MOX–Report No. 2/2015

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

Menafoglio, A.; Petris, G.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

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

Alessandra Menafoglio[1] and Giovanni Petris[2]

[1]MOX, Department of Mathematics, Politecnico di Milano, Italy
[2]Department of Mathematical Sciences, University of Arkansas, USA
alessandra.menafoglio@polimi.it
GPetris@uark.edu

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

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) 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 field valued in any separable Hilbert space, allowing for the analysis of a broad range of object data, such as curves, surfaces or images. These authors rely on the notion of inner product and norm to define global definitions of spatial dependence (i.e., trace-variography), and accordingly perform predictions. Amongst the possible applications of this geometric perspective we cite the works of (Menafoglio et al., 2014a,b). These authors analyze a set of particle-size densities by embedding the problem within the Hilbert space equipped with the Aitchison geometry proposed in (Egozcue et al., 2006; van den Boogaart et al., 2014). Such a geometrical approach enables them to employ the Kriging theory of Menafoglio et al. (2013), while properly accounting for the data constraints.

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 here 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) or that for Hilbert Data (Menafoglio et al., 2013).

Our perspective aims to constitute a generalization of the formulation by Nerini et al. (2010), who consider the problem of finding the best unbiased predictor over the class of linear Hilbert-Schmidt transformations of the observations, assumed to belong to a Reproducing Kernel Hilbert Space (RKHS). The RKHS-embedding is key to the well-posedness of the Kriging problem of Nerini et al. (2010), but still appears a too restrictive setting, as, for instance, the Hilbert space L^2 is not a RHKS, even though it is commonly employed in FDA.

The aim of this work is to establish an Operatorial Kriging theory able to fill this theoretical gap, relying upon the key notion of measurable linear transformation associated to a Gaussian measure (Mandelbaum, 1984; Luschgy, 1996). This broad class of operators includes linear Hilbert-Schmidt operators, and is here shown to allow for the Best Linear Unbiased (BLU) prediction in any finite- or infinite-dimensional separable Hilbert space.

The remaining part of this work is organized as follow. The theory of Gaussian measures on Hilbert spaces upon which we ground our developments is recalled in Section 2. Section 3 introduces the Operatorial Kriging theory for random fields under the assumption of known mean (i.e., Simple Kriging). Even though this assumption is often too restrictive in real-world applications, this case appears insightful, as it allows obtaining interesting interpretations in terms of conditional expectations. In Section 4 we focus on the case of unknown mean, in the stationary setting. Here, we propose an original formulation of the Operatorial Kriging problem and provide its explicit solution. The relation of our new theory with the existing literature works of Nerini et al. (2010); Menafoglio et al. (2013) is investigated in Section 5. Finally, Section 6 concludes the work.

2 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 σ-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 with $A' \in \mathcal{L}(\mathcal{H}_1, \mathcal{H})$ 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 with $\mu_{\mathcal{X}}$ 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 $\mathbb{E}[\|\mathcal{X}\|^2_{\mathcal{H}}] < \infty$. 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 to injective covariance operators. In this case, the following result holds.

Theorem 2 (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*

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

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

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

(iii) In both (1) *and* (2) *, the Hilbert-Schmidt norm of T is equal to*

$$\|L\|_{\mu_{\mathcal{X}}}^2 = \int_{\mathcal{H}_2} \|Lx\|_{\mathcal{H}_1}^2 \mu_{\mathcal{X}}(dx). \tag{3}$$

Finally, the following Corollary of Theorem 2 will be useful in the following.

Corollary 3 (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* (3)*. It is isometric to the space of Hilbert-Schmidt operators via the correspondence* (1) *and* (2)*.*

3 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

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

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 $C_{\mathcal{YZ}}$ 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 4 (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 $T : \mathcal{H}_2 \to \mathcal{H}_1$

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

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 injective. 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 $L : \mathcal{H}^n \to \mathcal{H}$

$$\mathbb{E}[\mathcal{X}_{\boldsymbol{s}_0} | \boldsymbol{\mathcal{X}}] = L\boldsymbol{\mathcal{X}} \tag{4}$$

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 $m_{\mathcal{X}_{\boldsymbol{s}}}$, (4) reads

$$\mathbb{E}[\mathcal{X}_{\boldsymbol{s}_0} | \boldsymbol{\mathcal{X}}] = m_{\mathcal{X}_{\boldsymbol{s}_0}} + L(\boldsymbol{\mathcal{X}} - m_{\boldsymbol{\mathcal{X}}}), \tag{5}$$

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 Best Linear Unbiased Predictor (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 5 (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 4 for stationary random fields.

4 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 6 (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 in (Nerini et al., 2010) for the particular case of Reproducing Kernel Hilbert Spaces (RKHS). These authors focus on linear predictors of the kind

$$\widehat{\mathcal{X}}_{\boldsymbol{s}_0} = \sum_{i=1}^{n} B_i \mathcal{X}_{\boldsymbol{s}_i}, \tag{6}$$

where $B_i : \mathcal{H} \to \mathcal{H}$ are Hilbert-Schmidt linear operators. In this context they derive Kriging equations, and provide an explicit solution for random processes valued in a k-dimensional Hilbert space ($k < \infty$) equipped with the L^2 inner product. Similar results are obtained in (Giraldo, 2009). Nevertheless, Nerini et al. (2010) acknowledge that the RKHS setting is quite restrictive, as, for instance, the Hilbert space L^2 is not a RHKS, 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 (6) cannot be, in general, the solution of Problem 6, even if it is for $\mathcal{H}$ RKHS. In the following paragraphs we show that Problem 6 is well-posed, instead.

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

$$\Lambda\, 1\, m = m \quad \text{for any} \quad m \in \mathcal{H}. \tag{7}$$

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

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

$$\Phi = \mathbb{E}\left[\|\mathcal{X}_{\boldsymbol{s}_0} - \Lambda \boldsymbol{\mathcal{X}}\|^2_{\mathcal{H}}\right] + 2\varphi_{\zeta}\,(\Lambda\, 1 - I), \tag{8}$$

where $I : \mathcal{H} \to \mathcal{H}$ is the identity operator and φ_{ζ} 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 (8), 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 $C_{\boldsymbol{\mathcal{X}}_0} : \mathcal{H}^{n+1} \to \mathcal{H}^{n+1}$, expressed in block form as

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

Moreover, 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 3, 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 (8) 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 ζ of φ_ζ

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

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

Hence, the objective functional reduces to

$$\Phi = tr(C_{\mathcal{X}_{\boldsymbol{s}_0}}) + tr(\Lambda C_{\boldsymbol{\mathcal{X}}} \Lambda') - 2\, tr(\Lambda C_{\boldsymbol{\mathcal{X}}\mathcal{X}_{\boldsymbol{s}_0}}) + 2\, tr(\zeta_1(\Lambda\, 1 - I)). \tag{9}$$

Kriging system To minimize functional (9) we compute its differential with respect to Λ and ζ_1.

$$\begin{aligned}
\Phi_\Lambda \quad &: \quad \mathcal{M}_{\boldsymbol{\mathcal{X}}} \ni h \mapsto \Phi_\Lambda(h) = 2\, tr(h(C_{\boldsymbol{\mathcal{X}}} \Lambda' - C_{\boldsymbol{\mathcal{X}}\mathcal{X}_{\boldsymbol{s}_0}} + 1\zeta_1)) \qquad (10) \\
\Phi_{\zeta_1} \quad &: \quad \mathcal{M}_{\mathcal{X}_0} \ni g \mapsto \Phi_{\zeta_1}(g) = 2\, tr(g(\Lambda\, 1 - I)).
\end{aligned}$$

Differentials (10) lead to the following Kriging system of operatorial equations:

$$\begin{cases} \Lambda C_{\boldsymbol{\mathcal{X}}} - C_{\mathcal{X}_{\boldsymbol{s}_0}\boldsymbol{\mathcal{X}}} + \zeta_1\, 1' = 0; \\ \Lambda\, 1 - I = 0, \end{cases} \tag{11}$$

where $1' : \mathcal{H}^n \to \mathcal{H}$ acts as $\mathcal{H}^n \ni (x_1, ..., x_n)^T \mapsto 1'x = \sum_{i=1}^n x_i$. System (11) admits the following matrix representation

$$\begin{pmatrix} \Lambda & \zeta_1 \end{pmatrix} \begin{pmatrix} C_{\boldsymbol{\mathcal{X}}} & 1 \\ 1' & 0 \end{pmatrix} = \begin{pmatrix} C_{\mathcal{X}_{s_0}\boldsymbol{\mathcal{X}}} & I \end{pmatrix}. \tag{12}$$

We note that system (12) is consistent with its finite dimensional counterpart.

An explicit solution to the Kriging system To derive an explicit solution of Kriging system (12), we exploit the following identity

$$\begin{pmatrix} C_{\boldsymbol{\mathcal{X}}} & 1 \\ 1' & 0 \end{pmatrix}^{-1} = \begin{pmatrix} C_{\boldsymbol{\mathcal{X}}}^{-1}[I - 1(1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1}1'C_{\boldsymbol{\mathcal{X}}}^{-1}] & C_{\boldsymbol{\mathcal{X}}}^{-1}1(1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1} \\ (1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1}1'C_{\boldsymbol{\mathcal{X}}}^{-1} & -(1'C_{\boldsymbol{\mathcal{X}}}^{-1}1)^{-1} \end{pmatrix}, \tag{13}$$

which can be proved by direct verification. Identity (13) leads to the following explicit solution of system (12)

$$\Lambda^* x = M^* x + L(x - 1\,M^* x), \quad x \in \mathcal{H}^n, \tag{14}$$

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 $T_L = C_{\mathcal{X}_{s_0}\boldsymbol{\mathcal{X}}} C_{\boldsymbol{\mathcal{X}}}^{-1/2}$. We recognize in expression (14) the same form as (5), since the operator L is the $\mu_{\boldsymbol{\mathcal{X}}}$-mlt defining the conditional expectation in (4). Moreover, M^*x plays the role of the mean appearing in (5): 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 7 (Operatorial Ordinary Kriging of the Mean)**.** *Given $\boldsymbol{\mathcal{X}} = (\mathcal{X}_{s_1}, ..., \mathcal{X}_{s_n})^T$ 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*

$$\mathbb{E}[\|\mathcal{X}_{s_0} - m^M\|^2_{\mathcal{H}}] \quad \textit{subject to} \quad \mathbb{E}[m^M] = m,$$

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

The solution of Problem 7 is obtained by following the same strategy introduced to solve Problem 6 (not shown).

Therefore, the operatorial Ordinary Kriging predictor $\mathcal{X}^*_{s_0} = \Lambda^*\boldsymbol{\mathcal{X}}$ is found by summing the estimated mean $m^* = M^*\boldsymbol{\mathcal{X}}$ to 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 $(\boldsymbol{\mathcal{X}} - 1M^*\boldsymbol{\mathcal{X}})$.

5 Operatorial Kriging as a unifying theory: finite dimensional approximations

In this Subsection, we focus on characterizing the existing Kriging formulations within the general framework here introduced. To this end, we introduce finite-dimensional approximations of the Operatorial Ordinary Kriging Predictor derived in Section 4.

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 8 (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.

5.1 A Cokriging solution

To derive a version of the discretized predictor, we note that the solution of Problem 8 can be expressed in the form (6), as 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 ι being an isometry, one has

$$\mathbb{E}[\|\mathcal{X}_{\boldsymbol{s}_0}^K - \mathcal{X}_{\boldsymbol{s}_0}^{\Lambda^K}\|_{\mathcal{H}}^2] = \mathbb{E}[\|\boldsymbol{\xi}_{\boldsymbol{s}_0} - \boldsymbol{\xi}_{\boldsymbol{s}_0}^*\|_{\mathbb{R}^K}^2], \tag{15}$$

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 with $\boldsymbol{\xi}_{\boldsymbol{s}}$ 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 Λ_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

$$\mathcal{X}_{\boldsymbol{s}_0}^{\Lambda, K} = \sum_{i=1}^n \sum_{j=1}^K \sum_{l=1}^K (\Lambda_i^K)_{jl} \xi_{\boldsymbol{s}_i, j} v_l, \tag{16}$$

and the unbiasedness constraint as

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

which reduces to

$$\sum_{i=1}^{n} (\Lambda_i^K)_{jl} = \begin{cases} 0, & j \neq l; \\ 1, & j = l. \end{cases} \tag{17}$$

We recognize in condition (17), predictor (16) and in the quadratic loss (15), the corresponding counterparts found in classical multivariate Ordinary Cokriging. Therefore, Problem 8 reduces to a multivariate Ordinary Cokriging (Cressie, 1993) of the 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

$$\begin{pmatrix} \mathbb{C}_{11} & \cdots & \mathbb{C}_{1,n} & \mathbb{I}_K \\ \vdots & \ddots & \vdots & \vdots \\ \mathbb{C}_{n1} & \cdots & \mathbb{C}_{nn} & \mathbb{I}_K \\ \mathbb{I}_K & \cdots & \mathbb{I}_K & 0 \end{pmatrix} \begin{pmatrix} \mathbb{L}_1 \\ \vdots \\ \mathbb{L}_n \\ \mathbb{Z} \end{pmatrix} = \begin{pmatrix} \mathbb{C}_{01} \\ \vdots \\ \mathbb{C}_{0n} \\ \mathbb{I}_K \end{pmatrix}, \tag{18}$$

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. Therefore, the explicit solution to the Kriging problem in RKHS proposed by Nerini et al. (2010) is found by embedding this result into a finite dimensional L^2 space.

5.2 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 (8) is equivalently found as solution of the following Problem.

Problem 9 (Ordinary Trace-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, in the trace sense, for* $\mathcal{X}_{\boldsymbol{s}_0}^K$, *i.e.,* $\mathcal{X}_{\boldsymbol{s}_0}^{\boldsymbol{\lambda},K*} = \sum_{i=1}^n \lambda_i^* \mathcal{X}_{\boldsymbol{s}_i}^K$, *where* $\lambda_1^*, ..., \lambda_n^* \in \mathbb{R}$ *minimize*

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

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

The solution of Problem 9 can be found through the methodology proposed in (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., through 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). Moreover, these authors prove that the unique solution of Problem 9 can be found by solving the linear system

$$\begin{pmatrix} C_{tr}(\mathbf{0}) & \cdots & C_{tr}(\boldsymbol{s}_1 - \boldsymbol{s}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{tr}(\boldsymbol{s}_n - \boldsymbol{s}_1) & \cdots & C_{tr}(\mathbf{0}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \zeta_{tr} \end{pmatrix} = \begin{pmatrix} C_{tr}(\boldsymbol{s}_0 - \boldsymbol{s}_1) \\ \vdots \\ C_{tr}(\boldsymbol{s}_0 - \boldsymbol{s}_n) \\ 1 \end{pmatrix}, \tag{19}$$

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 9, as opposed to the Operatorial Kriging predictor solving Problem 8, detailed in Subsection 5.1.

To accommodate the form of the Cokriging predictor in expression (16) and to ease the comparison, we rewrite the Trace-Kriging predictor $\mathcal{X}_{\boldsymbol{s}_0}^{\boldsymbol{\lambda},K}$ as

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

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 9 reads $\sum_{i=1}^n \lambda_i = 1$, which is equivalently expressed in terms of $(\Lambda_i^K)_{jl}$ as (17). Therefore, the solution of Problem 9 is an admissible solution of Problem 8. To prove the equivalence of the two solutions, one is left to prove that the solution of Problem 8 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.,

$$\text{rank} \begin{pmatrix} \xi_{\boldsymbol{s_1},1} & \cdots & \xi_{\boldsymbol{s_1},1} \\ \vdots & \ddots & \vdots \\ \xi_{\boldsymbol{s_n},1} & \cdots & \xi_{\boldsymbol{s_n},K} \end{pmatrix} = K \quad a.s. \tag{20}$$

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.

6 Conclusion and further research

In this work we established a novel theoretical framework for Object Oriented Operatorial Kriging, which is 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, it is possible to develop a comprehensive theory of spatial prediction in Hilbert spaces by relying on the notions on measurable linear transformations. Similar to the finite-dimensional context, 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, based on the same formal basis leading to the Operatorial Simple Kriging. Here, we focused on stationary Gaussian random fields. In this setting, we were able to formally and explicitly derive an Operatorial Ordinary Kriging predictor, based on the isometry between the space of measurable linear transformation and that of Hilbert-Schmidt operators.

3. We showed the unifying nature of our new theoretical framework through the notion of discretized kriging problem. The attained results appear particularly relevant, as (a) we recovered consistency with the predictor of (Nerini et al., 2010) when a finite-dimensional L^2 space is considered and (b) we proved that the best approximation of the Operatorial Ordinary Kriging problem in the space generated by the observations coincides almost surely with the Trace-Kriging obtained via trace-variography (Menafoglio et al., 2013). As such, our new setting includes the Kriging methods of FDA, besides those of CoDa and multivariate analysis. The generalization of these results to Banach space constitutes an on-going research along this line.

4. The attained theoretical results are key to address a number of computational issues in both functional and multivariate settings. Based on the results in Subsection 5.2 – which hold in most practical situations –, one can actually compute the K-dimensional approximation of the Operatorial Ordinary Kriging predictor equivalently by (18) or (19). We remark that the difference in terms of complexity is relevant: solving Trace-Kriging system (19) has a complexity $\mathcal{O}(n^3)$, as opposed to the complexity $\mathcal{O}(n^3(K+1)^3)$ of solving the Cokriging system (18). Moreover, the optimal linear predictor can be obtained by relying on trace-covariography only, instead of the complete characterization of the cross-covariance operator required in system (18). We remark that these observations hold for any Hilbert space, an particularly for the Euclidean space $\mathbb{R}^p$, $1 \leq p < \infty$, where multivariate geostatistics is performed. Hence, our results – although worked out in a complete generality – appears to be influential even in the classical setting. In fact, the trace-covariography is posed as a convenient alternative to the linear model of coregionalization which is commonly employed to fit the cross-variograms in multivariate geostatistics, guaranteeing the same degree of precision on the results. These observations open new and relevant perspectives for the Cokriging of large and high-dimensional datasets, which is one of the most challenging topic in modern geostatistics.

References

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

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

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

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

G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions.* Cambridge University Press, second edition, 2014.

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

P. Delicado. Dimensionality reduction when data are density functions. *Computational Statistics & Data Analysis*, 55(1):401 – 420, 2011.

P. Delicado, R. Giraldo, C. Comas, and J. Mateu. Statistics for spatial functional data. *Environmetrics*, 21(3-4):224–239, 2010.

J. J. Egozcue, J. L. Díaz-Barrero, and V. Pawlowsky-Glahn. Hilbert space of probability density functions based on aitchison geometry. *Acta Mathematica Sinica, English Series*, 22(4):1175–1182, Jul. 2006.

R. Giraldo. *Geostatistical Analysis of Functional Data.* PhD thesis, Universitat Politècnica da Catalunya, Barcellona, 2009.

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

H. Luschgy. Linear estimators and radonifying operators. *Theory of Probability & Its Applications*, 40(1):167–175, 1996. doi: 10.1137/1140017.

A. Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Zeitschrift fÃijr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(3):385–397, 1984. doi: 10.1007/BF00533743.

J. Steve Marron and AndrÃľs M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 2014. ISSN 1521-4036. doi: 10.1002/bimj.201300072. URL `http://dx.doi.org/10.1002/bimj.201300072`.

A. Menafoglio, P. Secchi, and M. Dalla Rosa. A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. *Electronic Journal of Statistics*, 7:2209–2240, 2013. doi: 10.1214/13-EJS843.

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

A. Menafoglio, P. Secchi, and A. Guadagnini. A class-kriging predictor for functional compositions with application to particle-size curves in heterogeneous aquifers. MOX-report 58/2014, Politecnico di Milano, 2014b.

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

J. Picard. *Séminaire de Probabilités XLIII*, chapter Representation Formulae for the Fractional Brownian Motion, pages 1–69. Springer, 2006.

J. Ramsay and B. Silverman. *Functional data analysis.* Springer, New York, second edition, 2005.

K. G. van den Boogaart, J. J. Egozcue, and V. Pawlowsky-Glahn. Bayes hilbert spaces. *Australian & New Zealand Journal of Statistics*, 56:171–194, 2014.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

63/2015 Pini, A.; Stamm, A.; Vantini, S.
Hotelling s T^2 in functional Hilbert spaces

64/2015 Menafoglio, A.; Petris, G.
Kriging for Hilbert-space valued random fields: the Operatorial point of view

62/2014 Andr, C.; Parolini, N.; Verani, M.
Using gambling simulators to foster awareness about gambling risks

61/2014 Taffetani, M.; Ciarletta, P.
Elasto-capillarity controls the formation and the morphology of beads-on-string structures in solid fibres

59/2014 Manzoni, A.; Pagani, S.; Lassila, T.
Accurate solution of Bayesian inverse uncertainty quantification problems using model and error reduction methods

57/2014 Giverso, C.; Verani, M.; Ciarletta P.;
Branching instability in expanding bacterial colonies

58/2014 Menafoglio, A.; Secchi, P.; Guadagnini, A.
A Class-Kriging predictor for Functional Compositions with Application to Particle-Size Curves in Heterogeneous Aquifers

60/2014 Signorini, M.; Micheletti, S.; Perotto, S.
CMFWI: Coupled Multiscenario Full Waveform Inversion for seismic inversion

55/2014 Antonietti, P. F.; Houston P.; Sarti, M.; Verani, M.
Multigrid algorithms for hp-version Interior Penalty Discontinuous Galerkin methods on polygonal and polyhedral meshes

56/2014 Antonietti, P. F.; Sarti, M.; Verani, M.; Zikatanov, L. T.
A uniform additive Schwarz preconditioner for the hp-version of Discontinuous Galerkin approximations of elliptic problems