

1

Mathematical foundations of functional Kriging in Hilbert spaces and Riemannian manifolds

Alessandra Menafoglio, Davide Pigoli and Piercesare Secchi

1.1

Introduction

Modern field studies yield diverse types of observations, in the form of highly heterogeneous and high-dimensional data. In this context, environmental observations are routinely available in the form of functional or distributional data. For instance, these kinds of variables are found in climatic investigations, where complex data are regularly collected at different sites in the study region. Examples are temperature profiles along the year, or the precipitation distributions, which are key to characterize and classify the domain of interest from the climatic viewpoint.

In these cases, the object of the analysis is often infinite-dimensional, i.e., it would need an infinity of point evaluations to be fully characterized. In some cases, it is constrained, e.g., when distributional data are concerned. In fact, the full interpretation and statistical treatment of such kinds of complex data poses relevant challenges for geoscience applications.

In this broad context, a relatively large body of recent literature has been devoted to the mathematical foundations of geostatistics for complex data, with particular reference to data embedded in Hilbert spaces and Riemannian manifolds. We here focus on the approach developed within the area of *Object Oriented Spatial Statistics* (O2S2, [1]), which was developed starting from the works [2, 3, 4]. The foundational idea of the approach is to interpret data as *objects*: the *atom* of the geostatistical analysis is the entire object, which is seen as an indivisible unit rather than a collection of features. In this view, the observations are interpreted as random points within a space of objects – called *feature space* – whose dimensionality and geometry should properly represent the data features and their possible constraints. The O2S2 approach follows the funding ideas of Object Oriented Data Analysis (OODA, [5]), and generalizes the theory of functional geostatistics developed from the seminal works of [6, 7, 8], mainly for functional data in L^2 .

Amongst the challenges related to the spatial analysis of complex data, we here

focus on the problems of spatial prediction. Similarly as in classical geostatistics [e.g. 9], in O2S2 the latter problem is addressed by formulating optimal unbiased predictors linear in the data. We here review the mathematical framework for kriging Hilbert and manifold data, in stationary or non-stationary settings, and discuss the estimators that can be used for the mean and the covariance structure.

The remaining of this Chapter is organized as follows. Section 1.2 introduces the main definitions and assumptions which may be formulated to perform a geostatistical analysis of Hilbert-space valued random fields. Section 1.3 describes a global approach to Kriging, interpreted as optimal linear combinations of the data. Here, we show an example of application to climate data, and thoroughly discuss on the importance of selecting an appropriate feature space for the analysis. In Section 1.4, we briefly review an alternative approach to Kriging, which arises when the predictor is interpreted in a more general sense, grounding on the theory of measurable linear transformations. Interestingly, this general theory allows to draw connections between several different formulations of functional Kriging available in the literature. Section 1.5 introduces the methodologies to perform geostatistical analysis of manifold-valued random fields, based on the local-approximation property of such spaces. Here, for illustration we consider the case of positive definite matrices, that is used to analyze and predict the field of covariance matrices between temperature and precipitations in a region of Canada.

1.2

Definitions and assumptions

We call $(\Omega, \mathfrak{F}, \mathbb{P})$ a probability space and \mathcal{H} a separable Hilbert space, endowed with operations $(+, \cdot)$, and an inner product $\langle \cdot, \cdot \rangle$. The space \mathcal{H} will indicate the feature space for the geostatistical analysis: we will consider the data as realizations of random points in \mathcal{H} . In several cases in this chapter, the space \mathcal{H} will represent a space whose elements are real-valued functions defined over a compact interval. Nevertheless the theory presented in this Chapter is entirely general, and may involve manifold data, as we shall show in Section 1.5.

In the following, we denote by χ a random element in \mathcal{H} , that is a measurable function defined on $(\Omega, \mathfrak{F}, \mathbb{P})$ and valued in \mathcal{H} , $\chi : \Omega \rightarrow \mathcal{H}$. We indicate a realization of χ – that is a non-random element of \mathcal{H} – with the symbol χ , i.e., $\chi = \chi(\omega)$, for $\omega \in \Omega$. We call $\mathcal{L}(\mathcal{H}, \mathcal{H}_1)$ the Banach space of continuous linear operators on \mathcal{H} in \mathcal{H}_1 . We say that two random elements χ_1, χ_2 are equivalent (indicated by $\chi_1 \stackrel{\mathcal{H}}{=} \chi_2$, or $\chi_1 = \chi_2$ for short) if $\chi_1 = \chi_2$ almost surely.

Given a set of locations s_1, \dots, s_n in a spatial domain $D \subset \mathbb{R}^d$ (usually $d = 2, 3$), we denote by $\chi_{s_1}, \dots, \chi_{s_n}$ the set of observations collected at these locations, that form our dataset of spatially dependent objects. As in classical geostatistics [e.g., 9], we assume this dataset to be a partial observation of a random field $\{\chi_s, s \in D\}$ on $(\Omega, \mathfrak{F}, \mathbb{P})$ in \mathcal{H} . The latter is defined as a collection of random elements χ_s , indexed by a continuous spatial vector s varying in D .

In this chapter, we will always assume that, for all $s \in D$, the element χ_s , satisfies

$\mathbb{E}[\|\chi_s\|^2] < \infty$. Under the latter assumption, one can define the expected value of the field in terms of Bochner integral as

$$m_s = \int_{\Omega} \chi_s(\omega) \mathbb{P}(d\omega), \quad s \in D. \quad (1.1)$$

In \mathcal{H} , the expected value (1.1) can be equivalently defined as the element m_s of \mathcal{H} such that, for any $x \in \mathcal{H}$, $\langle x, m_s \rangle = \mathbb{E}[\langle x, \chi_s \rangle]$. The elements m_s , $s \in D$, describe the first order structure of the field.

The second order structure can be fully characterized through the *spatial covariance function* \mathcal{C} , that associates with each pairs of locations (s_1, s_2) , the cross-covariance operator between the random elements at those locations, i.e.,

$$\begin{aligned} \mathcal{C} : D \times D &\rightarrow \mathcal{L}(\mathcal{H}, \mathcal{H}) \\ (s_1, s_2) &\mapsto \{\mathcal{C}(s_1, s_2) : \mathcal{H} \rightarrow \mathcal{H}, x \mapsto \mathbb{E}[\langle (\chi_{s_1} - m_{s_1}), x \rangle (\chi_{s_2} - m_{s_2})]\}. \end{aligned} \quad (1.2)$$

A global measure of spatial dependence is provided by the *trace-covariogram*, that is the real-valued function that associates with any pair of locations (s_1, s_2) the trace of the corresponding cross-covariance operator $\mathcal{C}(s_1, s_2)$, i.e. $C : D \times D \rightarrow \mathbb{R}$ such that

$$C(s_1, s_2) = \sum_{k=1}^{\infty} \langle \mathcal{C}(s_1, s_2) e_k, e_k \rangle = \mathbb{E}[\langle \chi_{s_1} - m_{s_1}, \chi_{s_2} - m_{s_2} \rangle], \quad (1.3)$$

where $\{e_k, k \in \mathbb{N}\}$ is any orthonormal basis of \mathcal{H} , and the second equality was proven in [2]. Note that the function C is well-defined, since $\mathcal{C}(s_1, s_2)$ is a trace-class Hilbert-Schmidt operator [10], thus the series converges absolutely for any orthonormal basis $\{e_k, k \geq 1\}$ of \mathcal{H} and the sum does not depend on the choice of the basis [11, Theorem 1.24]. The trace-covariogram fulfills all the properties of a classical covariogram (e.g., it is a symmetric, positive-definite function, [9]). It was defined in the context of L^2 data in [8], and then generalized to object data in any separable Hilbert space \mathcal{H} in [2].

The trace-covariogram is strictly related with a counterpart of the classical variogram, named *trace-variogram*, that is defined as the function $2\gamma : D \times D \rightarrow \mathbb{R}^+$ that maps any pair of locations (s_1, s_2) as

$$2\gamma(s_1, s_2) = \mathbb{E}[\|\chi_{s_1} - \chi_{s_2}\|^2] - \|m_{s_1} - m_{s_2}\|^2. \quad (1.4)$$

The trace-variogram fulfills classical properties, such as being a conditionally negative definite function [e.g., 2].

On these premises, definitions of stationarity can be stated for the random field $\{\chi_s, s \in D\}$. In particular, we here focus on definitions of second-order stationarity in a strong sense (Definition 1.2.1) and in a global sense (Definition 1.2.2). The interested reader can find weaker definitions of stationarity in [2].

Definition 1.2.1 *A process $\{\chi_s, s \in D\}$ is said to be strongly second-order stationary if the following conditions hold:*

- (i) $\mathbb{E}[\chi_s] = m$, for all $s \in D \subseteq \mathbb{R}^d$ (spatially constant mean);
- (ii) $\mathbb{E}[\langle \chi_{s_1} - m, \cdot \rangle (\chi_{s_2} - m)] = C(s_1 - s_2)$ for all $s_1, s_2 \in D \subseteq \mathbb{R}^d$ (spatial covariance function depending only on the increment vector).

Definition 1.2.2 A process $\{\chi_s, s \in D\}$ is said to be (globally) second-order stationary if the following conditions hold:

- (i) $\mathbb{E}[\chi_s] = m$, for all $s \in D \subseteq \mathbb{R}^d$ (spatially constant mean);
- (ii') $\mathbb{E}[\langle \chi_{s_1} - m, \chi_{s_2} - m \rangle] = C(s_1 - s_2)$ for all $s_1, s_2 \in D \subseteq \mathbb{R}^d$ (trace-covariogram depending only on the increment vector).

Second-order stationarity thus concerns a spatial homogeneity in the first and second order structure of the field. It should be noted that stationarity does not imply the existence of a directional homogeneity, which concerns the concept of isotropy instead. Indeed, a strongly second order stationary field is said to be isotropic if condition (ii) is reinforced by the following condition (iii)

- (iii) $\mathbb{E}[\langle \chi_{s_1} - m, \cdot \rangle (\chi_{s_2} - m)] = C(\|s_1 - s_2\|_d)$, for all $s_1, s_2 \in D \subseteq \mathbb{R}^d$, $\|\cdot\|_d$ being the norm on \mathbb{R}^d (spatial covariance function depending only on the distance between locations).

A globally second order stationary field is said to be isotropic if condition (ii') is reinforced by the following condition (iii')

- (iii') $\mathbb{E}[\langle \chi_{s_1} - m, \chi_{s_2} - m \rangle] = C(\|s_1 - s_2\|_d)$ for all $s_1, s_2 \in D$ (trace-covariogram depending only on the distance between locations).

Both strong and global second-order stationarity are of interest from the application-oriented viewpoint. Indeed, the methods introduced in Sections 1.3.1, 1.3.2 rely upon the assumption of global stationarity, while the methods devised in Section 1.4 assume the stronger condition of strong second-order stationarity (and Gaussianity). We finally remark that, although assuming isotropy greatly simplifies the notation, it should not be considered as essential for the development of the methods described in the next sections.

1.3

Kriging prediction in Hilbert space: a trace approach

A key goal of a typical geostatistical analysis is to perform spatial prediction at a target (unobserved) location. As long as one-dimensional Euclidean fields are concerned, classical geostatistics literature advocates the use of a kriging predictor, that is the Best Linear Unbiased Predictor (BLUP) $\chi_{s_0}^* = \sum_{i=1}^n \lambda_i^* \cdot \chi_{s_i}$, whose weights minimize the variance of prediction error under the unbiasedness constraint [e.g., 9]. In fact, in the scalar case, no ambiguity exists in the definition of linear predictor, as this can be equivalently interpreted either as a linear combination of the observations, or as a linear transformation of the vector of observations. Instead, when the

feature space \mathcal{H} is an infinite-dimensional Hilbert space, several possible definitions of kriging are available. In this section, we focus on the so-called *trace-approach*, that defines the kriging predictor as the best linear combination of the data, as presented in [8] for the stationary L^2 setting, and further generalized in [2] for possibly non-stationary Hilbert data.

1.3.1

Ordinary and Universal Kriging in Hilbert spaces

Given s_1, \dots, s_n in D and the observations of the field $\chi_{s_1}, \dots, \chi_{s_n}$ at these location, we look for the best linear unbiased predictor $\chi_{s_0}^* = \sum_{i=1}^n \lambda_i^* \cdot \chi_{s_i}$ for χ_{s_0} , where the weights $\lambda_1^*, \dots, \lambda_n^*$ solve the minimization problem

$$\min_{\lambda_1, \dots, \lambda_n \in \mathbb{R}} \mathbb{E} \left[\left\| \chi_{s_0} - \sum_{i=1}^n \lambda_i \cdot \chi_{s_i} \right\|^2 \right] \quad \text{subject to} \quad \mathbb{E} \left[\sum_{i=1}^n \lambda_i \cdot \chi_{s_i} \right] = \mathbb{E}[\chi_{s_0}]. \quad (1.5)$$

In the presence of second-order stationarity, one may employ an ordinary (trace-) kriging predictor, while for non-stationary data, universal (trace-) kriging may be employed instead. We here consider universal kriging in \mathcal{H} , following [2], since ordinary kriging is obtained as a special case.

We represent the elements of the field $\{\chi_s, s \in D\}$ as $\chi_s = m_s + \delta_s$, where m_s is the drift – which describes a possibly non-constant mean variation – whereas δ_s is assumed to be a globally second-order stationary and isotropic random field with zero-mean and trace-covariogram C . Following the approach of universal kriging for scalar data, we describe the drift term through a linear model with scalar regressors

$$m_s = \sum_{l=0}^L f_l(s) \cdot a_l, \quad s \in D, \quad (1.6)$$

where $f_0(s) = 1$ for all $s \in D$, $f_l, l = 1, \dots, L$, are known over the entire domain and $a_l, l = 0, \dots, L$ are (possibly unknown) coefficients in \mathcal{H} . Note that the stationary case is obtained when $L = 0$ as in that case the mean is spatially constant. Further, the spatial variation is assumed to be entirely captured by the regressors $\{f_l, l = 1, \dots, L\}$, since the coefficients do not depend on the location $s \in D$. Note that other approaches to model the non-stationarity of the mean are possible, e.g., based on (scalar or functional) covariates collected together with the data (i.e., the kriging with external drift proposed in [?], and discussed in Chapter ZZZ (Ignaccolo et al.)).

In our setting, the unbiasedness constraint in (1.5) reads

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

which is included in the minimization functional through $L+1$ Lagrange multipliers. Hence, problem (1.5) becomes that of minimizing, w.r.t. $\lambda_i, \zeta_l, i = 1, \dots, n, l =$

6 |

0, \dots, L,

$$\begin{aligned} \Phi = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\|s_i - s_j\|_d) + C(0) - 2 \sum_{i=1}^n \lambda_i C(\|s_i - s_0\|_d) \\ + 2 \sum_{l=0}^L \zeta_l \left(\sum_{i=1}^n \lambda_i f_l(s_i) - f_l(s_0) \right). \end{aligned} \quad (1.7)$$

Denote by $\Sigma \in \mathbb{R}^{n \times n}$ the variance-covariance matrix of the observations (in the trace sense), whose (i, j) -th element is $\Sigma_{i,j} = C(\|s_i - s_j\|_d)$ for $i, j = 1, \dots, n$, $C(\|s_i - s_j\|_d)$ appearing in (1.5). Indicate with $\mathbb{F} = (f_l(s_i)) \in \mathbb{R}^{n \times (L+1)}$ the design matrix of the linear model (1.6), by $\vec{\lambda} = (\lambda_1, \dots, \lambda_n)^T$ the vector of weights and $\vec{\zeta} = (\zeta_0, \dots, \zeta_L)^T$ the vector of Lagrange multipliers. Call $\vec{\sigma}_0 = (C(\|s_1 - s_0\|_d), \dots, C(\|s_n - s_0\|_d))^T$ the vector of (trace-) covariances between the observations and the target, and $\vec{f}_0 = (f_0(s_0), \dots, f_L(s_0))^T$ the vector of regressors at the target location (both appearing in (1.5)). With this notation, and under usual assumptions on the sampling design – namely Σ positive definite and \mathbb{F} of full rank – the functional (1.7) admits a unique minimum, found by solving the linear system

$$\begin{pmatrix} \Sigma & \mathbb{F} \\ \mathbb{F}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \vec{\lambda} \\ \vec{\zeta} \end{pmatrix} = \begin{pmatrix} \vec{\sigma}_0 \\ \vec{f}_0 \end{pmatrix}. \quad (1.8)$$

The latter system is easily found by equating to zero the differential of Φ w.r.t. the λ 's and the ζ 's.

The variance of the prediction error (i.e., the value of (1.7) at the optimum) is called *universal kriging variance*, and is obtained as

$$\begin{aligned} \sigma_{UK}^2(s_0) &= C(0) - \sum_{i=1}^n \lambda_i^* C(\|s_i - s_0\|_d) - \sum_{l=0}^L \zeta_l^* f_l(s_0) \\ &= \sum_{i=1}^n \lambda_i^* \gamma(\|s_i - s_0\|_d) + \sum_{l=0}^L \zeta_l^* f_l(s_0), \quad s_0 \in D. \end{aligned} \quad (1.9)$$

The latter quantifies the uncertainty of the prediction and can be used to build prediction bands (e.g., by using the Chebychev inequality). Nonetheless, it should be noted that it does not consider the variability of the possible estimate of the covariance structure, as the latter is assumed to be known over the entire construction. Instead, in most cases, the trace-covariogram needs to be estimated as well, and the estimated covariance is eventually plugged-in in (1.8). Classical geostatistics advocates the estimate of the (trace-)variogram in place of the (trace-)covariogram. The two functions are linked by the relation

$$2\gamma(h) = C(0) - C(h), \quad h \in D.$$

Estimators of the variogram are generally more robust than those of covariogram, hence preferred.

To estimate the variogram, a two step procedure is generally employed: (1) estimate an empirical variogram, and (2) fit a parametric model (e.g., spherical, exponential, Matérn) to the estimate at point (1), in order to guarantee that the properties of a valid variogram (e.g., conditional negative definiteness) are fulfilled. If global second-order stationarity and isotropy holds true, one can use the method-of-moment estimator to achieve point (1) [8, 2]

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{(i,j) \in N(h)} \|\chi_{s_i} - \chi_{s_j}\|^2, \quad (1.10)$$

where $N(h) = \{(i, j) \mid h - \Delta h \leq \|s_i - s_j\|_d \leq h + \Delta h\}$, and $|N(h)|$ is its cardinality. Estimator (1.10) provides an unbiased estimate of the trace-variogram 2γ only if the assumptions of stationarity and isotropy are verified. Otherwise, such estimator should not be employed, because it considers only the first term of the variogram definition in (1.4) (i.e., $\mathbb{E}[\|\chi_{s_1} - \chi_{s_2}\|^2]$), while it neglects the additional term $\|m_{s_1} - m_{s_2}\|$ (which is null under stationarity). When considering the non-stationary model here introduced, one can use estimator (1.10) on the residuals δ_{s_i} , $i = 1, \dots, n$, which are a partial observation of a globally-second order stationary and isotropic process. Nevertheless, these are usually latent, and can be estimated by difference once the drift has been assessed. Hence, although in principle one could perform universal kriging without having estimated the drift in advance, whenever the trace-covariogram is unknown, providing a good estimate of the drift is essential. Section 1.3.2 will be dedicated to this point. Instead, in case of stationarity (and isotropy), estimating the drift is not required for the purpose of performing spatial prediction.

1.3.2

Estimating the drift

The problem of estimating the drift for the spatial model here considered consists in estimating a linear model in the presence of spatially correlated random errors. Indeed, under model (1.6), the model for the data is

$$\chi_{s_i} = \sum_{l=0}^L f_l(s_i) \cdot a_l + \delta_{s_i}. \quad (1.11)$$

To set the notation, we call \mathcal{H}^n the Hilbert space $\mathcal{H} \times \dots \times \mathcal{H}$, with the inner product $\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^n \langle x_i, y_i \rangle$, for $\vec{x} = (x_1, \dots, x_n)^T \in \mathcal{H}^n$, $\vec{y} = (y_1, \dots, y_n)^T \in \mathcal{H}^n$. We denote by $\vec{\chi} = (\chi_{s_1}, \dots, \chi_{s_n})^T \in \mathcal{H}^n$, $\vec{a} = (a_0, \dots, a_L)^T \in \mathcal{H}^{L+1}$, and $\vec{\delta} = (\delta_{s_1}, \dots, \delta_{s_n})^T \in \mathcal{H}^n$. In this setting, model (1.11) can be expressed in matrix form as $\vec{\chi} = \mathbb{F}\vec{a} + \vec{\delta}$.

The theory of functional linear models was developed under the founding hypothesis of independent and identically distributed residuals. As a consequence, in the presence of correlated residuals, the ordinary least squares approach developed in that framework turns out to provide suboptimal results. To properly account for the

spatial dependence, a generalized least squares criterion (GLS) can be used instead [2]. The latter seeks to minimize the functional Mahalanobis distance between the observations and the evaluation of the drift at the sampled locations. That is, the GLS estimator for vector \vec{a} is found as the solution of the minimization problem

$$\min_{\vec{a} \in \mathcal{H}^{L+1}} \sum_{i=1}^n \|\Sigma^{-1/2} \circ (\vec{\mathcal{X}} - \mathbb{F} \vec{a})_i\|^2, \quad (1.12)$$

\circ indicating the matrix multiplication in \mathcal{H} , that is $[A\vec{x}]_i = \sum_{j=1}^m A_{ij} \cdot x_j$, for $i = 1, \dots, q$, with $\vec{x} \in \mathcal{H}^m$, $A \in \mathbb{R}^{q \times m}$. If $\text{rank}(\mathbb{F}) = L + 1 \leq n$ and $\text{rank}(\Sigma) = n$, problem (1.12) is well-posed and its unique solution is found as [2]

$$\hat{\vec{a}}^{GLS} = (\mathbb{F}^T \Sigma^{-1} \mathbb{F})^{-1} \mathbb{F}^T \Sigma^{-1} \circ \vec{\mathcal{X}}. \quad (1.13)$$

Estimator (1.13) was proved to be the best linear unbiased estimator for \vec{a} . However, it depends on the matrix Σ , hence on the trace-covariogram, which can be estimated only once the residuals have been assessed, and the latter residuals can be only computed by difference based on the drift estimate. To jointly assess the residuals and the trace-(co)variogram, one can then resort to an iterative algorithm, initialized, e.g., to the OLS estimate of the drift. Such algorithm usually converges within 5 iterations, although theoretical arguments on its convergence are yet to be proved. Having computed the drift and the trace-variogram, the universal kriging system can be solved by plugging-in the estimated variance-covariance matrix in (1.8).

1.3.3

An example: trace-variogram in Sobolev spaces

We here discuss through an example the key importance of the choice of the ambient space for the (geo)statistical analysis of functional or object data. Although there are cases in which a natural ambient space is available (e.g., suggested by dynamical equations governing the system), the choice of the feature space for the analysis is indeed a crucial modeling step. As a way of example, Figure 1.1 shows a set of spatially dependent curves, simulated from two random fields $\{\chi_s^{(m)}, s \in D\}$, $m = 1, 2$, $D = [0, 2] \times [0, 3]$. The latter fields were built in [2] by combining in different ways the same set of independent, zero mean, second-order stationary and isotropic scalar random fields $\{\xi_{s,j}, s \in D\}$, $j = 1, \dots, 7$, as

$$\chi_s^{(1)} = \sum_{k=1}^7 \xi_{s,k} e_k; \quad \chi_s^{(2)} = \sum_{k=19}^{25} \xi_{s,k-18} e_k, \quad (1.14)$$

where $\{e_k, k \geq 1\}$ denotes the Fourier basis. A detailed description of the simulation setting is provided in [2, , Supplementary material]. Figure 1.1 evidences the very different patterns displayed by the two groups of curves. Indeed, the realizations associated with the field $\{\chi_s^{(2)}, s \in D\}$ show a much higher amplitude variability (i.e., along the vertical direction) than those associated with the field $\{\chi_s^{(1)}, s \in D\}$.

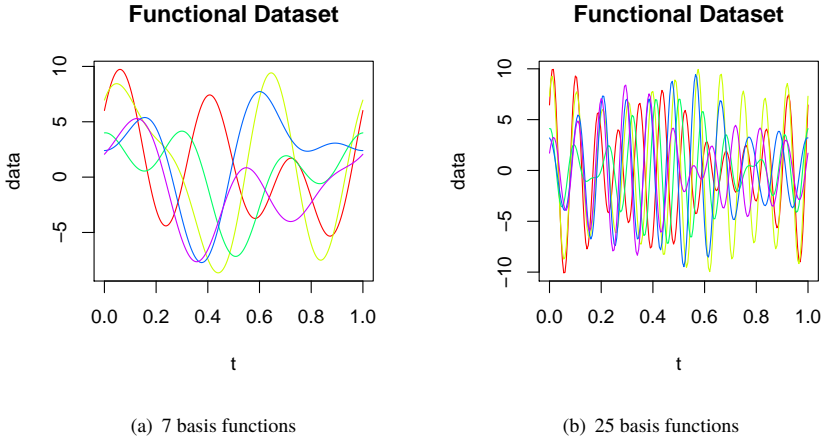


Figure 1.1 Spatially dependent curves simulated from the fields $\{\chi_s^{(1)}, s \in D\}$ (left panel) and $\{\chi_s^{(3)}, s \in D\}$, (right panel). Modified from [2].

This is due to the fact that the second field is built upon a higher order truncation of the basis, and involves only the 10th to 12th frequencies, while only the first 3 frequencies are included in the construction of the first field.

Despite these apparent diversities between the two fields, no difference exists in their spatial dependence structure if the fields are embedded in the space L^2 of square-integrable functions. Indeed, straightforward computations yield, for $m = 1, 2$,

$$2\gamma_{L^2}^{(m)}(s_i, s_j) = \mathbb{E}[\|\chi_{s_i}^{(m)} - \chi_{s_j}^{(m)}\|_{L^2}^2] = \sum_{k=1}^{N_m} \mathbb{E}[|\xi_{s_i, k}^{(m)} - \xi_{s_j, k}^{(m)}|^2] = \sum_{k=1}^7 2\gamma_{\xi_k}$$

γ_{ξ_k} indicating the variogram of the scalar field of coefficients $\{\xi_{s, k}, s \in D\}$, $k = 1, \dots, 7$. Instead, when modeling the data as objects in the Sobolev space H^1 – i.e., the space of functions in L^2 whose derivatives (in a weak sense) are in L^2 – one can capture the diverse behavior of the fields, through the geometry of the space. The latter choice entails the use of a norm based not only on point-wise evaluations, but also on the differential properties of the elements. In such a case, the variogram is indeed different in the two fields, being

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

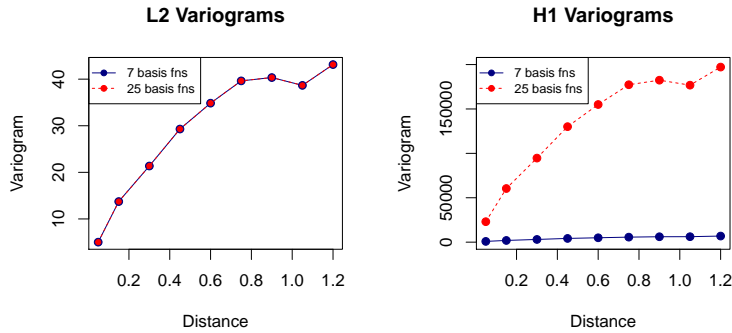


Figure 1.2 Empirical trace-variograms in L^2 (on the left) and H^1 (on the right). Modified from [2].

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

Figure 1.2 shows the empirical trace-variograms estimated in L^2 (Figure 1.2a) and in H^1 (Figure 1.2b) from a sample of 100 observations $\chi_{s_1}^{(m)}, \dots, \chi_{s_{100}}^{(m)}$ from each field $\{\chi_s^{(m)}, s \in D\}$, $m = 1, 2$, the same sites s_1, \dots, s_{100} being uniformly sampled in D . Although the shapes of the variograms are quite similar in the two cases, the orders of magnitude of the horizontal asymptotes — twice the global variance of the process — are significantly different. Indeed, the variogram $2\gamma_{H^1}^{(2)}$ (in red) assumes much higher values than $2\gamma_{H^1}^{(1)}$ (in blue), since the random field $\{\chi_s^{(2)}, s \in D\}$ has a much higher energy. Indeed, in dynamical system theory, the square of the Sobolev norm of the state (i.e., $\|\chi_s\|^2$) coincides with (twice) the energy of the system. Hence, the ambient space for geostatistical analysis not only provides the feature space for the object data, but implies a precise physical meaning for the measure of stochastic variability: the global variance represents (twice) the mean energy of the system, while the trace-variogram (twice) the mean energy of the increments between two states.

In conclusion, one should pay close attention to the choice of the feature space for the analysis. The latter should be guided by the dataset structure, the possible physical laws governing the system and by the purposes of the analysis.

1.3.4

An application to non-stationary prediction of temperatures profiles

We here show an example of application of the trace-approach to non-stationary environmental data. Following [2], the data we consider are daily mean temperature profiles, collected during 1980 at 27 locations in the Maritime Provinces of Canada

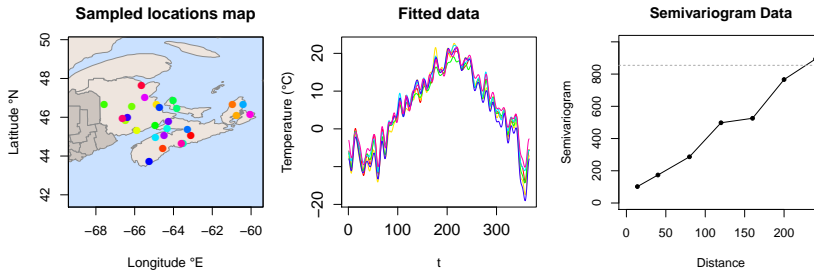


Figure 1.3 Canada's Maritime Provinces Temperatures dataset, year 1980. From left to right: map of Maritime Provinces and sampled locations; 6 fitted temperature curves; empirical estimate of the trace-variogram. Modified from [2].

(data source: Natural Resources of Canada; <http://atlas.nrcan.gc.ca/>). Raw data were smoothed by using a Fourier basis of 65 elements, obtaining the set of curves displayed in Figure 1.3.

For illustration purposes, we performed the geostatistical analysis in L^2 , using on the spatial domain a geodesic distance, since coordinates are given in latitude and longitude. The graphical inspection of the trace-semivariogram estimated from the data suggests the presence of a non-constant drift model. Indeed, the empirical estimate does not show any apparent finite sill (i.e., horizontal asymptote, see Figure 1.3, right panel). To select the drift model, we considered the polynomial models of degree 2 in the coordinates, and sought the one minimizing the kriging prediction error, assessed by leave-one-out cross-validation [2]. On this basis, we found as optimal model

$$m(s, t) = a_0(t) + a_1(t)y + a_2(t)x^2 + a_3(t)y^2 + a_4(t)xy,$$

for $s = (x, y) = (\text{Longitude}, \text{Latitude})$, $t \in \mathcal{T} = [0, 366]$.

Figure 1.4 displays the estimate of the drift for the days of summer solstice and the winter solstice. Note that we may have chosen any day of the year for such representation: the theory here presented allows obtaining joint and consistent estimates (and predictions) for all the days of the year. The maps in Figure 1.4 have a clear climatical interpretation, as they represent the presence of currents from the Ocean, whose circulation causes a change in the gradient of temperature along the year. In Figure 1.5 we represent the universal kriging maps for the same days considered in Figure 1.4. From the analysis of the maps one can conclude that the drift term tends to drive the estimates in the colder seasons (panel of December), whereas during the summer season (panel of June), the temperature map displays evident local patterns, due to the peculiar geographical configuration of the area – particularly for the Bay of Fundy. Although the spatial patterns evidenced during the year tend to be different, the universal kriging predictor allows to properly capture them, thanks to its flexibility.

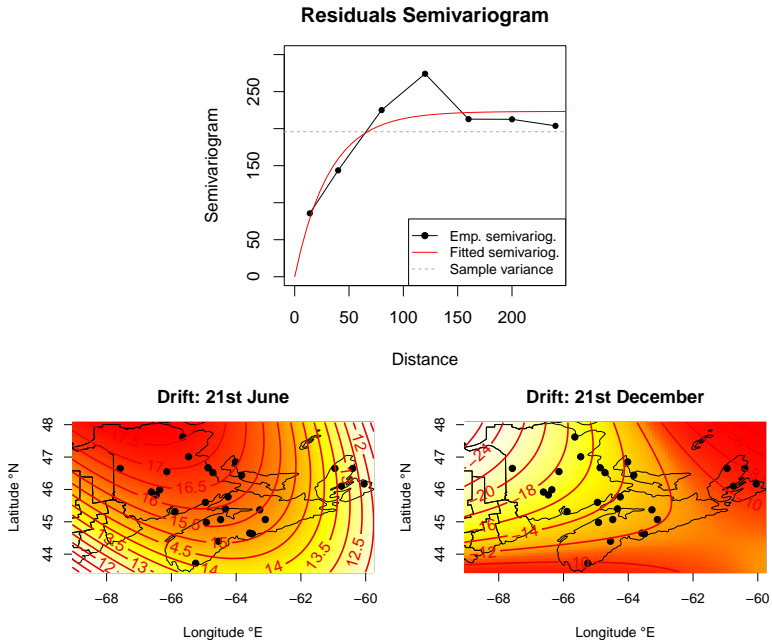


Figure 1.4 Estimated trace-semivariogram from the residuals (top panel) and estimated drift for the Summer Solstice (21st June; bottom-left panel) and the Winter Solstice (21st December; bottom-right panel). The drift maps are extracted from the drift temperature profiles estimated in L^2 . In the top panel, (geodesic) distances are given in kilometers.

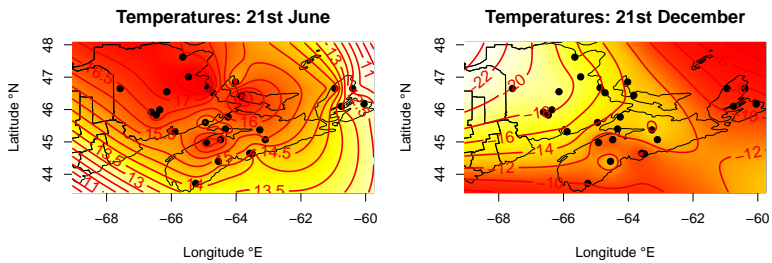


Figure 1.5 Universal kriging maps for the Summer Solstice (21st June) and the Winter Solstice (21st December), extracted from the temperature profiles predicted via universal kriging in L^2 .

1.4 An Operatorial viewpoint to kriging

In this section we briefly review a second approach to kriging, namely the operatorial ordinary kriging. Here, the kriging predictor is built upon a linear transformation of the vector of data $\mathcal{X}_{s_0}^\Lambda = \Lambda \vec{\mathcal{X}}$, for some linear operator $\Lambda : \mathcal{H}^n \rightarrow \mathcal{H}$.

The development of such classes of predictors is motivated by the observation that, despite its simplicity, predictor $\mathcal{X}_{s_0}^* = \sum_{i=1}^n \lambda_i \cdot \mathcal{X}_{s_i}$ does not provide, in general, the best linear unbiased transformation of the vector of observations. The operatorial viewpoint was firstly considered in [7] in Reproducing Kernel Hilbert Spaces (RKHSs), that are functional spaces whose elements are more regular than general Hilbert spaces (see also Chapter WWW (Nerini et al)). These authors addressed the problem of finding the best predictor over the class of linear unbiased Hilbert-Schmidt transformations of the observations, of the form $\mathcal{X}_{s_0}^B = \sum_{i=1}^n B_i \mathcal{X}_{s_i}$, where $B_i : \mathcal{H} \rightarrow \mathcal{H}$ are linear Hilbert-Schmidt operators and \mathcal{H} a RKHS. Although all finite-dimensional Hilbert spaces are RKHS, other widely-used spaces, such as the space L^2 , need not be a RKHS. In [4], a more general theory was introduced based upon the idea of working with measurable linear transformations instead of linear Hilbert-Schmidt operators: in this particular setting, these authors showed that the operatorial kriging problem is well-posed for any separable Hilbert space \mathcal{H} .

We now formally introduce the latter predictor, and discuss its relation with the other kriging predictors here discussed.

Given two separable Hilbert spaces, $\mathcal{H}_1, \mathcal{H}_2$, we denote by L a Borel-measurable map from $(\mathcal{H}_2, \mathfrak{B}(\mathcal{H}_2))$ to $(\mathcal{H}_1, \mathfrak{B}(\mathcal{H}_1))$, $\mathfrak{B}(\mathcal{H}_i)$ being the Borel σ -algebra of \mathcal{H}_i , for $i = 1, 2$. Given a measure μ on $(\mathcal{H}_2, \mathfrak{B}(\mathcal{H}_2))$, we say that L is a *measurable linear transformation with respect to μ* (μ -mlt), if L is linear on a subspace $\mathcal{D}_L \in \mathfrak{B}(\mathcal{H}_2)$ with $\mu(\mathcal{D}_L) = 1$.

Given a set of locations s_1, \dots, s_n and the observation of the process at these locations, $\mathcal{X}_{s_1}, \dots, \mathcal{X}_{s_n}$, we consider the operatorial ordinary kriging predictor $\mathcal{X}_{s_0}^{\Lambda^*} = \Lambda^* \vec{\mathcal{X}}$ for \mathcal{X}_{s_0} . Here, $\Lambda^* : \mathcal{H}^n \rightarrow \mathcal{H}$ is a measurable linear transformation with respect to the law $\mu_{\vec{\mathcal{X}}}$ of $\vec{\mathcal{X}}$, and minimizes

$$\mathbb{E}[\|\mathcal{X}_{s_0} - \mathcal{X}_{s_0}^\Lambda\|^2] \quad \text{subject to} \quad \mathbb{E}[\mathcal{X}_{s_0}^\Lambda] = m, \quad (1.15)$$

over all $\Lambda : \mathcal{H}^n \rightarrow \mathcal{H}$ a $\mu_{\vec{\mathcal{X}}}$ -mlt and where in the objective functional $\mathcal{X}_{s_0}^\Lambda$ stands for $\mathcal{X}_{s_0}^\Lambda = \Lambda \vec{\mathcal{X}}$.

To tackle this problem, throughout the section we assume $\{\mathcal{X}_s, s \in D\}$ to be a Gaussian random field on $(\Omega, \mathfrak{F}, \mathbb{P})$ in $(\mathcal{H}, \mathfrak{B}(\mathcal{H}))$, that is, we assume that all its finite-dimensional laws are Gaussian in \mathcal{H} . Recall that a random element in \mathcal{H} is Gaussian if $\langle x, \mathcal{X} \rangle$ has a Gaussian distribution for every $x \in \mathcal{H}$. Note that this assumption is crucial for the validity of the results here presented, because a full characterization of the properties of mlts is only available under Gaussianity. We further assume that the field is strongly second-order stationary; we call m its (spatially-constant) mean, and \mathcal{C} its (stationary) spatial covariance function, defined as in (1.2).

Under this assumptions, the ordinary kriging problem can be shown to be well-posed [4]. To state such result, we need the following further notation. We call

$1 : \mathcal{H} \rightarrow \mathcal{H}^n$ the linear operator acting on $x \in \mathcal{H}$ as $1x = (x, \dots, x)^T$, and $1'$ its adjoint. We denote by $\mu_{\vec{\chi}_0} = N(\mathbf{m}_{\vec{\chi}_0}, \mathcal{C}_{\vec{\chi}_0})$ the law of the random vector $\vec{\chi}_0 = (\chi_{s_0}, \vec{\chi}^T)^T$ in \mathcal{H}^{n+1} , with expected value $\mathbf{m}_{\vec{\chi}_0} = (m, (1m)^T)^T$ and covariance operator $\mathcal{C}_{\vec{\chi}_0} : \mathcal{H}^{n+1} \rightarrow \mathcal{H}^{n+1}$. The latter can be expressed in block form as

$$\mathcal{C}_{\vec{\chi}_0} = \begin{pmatrix} \mathcal{C}_{\chi_{s_0}} & \mathcal{C}_{\chi_{s_0}\vec{\chi}} \\ \mathcal{C}_{\vec{\chi}\chi_{s_0}} & \mathcal{C}_{\vec{\chi}} \end{pmatrix}.$$

Here $\mathcal{C}_{\vec{\chi}}$ indicates the covariance operator of $\vec{\chi}$, i.e., $(\mathcal{C}_{\vec{\chi}}\vec{x})_i = \sum_{j=1}^n \mathcal{C}(s_i - s_j)x_j$, for $\vec{x} \in \mathcal{H}^n$, $i = 1, \dots, n$, and $\mathcal{C}_{\vec{\chi}\chi_{s_0}}$ is the cross-covariance operator between $\vec{\chi}$ and χ_{s_0} , i.e., $\mathcal{C}_{\vec{\chi}\chi_{s_0}}\vec{x} = \sum_{j=1}^n (s_0, s_j)x_j$, for $\vec{x} \in \mathcal{H}^n$.

The following Theorem 1.4.1 – proved in [4] – states that the operatorial ordinary kriging problem is well-posed.

Theorem 1.4.1 ([4]) *Under the previous assumptions and notation, (1.15) admits a unique minimizer $\chi_{s_0}^{\Lambda^*} = \Lambda^* \vec{\chi}$, where Λ^* is the $\mu_{\vec{\chi}}$ -mlt solving*

$$\begin{cases} \Lambda \mathcal{C}_{\vec{\chi}} - \mathcal{C}_{\chi_{s_0}\vec{\chi}} + \zeta_1 1' = 0; \\ \Lambda 1 - I = 0, \end{cases} \quad (1.16)$$

with $1x = (x, x, \dots, x)^T$, for $x \in \mathcal{H}$, $I : \mathcal{H} \rightarrow \mathcal{H}$ the identity operator and ζ_1 a μ_{χ_0} -mlt that represents the Lagrange multiplier associated with the unbiasedness constraint. Moreover, for $x \in \mathcal{H}^n$, one has

$$\Lambda^* x = M^* x + L(x - 1 M^* x), \quad (1.17)$$

where M^* is the $\mu_{\vec{\chi}}$ -mlt defined, for $x \in \mathcal{H}^n$, as $M^* x = (1' \mathcal{C}_{\vec{\chi}}^{-1} 1)^{-1} 1' \mathcal{C}_{\vec{\chi}}^{-1} x$, and L is the $\mu_{\vec{\chi}}$ -mlt of conditional expectation that acts on $x \in \mathcal{H}^n$ as $Lx = \mathcal{C}_{\chi_{s_0}\vec{\chi}} \mathcal{C}_{\vec{\chi}}^{-1} x$.

We refer the interested reader to [4] for the proof of Theorem 1.4.1. We note however that (1.16) can be expressed in matrix form as

$$\begin{pmatrix} \Lambda & \zeta_1 \end{pmatrix} \begin{pmatrix} \mathcal{C}_{\vec{\chi}} & 1 \\ 1' & 0 \end{pmatrix} = \begin{pmatrix} \mathcal{C}_{\chi_{s_0}\vec{\chi}} & I \end{pmatrix}, \quad (1.18)$$

which as the very same form as (1.8), but in an operatorial setting. Moreover, a second key element of Theorem (1.4.1) is the explicit expression for the optimal $\mu_{\vec{\chi}}$ -mlt Λ^* in (1.17). The latter involves two parts: the first related with an operatorial version of the generalized least squares estimator for the mean function, analogue to that described in Section 1.3.2; the second part exploits the operator of conditional expectation L of χ_{s_0} given $\vec{\chi}$, applied to the estimated residuals. As showed in [12, 13], the latter operator L is the $\mu_{\vec{\chi}}$ -mlt that allows to obtain the conditional expectation $\mathbb{E}[\chi_{s_0} | \vec{\chi}]$ when applied to the centred observations $(\vec{\chi} - 1m)$, under the assumption that the mean m is known, i.e.,

$$\mathbb{E}[\chi_{s_0} | \vec{\chi}] = m + L(\vec{\chi} - 1m). \quad (1.19)$$

Note that (1.19) has the very same form of the familiar expression of conditional expectation in the multivariate setting. Hence, (1.17) shows an interesting relation of the operatorial kriging predictor with the conditional expectation, which has a very analogous counterpart in the finite dimensional case [9]. Indeed, kriging coincides with the conditional expectation only when the mean is known (i.e., simple kriging [4]). In all other cases, it is a plug-in estimator that is built upon the conditional expectation, by employing the GLS estimate of the mean.

We finally mention that both the operatorial kriging predictor in RKHSs and the trace-kriging predictor can be seen as particular cases of the operatorial kriging predictor defined by Theorem 1.4.1. Indeed, the kriging predictor proposed in [7] can be found by embedding Theorem 1.4.1 in a RKHS. A particular case is then obtained for finite-dimensional Hilbert spaces, already explored by [7], that are interpreted as finite-dimensional approximations of the BLUP $\chi_{s_0}^{\Delta^*}$. Similarly, in the stationary Gaussian case, the trace-kriging predictor can be interpreted as the finite-dimensional approximation of the operatorial kriging predictor within the n -dimensional Hilbert space generated by the observations. We refer the reader to [4] for further details.

1.5

Kriging for manifold-valued random fields

While spatial statistics of functional data has recently received much attention, as proved by the many contributions in this book, the extension to the case of non Euclidean data is even a greater challenge because they do not belong to a vector space. [3] recently proposed a kriging procedure for data belonging to Riemannian manifolds, where local tangent space approximations can be used. Indeed, any Riemannian manifold admits an approximation based on a Hilbert tangent space, where linear geostatistical methods can be developed. Thus, it is possible to use the local geometry of the manifold to find a data-driven linearization, i.e. looking for the tangent space where the parametric model provides the best possible fit for the available data. Then, the spatial dependence can be modelled in the tangent space using the methods for Hilbert spaces described above. In this section, we describe the method introduced in [3] and we discuss some possible generalisation.

1.5.1

Residual kriging

We first need to introduce some definitions and notations to model data that take values in a Riemannian manifold. Let \mathcal{M} be a Riemannian manifold and, given a point P in \mathcal{M} , let \mathcal{H} be the tangent space at the point P , $\mathcal{H} = T_P\mathcal{M}$. This is a Hilbert space when equipped with a inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ in \mathcal{H} . Since our aim is to model the spatial variation in the local tangent space in P , we need a way to map elements of the tangent space to the manifold and viceversa. Thus, two important objects are the exponential map and its inverse, the logarithmic map. The exponential map is a smooth map from $T_P\mathcal{M}$ to \mathcal{M} , which is defined via the geodesics (the

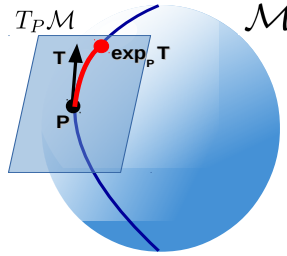


Figure 1.6 Visual representation of the tangent space in P on a sphere and of the exponential map of a vector T in the tangent space.

shortest paths between points on the manifold) passing through P : it maps a tangent vector $T \in T_P \mathcal{M}$ to an element in \mathcal{M} by travelling on the manifold, for a unit of time, along the geodesic starting in P in direction T . Indeed, under some technical assumptions on \mathcal{M} , for every pair $(P; T) \in \mathcal{M} \times T_P \mathcal{M}$, there is a unique geodesic curve $g(t)$ such that $g(0) = P$ and $g'(0) = T$. The exponential map of \mathcal{M} in P is defined as the point at $t = 1$ of this geodesic, i.e. $g(1)$. We indicate here with \exp_P the exponential map in P , and with \log_P its inverse. Figure 1.6 shows a visualisation of these concepts for the case of a sphere. More details on these definitions and on the properties of Riemannian manifolds can be found, e.g., in [14] and a detailed example for the case of the manifold of positive definite symmetric matrices is discussed in Section 1.5.2.

For a location s in the spatial domain D , we can now model the random element \mathcal{S}_s , taking value in \mathcal{M} , as

$$\mathcal{S}_s(\vec{a}, P) = \exp_P(A(\vec{f}(s); \vec{a}) + \delta_s). \quad (1.20)$$

Here, $A(\vec{f}(s); \vec{a})$ is a drift term defined in the tangent space \mathcal{H} , described by a linear model with $\vec{a} = (a_0, \dots, a_L)$ a vector of coefficients belonging to \mathcal{H} and $\vec{f}(s)$ a vector of scalar regressors:

$$A(\vec{f}(s); \vec{a}) = \sum_{l=0}^L f_l(s) \cdot a_l,$$

where $f_0(s) = 1$. Instead, $\{\delta_s, s \in D\}$ denotes a zero-mean globally second-order stationary and isotropic random field in the Hilbert space \mathcal{H} , with covariogram C and semivariogram γ .

Let now s_1, \dots, s_n be n locations in D , and let $\mathcal{S}_1, \dots, \mathcal{S}_n$ be the observations of (1.20) at these locations. The goals are to estimate the parameters P and \vec{a} in

model (1.20) and to perform spatial prediction at an unobserved location s_0 . We denote by $\Sigma \in \mathbb{R}^{n \times n}$ the covariance matrix of the array $\vec{\delta} = (\delta_{s_1}, \dots, \delta_{s_n})^T$ in \mathcal{H}^n , that is $\Sigma_{ij} = C(\|s_i - s_j\|_d^2)$, and call $\vec{R} \in \mathcal{H}^n$ the array of residuals $R_i = A(\vec{f}(s_i); \vec{a}) - \log_P(\mathbf{S}_i)$. To estimate (P, \vec{a}) accounting for the spatial dependence, the generalized least square (GLS) functional

$$(\hat{P}, \hat{\vec{a}}) = \operatorname{argmin}_{P \in \mathcal{M}, \vec{a} \in \mathcal{H}^{L+1}} \|\Sigma^{-1/2} \vec{R}\|_{\mathcal{H}^n}^2 \quad (1.21)$$

needs to be minimized. When Σ is known, problem (1.21) can be solved iteratively, by alternatively minimizing the GLS functional in (1.21) with respect to P given \vec{a} and to \vec{a} given P . The minimizer in \vec{a} given P can be explicitly determined as detailed in Section 1.3.2, i.e.,

$$\hat{\vec{a}}^{GLS}(P) = (\mathbb{F}^T \Sigma^{-1} \mathbb{F})^{-1} \mathbb{F}^T \Sigma^{-1} \circ \vec{Y}(P), \quad (1.22)$$

where $\mathbb{F} \in \mathbb{R}^{n \times (L+1)}$ is the design matrix, $\mathbb{F}_{il} = f_l(s_i)$, and $\vec{Y}(P)$ is the array $\vec{Y}(P) = (\log_P(\mathbf{S}_1), \dots, \log_P(\mathbf{S}_n))^T \in \mathcal{H}^n$. On the other hand, an expression for the minimizer in P given \vec{a} is not available, in general. The complexity of such minimization is problem dependent, and may require the development of specific optimization techniques.

When (P, \vec{a}) are known, it is possible to estimate Σ by estimating the semivariogram $\gamma(h)$, for example following the strategy of Section 1.3.2. That is (i) estimate the empirical semivariogram from the residuals as

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

and (ii) fit a parametric model to the empirical estimate to obtain a valid model. A good estimate of the spatial dependence (including the choice of the model semivariogram) is crucial in the applications. Note that, the tangent space being Hilbert, most of the existing methods in linear geostatistics can be used in this step of the analysis [see, e.g., 15, 16, and reference therein]. Since in practice both the parameters and the spatial dependence are unknown, there is the need to resort to a nested iterative algorithm, where the semivariogram is estimated from the residuals at the current iteration.

Estimated the parameters of model (1.20) as $(\hat{P}, \hat{\vec{a}}, \hat{\Sigma})$, the kriging prediction can be performed as follows. In the Hilbert space \mathcal{H} , the simple kriging predictor for δ_{s_0} is well-defined and it is obtained as $\sum_{i=1}^n \lambda_i^0 \hat{\delta}_{s_i}$, where $\hat{\delta}_{s_i}$ stands for the estimated residual at s_i , $\hat{\delta}_{s_i} = A(\vec{f}(s_i); \hat{\vec{a}}) - \log_{\hat{P}}(\mathbf{S}_i)$, and the vector of kriging weights $\vec{\lambda}_0 = (\lambda_1^0, \dots, \lambda_n^0)$ is found as $\vec{\lambda}_0 = \hat{\Sigma}^{-1} \vec{c}$, with $\vec{c} = (\hat{C}(\|s_1 - s_0\|_d), \dots, \hat{C}(\|s_n - s_0\|_d))^T$. The spatial prediction of \mathbf{S} at the target location s_0 is then

$$\hat{\mathbf{S}}_0 = \exp_{\hat{P}}(\hat{\vec{a}}_0^{GLS}(\hat{P})) + \sum_{l=1}^L \hat{\vec{a}}_l^{GLS}(\hat{P}) f_l(s_0) + \sum_{i=1}^n \lambda_i^0 \hat{\delta}_{s_i},$$

where $\vec{f}(s_0)$ is the vector of covariates given at the location s_0 .

1.5.2

An application to positive definite matrices

Positive definite matrices are an example of manifold-valued data, the modelling of positive definite matrices random field being relevant in applications such as Diffusion Tensor Imaging [17] and covariances between meteorological variables [3, 18]. The space $PD(p)$ of positive definite matrices of dimension p is a convex subset of $\mathbb{R}^{p(p+1)/2}$ but it is not a linear space: in general, a linear combination of elements of $PD(p)$ does not belong to $PD(p)$. The tangent space $T_P PD(p)$ to the manifold of positive definite symmetric matrices of dimension p in the point $P \in PD(p)$ can be identified with the space $Sym(p)$, the space of symmetric matrices of dimension p , which is linear and can be equipped with an inner product. A Riemannian metric in $PD(p)$ is then induced by the inner product in $Sym(p)$ and the choice of the inner product in the tangent space determines the form of the geodesic (i.e. the shortest path between two elements on the manifold) and thus the expression of the geodesic distance between two positive definite symmetric matrices. A possible choice for the Riemannian metric is generated by the scaled Frobenius inner product in $Sym(p)$, which is defined as $\langle A, B \rangle_P = \text{trace}(P^{-\frac{1}{2}} A^T P^{-1} B P^{-\frac{1}{2}})$, where $A, B \in Sym(p)$. This choice is very popular for covariance matrices, because it generates a distance which is invariant under affine transformation of the random variables. For every pair $(P, A) \in PD(p) \times Sym(p)$, there is then a unique geodesic curve $g(t)$ such that

$$\begin{aligned} g(0) &= P, \\ g'(0) &= A. \end{aligned}$$

When the Riemannian metric is generated by the scaled Frobenius inner product, the expression of the geodesic becomes

$$g(t) = P^{\frac{1}{2}} \exp(tP^{-\frac{1}{2}} A P^{-\frac{1}{2}}) P^{\frac{1}{2}},$$

where $\exp(C)$ indicates the exponential matrix of $C \in Sym(p)$. The exponential map of $PD(p)$ in P is defined as the point at $t = 1$ of this geodesic:

$$\exp_P(A) = P^{\frac{1}{2}} \exp(P^{-\frac{1}{2}} A P^{-\frac{1}{2}}) P^{\frac{1}{2}}.$$

Thus, the exponential map takes the geodesic passing through P with “direction” A and follows it until $t = 1$. The exponential map has an inverse which is called logarithmic map and is defined as

$$\log_P(D) = P^{\frac{1}{2}} \log(P^{-\frac{1}{2}} D P^{-\frac{1}{2}}) P^{\frac{1}{2}},$$

where $\log(C)$ is the logarithmic matrix of $C \in PD(p)$. The logarithmic map returns the tangent element A that allows the corresponding geodesic to reach D at $t = 1$.

With this structure, it is possible to apply the residual kriging method described above to positive definite matrix-valued data. [3] show the kriging interpolation for

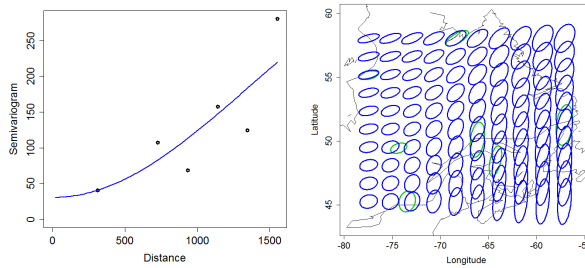


Figure 1.7 Left: empirical semivariogram (symbols) and fitted exponential model (solid line). The geodesic distances in the spatial domain are measured in km. Right: Ordinary kriging for the (temperature, precipitation) covariance matrix field for the month of January; green ellipses indicate original data. A covariance matrix S at location s is represented as an ellipse centered in s and with axis $\sqrt{\sigma_j} \vec{e}_j$, where $S \vec{e}_j = \sigma_j \vec{e}_j$ for $j = 1, 2$. Horizontal and vertical axes of the ellipses represent temperature and precipitation respectively. Modified from [3].

the covariance matrices between temperature and precipitation in Quebec, using data from Canadian meteorological stations made available by Environment Canada on the website <http://climate.weatheroffice.gc.ca>. The 7 meteorological stations where all monthly temperature and precipitation data are available from 1983 to 1992 are considered. For each station and for each month from January to December, these 10-year measures are used to estimate the 2×2 covariance matrix between temperature and precipitation, obtaining and separately analysing 12 datasets, each composed by $n = 7$ spatially dependent sample covariance matrices (with the previous notation, $n = 7$ and $p = 2$). [3] found out that the covariation between temperature and precipitation changes across the calendar year. We report here the results obtained for the month of January. Including only a constant term in the tangent space model (i.e., assuming that the matrix random field has a constant mean) leads to an estimate of the empirical semivariogram that suggests to move toward a non-stationary model, as it can be seen in Figure 1.7. The simplest drift model which guarantees the stationarity of the residuals is found to be the following linear model depending on longitude:

$$A(\text{Longitude}_i, \text{Latitude}_i) = a_0 + a_{\text{Long}} \text{Longitude}_i. \quad (1.23)$$

A possible meteorological interpretation is associated with the exposition of the region toward the sea, since model (1.23) accounts for the distance between the location of interest and the Atlantic Ocean. This is likely to influence temperatures, precipitations and their covariability. The estimates of the semivariogram and of the drift and the kriging prediction can be seen in Figure 1.8.

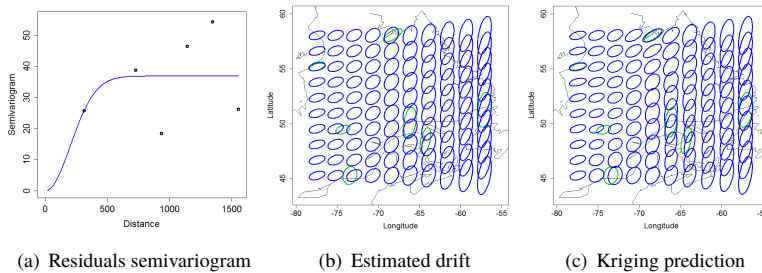


Figure 1.8 Kriging of the (temperature, precipitation) covariance matrix field during January, with a drift term depending on longitude. A covariance matrix S at location s is represented as an ellipse centered in s and with axis $\sqrt{\sigma_j} \vec{e}_j$, where $S \vec{e}_j = \sigma_j \vec{e}_j$ for $j = 1, 2$. Horizontal and vertical axes of the ellipses represent temperature and precipitation respectively. In subfigure (b) and (c) green ellipses indicate the data, blue ellipses the estimated drift and the kriging interpolation, respectively. In subfigure (a) the residual empirical semivariogram (symbols) and the fitted exponential model (solid line) are reported. The geodesic distances in the spatial domain are measured in km. Modified from [3].

1.5.3

Validity of the local tangent space approximation

The method introduced in [3] relies on a local Euclidean approximation and, albeit the choice of the best local approximation is data-driven, a natural question is if this is suitable to model the data at hand. The authors presented a simulation study to explore this issue in the case of positive-definite matrices, by evaluating the performance of the kriging predictor when data are generated from a model different from (1.20) (i.e. the local linear approximation is not valid). In particular, they generate a non stationary matrix field according to a probabilistic model with mean $G_s = \exp_P(A(\vec{f}(s); \vec{a}))$, where P and $A(\vec{f}, \vec{a})$ are set parameters. This random matrix field is obtained through the sample covariance matrices generated by the realizations of a Gaussian random vector field \vec{v} in the following way. Let $D \subset \mathbb{R}^2$ indicate the common spatial domain of two independent Gaussian random fields \mathbf{w}_s , \mathbf{y}_s , $s \in D$. Both random fields \mathbf{w}_s and \mathbf{y}_s have zero mean and Gaussian spatial covariance with decay $\phi = 0.1$, sill equal to 1 and zero nugget. For $s \in D$, the covariance matrix (between components) of the random vector field $\vec{v}_s = (G_s)^{\frac{1}{2}} (\mathbf{w}_s, \mathbf{y}_s)'$ is equal to G_s . Then, N independent realizations of the random vector field \vec{v} are generated and, for $s \in D$, the realization of the manifold-valued random field is given by the sample covariance matrices

$$S_s = \frac{1}{N-1} \sum_{k=1}^N (\vec{v}_{s,k} - \bar{\vec{v}}_s) (\vec{v}_{s,k} - \bar{\vec{v}}_s)^T \sim \text{Wishart}_2 \left(\frac{1}{N-1} G_s, N-1 \right),$$

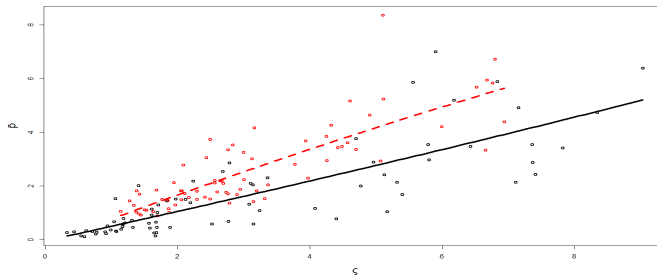
(1.24)

\bar{v}_s being the sample mean in $s \in D$. This simulation process is therefore defined on the manifold of positive definite symmetric matrices and the parameter N controls the variability of the the matrix random field \mathbf{S} in s . When N is large, the data will be concentrated around the drift (which satisfies the tangent space approximation). Therefore, N also controls the violation of the tangent space approximation. We want to evaluate the performance of the kriging procedure when applied to these simulated fields by comparison with the case when data are generated by model (1.20). Data are then generated on an equally spaced 10×10 grid, 15 locations are taken as known and the prediction error $\bar{p} = \frac{1}{85} \sum_{i=1}^{85} d(\mathbf{S}_{s_i}, \hat{\mathbf{S}}_{s_i})^2$ in the remaining 85 locations is measured. Here $d(\cdot, \cdot)$ denotes the Riemannian distance between two positive definite matrices [see 3]. This experiment is repeated with different values of the model parameters. Since the two models are controlled by different parameters, to compare them on the same footing we can measure the sample marginal variability, defined as $\varsigma = \frac{1}{100} \sum_{i=1}^{100} d(\mathbf{S}_{s_i}, G_{s_i})^2$, i.e. the variation of the realization of the field with respect to the true mean field G_s .

Figure 1.9 compares the performances of the kriging prediction when data are generated with model (1.24) and when data are generated with the tangent space model (1.20), for a range of values of the sample marginal variability.

This suggests that the higher the value of the sample marginal variability, the worse is the relative performance of the kriging predictor between the two cases. This is to be expected because a high dispersion on the manifold means that no tangent space can accurately describe the data. However, for low values of the sample marginal variability the performance of the kriging predictor in the two settings is comparable, supporting its robustness to the violation of the model provided that the tangent space approximation is able to describe the observations in a fairly accurate way. More details on this simulation study can be found in [3].

By way of example, Fig. 1.9b and 1.9c represent two realizations of the matrix generated from (1.24) for high and low values of N , respectively, i.e. low or high values of the sample marginal variability. It can be seen that values of sample marginal variability where the performance of the kriging predictor gets worse correspond to random fields too noisy to be of any use in applied scenarios. However, other examples of manifold-valued data may present cases where the tangent space approximation is not suitable and a kriging procedure defined directly on the manifold would be needed. Ordinary kriging for a stationary manifold-valued random field can be achieved with a weighted Fréchet mean. For example, [18] used this approach to estimate the mean from a spatially dependent sample. However, the optimal choice of the weights for the ordinary kriging predictor is still an open problem. When the field is non stationary the problem is even more complex since the non linear nature of the data does not allow the removal of a non-stationary mean function. A possibility currently investigated by the authors is to model the response field by segmenting the spatial domain into regions where the field can be assumed to be stationary and ordinary kriging prediction can be used. The challenge is of course the identification of the correct scale for these sub-regions and how to deal with the



(a) Empirical prediction error

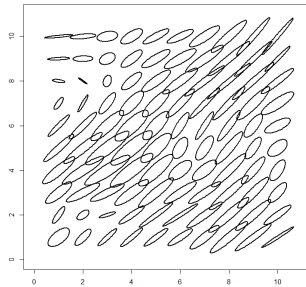
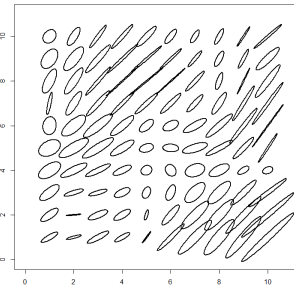
(b) Field realization with $\zeta = 2.9$ (c) Field realization with $\zeta = 6.7$

Figure 1.9 Panel (a): Empirical prediction error as a function of the sample marginal variability ζ , with a local polynomial smoothing added to help visual comparison, for data generated from the tangent space model (1.20) (black points and solid black line) and from procedure (1.24) (red points and dashed red line), both with Gaussian covariance function. Panel (b) and (c): Examples of simulated fields from procedure (1.24) for $N = 6$ (b) and $N = 4$ (c) and Gaussian covariance function, with the respective values of sample marginal variability ζ . Modified from [3].

discontinuities that may be introduced in the predicted field, e.g., via randomized approaches. Alternatively, one can think to extend to this setting locally stationary models in the same vein as [19].

References

- 1 Menafoglio, A. and Secchi, P. (2017) Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics. *European Journal of Operational Research*, **258** (2), 401–410.
- 2 Menafoglio, A., Secchi, P., and Dalla Rosa, M. (2013) A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. *Electronic Journal of Statistics*, **7**, 2209–2240.
- 3 Pigoli, D., Menafoglio, A., and Secchi, P. (2016) Kriging prediction for manifold-valued random fields. *Journal of Multivariate Analysis*, **145**, 117–131.
- 4 Menafoglio, A. and Petris, G. (2016) Kriging for Hilbert-space valued random fields: The operatorial point of view. *Journal of Multivariate Analysis*, **146**, 84–94, doi:10.1016/j.jmva.2015.06.012.
- 5 Marron, J.S. and Alonso, A.M. (2014) Overview of object oriented data analysis. *Biometrical Journal*, **56** (5), 732–753.
- 6 Goulard, M. and Voltz, M. (1993) Geostatistical interpolation of curves: A case study in soil science, in *Geostatistics Tróia '92*, vol. 2 (ed. A. Soares), Dordrecht: Kluwer Academic, pp. 805–816.
- 7 Nerini, D., Monestiez, P., and Manté, C. (2010) Cokriging for spatial functional data. *Journal of Multivariate Analysis*, **101** (2), 409–418.
- 8 Giraldo, R., Delicado, P., and Mateu, J. (2011) Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, **18** (3), 411–426.
- 9 Cressie, N. (1993) *Statistics for Spatial data*, John Wiley & Sons, New York.
- 10 Bosq, D. (2000) *Linear Processes in Function Spaces*, Springer, New York.
- 11 Zhu, K. (2007) *Operator Theory in Function Spaces*, American Mathematical Society, 2nd edn..
- 12 Mandelbaum, A. (1984) Linear estimators and measurable linear transformations on a Hilbert space. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **65** (3), 385–397.
- 13 Luschgy, H. (1996) Linear estimators and radonifying operators. *Theory of Probability & Its Applications*, **40** (1), 167–175.
- 14 Lee, J. (2012) *Introduction to Smooth Manifolds*, vol. 218, Springer Science & Business Media.
- 15 Diggle, P. and Ribeiro, P. (2007) *Model-based Geostatistics*, Springer.
- 16 Chiles, J.P. and Delfiner, P. (2009) *Geostatistics: modeling spatial uncertainty*, vol. 497, John Wiley & Sons.
- 17 Yuan, Y., Zhu, H., Lin, W., and Marron, J.S. (2012) Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74** (4), 697–719.
- 18 Pigoli, D. and Secchi, P. (2012) Estimation of the mean for spatially dependent data belonging to a riemannian manifold. *Electronic Journal of Statistics*, **6**, 1926–1942.
- 19 Dahlhaus, R. et al. (1997) Fitting time series models to nonstationary processes. *The annals of Statistics*, **25** (1), 1–37.