the probability measure on $(\mathcal{H}, \mathfrak{B}(\mathcal{H}))$ defined, for $\mathcal{A} \in \mathfrak{B}(\mathcal{H})$, as $\mu_{\mathcal{X}}(\mathcal{A}) = \mathbb{P}(\mathcal{X}^{-1}(\mathcal{A}))$.

Given a $\mathcal{H}$-valued random variable $\mathcal{X}$, we will always assume that $\mathbb{E}[\|\mathcal{X}\|_{\mathcal{H}}^2] < \infty$. In this setting, we define the expected value of $\mathcal{X}$ as

$$m_{\mathcal{X}} = \mathbb{E}[\mathcal{X}] = \int_{\mathcal{H}} x \mu_{\mathcal{X}}(dx),$$

where the integral is interpreted as a Bochner integral. In particular, for any $A \in \mathcal{L}(\mathcal{H}, \mathcal{H}_1)$, one has $\mathbb{E}[A(\mathcal{X})] = A\mathbb{E}[\mathcal{X}]$. Moreover, the covariance operator $C_{\mathcal{X}} : \mathcal{H} \to \mathcal{H}$ is defined, for every $x \in \mathcal{H}$, as

$$C_{\mathcal{X}} x = \mathbb{E}[\langle(\mathcal{X} - m_{\mathcal{X}}), x\rangle_{\mathcal{H}} (\mathcal{X} - m_{\mathcal{X}})].$$

A covariance operator is symmetric and positive definite. If $\mathcal{X}_1$ and $\mathcal{X}_2$ are $\mathcal{H}_1$- and $\mathcal{H}_2$-valued random variables, the cross-covariance operator $C_{\mathcal{X}_1 \mathcal{X}_2} \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ is defined as

$$C_{\mathcal{X}_1 \mathcal{X}_2} x_2 = \mathbb{E}[\langle \mathcal{X}_2 - m_{\mathcal{X}_2}, x_2 \rangle_{\mathcal{H}} (\mathcal{X}_1 - m_{\mathcal{X}_1})],$$

for every $x_2 \in \mathcal{H}_2$.

We say that a $\mathcal{H}$-valued random variable $\mathcal{X}$, with expected value $m_{\mathcal{X}}$ and covariance operator $C_{\mathcal{X}}$, has a Gaussian distribution — and we write $\mu_{\mathcal{X}} = N(m_{\mathcal{X}}, C_{\mathcal{X}})$ — if $\langle x, \mathcal{X}\rangle_{\mathcal{H}}$ has a Gaussian distribution for every $x \in \mathcal{H}$.

It is possible to associate to a given Gaussian measure $\mu_{\mathcal{X}}$ on a separable Hilbert space, another Hilbert space $\mathcal{H}_{\mu_{\mathcal{X}}} \subset \mathcal{H}$, which is called the Cameron-Martin space of $\mu_{\mathcal{X}}$ (Bogachev, 1998). The Cameron-Martin space coincides with the image of the operator $C_{\mathcal{X}}^{1/2}$.

We finally introduce the notion of measurable linear transformation (Luschgy, 1996) with respect to a given probability measure $\mu_{\mathcal{X}}$.

**Definition 1.** *(Mandelbaum (1984); Luschgy (1996)) A Borel measurable map $L : \mathcal{H}_2 \to \mathcal{H}_1$ is said to be a measurable linear transformation with respect to $\mu_{\mathcal{X}}$ ($\mu_{\mathcal{X}}$-mlt) if $L$ is linear on a subspace $\mathcal{D}_L \in \mathfrak{B}(\mathcal{H}_2)$ with $\mu_{\mathcal{X}}(\mathcal{D}_L) = 1$. A measurable linear transformation $L : \mathcal{H}_2 \to \mathbb{R}$ is called measurable linear functional ($\mu_{\mathcal{X}}$-mlf).*

In the following, we focus on measurable linear transformations with respect to Gaussian measures associated with injective covariance operators. In this case, the following result holds.

**Theorem 1 (Mandelbaum (1984)).** *(i) Let $L : \mathcal{H}_2 \to \mathcal{H}_1$ be $\mu_{\mathcal{X}}$-mlt, where $\mu_{\mathcal{X}} = N(m_{\mathcal{X}}, C_{\mathcal{X}})$ on $\mathcal{H}_2$. Then $L$ is linear on $\mathcal{H}_{\mu_{\mathcal{X}}}$ and the operator*

$$T = LC_{\mathcal{X}}^{1/2} : \mathcal{H}_2 \to \mathcal{H}_1 \tag{4}$$

*is Hilbert-Schmidt.*

*(ii) Let $T : \mathcal{H}_2 \to \mathcal{H}_1$ be Hilbert-Schmidt. Then there exists a unique (up to*

*$\mu_{\mathcal{X}}$-equivalence) $\mu_{\mathcal{X}}$-mlt $L : \mathcal{H}_2 \to \mathcal{H}_1$ such that*

$$L = TC_{\mathcal{X}}^{-1/2} \text{ on } \mathcal{H}_{\mu_{\mathcal{X}}}. \tag{5}$$

*(iii) In both* (4) *and* (5) *, the Hilbert-Schmidt norm of $T$ is equal to*

$$\|L\|_{\mu_{\mathcal{X}}}^2 = \int_{\mathcal{H}_2} \|Lx\|_{\mathcal{H}_1}^2 \mu_{\mathcal{X}}(dx). \tag{6}$$

Finally, the following Corollary of Theorem 1 will be useful in the following.

**Corollary 2 (Mandelbaum (1984)).** *The space $\mathcal{M}_{\mathcal{X}}$ of $\mu_{\mathcal{X}}$-mlt on $\mathcal{H}_2$ in $\mathcal{H}_1$ is a Hilbert space with the norm* (6)*. It is isometric to the space of Hilbert-Schmidt operators via the correspondence* (4) *and* (5)*.*

## 4. Spatial prediction in Hilbert Spaces via conditional expectations

In this Section we address the problem of spatial prediction in the presence of a partial observation of a Gaussian random field with known mean. We consider a $\mathcal{H}$-valued random field $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$, i.e., a collection of random variables on $(\Omega, \mathfrak{F}, \mathbb{P})$ in $\mathcal{H}$, indexed by a continuous spatial variable $\boldsymbol{s} \in D$. We here focus on Gaussian random fields. These are characterized by having all the finite dimensional laws Gaussian, i.e.,

$$\forall N > 0,\ \boldsymbol{s}_1, ..., \boldsymbol{s}_N \in D, \quad \boldsymbol{\mathcal{X}} = (\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_N}) \sim N(m_{\boldsymbol{\mathcal{X}}}, C_{\boldsymbol{\mathcal{X}}}).$$

Given $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ in $D$ and the observation of the random field $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ at these locations, $\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n}$, we aim to predict the unobserved element $\mathcal{X}_{\boldsymbol{s}_0}$ at the location $\boldsymbol{s}_0$. To ease the notation, hereafter in this Section we assume the mean function $m_{\mathcal{X}_{\boldsymbol{s}}} = \mathbb{E}[\mathcal{X}_{\boldsymbol{s}}]$ to be zero over the entire domain $D$.

We call $\mathcal{H}^n$ the Hilbert space $\mathcal{H} \times ... \times \mathcal{H}$, with the inner product $\langle x, y\rangle_{\mathcal{H}^n} = \sum_{i=1}^n \langle x_i, y_i\rangle_{\mathcal{H}}$, and $C_{\boldsymbol{\mathcal{X}}} \in \mathcal{L}(\mathcal{H}^n, \mathcal{B}^n)$ the covariance operator of $(\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n})^T \in \mathcal{H}^n$, which is defined, for $x = (x_1, ..., x_n)^T \in \mathcal{H}^n$, as

$$C_{\boldsymbol{\mathcal{X}}} x = \left( \sum_{j=1}^n C(\boldsymbol{s}_i, \boldsymbol{s}_j) x_j \right)_{i=1,\ldots,n}$$

where $C : D \times D \to \mathcal{L}(\mathcal{H}, \mathcal{H})$ is the Gaussian covariance function

$$\begin{aligned} C : D \times D &\to \mathcal{L}(\mathcal{H}, \mathcal{H}) \\ (\boldsymbol{s}_i, \boldsymbol{s}_j) &\mapsto \{C(\boldsymbol{s}_i, \boldsymbol{s}_j) : \mathcal{H} \to \mathcal{H}, \mathcal{H} \ni x \mapsto \mathbb{E}[\langle(\mathcal{X}_{\boldsymbol{s}_i} - m_{\mathcal{X}_{\boldsymbol{s}_i}}), x\rangle_{\mathcal{H}}(\mathcal{X}_{\boldsymbol{s}_j} - m_{\mathcal{X}_{\boldsymbol{s}_j}})]\}. \end{aligned}$$

Mandelbaum (1984) considers the problem of predicting a random element in a separable Hilbert space, given another random element in the same space, based on their joint (Gaussian) distribution. This author shows that the conditional expectation of the former given the latter is a measurable linear transformation and further derives the associated Hilbert-Schmidt operator. Luschgy (1996) considers a twofold generalization of the result of Mandelbaum (1984): (a) Banach-space valued Gaussian random elements are considered, and (b) the conditioning variable is allowed to be valued in a different space than the element to be predicted. For the purpose of our study, we here illustrate the general result of Luschgy (1996), embedded into the Hilbert space setting.

Hereafter we denote with $\mu_{\mathcal{Z}} = N(m_{\mathcal{Z}}, C_{\mathcal{Z}})$ the law of a random element $\mathcal{Z}$ in $\mathcal{H}_1$, with expected value $m_{\mathcal{Z}}$ and covariance operator $C_{\mathcal{Z}}$. Analogous notation is kept for the random element $\mathcal{Y}$ in $\mathcal{H}_2$. We call $C_{\mathcal{Y}\mathcal{Z}}$ the cross-covariance operator between $\mathcal{Y}$ and $\mathcal{Z}$. The following Theorem recalls the main result of Luschgy (1996) for the case of an injective covariance operator $C_{\mathcal{Z}}$.

**Theorem 3 (Luschgy (1996)).** *Let $\mathcal{Y}$ and $\mathcal{Z}$ be jointly Gaussian random vectors in $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. Assume that $m_{\mathcal{Y}} = m_{\mathcal{Z}} = 0$. Then*

$$\mathbb{E}[\mathcal{Y}|\mathcal{Z}] = L\mathcal{Z},$$

*where $L : \mathcal{H}_2 \to \mathcal{H}_1$ is the $\mu_{\mathcal{Z}}$-mlt*

$$L = TC_{\mathcal{Z}}^{-1/2}$$

*associated with the Hilbert-Schmidt operator $T : \mathcal{H}_2 \to \mathcal{H}_1$*

$$T = C_{\mathcal{Y}\mathcal{Z}} C_{\mathcal{Z}}^{-1/2}.$$

In our setting the result of Luschgy (1996) applies when interpreting the previous notation as follow. The random element $\mathcal{Z}$ is interpreted as the random vector $\boldsymbol{\mathcal{X}}$ on $\mathcal{H}^n$, with law $\mu_{\boldsymbol{\mathcal{X}}} = N(\mathbf{0}, C_{\boldsymbol{\mathcal{X}}})$. Hereafter, we assume $C_{\boldsymbol{\mathcal{X}}}$ to be invertible. The random element $\mathcal{Y}$ to be predicted is in our context $\mathcal{X}_{\boldsymbol{s}_0}$, and the cross-covariance operator $C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} \in \mathcal{L}(\mathcal{H}^n, \mathcal{H})$ between $\boldsymbol{\mathcal{X}}$ and $\mathcal{X}_{\boldsymbol{s}_0}$ is defined, for $x = (x_1, ..., x_n)^T \in \mathcal{H}^n$ as

$$C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} x = \sum_{j=1}^{n} C(\boldsymbol{s}_0, \boldsymbol{s}_j) x_j.$$

Therefore, the conditional expectation of $\mathcal{X}_{\boldsymbol{s}_0}$ given $\boldsymbol{\mathcal{X}}$ is obtained as the $\mu_{\boldsymbol{\mathcal{X}}}$-mlt $L : \mathcal{H}^n \to \mathcal{H}$

$$\mathbb{E}[\mathcal{X}_{\boldsymbol{s}_0}|\boldsymbol{\mathcal{X}}] = L\boldsymbol{\mathcal{X}} \tag{7}$$

with $L = TC_{\boldsymbol{\mathcal{X}}}^{-1/2}$ and

$$T = C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} C_{\boldsymbol{\mathcal{X}}}^{-1/2}.$$

We note that in case of a Gaussian random field with nonzero mean $m_{\mathcal{X}_{\boldsymbol{s}}}$, (7) reads

$$\mathbb{E}[\mathcal{X}_{\boldsymbol{s}_0}|\boldsymbol{\mathcal{X}}] = m_{\mathcal{X}_{\boldsymbol{s}_0}} + L(\boldsymbol{\mathcal{X}} - m_{\boldsymbol{\mathcal{X}}}), \tag{8}$$

with $m_{\boldsymbol{\mathcal{X}}} = (m_{\mathcal{X}_{\boldsymbol{s}_1}}, ..., m_{\mathcal{X}_{\boldsymbol{s}_n}})^T$.

We remark that the conditional expectation is an unbiased predictor and minimizes the mean squared prediction error $\mathbb{E}[\|\mathcal{X}_{\boldsymbol{s}_0} - f(\boldsymbol{\mathcal{X}})\|^2_{\mathcal{H}}]$, among all the measurable functions $f : \mathcal{H}^n \to \mathcal{H}$ (e.g., Luschgy, 1996). Therefore, for a Gaussian random field, the best spatial predictor — in the mean squared norm sense — coincides with the BLUP (i.e., the Simple Kriging predictor), if this is interpreted as the $\mu_{\boldsymbol{\mathcal{X}}}$-measurable linear transformation minimizing the mean squared prediction error. In this sense, similar to the finite-dimensional setting, the conditional expectation $\mathbb{E}[\mathcal{X}_{\boldsymbol{s}_0}|\boldsymbol{\mathcal{X}}]$ solves the following Simple Kriging problem in $\mathcal{H}$.

**Problem 1 (Operatorial Simple Kriging).** *Given $\boldsymbol{\mathcal{X}} = (\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n})^T$ and with the previous notation, find the BLUP for $\mathcal{X}_{\boldsymbol{s}_0}$, i.e., $\mathcal{X}^*_{\boldsymbol{s}_0} = \Lambda^* \boldsymbol{\mathcal{X}}$, where $\Lambda^* : \mathcal{H}^n \to \mathcal{H}$ is the $\mu_{\boldsymbol{\mathcal{X}}}$-mlt minimizing*

$$\mathbb{E}[\|\mathcal{X}_{\boldsymbol{s}_0} - \mathcal{X}^*_{\boldsymbol{s}_0}\|^2_{\mathcal{H}}] \quad \textit{subject to} \quad \mathbb{E}[\mathcal{X}^*_{\boldsymbol{s}_0}] = \mathbb{E}[\mathcal{X}_{\boldsymbol{s}_0}].$$

These observations motivates the introduction of a new Operatorial Ordinary Kriging formulation, which is addressed in Section 5 for stationary random fields.

## 5. An Operatorial Ordinary Kriging predictor for Hilbert-space valued random fields

In most real applications, the mean function of the random field which is partially observed is actually unknown. This renders the founding hypothesis of Simple Kriging too restrictive. In this Section, we address the problem of spatial prediction for Gaussian random fields with unknown mean, and we focus on the case of stationary processes.

Let $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ be a Gaussian random field on $(\Omega, \mathfrak{F}, P)$ in $\mathcal{H}$, with (unknown) mean function $m_{\mathcal{X}_{\boldsymbol{s}}}$ and Gaussian covariance function $C$. We assume that process $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ is strictly stationary, i.e.,

(i) $\mathbb{E}[\mathcal{X}_{\boldsymbol{s}}] = m$ for any $\boldsymbol{s} \in D$ (spatially constant mean);

(ii) $\mathbb{E}[\langle \mathcal{X}_{\boldsymbol{s}_i} - m, x\rangle_{\mathcal{H}}(\mathcal{X}_{\boldsymbol{s}_j} - m)] = C(\boldsymbol{h})$, for any $\boldsymbol{s}_i, \boldsymbol{s}_j \in D$, $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j$ (Gaussian covariance function depending only on the increment vector $\boldsymbol{h}$).

Given a set of locations $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ and the observation of the process at these locations, $\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n}$, we aim to predict $\mathcal{X}_{\boldsymbol{s}_0}$ via the Operatorial Ordinary Kriging predictor, i.e., to solve the following Problem.

**Problem 2 (Operatorial Ordinary Kriging).** *Given $\boldsymbol{\mathcal{X}} = (\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n})^T$ and with the previous notation, find the BLUP for $\mathcal{X}_{\boldsymbol{s}_0}$, i.e., $\mathcal{X}^*_{\boldsymbol{s}_0} = \Lambda^* \boldsymbol{\mathcal{X}}$, where $\Lambda^* : \mathcal{H}^n \to \mathcal{H}$ is a $\mu_{\boldsymbol{\mathcal{X}}}$-mlt and minimizes*

$$\mathbb{E}[\|\mathcal{X}_{\boldsymbol{s}_0} - \mathcal{X}^{\Lambda}_{\boldsymbol{s}_0}\|^2_{\mathcal{H}}] \quad \textit{subject to} \quad \mathbb{E}[\mathcal{X}^{\Lambda}_{\boldsymbol{s}_0}] = m,$$

*where $\mathcal{X}^{\Lambda}_{\boldsymbol{s}_0} = \Lambda \boldsymbol{\mathcal{X}}$, with $\Lambda : \mathcal{H}^n \to \mathcal{H}$ a $\mu_{\boldsymbol{\mathcal{X}}}$-mlt.*

A similar problem is addressed by Nerini et al. (2010) for the particular case of RKHSs, with focus on linear predictors of the kind (2). These authors derive Kriging equations, and provide an explicit solution for random processes valued in a $K$-dimensional $L^2$ space ($K < \infty$). Similar results are obtained by Giraldo (2009). Nevertheless, Nerini et al. (2010) acknowledge that the RKHS setting is quite restrictive, as the elements of a RKHS need to be sufficiently regular and, for instance, the Hilbert space $L^2$ is not a RKHS, even though it is commonly employed in FDA. Moreover, we note that the solution of the Simple Kriging problem in an infinite-dimensional Hilbert space generally is not a Hilbert-Schmidt linear operator, but a $\mu_{\boldsymbol{\mathcal{X}}}$-mlt. Therefore, a predictor of the form (2) cannot be, in general, the solution of Problem 2, even if it is for $\mathcal{H}$ RKHS. In the following paragraphs we show that Problem 2 is well-posed, instead.

To this end, we introduce the following notation. We call $\mu_{\boldsymbol{\mathcal{X}}_0} = N\left(\boldsymbol{m}_{\boldsymbol{\mathcal{X}}_0}, C_{\boldsymbol{\mathcal{X}}_0}\right)$ the law of the random vector $\boldsymbol{\mathcal{X}}_0 = \left(\mathcal{X}_{\boldsymbol{s}_0}, \boldsymbol{\mathcal{X}}^T\right)^T$ in $\mathcal{H}^{n+1}$, with expected value $\boldsymbol{m}_{\boldsymbol{\mathcal{X}}_0} = \left(m, (1\, m)^T\right)^T$ and covariance operator $C_{\boldsymbol{\mathcal{X}}_0} : \mathcal{H}^{n+1} \to \mathcal{H}^{n+1}$. The latter can be expressed in block form as

$$C_{\boldsymbol{\mathcal{X}}_0} = \begin{pmatrix} C_{\mathcal{X}_{\boldsymbol{s}_0}} & C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} \\ C_{\boldsymbol{\mathcal{X}}\mathcal{X}_{\boldsymbol{s}_0}} & C_{\boldsymbol{\mathcal{X}}} \end{pmatrix}.$$

The following Theorem states the first key result of this work.

**Theorem 4.** *Under the previous assumptions and notation, Problem 2 admits a unique solution $\mathcal{X}^*_{\boldsymbol{s}_0} = \Lambda^* \boldsymbol{\mathcal{X}}$, where $\Lambda^*$ is the $\mu_{\boldsymbol{\mathcal{X}}}$-mlt solving*

$$\begin{cases} \Lambda C_{\boldsymbol{\mathcal{X}}} - C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} + \zeta_1\, 1' = 0; \\ \Lambda\, 1 - I = 0, \end{cases} \tag{9}$$

*with $1\, x = (x, x, ..., x)^T$, for $x \in \mathcal{H}$, $I : \mathcal{H} \to \mathcal{H}$ the identity operator and $\zeta_1$ a $\mu_{\mathcal{X}_0}$-mlt and represents the Lagrange multiplier associated with the unbiasedness constraint. Moreover, for $x \in \mathcal{H}^n$, one has*

$$\Lambda^* x = M^* x + L(x - 1\, M^* x), \tag{10}$$

*where $M^*$ is the $\mu_{\boldsymbol{\mathcal{X}}}$-mlt defined, for $x \in \mathcal{H}^n$, as $M^* x = (1' C_{\boldsymbol{\mathcal{X}}}^{-1} 1)^{-1} 1' C_{\boldsymbol{\mathcal{X}}}^{-1} x$, and $L$ is the $\mu_{\boldsymbol{\mathcal{X}}}$-mlt of conditional expectation that solves Problem 1 and acts on $x \in \mathcal{H}^n$ as $Lx = C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} C_{\boldsymbol{\mathcal{X}}}^{-1} x$.*

In the following paragraphs we derive and comment the results stated in Theorem 4.

*Unbiasedness constraint.* To formulate the objective functional, we consider first the unbiasedness constraint. We define the operator $1 : \mathcal{H} \to \mathcal{H}^n$ acting on $x \in \mathcal{H}$ as $x \mapsto 1\, x = (x, x, ..., x)^T$. This enables to formulate the constraint as

$$\Lambda\, 1\, m = m \quad \text{for any} \quad m \in \mathcal{H}. \tag{11}$$

Here we have exploited the fact that, for a $\mu_{\boldsymbol{\mathcal{X}}}$-mlt $\Lambda$, $\mathbb{E}[\Lambda\boldsymbol{\mathcal{X}}] = \Lambda\mathbb{E}[\boldsymbol{\mathcal{X}}]$ (see e.g., Picard, 2006, p.64).

*Objective Functional.* Following the Lagrange multiplier method, we consider the following objective functional

$$\Phi = \mathbb{E}\left[\|\mathcal{X}_{\boldsymbol{s}_0} - \Lambda\boldsymbol{\mathcal{X}}\|^2_{\mathcal{H}}\right] + 2\varphi_\zeta\,(\Lambda\,1 - I), \tag{12}$$

where $I : \mathcal{H} \to \mathcal{H}$ is the identity operator and $\varphi_\zeta$ is a Lagrange multiplier, i.e., a functional acting on the space of $\mu_{\mathcal{X}_0}$-mlt.

To develop further the expression of functional (12), we call $P_0 : \mathcal{H}^{n+1} \to \mathcal{H}$, $P_n : \mathcal{H}^{n+1} \to \mathcal{H}^n$ the operators acting on $x = (x_0, x_1, ..., x_n)^T \in \mathcal{H}^{n+1}$ as $P_0 x = x_0$ and $P_n x = (x_1, ..., x_n)^T$, respectively. We note that both $P_0$ and $P_n$ are $\mu_{\boldsymbol{\mathcal{X}}_0}$-mlt.

In the light of Corollary 2, the space $\mathcal{M}_{\boldsymbol{\mathcal{X}}_0}$ of $\mu_{\boldsymbol{\mathcal{X}}_0}$-mlt from $\mathcal{H}^{n+1}$ in $\mathcal{H}$ is a Hilbert space if endowed with the inner product

$$\langle L_1, L_2\rangle_{\mathcal{M}_{\boldsymbol{\mathcal{X}}_0}} = \int_{\mathcal{H}^{n+1}} \langle L_1 x, L_2 x\rangle_{\mathcal{H}} \mu_{\boldsymbol{\mathcal{X}}_0}(dx), \quad L_1, L_2 \in \mathcal{M}_{\boldsymbol{\mathcal{X}}_0}.$$

Similarly, the space $\mathcal{M}_{\boldsymbol{\mathcal{X}}}$ of $\mu_{\boldsymbol{\mathcal{X}}}$-mlt from $\mathcal{H}^n$ into $\mathcal{H}$ is a Hilbert space if equipped with the inner product $\langle L_1, L_2\rangle_{\mathcal{M}_{\boldsymbol{\mathcal{X}}}} = \int_{\mathcal{H}^n} \langle L_1 x, L_2 x\rangle_{\mathcal{H}} \mu_{\boldsymbol{\mathcal{X}}}(dx)$, $L_1, L_2 \in \mathcal{M}_{\boldsymbol{\mathcal{X}}}$, and the space $\mathcal{M}_{\mathcal{X}_{\boldsymbol{s}_0}}$ of $\mu_{\mathcal{X}_{\boldsymbol{s}_0}}$-mlt from $\mathcal{H}$ into $\mathcal{H}$ is a Hilbert space with the inner product $\langle L_1, L_2\rangle_{\mathcal{M}_{\mathcal{X}_{\boldsymbol{s}_0}}} = \int_{\mathcal{H}} \langle L_1 x, L_2 x\rangle_{\mathcal{H}} \mu_{\mathcal{X}_{\boldsymbol{s}_0}}(dx)$, $L_1, L_2 \in \mathcal{M}_{\mathcal{X}_{\boldsymbol{s}_0}}$.

With this notation and denoting with $tr(\cdot)$ the trace operator, we can develop the first term of the objective functional (12) as

$$\begin{aligned}
&\mathbb{E}[\|\mathcal{X}_{\boldsymbol{s}_0} - \Lambda\boldsymbol{\mathcal{X}}\|^2_{\mathcal{H}}] = \mathbb{E}\left[\|\mathcal{X}_{\boldsymbol{s}_0}\|^2_{\mathcal{H}}\right] + \mathbb{E}\left[\|\Lambda\boldsymbol{\mathcal{X}}\|^2_{\mathcal{H}}\right] - 2\,\mathbb{E}\left[\langle \mathcal{X}_{\boldsymbol{s}_0}, \Lambda\boldsymbol{\mathcal{X}}\rangle_{\mathcal{H}}\right] = \\
&\quad = \langle P_0, P_0\rangle_{\mathcal{M}_{\boldsymbol{\mathcal{X}}_0}} + \langle \Lambda P_n, \Lambda P_n\rangle_{\mathcal{M}_{\boldsymbol{\mathcal{X}}_0}} - 2\,\langle P_0, \Lambda P_n\rangle_{\mathcal{M}_{\boldsymbol{\mathcal{X}}_0}} = \\
&\quad = \langle P_0 C^{1/2}_{\boldsymbol{\mathcal{X}}_0}, P_0 C^{1/2}_{\boldsymbol{\mathcal{X}}_0}\rangle_{HS} + \langle \Lambda P_n C^{1/2}_{\boldsymbol{\mathcal{X}}_0}, \Lambda P_n C^{1/2}_{\boldsymbol{\mathcal{X}}_0}\rangle_{HS} - 2\,\langle P_0 C^{1/2}_{\boldsymbol{\mathcal{X}}_0}, \Lambda P_n C^{1/2}_{\boldsymbol{\mathcal{X}}_0}\rangle_{HS} = \\
&\quad = tr(C_{\mathcal{X}_{\boldsymbol{s}_0}}) + tr(\Lambda C_{\boldsymbol{\mathcal{X}}}\Lambda') - 2\,tr(\Lambda C_{\boldsymbol{\mathcal{X}}\mathcal{X}_{\boldsymbol{s}_0}}),
\end{aligned}$$

where $\langle\cdot,\cdot\rangle_{HS}$ denotes the inner product in the space of Hilbert-Schmidt operators.

Moreover, we can express the Lagrange penalty in terms of the Riesz representative $\zeta$ of $\varphi_\zeta$

$$\begin{aligned}
\varphi_\zeta\,(\Lambda\,1 - I) &= \langle \zeta, (\Lambda\,1 - I)\rangle_{\mathcal{M}_{\mathcal{X}_0}} = \\
&= \langle \zeta C^{1/2}_{\mathcal{X}_0}, (\Lambda\,1 - I) C^{1/2}_{\mathcal{X}_0}\rangle_{HS} = \\
&= tr(\zeta_1(\Lambda\,1 - I)),
\end{aligned}$$

with $\zeta_1 = \zeta C_{\mathcal{X}_0}$.

Hence, the objective functional reduces to

$$\Phi = tr(C_{\mathcal{X}_{\boldsymbol{s}_0}}) + tr(\Lambda C_{\boldsymbol{\mathcal{X}}}\Lambda') - 2\,tr(\Lambda C_{\boldsymbol{\mathcal{X}}\mathcal{X}_{\boldsymbol{s}_0}}) + 2\,tr(\zeta_1(\Lambda\,1 - I)). \tag{13}$$

*Kriging system.* To minimize functional (13) we compute its differential with respect to $\Lambda$ and $\zeta_1$.

$$\begin{aligned} \Phi_\Lambda &: \mathcal{M}_{\boldsymbol{\mathcal{X}}} \ni h \mapsto \Phi_\Lambda(h) = 2\, tr(h(C_{\boldsymbol{\mathcal{X}}}\Lambda' - C_{\boldsymbol{\mathcal{X}}\mathcal{X}_{\boldsymbol{s}_0}} + 1\zeta_1)) \\ \Phi_{\zeta_1} &: \mathcal{M}_{\mathcal{X}_0} \ni g \mapsto \Phi_{\zeta_1}(g) = 2\, tr(g(\Lambda\, 1 - I)). \end{aligned} \tag{14}$$

Differentials (14) lead to the Kriging system of operatorial equations in (9), where $1' : \mathcal{H}^n \to \mathcal{H}$ acts as $\mathcal{H}^n \ni (x_1, ..., x_n)^T \mapsto 1'\, x = \sum_{i=1}^n x_i$. We note that system (9) admits the following matrix representation

$$\begin{pmatrix} \Lambda & \zeta_1 \end{pmatrix} \begin{pmatrix} C_{\boldsymbol{\mathcal{X}}} & 1 \\ 1' & 0 \end{pmatrix} = \begin{pmatrix} C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} & I \end{pmatrix}, \tag{15}$$

consistent with its finite dimensional counterpart.

*An explicit solution to the Kriging system.* To derive an explicit solution of Kriging system (15), we exploit the following identity

$$\begin{pmatrix} C_{\boldsymbol{\mathcal{X}}} & 1 \\ 1' & 0 \end{pmatrix}^{-1} = \begin{pmatrix} C_{\boldsymbol{\mathcal{X}}}^{-1}[I - 1(1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1}1'C_{\boldsymbol{\mathcal{X}}}^{-1}] & C_{\boldsymbol{\mathcal{X}}}^{-1}1(1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1} \\ (1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1}1'C_{\boldsymbol{\mathcal{X}}}^{-1} & -(1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1} \end{pmatrix}, \tag{16}$$

which can be proved by direct verification. Identity (16) leads to the explicit solution (10) of system (15), reported here for convenience

$$\Lambda^* x = M^* x + L(x - 1\, M^* x), \quad x \in \mathcal{H}^n,$$

where, for $x \in \mathcal{H}^n$, $M^* x = T_M C_{\boldsymbol{\mathcal{X}}}^{-1/2} x$ with $T_M = (1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1}1'C_{\boldsymbol{\mathcal{X}}}^{-1/2}$, and $Lx = T_L C_{\boldsymbol{\mathcal{X}}}^{-1/2} x$ with $T_L = C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} C_{\boldsymbol{\mathcal{X}}}^{-1/2}$. We recognize in expression (10) the same form as (8), since the operator $L$ is the $\mu_{\boldsymbol{\mathcal{X}}}$-mlt defining the conditional expectation in (7). Moreover, $M^* x$ plays the role of the mean appearing in (8): operator $1M^*$ is a (non-orthogonal) projection, which enables one to obtain the Best Linear Unbiased Estimator (BLUE), in the operatorial sense, of the mean vector $m_{\boldsymbol{\mathcal{X}}} = (m, ..., m)^T \in \mathcal{H}^n$. Indeed, operator $M^*$ is found by solving the following Problem.

**Problem 3 (Operatorial Ordinary Kriging of the Mean).** *Given* $\boldsymbol{\mathcal{X}} = (\mathcal{X}_{\boldsymbol{s}_1}, ..., \mathcal{X}_{\boldsymbol{s}_n})^T$ *and with the previous notation, find the BLUE for* $m$*, i.e.,* $m^* = M^* \boldsymbol{\mathcal{X}}$*, where* $M^* : \mathcal{H}^n \to \mathcal{H}$ *is* $\mu_{\boldsymbol{\mathcal{X}}}$*-mlt and minimizes*

$$\mathbb{E}[\|\mathcal{X}_{\boldsymbol{s}_0} - m^M\|^2_{\mathcal{H}}] \quad \textit{subject to} \quad \mathbb{E}[m^M] = m,$$

*where* $m^M = M\boldsymbol{\mathcal{X}}$*, with* $M : \mathcal{H}^n \to \mathcal{H}$ $\mu_{\boldsymbol{\mathcal{X}}}$*-mlt.*

The solution of Problem 3 is obtained by following the same strategy introduced to solve Problem 2 (not shown).

Therefore, the Operatorial Ordinary Kriging predictor $\mathcal{X}^*_{\boldsymbol{s}_0} = \Lambda^* \boldsymbol{\mathcal{X}}$ is found as the sum of the estimated mean $m^* = M^* \boldsymbol{\mathcal{X}}$ and the plug-in conditional expectation $L(\boldsymbol{\mathcal{X}} - 1M^*\boldsymbol{\mathcal{X}})$, which is obtained by applying the operator of conditional expectation $L$ to the estimated residuals $(\boldsymbol{\mathcal{X}} - 1M^*\boldsymbol{\mathcal{X}})$. A similar results is found in the finite-dimensional setting (e.g., Cressie, 1993), in confirmation of the consistency of our approach with that setting.

## 6. Operatorial Kriging as a unifying theory: finite-dimensional approximations

In this Section, we focus on characterizing the existing formulations of Kriging within the general framework here introduced. To this end, we introduce finite-dimensional approximations of the Operatorial Ordinary Kriging Predictor derived in Section 5.

We call discretization an operator $D_K \in \mathcal{L}(\mathcal{H}, \mathcal{H}_K)$, $K > 1$, such that

(i) $D_K(\mathcal{H}) = \mathcal{H}_K \subset \mathcal{H}$;

(ii) $\mathcal{H}_K$ is a $K$-dimensional Hilbert space ($K < \infty$).

For instance, given an orthonormal basis $\{e_k, k \geq 1\}$, a valid discretization $D_K^e$ is the projection into the space generated by the first $K$ elements of the basis. Hereafter, we denote with the superscript $K$ the quantities referring to a given discretization $D_K$.

Having fixed a discretization $D_K$, we consider the following Discretized Operatorial Ordinary Kriging Problem.

**Problem 4 (Discretized Operatorial Ordinary Kriging).** *Given* $\boldsymbol{\mathcal{X}}^K = (\mathcal{X}_{\boldsymbol{s}_1}^K, ..., \mathcal{X}_{\boldsymbol{s}_n}^K)^T$, $\mathcal{X}_{\boldsymbol{s}_i}^K = D_K \mathcal{X}_{\boldsymbol{s}_i}$, $i = 1, ..., n$, *and with the previous notation, find the BLUP for* $\mathcal{X}_{\boldsymbol{s}_0}^K$, *i.e.,* $\mathcal{X}_{\boldsymbol{s}_0}^{K*} = \Lambda^{K*}\boldsymbol{\mathcal{X}}^K$, *where* $\Lambda^{K*} : \mathcal{H}_K^n \to \mathcal{H}_K$ *is a* $\mu_{\boldsymbol{\mathcal{X}}^K}$*-mlt and minimizes*

$$\mathbb{E}[\|\mathcal{X}_{\boldsymbol{s}_0}^K - \mathcal{X}_{\boldsymbol{s}_0}^{\Lambda^K}\|_{\mathcal{H}}^2] \quad \textit{subject to} \quad \mathbb{E}[\mathcal{X}_{\boldsymbol{s}_0}^{\Lambda^K}] = m^K,$$

*where* $\mathcal{X}_{\boldsymbol{s}_0}^{\Lambda^K} = \Lambda^K \boldsymbol{\mathcal{X}}^K$, *with* $\Lambda^K : \mathcal{H}_K^n \to \mathcal{H}_K$ *a* $\mu_{\boldsymbol{\mathcal{X}}_K}$*-mlt.*

In the following, we show two possible solutions to the discretized problem, providing useful insights into the existing formulations of Kriging.

### *6.1. A Cokriging solution*

To derive a version of the discretized predictor, we note that the solution of Problem 4 can be expressed in the form (2), since any finite dimensional Hilbert space is a RKHS. Moreover, the image of $\mathcal{H}$ under $D_K$ is isometrically isomorphic to $\mathbb{R}^K$, by assumption. We call $\iota : \mathcal{H}_K \to \mathbb{R}^K$ such an isometric isomorphism. The operator $\iota$ being an isometry, one has

$$\mathbb{E}[\|\mathcal{X}_{\boldsymbol{s}_0}^K - \mathcal{X}_{\boldsymbol{s}_0}^{\Lambda^K}\|_{\mathcal{H}}^2] = \mathbb{E}[\|\boldsymbol{\xi}_{\boldsymbol{s}_0} - \boldsymbol{\xi}_{\boldsymbol{s}_0}^*\|_{\mathbb{R}^K}^2], \tag{17}$$

with $\boldsymbol{\xi}_{\boldsymbol{s}_0} = (\xi_{\boldsymbol{s}_0,1}, ..., \xi_{\boldsymbol{s}_0,K})^T = \iota\mathcal{X}_{\boldsymbol{s}_0}^K$ and $\boldsymbol{\xi}_{\boldsymbol{s}_0}^* = (\xi_{\boldsymbol{s}_0,1}^*, ..., \xi_{\boldsymbol{s}_0,K}^*)^T = \iota\mathcal{X}_{\boldsymbol{s}_0}^{\Lambda^K}$. Without loss of generality, hereinafter we denote by $\boldsymbol{\xi}_{\boldsymbol{s}}$ the vector of Fourier coefficients in $\boldsymbol{s}$, with respect to an orthonormal basis $\{v_j, j \geq 1\}$ of $\mathcal{H}_K$, i.e., $\boldsymbol{\xi}_{\boldsymbol{s}}^j = \langle \mathcal{X}_{\boldsymbol{s}}^K, v_j \rangle_{\mathcal{H}}$.

Further, we note that $\Lambda^K \boldsymbol{\mathcal{X}} = \sum_{i=1}^n \Lambda_i^K \mathcal{X}_{\boldsymbol{s}_i}$, where $\Lambda_i^K \in \mathcal{L}(\mathcal{H}_K, \mathcal{H}_K)$. Hence, each $\Lambda_i^K$, $i = 1, ..., n$, admits a matrix representation: for every $x \in \mathcal{H}_K$

$$\Lambda_i^K x = \sum_{j=1}^K \sum_{l=1}^K (\Lambda_i^K)_{jl} x_j v_l,$$

with $(\Lambda_i^K)_{jl} = \langle \Lambda_i^K v_j, v_l \rangle_{\mathcal{H}}$, $x_j = \langle x, v_j \rangle_{\mathcal{H}}$. Therefore, one can express the predictor as

$$\mathcal{X}_{\boldsymbol{s}_0}^{\Lambda,K} = \sum_{i=1}^n \sum_{j=1}^K \sum_{l=1}^K (\Lambda_i^K)_{jl} \xi_{\boldsymbol{s}_i,j} v_l, \tag{18}$$

and the unbiasedness constraint as

$$\Lambda^K 1 = I \quad \text{on} \quad \mathcal{H}_K,$$

which reduces to

$$\sum_{i=1}^n (\Lambda_i^K)_{jl} = \begin{cases} 0, & j \neq l; \\ 1, & j = l. \end{cases} \tag{19}$$

We recognize in condition (19), predictor (18) and in the quadratic loss (17), the corresponding counterparts found in classical multivariate Ordinary Cokriging. Therefore, Problem 4 reduces to a multivariate Ordinary Cokriging (Cressie, 1993) of the coefficient vectors $\iota\mathcal{X}_{\boldsymbol{s}_i}^K$, $i = 1, ..., n$. The matrices $\mathbb{L}_i = [(\Lambda_i^K)_{jl}]$ are thus found as solution of the following linear system

$$\begin{pmatrix} \mathbb{C}_{11} & \cdots & \mathbb{C}_{1,n} & \mathbb{I}_K \\ \vdots & \ddots & \vdots & \vdots \\ \mathbb{C}_{n1} & \cdots & \mathbb{C}_{nn} & \mathbb{I}_K \\ \mathbb{I}_K & \cdots & \mathbb{I}_K & 0 \end{pmatrix} \begin{pmatrix} \mathbb{L}_1 \\ \vdots \\ \mathbb{L}_n \\ \mathbb{Z} \end{pmatrix} = \begin{pmatrix} \mathbb{C}_{01} \\ \vdots \\ \mathbb{C}_{0n} \\ \mathbb{I}_K \end{pmatrix}, \tag{20}$$

where $\mathbb{C}_{ij}$ is the cross-covariance matrix between $\boldsymbol{\xi}_{\boldsymbol{s}_i}$ and $\boldsymbol{\xi}_{\boldsymbol{s}_j}$, $\mathbb{I}_K$ is the identity matrix in $\mathbb{R}^K$, and $\mathbb{Z}$ is the matrix of Lagrange multiplier. Hence, the explicit solution to the Kriging problem proposed by Nerini et al. (2010) is found by embedding this result into a finite-dimensional $L^2$ space. However, the present results can be employed in more general settings than just $L^2$. In Section 7, we discuss the relevance of this from the application viewpoint.

### *6.2. A Trace-Kriging solution*

We now focus on the case when the dimension $K$ of the discretized space is lower than or equal to the number $n$ of data, which is representative of most real applications. This case appears interesting, as the solution of the discretized problem can be significantly simplified. The aim of this Subsection is to prove that the finite dimensional approximation solving Problem (4) is equivalently found as solution of the following Problem.

**Problem 5 (Ordinary Trace-Kriging).** *Given* $\boldsymbol{\mathcal{X}}^K = (\mathcal{X}^K_{\boldsymbol{s}_1}, ..., \mathcal{X}^K_{\boldsymbol{s}_n})^T$, $\mathcal{X}^K_{\boldsymbol{s}_i} = D_K(\mathcal{X}_{\boldsymbol{s}_i})$, $i = 1, ..., n$, *and with the previous notation, find the BLUP, in the trace sense, for* $\mathcal{X}^K_{\boldsymbol{s}_0}$, *i.e.,* $\mathcal{X}^{\boldsymbol{\lambda},K*}_{\boldsymbol{s}_0} = \sum_{i=1}^n \lambda^*_i \mathcal{X}^K_{\boldsymbol{s}_i}$, *where* $\lambda^*_1, ..., \lambda^*_n \in \mathbb{R}$ *minimize*

$$\mathbb{E}[\|\mathcal{X}^K_{\boldsymbol{s}_0} - \mathcal{X}^{\boldsymbol{\lambda},K}_{\boldsymbol{s}_0}\|^2_{\mathcal{H}}] \quad \textit{subject to} \quad \mathbb{E}[\mathcal{X}^{\boldsymbol{\lambda}}_{\boldsymbol{s}_0}] = m^K,$$

*where* $\mathcal{X}^{\boldsymbol{\lambda},K}_{\boldsymbol{s}_0} = \sum_{i=1}^n \lambda_i \mathcal{X}^K_{\boldsymbol{s}_i}$, *with* $\lambda_1, ..., \lambda_n \in \mathbb{R}$.

The solution of Problem 5 can be found through the methodology proposed by Menafoglio et al. (2013), embedded into the finite-dimensional Hilbert space $\mathcal{H}_K$. These authors address the problem of the spatial prediction via linear combinations of the data by introducing global notions of spatial dependence and stationarity. Specifically, given a $\mathcal{H}$-valued random field $\{\mathcal{Y}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$, they propose to describe the spatial dependence through the trace-covariogram $C_{tr} : D \times D \to \mathbb{R}$, which is defined, under stationarity, as

$$C_{tr}(\boldsymbol{s}_i - \boldsymbol{s}_j) = \mathbb{E}[\langle \mathcal{Y}_{\boldsymbol{s}_i} - m, \mathcal{Y}_{\boldsymbol{s}_j} - m\rangle_{\mathcal{H}}], .$$

i.e., the trace of the associated cross-covariance operator $C_{\mathcal{Y}_{\boldsymbol{s}_i}\mathcal{Y}_{\boldsymbol{s}_j}}$ (Menafoglio et al., 2013, Proposition 3). This represents a generalization to Hilbert spaces of the notion of trace-variogram proposed by Giraldo et al. (2011). On this basis, Menafoglio et al. (2013) prove that the unique solution of Problem 5 can be found by solving the linear system

$$\begin{pmatrix} C_{tr}(\mathbf{0}) & \cdots & C_{tr}(\boldsymbol{s}_1 - \boldsymbol{s}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{tr}(\boldsymbol{s}_n - \boldsymbol{s}_1) & \cdots & C_{tr}(\mathbf{0}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \zeta_{tr} \end{pmatrix} = \begin{pmatrix} C_{tr}(\boldsymbol{s}_0 - \boldsymbol{s}_1) \\ \vdots \\ C_{tr}(\boldsymbol{s}_0 - \boldsymbol{s}_n) \\ 1 \end{pmatrix}, \tag{21}$$

where $\zeta_{tr} \in \mathbb{R}$ is a Lagrange multiplier, accounting for the unbiasedness constraint. Note that the optimal weights $\lambda^*_1, ..., \lambda^*_n$ can be uniquely determined through the trace-covariogram, without the need of specifying the entire structure of spatial dependence. Hereafter, we call Trace-Kriging predictor the solution of Problem 5, as opposed to the Cokriging predictor solving Problem 4, detailed in Subsection 6.1.

To accommodate the form of the Cokriging predictor in expression (18) and to ease the comparison, we rewrite the Trace-Kriging predictor $\mathcal{X}^{\boldsymbol{\lambda},K}_{\boldsymbol{s}_0}$ as

$$\mathcal{X}^{\boldsymbol{\lambda},K}_{\boldsymbol{s}_0} = \sum_{i=1}^n \lambda_i \mathcal{X}^K_{\boldsymbol{s}_i} = \sum_{i=1}^n \sum_{l=1}^K \lambda_i \xi_{\boldsymbol{s}_i,l} v_l = \sum_{i=1}^n \sum_{j=1}^K \sum_{l=1}^K (\Lambda_i)_{jl} \xi_{\boldsymbol{s}_i,j} v_l$$

with

$$(\Lambda_i^K)_{jl} = \begin{cases} 0, & j \neq l; \\ \lambda_i, & j = l, ..., K \end{cases}$$

for $i = 1, ..., n$.

In the stationary case, the unbiasedness constraint in Problem 5 reads $\sum_{i=1}^n \lambda_i = 1$, which is equivalently expressed in terms of $(\Lambda_i^K)_{jl}$ as (19). Therefore, the solution of Problem 5 is an admissible solution of Problem 4. To prove the equivalence of the two solutions, one is left to prove that the solution of Problem 4 admits the form $\mathcal{X}_{\boldsymbol{s}_0}^{\boldsymbol{\lambda},K} = \sum_{i=1}^n \lambda_i \mathcal{X}_{\boldsymbol{s}_i}^K$, with $\lambda_1, ..., \lambda_n \in \mathbb{R}$. To this end, we recall that a system of $K$ linearly independent vectors $\boldsymbol{x}_1, ..., \boldsymbol{x}_K$ in $\mathbb{R}^K$, with $\boldsymbol{x}_i = (x_{i1}, ..., x_{iK})$, $i = 1, ..., K$, constitutes a basis of $\mathbb{R}^K$, that is, for every $\boldsymbol{y} \in \mathbb{R}^K$, there exist $\lambda_1, ..., \lambda_K$ such that $\boldsymbol{y} = \sum_{i=1}^K \lambda_i \boldsymbol{x}_i$. Moreover, given $n$ vectors $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ in $\mathbb{R}^K$, these constitute a basis of $\mathbb{R}^K$ if and only if $\text{rank}(\mathbb{X}) = K$, where $\mathbb{X}_{ik} = x_{ik}$. The elements $\mathcal{X}_{\boldsymbol{s}_1}^K, ..., \mathcal{X}_{\boldsymbol{s}_n}^K$ form almost surely a basis of $\mathcal{H}_K$, or, equivalently, $\boldsymbol{\xi}_{\boldsymbol{s}_1}, ..., \boldsymbol{\xi}_{\boldsymbol{s}_n}$, with $\boldsymbol{\xi}_{\boldsymbol{s}_i} = \iota \mathcal{X}_{\boldsymbol{s}_i}^K$, $i = 1, ..., n$, constitute almost surely a basis of $\mathbb{R}^K$ i.e.,

$$\text{rank} \begin{pmatrix} \xi_{\boldsymbol{s_1},1} & \cdots & \xi_{\boldsymbol{s_1},1} \\ \vdots & \ddots & \vdots \\ \xi_{\boldsymbol{s_n},1} & \cdots & \xi_{\boldsymbol{s_n},K} \end{pmatrix} = K \quad a.s. \tag{22}$$

Hence, the optimal predictor $\mathcal{X}_{\boldsymbol{s}_0}^{K*} = \Lambda^{K*} \boldsymbol{\mathcal{X}}^K$ is almost surely expressed as $\mathcal{X}_{\boldsymbol{s}_0}^{\boldsymbol{\lambda},K*} = \sum_{i=1}^n \lambda_i^* \mathcal{X}_{\boldsymbol{s}_i}^K$, where $\lambda_1^*, ..., \lambda_n^*$ satisfy

$$\begin{pmatrix} \xi_{\boldsymbol{s_0},1}^* \\ \vdots \\ \xi_{\boldsymbol{s_0},K}^* \end{pmatrix} = \begin{pmatrix} \lambda_1^* & \cdots & \lambda_n^* \end{pmatrix} \begin{pmatrix} \xi_{\boldsymbol{s_1},1} & \cdots & \xi_{\boldsymbol{s_1},1} \\ \vdots & \ddots & \vdots \\ \xi_{\boldsymbol{s_n},1} & \cdots & \xi_{\boldsymbol{s_n},K} \end{pmatrix}.$$

Therefore, the Kriging predictor of Menafoglio et al. (2013) is here interpreted as the best approximation of the Operatorial Kriging predictor within the finite-dimensional Hilbert space generated by the observations.

## 7. Discussion: the application viewpoint

In this Section, we discuss the relevance of our results of Sections 5 and 6 from the application viewpoint, at the crossing between functional and high-dimensional statistics. We note that, from a general viewpoint, these disciplines frequently benefit of their deep links and interactions. Indeed, providing novel and general frameworks for the statistical analysis of infinite-dimensional data has been of interest to address a number of issues in the so called "large $p$ small $n$" problems, related to the curse of dimensionality. For instance, Chen et al. (2011) propose to adopt a FDA viewpoint to perform regression in the presence of high-dimensional predictors, and show that their approach offers a sensible alternative to classical regression models, that often fail in this setting. Viceversa, well-known methods that are commonly employed in high-dimensional statistics can be inspiring to address similar problems arising in the FDA setting. This is the case, e.g., of Aneiros and Vieu (2014), who adapt multivariate LASSO techniques to face the problem of functional variable selection.

In this work, we derived the formal expression of the Operatorial Ordinary Kriging predictor as (10), inspired by the similar results that are available in the finite-dimensional setting. Further, we showed that: (i) for any Hilbert space embedding and any dimension $K$ of the discretization, one can approximate the Operatorial Ordinary Kriging predictor of Theorem 4 via the multivariate Cokriging predictor based on the $K$-dimensional vectors $\iota\mathcal{X}_{\boldsymbol{s}_i}^K$, $i = 1, ..., n$ (e.g., the Fourier coefficients with respect to an orthonormal basis); and (ii) as long as $K \leq n$, one can actually compute a $K$-dimensional approximation of the Operatorial Ordinary Kriging predictor of Theorem 4, equivalently by (20) or (21). Nevertheless, in the case (ii), the difference in terms of complexity is relevant: solving Trace-Kriging system (21) has a complexity $\mathcal{O}(n^3)$, as opposed to the complexity $\mathcal{O}(n^3(K+1)^3)$ of solving the Cokriging system (20). Moreover, the former can be solved by relying on trace-covariography only, while the latter requires the complete characterization of the discretized cross-covariance operator. This is particularly relevant, since in most applications the covariance operator is unknown. As a matter of fact, modeling a multivariate covariance structure as that appearing in (15) requires the estimation and fitting of covariograms and cross-covariograms of the elements of $\iota\mathcal{X}_{\boldsymbol{s}_i}^K$, $i = 1, ..., n$, e.g., via a linear model of coregionalization (LMC, Cressie, 1993). This is a crucial but delicate point of the geostatistical analysis, especially when high-dimensional vectors are concerned. Instead, modeling a trace-covariogram requires the same effort as the estimation of a one-dimensional covariogram (i.e., first provide an empirical estimate, then fit a parametric model, e.g., spherical or Matérn, see Menafoglio et al., 2013, Section 2.3). Therefore, providing alternative yet consistent approaches to Kriging is of key importance in the geosciences.

We notice that these observations hold for any Hilbert space, and particularly for Euclidean spaces $\mathbb{R}^p$, $1 \leq p < \infty$, where multivariate and high-dimensional geostatistics is performed. Even in that setting, the trace-covariography is posed as a convenient alternative to the widely-used Cokriging approach, that still guarantees the same degree of precision, as long as the dimension $p$ does not exceed the number $n$ of available data — which is the case of most multivariate geostatistical analyses. Notice that even in $\mathbb{R}^p$, with $p > n$, one may consider a discretization strategy as in Section 6, that is, first perform a dimensionality reduction (e.g., via Principal Component Analysis), and then proceed with the geostatistical analysis in the discretized space $\mathbb{R}^K$, with $K \leq n$, e.g., via Trace-Kriging.

In all the cases for which a discretization of order $K \leq n$ is considered inappropriate, the Cokriging approach of Subsection 6.1 can be employed. Here, close attention should be paid to the modeling and estimation of the covariance structure of the field. To this end, a LMC could be employed, or simplifying assumptions may be considered in order to ease the estimation procedure. Amongst these, Markov models may be introduced for the elements of $\iota\mathcal{X}_{\boldsymbol{s}_i}^K$, $i = 1, ..., n$, as in multivariate co-located Kriging (e.g., Journel, 1999).

We finally remark that the availability of explicit solutions to the (discretized) Kriging problem that are valid in any Hilbert space provides an important source of flexibility in a functional geostatistical analysis. For instance, one may want to explicitly account for differential information through the use of a Sobolev space embedding for the data (see, e.g., Menafoglio et al., 2013). Further, even data constraints can be accounted for in the Hilbert space setting, provided that an appropriate geometry is considered. For instance, Menafoglio et al. (2014a,b) deal with a georeferenced dataset of probability density functions in the form of particle-size densities and employ a Hilbert space embedding for this kind of data, namely, the embedding into the Bayes Hilbert space (Egozcue et al., 2006; van den Boogaart et al., 2014), whose geometry account for the key properties of distributions.

## 8. Conclusion and further research

In this work we established a novel theoretical framework for Operatorial Ordinary Kriging, grounded on the theory of Gaussian processes in Hilbert spaces. Our research led to the following key conclusions.

1. Under the assumption of stationarity and known mean, a comprehensive theory of spatial prediction in Hilbert spaces can be developed by relying on the notion of measurable linear transformation. This setting allows to derive the formal relation between the Operatorial Simple Kriging predictor and the conditional expectation of a Gaussian measure.

2. We addressed the problem of Kriging in case of unknown mean, focusing on stationary Gaussian random fields. We formalized the Operatorial Ordinary Kriging Problem and proved its well-posedness, deriving an explicit expression for the optimal predictor (Theorem 4). Possible extensions include the non-stationary case, i.e., the Operatorial Universal Kriging.

3. We showed the unifying nature of our new theoretical framework through the notion of Discretized Kriging problem. We showed that our new setting includes the available Kriging predictor in FDA, besides those of multivariate geostatistics. Further, we proved that the Trace-Kriging predictor based on trace-covariography (Menafoglio et al., 2013) provides the best approximation of the Operatorial Ordinary Kriging predictor in the space generated by the observations. The generalization of these results to Banach spaces is an on-going research along this line.

4. The attained results are key to address a number of computational issues in both functional and multivariate settings. Indeed, in most field studies, one can provide an efficient alternative to the Cokriging solution of Subsection 6.1, by addressing the problem via the Trace-Kriging approach of Subsection 6.2, with a relevant gain in terms of computational and modeling complexity. These observations open new and relevant perspectives for the Kriging of large and high-dimensional datasets, which is one of the most challenging topics in modern geostatistics.

## References


Aneiros G, Vieu P. Variable selection in infinite-dimensional problems. Statistics & Probability Letters 2014;94(0):12 – 20. doi:`http://dx.doi.org/10.1016/j.spl.2014.06.025`.

Bogachev V. Gaussian measures. American Mathematical Society, 1998.

Bongiorno E, Goia A, Salinelli E, Vieu P. An overview of IWFOS'2014. In *Contributions in infinite-dimensional statistics and related topics*; Esculapio, Bologna. p. 1–6.

Bosq D. Linear Processes in Function Spaces. Springer, New York, 2000.

Caballero W, Giraldo R, Mateu J. A universal kriging approach for spatial functional data. Stochastic Environmental Research and Risk Assessment 2013;27:1553–63.

Chen K, Chen K, Müller HG, Wang JL. Stringing high-dimensional data for functional analysis. Journal of the American Statistical Association 2011;106(493):275–84. doi:`10.1198/jasa.2011.tm10314`.

Cressie N. Statistics for Spatial data. John Wiley & Sons, New York, 1993.

Cuevas A. A partial overview of the theory of statistics with functional data. Journal of Statistical Planning and Inference 2014;147(0):1–23.

Da Prato G, Zabczyk J. Stochastic equations in infinite dimensions. 2nd ed. Cambridge University Press, 2014.

Delicado P. Functional k-sample problem when data are density functions. Computational Statistics 2007;22:391–410.

Delicado P. Dimensionality reduction when data are density functions. Computational Statistics & Data Analysis 2011;55(1):401–20.

Delicado P, Giraldo R, Comas C, Mateu J. Statistics for spatial functional data: some recent contributions. Environmetrics 2010;21(3-4):224–39.

Egozcue JJ, Díaz-Barrero JL, Pawlowsky-Glahn V. Hilbert Space of Probability Density Functions Based on Aitchison Geometry. Acta Mathematica Sinica, English Series 2006;22(4):1175–82.

Ferraty F, Vieu P. Nonparametric functional data analysis : theory and practice. Springer, New York, 2006.

Giraldo R. Geostatistical Analysis of Functional Data. Ph.D. thesis; Universitat Politècnica da Catalunya; Barcellona; 2009.

Giraldo R, Delicado P, Mateu J. Ordinary kriging for function-valued spatial data. Environmental and Ecological Statistics 2011;18(3):411–26.

Horváth L, Kokoszka P. Inference for Functional Data with Applications. Springer Series in Statistics. Springer, 2012.

Journel AG. Markov models for cross-covariances. Mathematical Geology 1999;31(8):955–64.

Luschgy H. Linear Estimators and Radonifying Operators. Theory of Probability & Its Applications 1996;40(1):167–75. doi:10.1137/1140017.

Mandelbaum A. Linear estimators and measurable linear transformations on a Hilbert space. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 1984;65(3):385–97. doi:10.1007/BF00533743.

Marron JS, Alonso AM. Overview of object oriented data analysis. Biometrical Journal 2014;56(5):732–53. URL: http://dx.doi.org/10.1002/bimj.201300072. doi:10.1002/bimj.201300072.

Menafoglio A, Guadagnini A, Secchi P. A Kriging Approach based on Aitchison Geometry for the Characterization of Particle-Size Curves in Heterogeneous Aquifers. Stochastic Environmental Research and Risk Assessment 2014a;28(7).

Menafoglio A, Secchi P, Dalla Rosa M. A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. Electronic Journal of Statistics 2013;7:2209–40. doi:10.1214/13-EJS843.

Menafoglio A, Secchi P, Guadagnini A. A Class-Kriging predictor for Functional Compositions with Application to Particle-Size Curves in Heterogeneous Aquifers. MOX-report 58/2014; Politecnico di Milano; 2014b. https://www.mate.polimi.it/biblioteca/add/qmox/58-2014.pdf.

Nerini D, Monestiez P, Manté C. Cokriging for spatial functional data. Journal of Multivariate Analysis 2010;101(2):409–18.

Picard J. Séminaire de Probabilités XLIII; Springer. p. 1–69.

Pini A, Stamm A, Vantini S. Hotelling's $T^2$ in functional Hilbert spaces. MOX-report 1/2015; Politecnico di Milano; 2015. https://www.mate.polimi.it/biblioteca/add/qmox/1-2015.pdf.

Ramsay J, Silverman B. Functional data analysis. 2nd ed. Springer, New York, 2005.

Sangalli LM, Secchi P, Vantini S, Vitelli V. K-mean alignment for curve clustering. Computational Statistics & Data Analysis 2010;54(5):1219 –33.

Srivastava A, Wu W, Kurtek S, Klassen E, Marron JS. Registration of Functional Data Using Fisher-Rao Metric. Technical Report; 2011. ArXiv:1103.3817v2.

Tarpey T, Kinateder K. Clustering functional data. Journal of Classification 2003;20:93–114.

van den Boogaart KG, Egozcue J, Pawlowsky-Glahn V. Bayes hilbert spaces. Australian & New Zealand Journal of Statistics 2014;56:171–94.

Vantini S. On the definition of phase and amplitude variability in functional data analysis. TEST 2012;21(4):676–96.