

Statistical analysis of complex and spatially dependent data: a review of Object Oriented Spatial Statistics

Alessandra Menafoglio^a, Piercesare Secchi^{a,1}

^a*MOX-Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, Italy*

Abstract

We review recent advances in Object Oriented Spatial Statistics, a system of ideas, algorithms and methods that allows the analysis of high dimensional and complex data when their spatial dependence is an important issue. At the intersection of different disciplines – including mathematics, statistics, computer science and engineering – Object Oriented Spatial Statistics provides the right perspective to address key problems in varied contexts, from Earth and life sciences to urban planning. We illustrate a few paradigmatic methods applied to problems of prediction, classification and smoothing, giving emphasis to the key ideas Object Oriented Spatial Statistics relies upon.

Keywords: object oriented data analysis, kriging for object data, bagging
Voronoi algorithm, spatial regression models with differential regularization
2010 MSC: 62H11, 62M30

1. Introduction: a few paradigmatic problems

Object Oriented Spatial Statistics (O2S2) is a new branch of statistics, that aims to provide a unifying viewpoint to address a variety of application-oriented challenges of modern spatial statistics. In O2S2 the “atom” of the analysis is typically a complex data point (such as a curve or a surface), spatially distributed. The object oriented viewpoint to the statistical treatment of these types of data consists in considering as building block of the analysis the whole data point, regardless of its complexity. The datum is interpreted as a point within an appropriate space of objects (called *feature space*), which should properly represent all the salient data features through its geometry. Implicit or explicit formalization of the notion that nearby data provide similar information then enables one to develop statistical procedures that take full advantage of the information content embedded within the data for the purpose of modeling, classifying or predicting spatial data. As such O2S2 is part of Object

*Corresponding author

Email address: piercesare.secchi@polimi.it (Piercesare Secchi)

Oriented Data Analysis, the seminal name chosen by Wang & Marron (2007) to baptize a system of ideas according to which new statistical tools are developed urged by the analysis of populations of complex objects. Object Oriented Data Analysis requires a strong interplay of statistics with other scientific disciplines, including maths (analysis, scientific computing, geometry and algebra, operational research), engineering, scientific communication design, computer graphics, computer science, and information technology (see Marron & Alonso (2014) for a recent review on Object Oriented Data Analysis).

Nowadays, the analysis of complex spatial objects plays a key role in a variety of data-driven engineering and geoscience applications. Even though most methods in OODA rely on the key assumption that the observations are independent and identically distributed, collections of spatially distributed data are increasingly available in field studies. O2S2 then meets the need of analysing populations of spatially dependent object data. Here, we propose a first paradigmatic example in O2S2, to engage the reader in the object oriented viewpoint to spatial statistics. This example represents the *fil rouge* of the present review, since it fully expresses our viewpoint on the way the research on O2S2 has been motivated, interpreted and developed in recent years.

Example 1. *Figure 1 displays a representation of a heterogenous aquifer system at the Lauswiesen site, in the Neckar river valley, near the city of Tübingen (Germany). Data at the site consist of particle-size densities (PSDs) measured at 406 locations along 12 boreholes. PSDs describe the local distributions of grain sizes within the aquifer system. From the mathematical viewpoint, a particle-size density is a probability density function, associated with the distribution of particle sizes within a given soil sample. As such, available data consist of a set of constrained curves, spatially distributed. The statistical characterization of PSDs plays a key role in the classification of soil types, for inferring hydraulic parameters (e.g., porosity and hydraulic conductivity), and reconstructing the internal architecture of the groundwater system. In this vein, the analysis of PSDs may be concerned with, e.g., (a) the classification of PSDs to identify geomaterials at the site, and (b) the spatial prediction and stochastic simulation of the PSDs at unsampled locations of the system (Menafoglio et al., 2014, 2016b,c).*

In Example 1, each data point is a complex object: here, critical elements of complexity are the data dimensionality and constraints. Indeed, PSDs are curves, thus infinite-dimensional data: they need an infinity of point evaluations to be fully characterized. Furthermore, PSDs are distributional data, hence constrained to be positive and integrate to unity. This cannot be just neglected: capturing the complexity of the objects is key to precisely model and express the information content embedded within data, which will then be used for exploration, inference and prediction with the final aim of accruing knowledge.

Nevertheless, the spatial domain of Example 1 is relatively simple: the aquifer can be thought of as a three-dimensional Euclidean parallelepiped, of moderate size. In this setting, one can imagine to extend to object data the

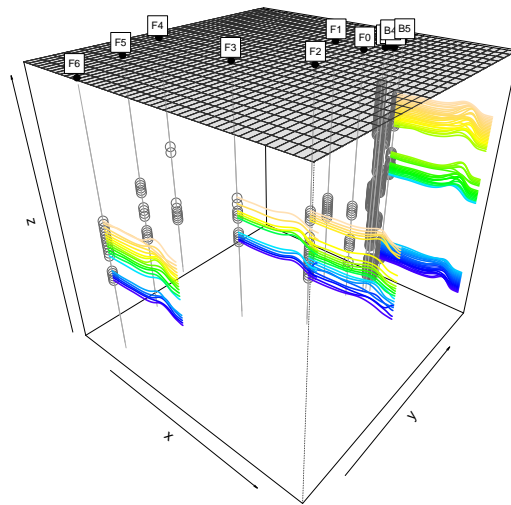


Figure 1: Three-dimensional representation of particle-size densities at the Lauswiesen site. Grey points represent measurement locations, colored curves represent a subset of the dataset of PSDs. Colors indicate the ordering of the curves along the borehole

ensemble of model-based methodologies that are widely-employed in classical geostatistics to deal with the spatial dependence (e.g., Kriging, Cressie, 1993).

In fact, one can readily envision situations of complex objects observed over a more complex spatial domain. Here, complexity may be associated with the size or the texture of the area of interest. For instance, data observed over very large regions pose challenges related to the practical impossibility of employing approaches based on global models for spatial dependence. Similarly, local models need to be employed to account for a very strong, irregular or sudden variability induced by, e.g., the texture of the domain. An instance of this is found in the study of vehicular traffic in urban areas: here, the road system induces a very fine texture. The spatial variability can be then expressed through local models of diffusion, governed by spatially varying tensors (see, e.g., Della Rossa et al., 2010). In some cases, the process could develop over a complex and evolving system such as a spatial network (Reggiani et al., 2001). In all these cases, local models or fully non-parametric approaches appear to be more appropriate than those inspired by classical geostatistics. An example in this class, based on a computationally intensive, explorative approach, will be illustrated in Section 5.

A different kind of complexity is that pertaining to domains with holes or highly irregular borders determined by, e.g., geographical constraints. For instance, in the setting of Example 1, one may think at the Neckar river as a peculiar boundary of the aquifer domain, and particular geological formations within the aquifer as holes in the domain. Part of the modern literature on spatial statistics – which in our view partially shares the approach of O2S2 – has focused on developing methods for dealing with these issues. To get closer to the topic, we now present an example where relatively simple data over a complex domain are concerned.

Example 2. *Sangalli et al. (2013) analyze census data in the Island of Montréal. The data consists of population densities (scalar observations) collected at a limited number of spatial locations over the region, depicted in Figure 2. The Island of Montréal is located between two rivers – Saint Lawrence and Rivière de Prairies – which form natural geographical constraints. Additionally, the harbour and the public parks are areas where people cannot live. Hence the spatial domain is here defined by irregular boundaries and it also has holes. In this case, the goal of the analysis is to reconstruct the target surface (density over the spatial domain) by (a) including a set of covariates and possible prior knowledge on the phenomenon (e.g., a set of known differential equations governing the phenomenon) and (b) properly accounting for the topology of the spatial domain and the shape of its boundary.*

Note that highly textured spatial regions, characterized by irregular boundaries and holes, are inevitably associated with boundary conditions: for instance, the population density of Example 2 needs to be zero in correspondence of the uninhabited areas (e.g., the harbour or some parts of the river banks marked with red in Figure , as noted by Sangalli et al. (2013). Furthermore, observations of the phenomenon might be also repeated in time, or other kinds of more

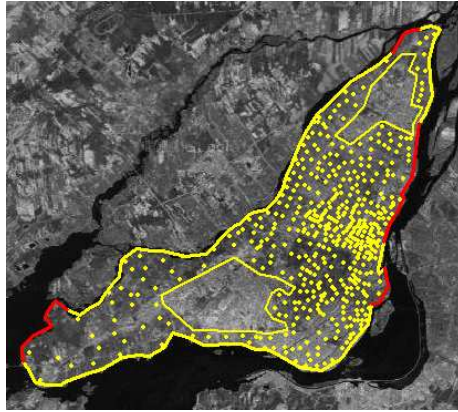


Figure 2: Island of Montréal and sampling sites for the census data of Example 2. Red lines along the boundary of the region denote places where population density needs to be zero. (Figure modified from Sangalli et al. (2013), courtesy of Laura Sangalli).

complex objects might be observed at the measurement locations. This projects us to the very cutting-edge of O2S2, where methods which jointly treat both the object and the domain complexities are yet to be developed.

For the sake of completeness in the present overview on spatial object data, we finally mention the viewpoint of those who interpret spatial patterns as object data (Mateu et al., 2015). Here, data are a discrete collection of (possibly linked) spatial points, such as point patterns, trees, graphs or networks. This finds application, e.g., in the treatment of spike train datasets, which is relevant in neuroscience (see, e.g., Brown et al., 2004). For instance, one may want to analyze profiles of neuronal activation (i.e., random linked points) measured on different patients in response to a stimulus, in order to identify healthy/unhealthy conditions. Additional examples of these types of data are available in environmental studies: Mateu et al. (2015) deal with plant communities with high biodiversity, aiming at the classification of plant species according to their spatial structure and dislocation. Unlike traditional analyses of point patterns, in the object oriented approach the attention is posed on revealing similarities and correlations between entire spatial objects, rather than characterizing the interactions between single points within a given object. Note that here the target objects are independent observations of random spatial domains, rather than spatially dependent objects distributed over a deterministic domain. This case is somehow different from the previous examples and may be deemed part of the general setting of OODA rather than O2S2, which will be the core of this survey instead.

The remaining part of the work is organized as follows. Section 2 provides a brief review of recent approaches to the analysis of complex data such as those of Example 1. Section 3 serves as introduction to the following methodological sections. Section 4 describes a model-based methodology which provides a paradigmatic illustration of the way classical geostatistical methods can be

extended to object data. Section 5 introduces a flexible fully non-parametric method to deal with large or textured domains, whereas Section 6 illustrates a method to treat data as those in Example 2. Finally, Section 7 concludes the review, discussing some of the present open issues in O2S2.

2. Recent approaches to the analysis of object data

The statistical analysis of spatial data has a long history and dates back to the beginning of the 20th century. Analysis of geostatistical, lattice and areal data, as well as point processes, is grounded in a rich body of methods and algorithms, when data are either univariate or multivariate (see, e.g., Cressie, 1993; Gelfand et al., 2010, for a comprehensive introduction to the topic and a recent collection of reviews). Despite the richness of the available literature, classical techniques prove to be somehow inadequate when data objects are high-dimensional, constrained, or distributed over a complex spatial domain.

For instance, the particle-size densities (PSDs) introduced in Example 1 pose challenges related to the dimensionality of the data and their constraints. Indeed, for each spatial point (soil sample) within the aquifer domain, the actual datum is a curve, hence an infinite dimensional object (or high-dimensional if it is discretely observed). Spatial statistical methods typically rely on the assumption that the dimensionality of each datum does not exceed the size of the sample. Thus, in the presence of high- or infinite-dimensional data, all those methods inevitably turn out to be unsatisfactory (i.e., curse of dimensionality).

A novel viewpoint to the analysis of infinite dimensional data has been recently fostered by the fast and extensive development of functional data analysis (FDA, Ramsay & Silverman, 2005). This is a recent branch of statistics that aims to develop explorative and inferential methods for the analysis of data in the form of functions (i.e., functional data), such as curves or surfaces. In recent years, FDA has become increasingly popular as an approach to face relevant operational research problems (see, e.g., Martin-Barragan et al., 2014; Laukaitis & Račkauskas, 2005; Laukaitis, 2008). The idea behind FDA is to consider the curves as a unique entity, and analyze these within an appropriate, functional space. In this sense, FDA is part of Object Oriented Data Analysis, since both approaches share a common viewpoint in understanding the curve as the building-block of the analysis, regardless of its dimension or complexity.

The functional approach to the analysis of georeferenced data was firstly introduced as an alternative to the more classical methods of space-time geostatistics, that is a class of advanced statistical methodologies devoted to the analysis and prediction of georeferenced data featured by a temporal dynamic (e.g., Cressie & Wikle, 2011; Lohmann et al., 2016, and references therein). From the point of view of applications, a justification of the functional approach in this setting comes from the observation that the sample design over time is typically much finer than that in space. Indeed, one typically observes the dynamic of the phenomenon over time along a fine grid of instants (e.g., hourly or daily data), whereas only few given locations are considered for the spatial sampling. Moreover, prior information on the time dynamic might dictate an appropriate

model and analytical regularity conditions for its representation, for instance in terms of a differential equations (Ramsay & Silverman, 2005). The FDA approach then allows to simplify the modeling of this type of data, by capturing the time dynamics through the functional form of the curve. Note that this strategy not only allows to cope with the curse of dimensionality that plagues space-time geostatistics, but is apt to treat any kind of abscissa, not just time.

In this setting, a relatively large body of literature has been developed in FDA to address several problems of spatial statistics, such as estimation of the mean and principal component analysis (Horváth & Kokoszka (2012, Chap. 17), Liu et al. (2014)), clustering (Giraldo et al., 2012; Ignaccolo et al., 2008; Romano et al., 2010, 2015), uncertainty modeling (Ignaccolo et al., 2015) and estimation of spatial auto-regressive processes (Ruiz-Medina, 2012). Particular attention has been paid to the problem of spatial prediction, especially to the development of novel notions of functional Kriging, both in the stationary (see, e.g., Delicado et al., 2010, for a recent review) and in the non-stationary setting (e.g., Caballero et al., 2013; Ignaccolo et al., 2014). This theory was widely applied, e.g., in climatology (prediction of daily temperature profiles, see Delicado et al., 2010, and references therein), oceanography (temperature profiles along depth of the ocean, see Nerini et al., 2010), air quality monitoring (Ignaccolo et al., 2014), wireless sensors networks (Lee et al., 2015), petroleum system modeling and production forecast (Menafoglio et al., 2016a).

Although in principle functional geostatistics is concerned with the treatment of *functional* data in its general acceptance, most FDA methods are designed to deal with the particular case of square-integrable functions (i.e., elements of the functional space L^2). Even though actual functional observations can often be embedded in this space, developing ad-hoc techniques for L^2 rather than for more general spaces of objects, has a number of limitations. For instance, the PSDs of Example 1 are distributional data, thus a kind of functional data with peculiar constraints: these modify the geometry of the sample space and are thus crucial for the analysis. Furthermore, as we already mentioned, one may want to account for the dynamic of the system, as described by a set of differential equations in an appropriate functional space (e.g., a Sobolev space, Azzimonti et al., 2015). In general, approaches developed in functional geostatistics do not allow to (i) jointly consider both the point-wise and the differential information embedded within the functions, (ii) account for irregularly shaped domains or boundary conditions, (iii) properly account for data constraints that requires particular geometrical approaches, and (iv) handle other kinds of objects, such as shapes, covariance matrices or spatial patterns. Developing an object oriented viewpoint to spatial statistics then comes as a response to the need of treating all these issues jointly, within a unifying and mathematically sound framework.

3. Methods in Object Spaces

In O2S2, the data point – whether it is a curve, a constrained vector or a shape – is seen as an object belonging to an appropriate (finite- or infinite-

dimensional) feature space. The geometrical structure of the feature space determines the kind of operations and analyses that one is allowed to perform on the data. For instance, if the feature space is an Euclidean space of finite dimension (i.e., data are vectors of dimension p , lying in \mathbb{R}^p), one can consider the ensemble of techniques available in classical multivariate spatial statistics. When infinite-dimensional data (e.g., curves) are concerned and the feature space is a vector space with a Hilbert structure (i.e., an inner product is available and the space is complete), one can imagine to extend all those statistical methods that relies upon the concepts of angles, distances and projections, because these are well-defined in the feature space. Most of this work is concerned with the latter setting, as it is the most common in applications. Nevertheless, we will provide insights into more complex features spaces, such as Riemannian manifolds. These allow treating, e.g., shapes and covariance matrices, but require the introduction of complex mathematical constructions relying upon non-Euclidean geometries.

The next Sections review three methodologies in O2S2 which follow quite different approaches. We first introduce a generalization to object data of a classical prediction technique in spatial statistics, Kriging (e.g., Cressie, 1993). We then illustrate a fully non-parametric method allowing to treat the spatial dependence in the presence of large or textured domains through a computationally intensive algorithm. Finally, we describe a semi-parametric method that allows accounting for domains with holes and boundary conditions, such as those described in Example 2.

4. Kriging for Object Data

Kriging is a widely-employed geostatistical technique that allows to perform linear spatial prediction from a set of spatially distributed data. From the time of the review of Kleijnen (2009), the geostatistical literature on Kriging has taken several step forward in the direction of object data, as we shall see in this section. Even though some advanced mathematical construction is needed to develop Kriging in the generality of object data, most of the results detailed hereafter closely resemble very analogue developments, well-known in the classical setting.

Let us denote by $\mathcal{X}_{\mathbf{s}}$ a random object referred to the location \mathbf{s} in the spatial domain D , which is usually a subset of \mathbb{R}^d , with $d = 2, 3$. In Example 1, D is the three-dimensional aquifer domain ($D \subset \mathbb{R}^3$), whereas $\mathcal{X}_{\mathbf{s}}$ denotes a random PSD. Given n measurement locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, and the observations $\mathcal{X}_{\mathbf{s}_1}, \dots, \mathcal{X}_{\mathbf{s}_n}$ at these locations, one may want to predict an unobserved object $\mathcal{X}_{\mathbf{s}_0}$ at a new location \mathbf{s}_0 in D (e.g., the PSDs at a new borehole). Note that the aim is to predict the entire object, not only some of its features. In this sense, an object oriented geostatistical viewpoint is considered.

Call H the feature space the objects $\mathcal{X}_{\mathbf{s}}$ belongs to: H should represent through its geometry all the data characteristics that one is willing to consider. For instance, for the PSDs in Example 1 and Figure (1), one should consider a feature space whose points are probability density functions (PDFs), or, more generally, distributions. Hereafter, the focus is posed on feature spaces that are

Hilbert spaces, that is, complete spaces with respect to a vector structure $(+, \cdot)$ and an inner product $\langle \cdot, \cdot \rangle$ (with induced norm $\|f\| = \sqrt{\langle f, f \rangle}$, for $f \in H$). Informally, Hilbert spaces behave like Euclidean spaces, and may have finite or infinite dimension. In the object oriented view, $\mathcal{X}_{\mathbf{s}_1}, \dots, \mathcal{X}_{\mathbf{s}_n}$ are interpreted as a partial observation of a random field in the feature space H , $\{\mathcal{X}_{\mathbf{s}}, \mathbf{s} \in D\}$. Modeling the main distributional features of the field (e.g., mean, covariance) is thus a key point of the spatial analysis.

The mean term $m_{\mathbf{s}}$ of $\mathcal{X}_{\mathbf{s}}$ is usually represented by the linear model in H

$$m_{\mathbf{s}} = \sum_{l=0}^L f_l(\mathbf{s}) \cdot a_l, \quad (1)$$

where

$\{a_l, l = 0, \dots, L\}$ are coefficients in H , independent of the spatial location \mathbf{s} ;

$f_0(\mathbf{s}) = 1$ for all $\mathbf{s} \in D$ and $\{f_l(\mathbf{s}), l = 1, \dots, L\}$ are known scalar regressors;

$(+, \cdot)$ are operations in the sense of the geometry of H .

The coefficient a_l quantifies the effect of a unit variation of the spatial regressor $f_l(\mathbf{s})$ – when all the others are fixed – on the mean value of the process, $l = 1, \dots, L$, whereas the coefficient a_0 represents a “functional intercept”, in the sense that it corresponds to the (functional) mean value of the response when all the regressors $f_l(\mathbf{s}), l = 1, \dots, L$, are null. This model was proposed by Menafoglio et al. (2013), and, concurrently, for the L^2 case by Caballero et al. (2013). It includes the model with spatially constant mean, studied in FDA by Giraldo et al. (2011, 2010a); Gromenko et al. (2012) (see also Delicado et al., 2010, for a recent review). Note that model (1) is quite general, and allows to precisely describe any separable drift term (Menafoglio et al., 2013). Separability of the drift term is generally considered as a viable assumption, and is widely employed in the recent literature on space-time analysis (e.g., Augustin et al., 2013; Bernardi et al., 2016; Cressie & Wikle, 2011; Marra et al., 2012; Wood et al., 2008). Furthermore, model (1) can be used to account for cluster information (Menafoglio et al., 2016c), or to model an external drift, as proposed by Ignaccolo et al. (2014) in the FDA context. Note that model (1) is assumed to be valid over the entire spatial domain D (in this sense, it is a global model for the drift).

If H were the space of real numbers ($H = \mathbb{R}$), one could describe the spatial dependence through the *covariogram*, which is defined, for any pair of locations $(\mathbf{s}_i, \mathbf{s}_j) \in D \times D$, as the covariance of the variables at these locations, i.e.,

$$C_{\mathbb{R}}(\mathbf{s}_i, \mathbf{s}_j) = \mathbb{E}_{\mathbb{R}}[(\mathcal{X}_{\mathbf{s}_i} - m_{\mathbf{s}_i})(\mathcal{X}_{\mathbf{s}_j} - m_{\mathbf{s}_j})].$$

The latter is a measure of second-order dependence widely-employed in geostatistics (Cressie, 1993). Menafoglio et al. (2013) propose to define a similar measure of second-order dependence for objects in H , called *trace-covariogram*.

This is the real valued function C_H , defined for $(\mathbf{s}_i, \mathbf{s}_j)$ in $D \times D$ as

$$C_H(\mathbf{s}_i, \mathbf{s}_j) = \mathbb{E}_H[\langle \mathcal{X}_{\mathbf{s}_i} - m_{\mathbf{s}_i}, \mathcal{X}_{\mathbf{s}_j} - m_{\mathbf{s}_j} \rangle]. \quad (2)$$

The trace-covariogram describes, in a global sense, the covariation between couples of objects of the field, and thus measures the second-order spatial dependence of the process. Intuitively, as the distance between locations increases, the spatial dependence between the associated objects vanishes and thus the absolute value of the corresponding trace-covariogram decreases toward zero (stationarity and ergodicity conditions apply). Note that this definition is consistent with that of $C_{\mathbb{R}}$, as in \mathbb{R} the inner product is simply the product between scalars and the trace-covariance is just the usual covariance. The definition is also consistent with that given in functional geostatistics by Giraldo et al. (2011) for data in L^2 . In this case, the trace-covariogram reads

$$C_{L^2}(\mathbf{s}_i, \mathbf{s}_j) = \int_{\mathcal{T}} \mathbb{E}_H[(\mathcal{X}_{\mathbf{s}_i}(t) - m_{\mathbf{s}_i}(t))(\mathcal{X}_{\mathbf{s}_j}(t) - m_{\mathbf{s}_j}(t))]dt = \int_{\mathcal{T}} C_{\mathbb{R}}(\mathbf{s}_i, \mathbf{s}_j; t)dt, \quad (3)$$

with $C_{\mathbb{R}}(\mathbf{s}_i, \mathbf{s}_j; t)$ the point-wise covariogram, defined for the point evaluations at t in \mathcal{T} of the functions $\mathcal{X}_{\mathbf{s}}$.

Methods in object oriented geostatistics usually assume that the function C_H only depends on the increment between locations (i.e., stationarity of centered observations $\mathcal{X}_{\mathbf{s}_i} - m_{\mathbf{s}_i}$, or residuals). Hereafter, we focus on this case, and express C_H as a function of $\mathbf{s}_i - \mathbf{s}_j$.

Based on (1) and (3), prediction of the unobserved object $\mathcal{X}_{\mathbf{s}_0}$ can be performed through a Universal Kriging predictor. This is the best linear unbiased predictor $\mathcal{X}_{\mathbf{s}_0}^* = \sum_{i=1}^n \lambda_i^* \cdot \mathcal{X}_{\mathbf{s}_i}$, in the sense that the scalar weights $\lambda_1^*, \dots, \lambda_n^*$ are determined as to minimize the mean square error in H (MSE_H) subject to the unbiasedness constraint, i.e.,

$$\begin{aligned} \min_{\lambda_1, \dots, \lambda_n} \quad & \mathbb{E}_H \left[\left\| \mathcal{X}_{\mathbf{s}_0} - \sum_{i=1}^n \lambda_i \cdot \mathcal{X}_{\mathbf{s}_i} \right\|^2 \right] \\ \text{subject to} \quad & \mathbb{E}_H \left[\sum_{i=1}^n \lambda_i \cdot \mathcal{X}_{\mathbf{s}_i} \right] = \mathbb{E}_H [\mathcal{X}_{\mathbf{s}_0}]. \end{aligned} \quad (4)$$

Note that, geometrically, MSE_H is the expected square distance between the prediction and the target, measured according to the metric of H . Object oriented Kriging thus assumes the same geometrical interpretation as in one-dimension (see, e.g., Chilès & Delfiner, 1999, Chap. 3). Furthermore, even though the Kriging problem is formally posed in H (recall, the space H may be of any dimension), the particular form of the predictor – as a linear combination of the data – and the geometrical properties of H – as a Hilbert space – allow finding the optimum with standard techniques. Specifically (Menafoglio et al., 2013): (a) the MSE_H and the unbiasedness constraints are expressed in terms of the trace-covariogram C_H , the regressors $\{f_l(\mathbf{s})\}$ and the weights λ_i , (b)

Lagrange multipliers are employed to account for the unbiasedness constraints, and (c) a Lagrange functional is differentiated to find the optimum. The vector of optimal weights $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_n^*)^T$ is eventually found as

$$\boldsymbol{\lambda}^* = \Sigma^{-1} [\boldsymbol{\sigma}_0 + \mathbb{F}(\mathbb{F}'\Sigma^{-1}\mathbb{F})^{-1}(\mathbf{f}_0 - \mathbb{F}'\Sigma^{-1}\boldsymbol{\sigma}_0)], \quad (5)$$

where

$\Sigma \in \mathbb{R}^{n,n}$ is the variance covariance matrix of the observations, in the sense of the measure of dependence defined through the trace-covariogram C_H : $\Sigma_{ij} = C_H(\mathbf{s}_i - \mathbf{s}_j)$, $i, j = 1, \dots, n$;

$\boldsymbol{\sigma}_0 \in \mathbb{R}^n$ is the vector of trace-covariances between $\mathcal{X}_{\mathbf{s}_0}$ and the observations: $[\boldsymbol{\sigma}_0]_i = C_H(\mathbf{s}_0 - \mathbf{s}_i)$, $i = 1, \dots, n$;

$\mathbb{F} \in \mathbb{R}^{n,(L+1)}$ is the design matrix of model (1): $\mathbb{F}_{il} = f_l(\mathbf{s}_i)$, $i = 1, \dots, n$, $l = 0, \dots, L$;

$\mathbf{f}_0 \in \mathbb{R}^{(L+1)}$ collects the values of the regressors at the target location \mathbf{s}_0 (assumed to be known): $[\mathbf{f}_0]_l = f_l(\mathbf{s}_0)$, $l = 0, \dots, L$.

Note that a very analogue result is found in the classical setting (see, e.g., Cressie, 1993, p. 154). The Universal Kriging prediction at \mathbf{s}_0 is then found as $\mathcal{X}_{\mathbf{s}_0}^* = \sum_{i=1}^n \lambda_i^* \cdot \mathcal{X}_{\mathbf{s}_i}$, and a measure of the uncertainty associated with the point estimate is provided by the Universal Kriging trace-variance, that is the value of the MSE_H attained at the optimum of problem (4) (i.e., for $\boldsymbol{\lambda}^*$). The latter can be computed explicitly by combining the expression of the objective functional with that of $\boldsymbol{\lambda}^*$ in (5).

The results illustrated so far assume that the covariance structure is known. This is hardly verified in applications. An estimate $\hat{\Sigma}$ of the covariance matrix Σ is usually plugged-in in the above expressions, and $\mathcal{X}_{\mathbf{s}_0}^*(\hat{\Sigma})$ is eventually used for the purpose of prediction. Most of the literature advocates a two step procedure to estimate C_H , that is (i) use a method of moments estimator to provide a raw estimate of the trace-covariogram (or the associated trace-variogram, e.g., Menafoglio et al., 2013), (ii) fit a parametric model (e.g., spherical, Matérn) to the obtained estimate, by least squares. Note that least square methods are usually preferred to maximum likelihood methods in the object oriented setting, since the likelihood might not be defined in the space H (e.g., if it is infinite dimensional); in fact, developing more effective estimation procedures is still an open problem.

Although expression (5) does not depend on the (unknown) coefficients a_l , trace-covariogram estimation requires to estimate the spatial mean (1) whenever the mean function $m_{\mathbf{s}}$ is not spatially constant. Indeed in practice, to estimate C_H one employs the identity $C_H(\mathbf{h}) = C_H(\mathbf{0}) - \gamma_H(\mathbf{h})$, that relates the trace-covariogram with the trace-semivariogram

$$\gamma_H(\mathbf{s}_i - \mathbf{s}_j) = \frac{1}{2} [\mathbb{E}[\|\mathcal{X}_{\mathbf{s}_i} - \mathcal{X}_{\mathbf{s}_j}\|^2] - \|m_{\mathbf{s}_i} - m_{\mathbf{s}_j}\|^2], \quad \mathbf{s}_i, \mathbf{s}_j \in D. \quad (6)$$

The latter function is proportional to the global variance of increments of the process, and approaches the variance of the process when the modulus of its argument diverges, i.e., $C_H(\mathbf{0}) = \lim_{|\mathbf{h}| \rightarrow \infty} \gamma_H(\mathbf{h})$ (stationarity and ergodicity conditions apply). If the mean function $m_{\mathbf{s}}$ is spatially constant, estimation of the trace-covariogram can be then obtained by fitting a parametric valid model (e.g., spherical, Matérn) to the empirical trace-semivariogram

$$\hat{\gamma}_H(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} \|\mathcal{X}_{\mathbf{s}_i} - \mathcal{X}_{\mathbf{s}_j}\|^2, \quad (7)$$

where $N(\mathbf{h})$ is the set of pairs of locations $(\mathbf{s}_i, \mathbf{s}_j)$ whose separating vector is \mathbf{h} (or within $\mathbf{h} \pm \Delta\mathbf{h}$, with $\Delta\mathbf{h}$ a given tolerance), and $|N(\mathbf{h})|$ is its cardinality. Note that, when $H = L^2$, estimator (7) reads (Giraldo et al., 2011)

$$\hat{\gamma}_{L^2}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} \int_{\mathcal{T}} (\mathcal{X}_{\mathbf{s}_i}(t) - \mathcal{X}_{\mathbf{s}_j}(t))^2 dt,$$

that can be computed from data via numerical quadrature schemes (e.g., trapezoidal rule). When the mean functions $m_{\mathbf{s}}$ is not spatially constant, estimator (7) cannot be directly used to provide a meaningful estimate of the trace-covariogram (it misses to estimate the second term in (6)), but should be applied to the (estimated) residuals $\hat{\delta}_{\mathbf{s}} = \mathcal{X}_{\mathbf{s}} - \sum_{l=0}^L f_l(\mathbf{s}) \cdot \hat{a}_l$ instead. In fact, likewise in the classical setting (Chilès & Delfiner, 1999), a good estimate of the drift (i.e., of the parameters $\{a_l, l = 0, \dots, L\}$) is then crucial for the statistical analysis. For this purpose, one may employ, e.g., Generalized Least Squares. Having estimated $\{a_l, l = 0, \dots, L\}$, one could optionally consider to perform Kriging of the residuals $\hat{\delta}_{\mathbf{s}}$ (i.e., Residual Kriging) instead of Universal Kriging, taking care of setting the residuals mean to zero in the unbiasedness constraint appearing in (4).

We finally remark that the predictor $\mathcal{X}_{\mathbf{s}_0}^*(\hat{\Sigma})$ is generally characterized by a higher uncertainty than that described by the Universal Kriging trace-variance. For the purpose of uncertainty assessment, resampling methods can be considered. Recent works in this direction employ semi-parametric bootstrap (Franco-Villoria & Ignaccolo, 2015; Pigoli et al., 2016), or conditional stochastic simulation (Menafoglio et al., 2016b).

Example 1 (continued). *Based on (Menafoglio et al., 2014, 2016c), we describe the results of Kriging for the dataset of PSDs of Example 1, displayed in Figure 1. Since the data are PDFs, Menafoglio et al. (2014, 2016c) propose to consider as feature space the Bayes Hilbert space $A^2(\mathcal{T})$ of functional compositions on $\mathcal{T} = [t_m, t_M]$ (t_m, t_M denoting the minimum and maximum observable grain-size), introduced by Egozcue et al. (2006); van den Boogaart et al. (2014). $A^2(\mathcal{T})$ is a space of functions with constant integral (such as PDFs), equipped with the generalization to the infinite-dimensional setting of the Aitchison geometry for compositional data (Pawlowsky-Glahn et al., 2015), which is widely employed in the multivariate setting to describe PDFs. The space $A^2(\mathcal{T})$ is a*

Hilbert space if equipped with the following operations of perturbation (+) and powering (\cdot)

$$(f + g)(t) = \frac{f(t)g(t)}{\int_{\mathcal{T}} f(s)g(s) ds}; \quad (\alpha \cdot f)(t) = \frac{f(t)^\alpha}{\int_{\mathcal{T}} f(s)^\alpha ds}$$

and the inner product

$$\langle f, g \rangle = \frac{1}{2|\mathcal{T}|} \int_{\mathcal{T}} \int_{\mathcal{T}} \ln \frac{f(t)}{f(s)} \ln \frac{g(t)}{g(s)} dt ds, \quad f, g \in A^2(\mathcal{T}).$$

The key point here is that the geometry of the space is precisely designed to account for the inherent features of the data (for further details see Egozcue et al., 2013; van den Boogaart et al., 2014). This would not be the case if one just neglected the data constraint and worked in $L^2(\mathcal{T})$. Note that other choices of the feature space would be sensible as well for these distributional data. Indeed, one could summarize the underlying distribution through the cumulative distribution function or the quantile function instead of the PDFs, and then embed these quantities in an appropriate (Hilbert) space. In this sense, we here focus on one of the possible object oriented geostatistical analyses that could be performed on this dataset; we refer the reader to Section 7 for further discussion on the choice of the feature space.

Figure 3 represents the results of Kriging of the PSDs, based on the work of Menafoglio et al. (2016c). Here, information regarding the presence of three groups of data (associated with as many soil types) was included via linear model (1), through binary variables. Unlike traditional approaches that focus on some selected features of the PSDs (e.g., quantiles of the distribution), object oriented Kriging allows to project the entire information embedded within the data – including their grouping structure – to unsampled locations in the system. This can be used for, e.g., global quantile assessment and hydrofacies characterization in Monte Carlo simulations.

For the sake of completeness, we here mention some recent works that stand at the very forefront of object oriented Kriging. Extensions of the methodology here described have been considered, e.g., by Menafoglio & Petris (2016), who develop a linear operatorial predictor, based on optimal linear operators rather than optimal scalar weights. Further extensions in the direction of object data in Riemannian manifolds are proposed by Pigoli et al. (2016). Informally, Riemannian manifolds are curved spaces that admit local approximations via Hilbert tangent spaces. As a key element of innovation, Pigoli et al. (2016) propose to locally approximate the data through the Hilbert tangent space, and accordingly employ the methodology here described for the purpose of prediction. Finally, much effort has been recently devoted to the extension of Kriging to more general feature spaces, like Banach spaces or Hadamard metric spaces. The research field is still very open, and its development involves advanced and convoluted mathematical constructions. Generally, in these settings Kriging solutions are hardly obtainable in explicit form, and sub-optimal solutions might

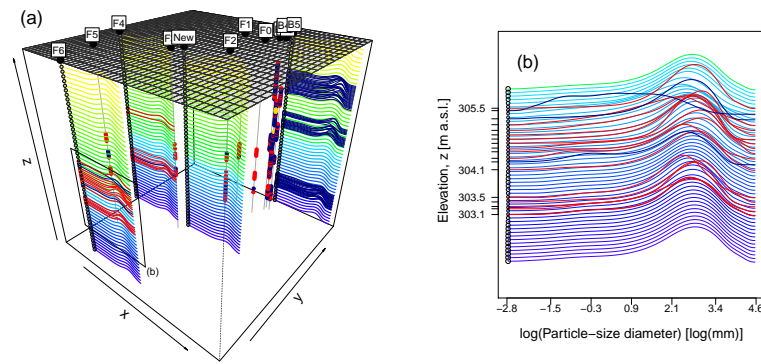


Figure 3: Prediction of PSDs via Kriging for object data: (a) results at boreholes B5, F4 and F6 and at an undrilled location (“New”) having coordinates (3508600,5377670). (b) Vertical distribution of kriged PSDs, for the samples at elevations $301 \leq z \leq 306$ m above sea level (a.s.l.), at borehole F6. In both panels: colors of the solid curves indicate the ordering of the data along depth; the size of the symbols is proportional to the Kriging variance; smoothed data are represented with solid curves colored according to the soil type (Figure modified from Menafoglio et al. (2016c))

be employed; recent results in this direction are presented in Gouet et al. (2015).

5. Bagging Voronoi: an efficient algorithm for the statistical treatment of complex spatial data

The methods discussed so far are based on global models for the mean and the spatial dependence, and accordingly rely upon the assumption that a certain *regularity* characterizes the spatial field (i.e., linear model for the drift, residual stationarity). Even though this assumption may be viable in field scale studies (e.g., Example 1), in the presence of very large or textured spatial domains it might be too restrictive. Several probabilistic constructions have been proposed in geostatistics to model non-stationary spatial (or space-time) processes (see Sampson, 2010, for a recent review), e.g., smoothing and weighted-average methods, basis function methods, process convolutions or spatially varying parameters (see, e.g., Fouedjio et al., 2016; Guttorp & Sampson, 1994; Guttorp & Schmidt, 2013; Paciorek & Schervish, 2006; Sampson et al., 2001; Stein, 2005). However it is hard to reformulate these methodologies and generalize them to the context of object data analysis. A powerful alternative is represented by non-parametric and computationally intensive approaches, that allow avoiding strong distributional assumptions (e.g., Gaussianity) or parametric models for the spatial dependence. In this setting, the Bagging Voronoi (BV) algorithm (Secchi et al., 2013) is a notable instance of a fully non-parametric and flexible method that allows working with general types of object data.

BV has been originally conceived to cluster functional data observed over a lattice D_0 . Nevertheless, the algorithm can be applied without substantial modifications to general object data, and for a variety of target statistical analyses (e.g., dimensionality reduction or joint clustering and alignment of data Secchi et al., 2015; Manfredini et al., 2015; Abramowicz et al., 2016). For ease of exposition, we here focus on an Euclidean spatial domain D_0 , although more complex situations may be considered as well; the reader is referred to Cressie (1993, Part II) for an introduction to classical spatial statistics for lattice data.

We call $\mathcal{X}_{\mathbf{s}}$ the observation at $\mathbf{s} \in D_0$: in most applications, observations are given over the entire lattice D_0 . Suppose for the moment that the aim of the analysis is to cluster the observations, i.e., unsupervised classification. If data were independent, one could employ an extension to object data of classical methods, for instance k-means (see Tarpey & Kinateder, 2003, for an extension of k-means to functional data). To include the spatial information in the procedure one might want to adapt the algorithm to explicitly account for spatial dependence (e.g., by modeling a trace-covariogram, Menafoglio et al., 2016c). The idea behind BV is however different and follows a *divide et impera* approach. This avoids modifying the target statistical method (e.g., k-means), by applying it to a batch of auxiliary data for which spatial dependence is negligible. BV consists of two steps, *bootstrap* and *aggregation*, that are detailed in the following paragraphs.

Bootstrap. Aim of this step is to (i) build a batch of B auxiliary datasets characterized by a negligible spatial dependence, and (ii) apply the target statistical analysis to each of these datasets. More precisely, the following steps are repeated B times. First the spatial domain D_0 is partitioned in n subregions (*neighborhoods*): n sites (*nuclei*) in D_0 are randomly sampled, and the remaining sites are assigned to the nearest nucleus (i.e., a Voronoi tessellation is built). Second, associated with each nucleus a local representative is computed as a summary of the data within the corresponding neighborhood (e.g., via a weighted average or the use of kernels, Genton (2002)). These steps actually weaken the spatial dependence: indeed, computation of representatives via geographically weighted average (or spatial smoothing) acts as a filter of the correlated random noise, whereas the spatial dependence of the local representatives tends to be negligible when neighborhoods are large, as the local representatives tend to be spread across the domain. The target analysis is thus applied to the dataset of n local representatives by neglecting their spatial information, and the result is stored.

Aggregation. The results of the B analyses performed on the auxiliary datasets are aggregated in order to obtain the final result. Here, the kind of aggregation which can be used is problem specific. For instance, in the case of clustering, at each iteration of the bootstrap step, every datum is assigned the cluster label pertaining to the nucleus of the neighborhood it belongs to. For the final cluster label one may employ a majority vote and assign to each datum the label obtained more often along the B analyses.

Note that two parameters need to be initialized: the number of auxiliary analyses B , and the number of nuclei n . Parameter B controls the “strength” of the final result: the higher the number B of “weak” analyses performed, the “stronger” the basis upon which the final result is obtained. Parameter n controls the size of the auxiliary datasets as well as the mean size of the neighborhoods. Thus, high values of n are associated with high complexity in each weak analysis and with small neighborhoods. Note that the latter directly reflects also on (a) the spatial correlation among representatives (higher when the neighborhood size decreases) and (b) the number of data upon which the local representative are computed. In fact, a bias variance trade-off exist when setting parameter n , and this should be carefully evaluated (see Secchi et al., 2013, Section 3.2). Several criterion, mainly empirical, can be employed to set the parameters, e.g., minimization of the entropy (Secchi et al., 2013), or of the total average variance (Secchi et al., 2015).

6. Smoothing spatial objects with a differential penalization

We finally illustrate an approach recently proposed to face the problem of textured spatial domains D_0 characterized by holes or irregular boundaries, as those described in Example 2. For clarity of exposition we again focus on a Euclidean spatial domain; extensions to non-Euclidean domains are discussed

at the end of the section. In this setting, the available literature mostly focuses on real-valued responses (i.e., $H = \mathbb{R}$ with the notation of Section 4), with a few extensions to deal with spatio-temporal processes. Extensions dealing with object data are still at the very cutting edge of the research on this field.

A semi-parametric model is here considered for the observations

$$\mathcal{X}_{\mathbf{s}_i} = \sum_{l=0}^L f_l(\mathbf{s}_i) \cdot a_l + g(\mathbf{s}_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (8)$$

where

$a_l, l = 0, \dots, L$, are coefficients in \mathbb{R} (as in (1), with $H = \mathbb{R}$);

$f_l(\mathbf{s}_i), l = 0, \dots, L$, are known regressors (covariates) observed at point \mathbf{s}_i (as in (1), with $H = \mathbb{R}$);

g is a twice differentiable real-valued function;

$\varepsilon_i, i = 1, \dots, n$, are *independent* errors with zero mean and variance σ^2 .

We note that two main differences exist between this model and the Universal Kriging model detailed in Section 4. Model (8) is semi-parametric, in the sense that the term g is a non-parameterized spatial effect on the response, assumed to be deterministic. This term is assumed to capture the entire spatial dependence, and the regularity (e.g., smoothness) of the field is implicitly accounted for through the functional properties of g . The random part of the model is here simpler: the ε 's are assumed to be independent, the residuals of the Universal Kriging approach are spatially correlated instead.

In this context, estimation of the surface g and of the parameters $\mathbf{a} = (a_0, \dots, a_L)^T$, can be performed by minimizing a penalized sum-of-square-error (penalized SSE) functional

$$J_\lambda(\mathbf{a}, g) = \sum_{i=1}^n \left(\mathcal{X}_{\mathbf{s}_i} - \sum_{l=0}^L f_l(\mathbf{s}_i) \cdot a_l - g(\mathbf{s}_i) \right)^2 + \lambda \mathcal{P}(g), \quad (9)$$

where $\mathcal{P}(g)$ is a penalty that controls the functional properties of the estimated surface g (e.g., the curvature), and $\lambda > 0$ a parameter that set the importance of the penalization with respect to the sum-of-square-error term. Popular choices for $\mathcal{P}(g)$ account for the second derivative (i.e., curvature) of the surface g , to control its roughness. For instance, Ramsay (2002) proposes to consider (in a model without covariates)

$$\mathcal{P}(g) = \int_D \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial x \partial y} + \frac{\partial^2 g}{\partial y^2} \right)^2, \quad \mathbf{s} = (x, y) \in \mathbb{R}^2,$$

while Wood et al. (2008); Sangalli et al. (2013) consider the Laplacian

$$\mathcal{P}(g) = \int_D (\Delta g)^2 = \int_D \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right)^2, \quad \mathbf{s} = (x, y) \in \mathbb{R}^2.$$

The penalization can express a prior knowledge on the phenomenon. For instance, if a set of differential equations are known to govern the system – at least in an ideal setting of the problem – one can consider as penalization the misfit of the estimated surface from the solution of those equations (see Azzimonti et al., 2015).

Note that this approach is quite different from that illustrated in Section 4. Indeed, the first term of $J_\lambda(\mathbf{a})$ is the sum of square errors between the model and the observations: in this sense, $J_\lambda(\mathbf{a})$ is conceived as in an estimation problem rather than in a prediction problem, even though the estimated surface is eventually used for prediction purposes at unsampled locations. In addition, the penalization represents a weaker constraint than the unbiasedness constraint in problem (4), and its strength is governed by parameter λ . The latter controls also the bias associated with the resulting estimators. For an interesting discussion on the bias-variance trade-off in smoothing with regularization see Hastie et al. (2009, Chap. 5.4).

In general, minimization of functional (9) cannot be performed explicitly. Amongst the possible approaches, particularly powerful are those based on spatial basis approximations (e.g Wood et al., 2008), especially finite elements methods based on triangulations of the spatial domain – i.e., partition of the domain in triangles (Ramsay, 2002; Sangalli et al., 2013; Azzimonti et al., 2015; Bernardi et al., 2016). These procedures enable one to cope with irregularities in the borders and texture of the domain; the numerical approximation finally allows to minimize the functional (9) by solving a highly sparse linear system. These methods can also deal with the presence of boundary conditions (Dirichlet, Neumann, Robin), as in the case of geographical constraints (see Example 2). They are included in the minimization problem by restricting the class of functions g to those fulfilling the boundary conditions. The numerical approximation of the problem is then performed as to guarantee that these conditions are satisfied (e.g., via finite elements methods).

For the sake of completeness, we have to mention that much effort has been recently spent to extend the perspective described in this section to spatial domains more complex than Euclidean. Amongst the models aimed to handle data distributed over non-planar domains, we cite the works devised for specific Riemannian manifolds such as spheres, hyperspheres or sphere-like domains (domains representable via spherical coordinates), e.g., Wahba (1981); Alfeld et al. (1996); Baramidze et al. (2006); Jun & Stein (2007); Lindgren et al. (2011); Gneiting (2013). Remarkable works in the direction of treating complex domains are Ettinger et al. (2015); Dassi et al. (2015); Lila et al. (2016a); Wilhelm et al. (2015), which extend the seminal approach of Sangalli et al. (2013) to manifolds, such as the cerebral cortex of the brain, by combining methods from numerical analysis (e.g., finite elements) and differential geometry.

7. Discussion

Choice of the feature space. In any (geo)statistical analysis based on an object oriented viewpoint, the choice of the feature space plays a key role. This has been widely recognized in the literature on FDA (see, e.g., Ferraty & Vieu, 2006, Chapter 3). Although most of the earliest papers on functional geostatistics embedded data in $L^2(\mathcal{T})$ (recall, $L^2(\mathcal{T})$ denotes the space of square-integrable functions on \mathcal{T}), the development of the object oriented approach to geostatistics dramatically increased the awareness on the importance of the feature space (see Menafoglio et al., 2013, p. 2214). Granted that in applications concerning unconstrained functional data it is often tempting to set the feature space $H = L^2(\mathcal{T})$, this implies that the differential properties of the data will not be explicitly accounted for in the analysis, since the $L^2(\mathcal{T})$ geometry is based on point-wise operations rather than differential information (i.e., the data derivatives). Menafoglio et al. (2013) argue that a normed space H different from $L^2(\mathcal{T})$ should be chosen when the point-wise and differential properties of the data carry valuable intelligence. For instance, one could consider Sobolev spaces, i.e., Hilbert spaces whose norms jointly account for point-wise and differential information. The proper setting of the feature space H becomes however decidedly relevant for the analysis of constrained functional data, as for the PSDs in Example 1. Here, the geometry of the feature space has to precisely represent the functional form of the data (i.e., functional compositions). Note that the choice of the feature space is crucial even for multivariate data in the presence of constraints. For instance, in the presence of multivariate compositions (i.e., positive vectors summing to a constant, Aitchison, 1982), one should adopt an appropriate geometry such as that of Aitchison in the simplex (Billheimer et al., 2001; Pawlowsky-Glahn & Egozcue, 2001). The reader is referred to (Pawlowsky-Glahn & Egozcue, 2016) for a recent review on geostatistics for compositional data.

O2S2 vs space-time geostatistics. As mentioned in Section 2, some of the problems described in the present review, and particularly those in Sections 4 and 6, could be handled by a spatio-temporal geostatistical approach, for which a large body of literature is available (e.g., Gneiting et al., 2006; Jun & Stein, 2008). Even though the potential of O2S2 is clear when dealing with complex or constrained data as the PSDs in Example 1, understanding the formal relation, as well as strength and weaknesses of the two approaches is yet to be fully investigated. This could also foster the use of object oriented spatial statistics to face relevant operational research issues, such as those related with simulations and meta-modeling of complex systems, that are classically dealt with via Kriging methods (e.g., Kleijnen, 2009; Kleijnen & Mehdad, 2014). Finding the formal connections between (space and space-time) geostatistics and spatial smoothing has stimulated extensive research since the work by Watson (1984), and it still is an open problem within the theoretical framework based on differential penalization discussed in Section 6.

Another issue which is still open concerns the possible interplay of the three

methodologies discussed in Sections 4, 5 and 6. For instance, combining object oriented Kriging and methods based on differential penalizations would enable the benefits and the strengths of both approaches. On the other hand, the trace-covariogram of Section 4 may represent a tool to set the dimension of the Voronoi tessellations used in Section 5, following the intuition that the neighborhood system should be such as to minimize the dependence among representatives, which in turn can be described by a trace-covariogram function. The BV strategy might be then combined with Universal Kriging to build up a concise methodology allowing to deal with severe global non-stationarity.

O2S2 for Large datasets. One of the hottest topic in modern spatial statistics regards the treatment of large datasets (e.g., Sun et al., 2012). O2S2 provides a general framework to treat large amounts of data, as shown by Secchi et al. (2015), who analyze space-time mobile-phone data through a Bagging Voronoi approach. In general, high-dimensional spatial statistics may benefit from O2S2, as the original viewpoint of the latter can provide efficient strategies to handle classical problems. An instance of this is given by the object oriented Kriging of Section 4, which is posed as a convenient and efficient alternative to Cokriging, in high-dimensional problems (Menafoglio & Petris, 2016).

Software. Few pieces of software are available for O2S2; most of these are developed in R (R Core Team, 2013). Amongst these, the R package `geofd` (Giraldo et al., 2010b) has been developed within the context of functional geostatistics; it allows to perform a geostatistical analysis of a functional dataset of square-integrable functions (i.e., setting the feature space to $H = L^2$) according to the approach introduced in Section 4, for the case of constant spatial mean. Voronoi tessellations as those needed to implement a Bagging Voronoi approach can be obtained through, e.g., the R package `deldir` (Turner, 2015). The package `fdapde` by Lila et al. (2016b) implements smoothing with differential penalization as described in Sangalli et al. (2013), and also its extension to the space-time context (Bernardi et al., 2016). For the models with differential penalizations described by Augustin et al. (2013); Marra et al. (2012), alternative implementations can be considered, e.g., by exploiting the package `mgcv` (Wood, 2006). An R package for O2S2 is going to be released shortly by the authors of the present paper.

Acknowledgements. The authors gratefully acknowledge the Editor and the anonymous Referees for their valuable comments, which contributed to improve the paper.

References

- Abramowicz, K., Arnvist, P., Secchi, P., Luna, S. S. d., Vantini, S., & Vitelli, V. (2016). Clustering misaligned dependent curves applied to varved lake sediment for climate reconstruction. *Stochastic Environmental Research and Risk Assessment*, (pp. 1–15). doi:10.1007/s00477-016-1287-6.

- Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *44*, 139–177.
- Alfeld, P., Neamtu, M., & Schumaker, L. (1996). Fitting scattered data on sphere-like surfaces using spherical splines. *Journal of Computational and Applied Mathematics*, *73*, 5–43.
- Augustin, N. H., Trenkel, V. M., Wood, S. N., & Lorance, P. (2013). Space-time modelling of blue ling for fisheries stock management. *Environmetrics*, *24*, 109–119.
- Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M., & Nobile, F. (2015). Blood Flow Velocity Field Estimation Via Spatial Regression With PDE Penalization. *Journal of the American Statistical Association*, *110*, 1057–1071.
- Baramidze, V., Lai, M., & Shum, C. (2006). Spherical splines for data interpolation and fitting. *SIAM J Sci Comput*, *28*, 241–259.
- Bernardi, M. S., Mazza, G., Ramsay, J. O., & Sangalli, L. M. (2016). A penalized regression model for spatial functional data with application to the analysis of the production of waste in venice province. *Stochastic Environmental Research and Risk Assessment*, . doi:10.1007/s00477-016-1237-3.
- Billheimer, D., Guttorp, P., & Fagan, W. F. (2001). Statistical interpretation of species composition. *Journal of the American statistical Association*, *96*, 1205–1214.
- Brown, E. N., Kass, R. E., & Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience*, *7*, 456–461.
- Caballero, W., Giraldo, R., & Mateu, J. (2013). A universal kriging approach for spatial functional data. *Stochastic Environmental Research and Risk Assessment*, *27*, 1553–1563.
- Chilès, J. P., & Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York.
- Cressie, N. (1993). *Statistics for Spatial data*. John Wiley & Sons, New York.
- Cressie, N., & Wikle, K. (2011). *Statistics for Spatio-Temporal Data*. Wiley.
- Dassi, F., Ettinger, B., Perotto, S., & Sangalli, L. M. (2015). A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Applied Numerical Mathematics*, *90*, 111 – 131.
- Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data. *Environmetrics*, *21*, 224–239.

- Della Rossa, F., D'Angelo, C., & Quarteroni, A. (2010). A distributed model of traffic flows on extended regions. *Networks and Heterogeneous Media*, 5, 525–544.
- Egozcue, J., Díaz-Barrero, J. L., & Pawlowsky-Glahn, V. (2006). Hilbert space of probability density functions based on Aitchison geometry. *Acta Mathematica Sinica, English Series*, 22, 1175–1182.
- Egozcue, J., Pawlowsky-Glahn, V., Tolosana-Delgado, R., Ortego, M., & van den Boogaart, K. (2013). Bayes spaces: use of improper distributions and exponential families. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 107, 475–486.
- Ettinger, B., Perotto, S., & Sangalli, L. (2015). Spatial regression models over two-dimensional manifolds. *Biometrika*, 103, 71–88. doi:10.1093/biomet/asv069.
- Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis : theory and practice*. Springer, New York.
- Fouedjio, F., Desassis, N., & Rivoirard, J. (2016). A generalized convolution model and estimation for non-stationary random functions. *Spatial Statistics*, 16, 35 – 52. doi:<http://dx.doi.org/10.1016/j.spasta.2016.01.002>.
- Franco-Villoria, M., & Ignaccolo, R. (2015). *Bootstrap based uncertainty bands for prediction in functional kriging*. arXiv:1505.06966.
- Gelfand, A., Diggle, P. J., Fuentes, M., & Guttorp, P. (2010). Handbook of spatial statistics. Handbooks of Modern Statistical Methods. Chapman & Hall/CRC.
- Genton, M. G. (2002). Classes of kernels for machine learning: a statistics perspective. *The Journal of Machine Learning Research*, 2, 299–312.
- Giraldo, R., Delicado, P., & Mateu, J. (2010a). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 66–82.
- Giraldo, R., Delicado, P., & Mateu, J. (2010b). *geofd: a package for prediction for functional data*. URL: <http://code.google.com/p/geofd>.
- Giraldo, R., Delicado, P., & Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, 18, 411–426.
- Giraldo, R., Delicado, P., & Mateu, J. (2012). Hierarchical clustering of spatially correlated functional data. *Statistica Neerlandica*, 66, 403–421. URL: <http://dx.doi.org/10.1111/j.1467-9574.2012.00522.x>. doi:10.1111/j.1467-9574.2012.00522.x.
- Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19, 1087–1500.

- Gneiting, T., Genton, M., & Guttorp, P. (2006). Statistical methods for spatio-temporal systems. chapter Geostatistical space-time models, stationarity, separability, and full symmetry. (pp. 151–175). Chapman & Hall / CRC press volume 107 of *Monographs On Statistics and Applied Probability*.
- Gouet, R., Lopez, A., & Ortiz, J. M. (2015). Geodesic kriging in the wasserstein space. In *Proceedings of the 17th Annual Conference of the International Association for Mathematical Geosciences, Freiberg, September 5-13, 2015*.
- Gromenko, O., Kokoszka, P., Zhu, L., & Sojka, J. (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Annals of Applied Statistics*, 6, 669–696.
- Guttorp, P., & Sampson, P. D. (1994). 20 methods for estimating heterogeneous spatial covariance functions with environmental applications. In *Environmental Statistics* (pp. 661 – 689). Elsevier volume 12 of *Handbook of Statistics*.
- Guttorp, P., & Schmidt, A. M. (2013). Covariance structure of spatial and spatiotemporal processes. *Wiley Interdisciplinary Reviews: Computational Statistics*, (pp. 279–287).
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of Statistical Learning*. (2nd ed.). Springer.
- Horváth, L., & Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer.
- Ignaccolo, R., Franco-Villoria, M., & Fassò, A. (2015). Modelling collocation uncertainty of 3D atmospheric profiles. *Stochastic Environmental Research and Risk Assessment*, 29, 417–429. doi:10.1007/s00477-014-0890-7.
- Ignaccolo, R., Ghigo, & Giovenali, E. (2008). Analysis of air quality monitoring networks by functional clustering. *Environmetrics*, 19, 672–686. doi:10.1002/env.946.
- Ignaccolo, R., Mateu, J., & Giraldo, R. (2014). Kriging with external drift for functional data for air quality monitoring. *Stochastic Environmental Research and Risk Assessment*, 28, 1171–1186. doi:10.1007/s00477-013-0806-y.
- Jun, M., & Stein, M. (2007). An approach to producing space-time covariance functions on spheres. *Technometrics*, 49, 468–479.
- Jun, M., & Stein, M. L. (2008). Nonstationary covariance models for global data. *Ann. Appl. Stat.*, 2, 1271–1289.
- Kleijnen, J. P. (2009). Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, 192, 707 – 716.

- Kleijnen, J. P., & Mehdad, E. (2014). Multivariate versus univariate kriging metamodels for multi-response simulation models. *European Journal of Operational Research*, *236*, 573 – 582. doi:<http://dx.doi.org/10.1016/j.ejor.2014.02.001>.
- Laukaitis, A. (2008). Functional data analysis for cash flow and transactions intensity continuous-time prediction using hilbert-valued autoregressive processes. *European Journal of Operational Research*, *185*, 1607 – 1614. doi:<http://dx.doi.org/10.1016/j.ejor.2006.08.030>.
- Laukaitis, A., & Račkauskas, A. (2005). Functional data analysis for clients segmentation tasks. *European Journal of Operational Research*, *163*, 210 – 216. doi:<http://dx.doi.org/10.1016/j.ejor.2004.01.010>.
- Lee, D., Zhu, Z., & Toscas, P. (2015). Spatio-temporal functional data analysis for wireless sensor networks data. *Environmetrics*, *26*, 354–362. doi:10.1002/env.2344.
- Lila, E., Aston, J., & Sangalli, L. (2016a). *Smooth Principal Component Analysis over two-dimensional manifolds with an application to Neuroimaging*. arXiv:1601.03670. URL: <http://arxiv.org/abs/1601.03670>.
- Lila, E., Sangalli, L. M., Ramsay, J., & Formaggia, L. (2016b). *fdaPDE: Regression with Partial Differential Regularizations, using the Finite Element Method*. URL: <http://CRAN.R-project.org/package=fdaPDE> r package version 0.1-1.
- Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*, 423–498.
- Liu, C., Ray, S., & Hooker, G. (2014). *Functional Principal Components Analysis of Spatially Correlated Data*. arXiv:1411.4681v1.
- Lohmann, T., Hering, A. S., & Rebennack, S. (2016). Spatio-temporal hydro forecasting of multireservoir inflows for hydro-thermal scheduling. *European Journal of Operational Research*, *255*, 243 – 258. doi:<http://dx.doi.org/10.1016/j.ejor.2016.05.011>.
- Manfredini, F., Pucci, P., Secchi, P., Tagliolato, P., Vantini, S., & Vitelli, V. (2015). Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the milan urban region. In A. M. Paganoni, & P. Secchi (Eds.), *Advances in Complex Data Modeling and Computational Methods in Statistics Contributions to Statistics* (pp. 133–147). Springer International Publishing.
- Marra, G., Miller, D. L., & Zanin, L. (2012). Modelling the spatiotemporal distribution of the incidence of resident foreign population. *Statistica Neerlandica*, *66*, 133–160.

- Marron, J. S., & Alonso, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal*, *56*, 732–753. doi:10.1002/bimj.201300072.
- Martin-Barragan, B., Lillo, R., & Romo, J. (2014). Interpretable support vector machines for functional data. *European Journal of Operational Research*, *232*, 146 – 155. doi:http://dx.doi.org/10.1016/j.ejor.2012.08.017.
- Mateu, J., Schoenberg, F., Frederic, P., Diez, D. M., González, J. A., & Lu, W. (2015). On measures of dissimilarity between point patterns: Classification based on prototypes and multidimensional scaling. *Biometrical Journal*, *57*, 340–358. doi:10.1002/bimj.201300150.
- Menafoglio, A., Grujic, O., & Caers, J. (2016a). Universal kriging of functional data: trace-variography vs cross-variography? Application to forecasting in unconventional shales. *Spatial Statistics*, *15*, 39–55.
- Menafoglio, A., Guadagnini, A., & Secchi, P. (2014). A Kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. *Stochastic Environmental Research and Risk Assessment*, *28*, 1835–1851.
- Menafoglio, A., Guadagnini, A., & Secchi, P. (2016b). Stochastic Simulation of Soil Particle-Size Curves in Heterogeneous Aquifer Systems through a Bayes space approach. *Water Resources Research*, (pp. 1–38). doi:10.1002/2015WR018369. In press.
- Menafoglio, A., & 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.
- Menafoglio, A., Secchi, P., & Dalla Rosa, M. (2013). A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. *Electronic Journal of Statistics*, *7*, 2209–2240.
- Menafoglio, A., Secchi, P., & Guadagnini, A. (2016c). A Class-Kriging predictor for Functional Compositions with application to particle-size curves in heterogeneous aquifers. *Mathematical Geosciences*, *48*, 463 – 485. doi:10.1007/s11004-015-9625-7.
- Nerini, D., Monestiez, P., & Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis*, *101*, 409–418.
- Paciorek, C. J., & Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, *17*, 483–506.
- Pawlowsky-Glahn, V., & Egozcue, J. (2001). Geometric approach to statistical analysis in the simplex. *Stochastic Environmental Research and Risk Assessment*, *15*, 384–398.

- Pawlowsky-Glahn, V., Egozcue, J., & Tolosana-Delgado, R. (2015). *Modelling and analysis of compositional data*. John Wiley & Sons, Ltd.
- Pawlowsky-Glahn, V., & Egozcue, J. J. (2016). Spatial analysis of compositional data: A historical review. *Journal of Geochemical Exploration*, *164*, 28–32.
- Pigoli, D., Menafoglio, A., & Secchi, P. (2016). Kriging prediction for manifold-valued random field. *Journal of Multivariate Analysis*, *145*, 117–131.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: <http://www.R-project.org/>.
- Ramsay, J., & Silverman, B. (2005). *Functional data analysis*. (2nd ed.). Springer, New York.
- Ramsay, T. O. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 307–319.
- Reggiani, A., Nijkamp, P., & Sabella, E. (2001). New advances in spatial network modelling: Towards evolutionary algorithms. *European Journal of Operational Research*, *128*, 385 – 401. doi:[http://dx.doi.org/10.1016/S0377-2217\(00\)00079-5](http://dx.doi.org/10.1016/S0377-2217(00)00079-5).
- Romano, E., Balzanella, A., & Verde, R. (2010). Clustering spatio-functional data: A model based approach. In H. Locarek-Junge, & C. Weihs (Eds.), *Classification as a Tool for Research Studies in Classification, Data Analysis, and Knowledge Organization* (pp. 167–175). Springer Berlin Heidelberg. doi:10.1007/978-3-642-10745-0_17.
- Romano, E., Mateu, J., & Giraldo, R. (2015). On the performance of two clustering methods for spatial functional data. *AStA Advances in Statistical Analysis*, *99*, 467–492. doi:10.1007/s10182-015-0253-9.
- Ruiz-Medina, M. D. (2012). Spatial functional prediction from spatial autoregressive Hilbertian processes. *Environmetrics*, *23*, 119–128. doi:10.1002/env.1143.
- Sampson, P. (2010). Handbook of spatial statistics. Handbooks of Modern Statistical Methods chapter Constructions for Nonstationary Spatial Processes. (pp. 119–130). Chapman & Hall/CRC.
- Sampson, P., Damian, D., & Guttorp, P. (2001). Advances in modeling and inference for environmental processes with nonstationary spatial covariance. In P. Monestiez, D. Allard, & R. Froidevaux (Eds.), *geoENV III Geostatistics for Environmental Applications* (pp. 17–32). Springer Netherlands volume 11 of *Quantitative Geology and Geostatistics*. URL: http://dx.doi.org/10.1007/978-94-010-0810-5_2. doi:10.1007/978-94-010-0810-5_2.

- Sangalli, L. M., Ramsay, J. O., & Ramsay, T. O. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*, 681–703.
- Secchi, P., Vantini, S., & Vitelli, V. (2013). Bagging voronoi classifiers for clustering spatial functional data. *International Journal of Applied Earth Observation and Geoinformation*, *22*, 53 – 64. Spatial Statistics for Mapping the Environment.
- Secchi, P., Vantini, S., & Vitelli, V. (2015). Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of Milan (with discussion). *Statistical Methods & Applications*, *24*, 279–300.
- Stein, M. (2005). *Nonstationary spatial covariance functions*. Technical Report. URL: <http://www-personal.umich.edu/~jizhu/jizhu/covar/Stein-Summary.pdf>.
- Sun, Y., Li, B., & Genton, M. G. (2012). Geostatistics for large datasets. In E. Porcu, J. M. Montero, & M. Schlather (Eds.), *Space-Time Processes and Challenges Related to Environmental Problems* chapter 3. (pp. 55–77). Springer volume 207.
- Tarpey, T., & Kinateder, K. (2003). Clustering functional data. *Journal of Classification*, *20*, 93–114.
- Turner, R. (2015). *deldir: Delaunay Triangulation and Dirichlet (Voronoi) Tessellation*. URL: <https://www.stat.auckland.ac.nz/~rolf/>.
- van den Boogaart, K. G., Egozcue, J., & Pawlowsky-Glahn, V. (2014). Bayes Hilbert spaces. *Australian & New Zealand Journal of Statistics*, *56*, 171–194.
- Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, *2*, 5–16.
- Wang, H., & Marron, J. S. (2007). Object oriented data analysis: Sets of trees. *The Annals of Statistics*, *35*, 1849–1873.
- Watson, G. (1984). Smoothing and interpolation by kriging and with splines. *Journal of the International Association for Mathematical Geology*, *16*, 601–615.
- Wilhelm, M., Dedè, L., Sangalli, L., & Wilhelm, P. (2015). Igs: an isogeometric approach for smoothing on surfaces. *Computer Methods in Applied Mechanics and Engineering*, . In press.
- Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.
- Wood, S. N., Bravington, M. V., & Hedley, S. L. (2008). Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*, 931–955.