# Geostatistics: Unde venis et quo vadis?

## JOSÉ-MARÍA MONTERO

University of Castilla-La Mancha, Faculty of Law and Social Sciences, Campus of Toledo, Cobertizo San Pedro Mártir, s/n, 45071 Toledo, Spain. E-mail: jose.mlorenzo@uclm.es

#### ABSTRACT

Spatial and spatio-temporal data are not new. They have always been here. However, until fairly recently the scientific community had at its disposal neither the theory nor the technology to deal with them. It was thus unable to take advantage of the spatial or spatio-temporal dependencies existing in a large number of phenomena when it comes to estimating, predicting, mapping, experimental designing, etc. This article first addresses the adverse consequences of failing to account for such dependencies. Then, it traces the route travelled by geostatistics from its beginnings to the present day, providing the reader with the rudiments of kriging, the technique geostatistics uses for inferential purposes. However, this article is not only an overview. It poses a number of criticisms, alternatives, open questions and future avenues of research in its different phases: spatial-only, spatio-temporal and functional. Initially having arisen in the Earth sciences, and having been applied in many natural sciences, in recent years geostatistics has stepped into the sphere of social sciences and particularly applied economy.

Keywords: Geostatistics, Kriging, Space, Space-Time, Functional Kriging.

# Geoestadística: ¿De dónde vienes y a dónde vas?

#### RESUMEN

Los datos espaciales o espacio-temporales no son nuevos. Han estado siempre ahí. Sin embargo, hasta hace pocos años no se ha desarrollado la teoría y la tecnología para que la comunidad científica pueda trabajar con ellos y sacar provecho de las dependencias espaciales o espacio-temporales que presentan un buen número de fenómenos de la realidad a la hora de llevar a cabo tareas de estimación, predicción, mapeo, diseño experimental, etc. Este artículo pone de manifiesto, primeramente, las perversas consecuencias que se derivan de ignorar tales dependencias. A continuación, recorre el camino seguido por la geostadística desde sus inicios hasta el presente, proporcionando al lector los rudimentos básicos del kriging, la técnica que usa la geoestadística para llevar a cabo labores inferenciales. Sin embargo, este artículo no es sólo una visión panorámica de la evolución de la geoestadística hasta nuestros días. También plantea un buen número de críticas, alternativas, cuestiones que todavía permanecen abiertas y futuras líneas de investigación en cada uno de los estadios de la disciplina: espacial, espacio-temporal y funcional. Nacida en el ámbito de las ciencias de la Tierra y aplicada en muchas de las ciencias de la naturaleza, en la actualidad la geoestadística ha encaminado sus pasos hacia las ciencias sociales, y en particular hacia la economía aplicada.

Palabras Clave: Geoestadística, kriging, espacio, espacio-tiempo, kriging funcional.

JEL Classification: C14, C21, C23, C31, C33, C55

Artículo recibido en septiembre de 2017 y aceptado en noviembre de 2017 Artículo disponible en versión electrónica en la página www.revista-eea.net, ref. ə-36105

# 1. SPATIAL DATA AND SPATIAL STATISTICS

Spatial data are not new; they have always been here. The first manifestations of statistics for spatial data arose in the form of maps. A famous example is that of Halley (1686), who superimposed onto a map of land forms directions of trade winds and monsoons between and near the tropics, and attempted to assign a physical cause to this phenomenon. However, although data that are spatial in nature have been frequently used in the literature, the spatial aspects of empirical information have not been taken into account by the scientific community until recent times.

Spatial data arise in a myriad of fields and applications<sup>1</sup> and, consequently, there is also a myriad of spatial data types, structures and scenarios. Thus, an exhaustive classification of spatial data is a difficult challenge, and this is why the scientific community have opted to embrace the general, simple and useful classification of spatial data provided by Cressie (1993, pp. 8–13). That classification is based on the nature of the spatial domain under study, according to which the spatial data can be geostatistical data, lattice data or point patterns.

Following Cressie (1993, 2015), let  $\mathbf{s} \in \mathbb{R}^d$  be a generic location in a *d*dimensional Euclidean space and  $\{Z(\mathbf{s}):\mathbf{s} \in \mathbb{R}^d\}$  a spatial random function (RF), Z denoting the attribute of interest. Geostatistical data arise when the domain under study is a fixed set *D* that is continuous. That is, (i)  $Z(\mathbf{s})$  can be observed at any point of the domain (continuous); and (ii) the points in *D* are non-stochastic (fixed, *D* is the same for all the realizations of the spatial RF). From (i) it is clear that geostatistical data are identified with spatial data with continuous variation (the spatial process is indexed over a continuous space).

Lattice data arise when: (i) The domain under study D is discrete, that is,  $Z(\mathbf{s})$  can be observed in a number of fixed locations that can be enumerated. These locations can be points or regions, but they are usually ZIP codes, census tracks, neighborhoods, provinces, etc., and in most cases the data are spatially aggregated over these areal regions. Although these regions can be regularly shaped, their shape is usually irregular, and this, together with the spatially aggregated nature of the data, is why lattice data are also called regional data. And (ii) the locations in D are non-stochastic. Of course, a core concept in

<sup>&</sup>lt;sup>1</sup> Mining, Petroleum Industry, Geology, Oceanography, Meteorology, Soil Cartography, Soil Science, Hydrology and Biohydrology, Forestry, Ecology, Geometallurgy, Plant Pathology, Landscape Ecology, Agriculture, Epidemiology, Entomology, Environmental Sciences, Contaminated Soil Remediation, Public Health, Medical Geography, Criminology, Commerce, Military Planning, Real Estate Pricing, Applied Economics and Developing of Spatial Networks, to mention only some of them.

lattice data analysis is neighborhood. Unlike geostatistical data, lattice data can be exhaustively observed and in this case, prediction makes no sense. However, smoothing and clustering become particularly important when dealing with this type of spatial data.

Unlike with geostatistical and lattice data, in point pattern data the domain is not fixed, but random (discrete or continuous). Point patterns arise when the attribute under study is the location of events (observations). Thus, the domain is random and the observation points do not depend on the researcher. The main goal of point pattern analysis is to determine whether the location of events tends to exhibit a systematic pattern over the area under study or they are randomly distributed. More specifically, we are interested in analyzing whether the location of events is completely spatially random (the location where events occur is not affected by the location of other events), uniform or regular (every point is as far from all of its neighbors as possible) or clustered or aggregated (the location of events is concentrated in clusters). Other interesting questions in point pattern analysis include: how does the intensity of point patterns vary over an area? Over what spatial scales do patterns exist?

The above refers to merely spatial data, but in recent years spatio-temporal data analysis has become a promising research area in a wide variety of scientific disciplines. In the spatio-temporal context, the observed data are viewed as partial realizations of a spatio-temporal RF  $\{Z(\mathbf{s},t):\mathbf{s} \in \mathbb{R}^d, t \in T\}$  which spreads out in space and evolves in time. Thus, spatio-temporal data simultaneously capture spatial and temporal aspects of data.

According to Montero et al. (2015), examples of geostatistical data include the concentrations of air pollution in a city, precipitation or air temperature values in a country, etc. At least theoretically, the concentrations of a specific pollutant could be measured at any location of the city (the same can be said for measurements of precipitations or air temperatures across a country). However, in practice, an exhaustive observation of the spatial process is not possible and so the spatial process is observed at a set of locations (for example, in the case of air pollution, at the points where the monitoring stations are located). Then, based on these observed values, geostatistical analysis reproduces the behavior of the spatial process across the entire domain of interest. Sometimes the goal is not so ambitious; instead, the objective is limited to prediction at one or a few non-observed points or the estimation of an average value over small areas, or over the whole area under study. Geostatistics takes advantage of the spatial (or spatio-temporal) dependencies existing in the data, which are captured by the basic tool in geostatistics, the semivariogram, and uses this information to achieve the above goals.

Some examples of lattice data include the unemployment rate by states, crime data by counties, agricultural yields in plots, average housing prices by provinces, etc. Examples of point patterns are the location of fires in an area, the location of trees in a forest or the location of nests in a breeding colony of birds, etc.

# 2. THE NEED FOR GEOSTATISTICS

Classical statistics is based on the independence of the observed values, which are considered as independent realizations of the same random variable. Data used in classical statistics are typically summarized by means, medians, variance, skewness, excess of kurtosis, box-plots, etc. However, when data are spatial in nature, the hypothesis of independence is not really acceptable and the First Law of Geography, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970) applies. In this case, the summary must include the empirical semivariogram (or alternatively the spatial covariance), which is the instrument geostatistics uses to capture the spatial correlation existing in the data. In addition, the spatial location of the observed data becomes particularly relevant.

Ignoring the spatial dependencies in the data and using classical statistical methods with spatial data has adverse consequences for estimation, prediction and experimental design. By way of example, we focus on estimation in the spatial case but the consequences of using classical statistics with spatio-temporal data are the same in the cases of prediction and experimental design (see Cressie, 1993).

Let  $\{Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n)\}$  be *n* identically distributed observations measured at spatial locations  $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$ . More specifically, suppose that they follow a Gaussian distribution with unknown mean,  $\mu$ , and known variance,  $\sigma_0^2$ , and that covariances between the observed locations are positive and diminish with the distance *h* between them according to the covariance function:  $C(h) = \sigma_0^2 \left[ 1 - \left(\frac{3}{2}\frac{h}{110} - \frac{1}{2}\frac{h^3}{110^3}\right) \right]$ . Suppose that our aim is to estimate the unknown mean and that we ignore the existing spatial correlation. In such a situation, the sample mean,  $\overline{Z}$ , would be the estimator for  $\mu$ , so that  $E(\overline{Z}) = \mu$ and  $V(\overline{Z}) = \sigma_0^2/n$ . But, in the presence of spatial correlation, although  $\overline{Z}$ remains unbiased it no longer has a variance of  $\sigma_0^2/n$ , but rather  $V(\overline{Z}) = \frac{\sigma_0^2}{n} \left(1 + \frac{2}{n}\sum_{i<j}^n \left[1 - \left(\frac{3}{2}\frac{h}{110} - \frac{1}{2}\frac{h^3}{110^3}\right)\right]\right)$  (see calculations in Montero *et al.*, 2015), which is greater than  $\sigma_0^2/n$  in the typical case of positive correlation. Therefore, the use of the classical estimator of  $\mu$  when dealing with spatial data results in the underestimation of  $V(\overline{Z})$ , which has unfortunate consequences for inference about  $\mu$ : (i) The classical confidence intervals for a specific confidence level will be narrower than they really are; or, in other words, the confidence level of classical intervals is greater than it really is. When testing, the *p*-values will be larger than they really are, which will lead to undesirable rejections of the null hypothesis; in addition, the power of the tests will be overstated (examples can be seen in Cressie, 1993, 2015; Shabenberger and Gotway, 2005; Montero *et al.*, 2015).

Summarizing, classical statistics has important limitations when dealing with spatial (or spatio-temporal) data, an issue which is becoming ever more common in empirical research (especially, in recent times, in the field of applied economics), and new instruments are needed to overcome these limitations. Geostatistics is one of the disciplines that provides these instruments (spatial econometrics and local modeling are other disciplines dealing with the spatial, or spatio-temporal, aspects of data).

# 3. THE ORIGINS: SPATIAL-ONLY GEOSTATISTICS

# **3.1.** The univariate case: Kriging<sup>2</sup>

Following Chilès and Delfiner (1999), geostatistics aims at providing quantitative descriptions of variables distributed in space or in time and space, and provides the practitioners with a methodology to quantify spatial (or spatiotemporal) uncertainty. The term "geostatistics" was first introduced by Matheron (1962). As stated in Chilès and Delfiner (1999), the term "statistics" comes into play because probability distributions are the meaningful way to represent the range of possible values of a quantity of interest. In addition, a statistical model is well suited to the apparent randomness of spatial variations. The prefix "geo" emphasizes the spatial aspect of the problem. Originally, it referred to the domain of the Earth Sciences but, as seen in section 1, over time the number of domains geostatistics has been applied to has increased enormously. Matheron also coined the term "regionalized variable" to designate a numerical function depending on a continuous spatial index and combining high irregularity of detail with spatial correlation. Therefore, geostatistics can also be defined as the application of probabilistic methods to regionalized variables. In other words, geostatistics deals with regionalized phenomena, which are those that extend across space and present an organization or structure. By space it is understood that we are

<sup>&</sup>lt;sup>2</sup> We focus on linear kriging. Nonlinear kriging procedures are also of interest, but their analysis goes beyond the scope of this article.

referring to geographical space; however, in recent research, economic, social, and other more sophisticated types of space are considered (for example, in Fernández-Avilés *et al.*, 2012, and Fernández-Avilés and Montero, 2016, economic and financial spaces have been constructed).

The instrument geostatistics uses to predict (in the sense of interpolation) the value of a random value in an unobserved location is called kriging, in honor of Professor and Mining Engineer, Daniel Gerhardus Krige (who passed away on 2nd March 2013 at the age of 93). The seminal papers by Krige, published in the Journal of Chemical, Metallurgical and Mining Society of South Africa, led to the celebrated research by Georges Matheron and his team in France, on regionalized variables; Matheron named the new method of linear estimation of the regionalized variables using a spatial model "Kriging", in recognition of the distinguished pioneering work of Krige.

Let  $\{Z(\mathbf{s}_1), Z(\mathbf{s}_2), ..., Z(\mathbf{s}_n)\}$  be *n* observations measured at spatial locations  $\{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n\}$ . Let  $\mathbf{s}_0$  be an unobserved location where we want to predict  $Z(\mathbf{s}_0)$ .

The kriging predictor is  $\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)$ , where  $\lambda_i$ , i = 1, ..., n, are the kriging

weights obtained by imposing on the prediction error the classical conditions of unbiasedness and minimum variance. We may often be interested in block prediction; that is, our aim is to predict the average value of the RF being studied in a specific area, a block, with the observation support being points or also blocks. However, the bases of the procedure are the same as for kriging based on point observations.

How does kriging estimate the weights used to obtain optimal predictions at unobserved locations? Roughly speaking, the procedure involves two stages. First, the structure of the spatial dependence or correlation present in the realization observed must be represented in a function. The fact that it accounts for this structure is the main advantage of using kriging over other spatial interpolation techniques (inverse distance method, splines and polynomial regression, among others). Another important advantage is that kriging makes it possible to quantify how accurate the predictions are using the prediction errors. This stage is known in the geostatistics literature as structural analysis and is a key issue in the subsequent process of optimal (kriging) prediction, as the success of the kriging methods is based on the functions yielding information about the spatial dependence detected.

The functions referred to above are covariance functions (also called covariograms) and semivariograms. They must meet a series of requisites (see Montero *et al.*, 2015 for details), otherwise, they may produce nonsensical results. However, in practice, the empirical covariance functions and semivariograms

constructed from the observed realizations do not usually satisfy such requisites. For this reason, one of the theoretical models (also called the valid models) that do comply must be fitted to it (Chilès and Delfiner, 1999, present a long list of such models).

As stated in Webster and Oliver (2001, p. 128), choosing models and fitting them to the empirical covariance function or semivariogram plot, are key and arduous tasks that are among the most controversial topics in geostatistics. The fitting process can be done (i) by eye (fitting by eye, manual fitting, fitting at first glance), using visual and graphical methods, or (ii) using statistical procedures. A combination of both is recommended. Manual fitting may seem, initially, like a strange tradition but, as stated in Wackernagel (2003, p. 49), it does not generally have a particularly relevant impact on how well the semivariogram function fits the sequence of points in the empirical semivariogram. What is really crucial is the type of continuity assumed for the regionalized variable and the stationarity hypothesis associated with the RF. These assumptions have many more implications than the way the theoretical function is fitted to the empirical counterpart (see Matheron, 1989, for a thorough discussion). Armstrong and Wackernagel (1988, pp. 53-4) state that the analytic form of the model does not matter very much as long as the major features of the phenomenon are respected.

Unfortunately, most practitioners use automatic fitting, which rarely provides good results. For example, the fitting near the origin is much more important than the fitting at large distances from it, because the behavior of the semivariogram near the origin indicates the degree of irregularity the RF shows; there may be anisotropies, etc. Finally, an understanding of the phenomenon under study is very welcome for this purpose, though the introduction of auxiliary hypotheses carries with it the risk that kriging loses its optimality.

The second stage of kriging is the solution of the kriging equations, which provides the estimation of the kriging weights and the prediction variance. They result from imposing on the prediction error zero expectation and minimum variance. When the RF of interest is second-order stationary (with a constant but unknown mean) or intrinsically stationary, the resulting kriging equations are called ordinary kriging (OK) equations and are as follows (in terms of the semivariogram):

$$\begin{cases} \sum_{j=1}^{n} \lambda_{j} \gamma \left( \mathbf{s}_{i} - \mathbf{s}_{j} \right) + \alpha = \gamma \left( \mathbf{s}_{i} - \mathbf{s}_{0} \right), & \forall i = 1, \dots, n \\ \\ \sum_{i=1}^{n} \lambda_{i} = 1 \end{cases}, \quad (1)$$

with prediction variance:

*c* 

$$\sigma_{OK}^{2}(\mathbf{s}_{0}) = V(Z^{*}(\mathbf{s}_{0}) - Z(\mathbf{s}_{0})) = \sum_{i=1}^{N} \lambda_{i} \gamma(\mathbf{s}_{i} - \mathbf{s}_{0}) + \alpha \quad (2)$$

In the above equations, the semivariogram terms  $\gamma(\cdot)$  depend only on the distance between the pairs of locations and not on their specific location (and on the orientation in the case of isotropic kriging), and  $\alpha$  is a Lagrange multiplier.

However, it is difficult to argue that the RF under study is stationary or intrinsically stationary across a large area (or even in small ones). In particular, the assumption that the unknown mean of the RF is constant across the domain under study is usually not the case in reality. However, given its "simplicity", OK is often overused. In the presence of a spatial drift evolving with different shapes across the domain under study, which results in a non-stationary situation, alternatives to OK must be used. The traditional alternative is universal kriging (UK), which involves a parametric drift on the spatial coordinates operating across the domain of interest whose order is predetermined, and uses the semivariogram to capture the spatial dependencies in the errors (that is, the difference between the observations and the values of the drift at the observed locations) instead of in the observed values. The UK equations are as follows:

$$\sum_{j=1}^{n(\mathbf{s}_0)} \lambda_j \gamma_e(\mathbf{s}_i - \mathbf{s}_j) + \sum_{h=1}^{p} \alpha_h f_h(\mathbf{s}_i) = \gamma_e(\mathbf{s}_i - \mathbf{s}_0), \quad \forall i = 1, ..., n(\mathbf{s}_0) \\ \sum_{i=1}^{n(\mathbf{s}_0)} \lambda_i f_h(\mathbf{s}_i) = f_h(\mathbf{s}_0), \quad \forall h = 1, ..., p$$
(3)

Note that UK does not take into account all the observations but a number of them,  $n(\mathbf{s}_0)$ , around the prediction point. Note also that  $\sum_{h=1}^{p} \alpha_h f_h(\mathbf{s}_i)$  represents the local drift in the surroundings of  $\mathbf{s}_i$ , where  $\{f_h(\mathbf{s}), h = 1, ..., p\}$  are p known linearly independent parametric functions (monomials of the spatial coordinates; given the polynomial character of the drift,  $f_1(\mathbf{s}_0) = 1$ ),  $\alpha_h$  are constant coefficients obtained with a moving neighborhood which can differ (and in reality does differ) from one neighborhood to another, and p is the number of terms employed in the approximation to the drift. It is important to highlight that the local drift is not explicitly estimated; It is also very important to note that the semivariogram terms,  $\gamma_e$ , are the residual semivariogram terms and not observed semivariogram terms, which constitutes a severe limitation of the procedure. The prediction variance is also given in terms of  $f_h(\mathbf{s}_i)$  and  $\gamma_e$ :

$$V\left(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right) = \sum_{i=1}^{n(s_0)} \lambda_i \gamma_e(\mathbf{s}_i - \mathbf{s}_0) + \sum_{h=1}^p \alpha_h f_h(\mathbf{s}_0).$$
(4)

Since residuals are not observable and consequently their empirical semivariogram cannot be constructed, some "solutions" to the  $\gamma_e$  – problem can be found in the geostatistical literature. The most popular ones are: (i) to assume that  $\gamma_{z} \approx \gamma_{e}$  in a small neighborhood of the prediction site, given that in this case the drift cannot significantly change. (ii) Iterated universal kriging (IUK), which consists of applying UK with the semivariogram of the data (thus assuming that  $\gamma_z \approx \gamma_e$ ) and obtaining the initial weights to estimate the drift and the residuals. Using these residuals as a basis, a valid semivariogram is fitted to its empirical counterpart and then the UK equations are computed and new weights for the drift are obtained. This process is repeated iteratively until convergence. (iii) Kriging with external drift (KED) instead of using monomials on the spatial coordinates, the drift is defined explicitly by using auxiliary covariates that are known at the same points where the observed RF has been observed as well as at the prediction point (exhaustive information; in a heterotopic situation, cokriging is recommended). (iv) Residual kriging (RK), also known as regression kriging or kriging after detrending, which assumes that the shape of the drift is known (based on previous knowledge). In RK, the drift is first externally estimated; second, the residuals, supposedly stationary, are obtained by differences; third, a theoretical semivariogram is fitted to the empirical semivariogram of the residuals; fourth, OK of the residuals is performed, assuming a zero mean for the residuals (the situation where the mean is constant and known is called simple kriging); and finally, the prediction is obtained by adding the estimates of the drift and the residual at the estimation point. (v) Kriging with total drift (KTD), which can be considered an extension of KED whereby the coordinates are included in the set of auxiliary covariates (see Montero and Fernández-Avilés, 2017a). However, the above approaches are also subject to theoretical objections. Other alternatives frequently used in practice are the median polish kriging and iterative RK. The algorithms for the above alternatives can be seen in Montero et al. (2015) and the references therein.

As pointed out in Hengl *et al.* (2003), it is important to emphasize that KU, KR and KDE (and also KTD) describe practically the same generic method and should provide the same predictions and prediction errors if the inputs that feed them are the same. They only differ in the methodological steps. In KED, the semivariogram matrix is extended with auxiliary covariates, so that the universality conditions are integrated into the kriging equations. The challenge is to obtain a satisfactory residual semivariogram in the presence of a drift. In RK, the difficulty is to obtain unbiased regression coefficients in the presence of spatially autocorrelated residuals. However, RK has the advantage that it can be easily combined with stratification, generalized additive models, regression trees,

etc. Another advantage is the separate estimation of the drift and the residual component, whereas in UK and KED, both are estimated simultaneously.

A word of caution: KED (like UK) is more computation intensive than RK, which is why it is common in research (as well as in software packages) to make different assumptions and skip some computational steps that lead to RK. Obviously, these short cuts might be attractive for practical applications, but are sub-optimal statistically speaking.

In short, the presence of a non-constant drift results in problems that have yet to be resolved. It should be highlighted that since the drift usually has a marked local character that is difficult to capture through global parametric specifications, kriging with non-parametrically specified drift (internal, external or total) constitutes an interesting research avenue. This way, the local character of the drift can be captured by a global drift. In my opinion, the non-parametric spatial drifts obtained by using penalized B-splines are a promising approach. Moreover, I recommend the transformation of the drift model into a mixed model, which allows for the simultaneous estimation of the drift parameters and the smoothness coefficient (see Montero *et al.*, 2017b, for details on this approach in the field of spatial econometrics). It is of note that in the non-stationary cases, UK is rarely acceptable since it introduces an a priori parametric specification that is difficult to justify in almost all cases. The non-parametric drifts overcome this limitation. It is also very important to remark that in the prediction of the drift and residual components, the attention must be focused on the former because the drift tends to account for 90-95% of the total prediction value.

Finally, a word on the estimation of the (constant) mean across an area of interest. As outlined in section 2, in the presence of spatial dependencies, the traditional estimator of such a mean should be substituted with the kriging-of-the-mean estimator (see Montero *et al.*, 2015, for details). Otherwise, the variance of the estimator will be underestimated, with adverse inferential consequences.

#### 3.2. The multivariate extension: Cokriging

Cokriging is the name given to the multivariate version of kriging. Cokriging not only uses the information on the RF of interest but also information on other auxiliary RFs strongly correlated with it. Thus, it constitutes a good alternative to univariate kriging. As stated in Montero *et al.* (2009), cokriging is a technique designed to benefit from "borrowing power" in cases where the RF of interest is scarce and/or expensive to collect, but there are other related and more abundant RFs that can provide additional information useful for estimating the spatial RF of direct interest. By way of example, if we are interested in predicting the price of a commercial property in a specific location, we can perform a kriging prediction based on the prices of commercial properties in the surrounding area of the location of interest. Unfortunately, however, information on the price of Cokriging allows the use of additional RFs such as housing prices, office prices, etc. (for which abundant information is at our disposal) which potentially provide valuable information to compensate for the relative scarcity of commercial property transactions. Cokriging thus incorporates the optimal spatial prediction capabilities found in kriging, as well as the ability to exploit the commonalities in spatial structure shared by the main and the auxiliary RFs. Cokriging requires homotopy (the main and the auxiliary variables have been observed at the same locations) or partial heterotopy (there are only a number of sites where we have information on both the main and the auxiliary RFs). In the case of total heterotopy, cokriging cannot be used (see Wackernagel, 2003). Note that there is an important difference with the kriging methods using secondary information (such as KED, for example), because these methods require exhaustive secondary information (meaning that the secondary information is available at all primary data locations and at all locations where the main RF is going to be predicted), while cokriging algorithms allow for non-exhaustive secondary information. Another relevant difference is the approach: whereas cokriging tries to "enlarge" the sample observed via auxiliary RFs correlated with the RF of interest, alternatives like kriging with KED or RK, for example, try to "explain" the behavior of the main RF, or to "clean" it from the impact of factors other than the spatial dependencies in the RF itself.

Let  $\mathbf{Z} = (Z_1, Z_2, ..., Z_m)'$  be a vector of correlated RFs. If we aim to predict the value of  $Z_1$  at a specific location  $\mathbf{s}_0$ , the cokriging predictor is given by a weighted linear combination of the data values  $Z_i$ , j = 1, ..., m, located at sampled points in the neighborhood of  $\mathbf{s}_0$ :

$$\hat{Z}_1(\mathbf{s}_0) = \sum_{j=1}^m \sum_{\alpha=1}^{n_j} \lambda_\alpha^j Z_j(\mathbf{s}_\alpha^j) \,.$$
(5)

The weights  $\lambda_{\alpha}^{j}$ ,  $\alpha = 1, ..., n_{j}$ , j = 1, ..., m, as is the case for kriging, are calculated to ensure that the estimator is optimal, in the sense that it is unbiased and minimizes the error-variance (see details, for example, in Wackernagel, 2003).

The extension of the kriging predictor to the cokriging case is not a problem. The real problem is that the multivariate extension of kriging requires not only direct- but also cross-semivariograms to capture the pattern of spatial dependencies existing within the RFs and between them. In other words, cokriging requires a model of the coregionalization matrix that has an acceptable mathematical fit to the empirical cross-semivariogram matrix. The selection of this model is crucial in multivariate geostatistics; however, it is not an easy task (in fact it is a challenge that goes beyond the scope of this article), and this is one of the reasons (maybe the main reason) why cokriging is not often used (see Goovaerts, 1997, for basic details on the coregionalization model). In general, with *m* RFs, *m* direct semivariograms and  $m \times (m-1)$  cross-semivariograms must be estimated to guarantee positive definiteness.

In addition, if the appropriate model of coregionalization has been obtained (which is practically impossible given that traditional coregionalized models ignore the nonlinear dependencies between RFs, see Leuangthong and Deutsch, 2003), the cokriging equations (whatever the type: simple, ordinary, universal...) are so big that when dealing with large databases the "big n problem" arises and the computational burden becomes prohibitive. By way of example, in the case of ordinary cokriging (a simple case) and only one auxiliary RF, the system of cokriging equations is as follows:

$$\begin{cases} \sum_{\beta=1}^{n_1} \lambda_{\beta}^1 \gamma_{j1} (\mathbf{s}_{\alpha}^j - \mathbf{s}_{\beta}^1) \\ + \sum_{\beta=1}^{n_2} \lambda_{\beta}^2 \gamma_{j2} (\mathbf{s}_{\alpha}^j - \mathbf{s}_{\beta}^2) + \omega_j = \gamma_{j1} (\mathbf{s}_{\alpha}^j - \mathbf{s}_0), \quad \forall j = 1, 2; \quad \forall \alpha = 1, ..., n_j, \quad (6) \\ \sum_{\alpha=1}^{n_j} \lambda_{\alpha}^j = \delta_{1j} = \begin{cases} 1 & if \quad j = 1 \\ 0 & if \quad j = 2 \end{cases} \end{cases}$$

where  $\omega_j$ , j = 1, 2, are Lagrange multipliers corresponding to the constraints needed to guarantee the non-bias condition of the cokriging estimator. Note that, in the above case, where m = 2, when  $n_1 = n_2 = N = 500$ , the number of cokriging equations is  $m^2 (N+1) = 2004$ . In the case of a greater number of observations, the number of resulting equations would increase considerably.

In conclusion, cokriging is tedious due to (i) the simultaneous modeling of a large number of direct- and cross-semivariograms; although the linear model of coregionalization guarantees the positive definiteness of matrices, the restrictions imposed on it usually result in poor semivariogram fitting which deprives cokriging of some of its possible advantages over kriging; and (ii) the "big n problem" when dealing with massive databases. This is the reason why a number of alternatives based on the reduction of the dimensionality of the problem and the kriging of the orthogonal factors have been proposed and frequently used. They include: (i) Finding the principal component which captures the highest quantity of the variance of the auxiliary variables involved in the problem and use it as the only auxiliary variable in cokriging (see Martinez-Martinez *et al.*, 2008); (ii) The transformation of the spatially cross-correlated variables into a set of orthogonal factors which can be separately estimated using univariate

techniques; eventually, the results can be back-transformed into the original data space (see Wackernagel, 2003, for example). This approach includes the Stepwise Conditional Transformation (SCT) and the Min/Max Autocorrelation Factors (MAF) procedure. With SCT, each variable is transformed sequentially and conditionally to previous variables into an independent multivariate Gaussian distribution. Since each transformed variable is independent, they may be simulated separately and then back-transformed to re-impose their multivariate relationships. This approach produces normally distributed uncorrelated scores at zero lag distance (although it suffers from ordering issues associated with the transformation sequence of variables) (see Leuangthong and Deutsch, 2003, for details). MAF allows the creation of a set of factors that are uncorrelated at two lag separation vectors: the zero vector and a vector that can be freely chosen by the user. As in principal component analysis (PCA), each original variable is a linear combination of the factors (Rondon, 2012). And (iii) nonlinear principal component analysis (NLPCA) (see Linting et al., 2007, for details), which has recently been combined with pair-copula models to model the univariate nonlinear spatial dependencies once the existing nonlinear multivariate dependence between spatial variables has been removed using NLPCA (Musafer et al., 2017).

Independent component analysis (ICA), which tries to find mutually independent factors from a linear combination of original variables, is among the blind source separation methods that have been frequently used (see Hyvarinen *et al.*, 2001, for details).

In practice, it is extremely difficult to obtain orthogonal factors at all lag distances (it could even be said that it is practically impossible), and this has resulted in the popularization of methods aimed at obtaining approximately orthogonal factors at several lag distances. This approach includes the fitting of matrix-valued semivariogram models by simultaneous diagonalization (Xie and Myers, 1995), the joint diagonalization of correlation matrices by using Newton methods (Joho and Rahbar, 2002), the uniformly weighted exhaustive diagonalization with Gauss iterations (U-WEDGE) by Mueller and Ferreira (2012) and the Minimum Spatial Cross-correlation (MSC) procedure recently introduced by Sohrabian and Tercan (2014), which can be applied even in approximate orthogonalization of matrices with non-invertible subspaces.

As can be seen, again the difficulties in appropriately capturing the directand cross-spatial dependencies existing in correlated RFs is the main obstacle cokriging has to overcome to produce reliable predictions. Something similar can be said of the alternative approaches outlined above based on the kriging of the orthogonalized factors (MSC-kriging, IC-Kriging, MAF-Kriging, etc.). Consequently, open questions remain in the field of multivariate geostatistics (for example, which has a more adverse effect on prediction: the limitations of cokriging in capturing the structure of the direct- and cross-dependencies or the fact that in the factor approach the factors still exhibit some cross-correlations at several lag distances?) and much work is still needed. Therefore, it is no surprise that multivariate geostatistical prediction is a hot topic in the literature of a large number of scientific disciplines.

## 4. RECENT TIMES: SPATIO-TEMPORAL GEOSTATISTICS

The study of spatio-temporal variability is a relatively new area within statistics. There has been a growing realization in the last 15 years that knowing where and when data were observed could help enormously in answering the substantive questions that precipitated their collection. Cressie (1993), perhaps the leading academic in geostatistics, devoted only three pages to the analysis of spatio-temporal data. Currently, one can find in the literature summaries of the main spatio-temporal modelling techniques and a wide range of practical applications.

The extension of the spatial-only to the spatio-temporal context is a natural extension that, at least from the theoretical perspective, should not pose substantive problems. In the space-time dimension  $\{Z(\mathbf{s},t), \mathbf{s} \in D, t \in T\}$ , with  $D \subset \mathbb{R}^2$  and  $t \subset \mathbb{R}$ , denotes the spatio-temporal RF whose value is observed at a set of *n* spatio-temporal locations  $\{Z(\mathbf{s}_1,t_1),...,Z(\mathbf{s}_n,t_n)\}$ . The objective is to predict an unknown point value  $Z(\mathbf{s}_0,t_0)$  at a non-observed location  $(\mathbf{s}_0,t_0)$ . For this purpose, the spatio-temporal kriging (STK) predictor,  $Z^*(\mathbf{s}_0,t_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i,t_i)$ , is used. As in the spatial-only case, the weights,  $\lambda_i$ , are obtained from the STK

used. As in the spatial-only case, the weights,  $\lambda_i$ , are obtained from the STK equations, which depend on the degree of stationarity attributed to the RF that supposedly generates the observed realization. For example, in the case of second-order or intrinsic stationarity the STK are as follows:

$$\begin{cases} \sum_{i=1}^{n} \lambda_{j} \gamma \left( \mathbf{s}_{i} - \mathbf{s}_{j}, t_{i} - t_{j} \right) + \alpha = \gamma \left( \mathbf{s}_{i} - \mathbf{s}_{0}, t_{i} - t_{0} \right), & \forall i = 1, \dots, n \\ \sum_{i=1}^{n} \lambda_{i} = 1 & , \end{cases}$$

$$(7)$$

with prediction (kriging) variance:

$$V\left(Z^*\left(\mathbf{s}_0, t_0\right) - Z\left(\mathbf{s}_0, t_0\right)\right) = \sum_{i=1}^n \lambda_i \gamma\left(\mathbf{s}_i - \mathbf{s}_0, t_i - t_0\right) + \alpha \quad , \tag{8}$$

where  $\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$  is the semivariogram selected to capture the structure of the spatio-temporal dependencies existing in the phenomenon under study (in

the second-order case the covariogram  $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$  can also be used) and  $\alpha$  is a Lagrange multiplier associated with the non-bias condition.

As can be seen, there does not appear to be any problem in the extension of spatial-only kriging to STK. However, this is not the case in reality. Below, we focus on the three main limitations of this extension.

First, as in the spatial-only case, STK strongly depends on the choice of the autocovariance associated with the space-time random field under study. This is not a problem in spatial-only geostatistics. It could be said that there is a "standard" list of ones that guarantee positive definiteness, covering almost all practical situations. However, in the spatio-temporal context such a list is a challenge, because the construction of non-separable spatio-temporal covariance or semivariogram structures (the space-time interaction is crucial for prediction) associated with stationary or non-stationary, isotropic or anisotropic random fields, which meet the requirements to be a valid covariance or semivariogram model is not an easy task (it is not easy to prove the positive definiteness of a spatio-temporal dependence structure). Even the construction and visualization of an empirical space-time semivariogram can often be considered a challenging task (see example 6.2 in Montero *et al.*, 2015). The original separable, isotropic and stationary covariance models (the metric model by Dimitrakopoulos and Luo, 1994; the sum model introduced by Rouhani and Hall, 1989; the product model by Rodriguez-Iturbe and Mejia, 1974, and De Cesare et al., 1997; the product-sum model popularized by De Iaco et al., 2001, 2002a; and the integrated product and product-sum models also popularized by De Iaco et al., 2002b, among other initiatives) imply a number of restrictive conditions that makes them infeasible options for use in real applications. Therefore, more complex and realistic covariance models were needed to improve predictive performance. Thus, it is no surprise that over the past decade a great effort has been made to overcome separability, stationarity, isotropy and full symmetry in the spatio-temporal structures capturing the spatio-temporal dependencies existing in the phenomena under study. Examples of such efforts include the models proposed by Cressie and Huang (1999), Gneiting (2002a), Ma (2002, 2003b, 2005a, 2005b), Stein (2005) and Porcu and Mateu (2007) to take into account the space-time interaction; the specifications by Ma (2003a) and Gneiting et al. (2007) to overcome full symmetry; the mixture-based Bernstein zonally anisotropic covariance functions by Porcu et al. (2006) and the nonseparable models constructed by using quasi-arithmetic functionals (Porcu et al., 2009), which go beyond isotropy; and the non-stationary models developed by Fuentes and Smith (2001), Chen et al. (2006), Ma (2002, 2003c) and Porcu and Mateu (2007), among others.

Unfortunately, there has been a dramatic decline in research on the topic in the present decade and it is now rare to find articles with titles like "A new family of spatio-temporal covariance functions...", as was common in the past decade. The related research has moved on to the field of cross-covariance functions in the multivariate space-time context (e.g., Ip and Li, 2016, 2017; Bourotte *et al.*, 2016) and to the topic of covariance functions on the sphere (Hitczenko and Stein, 2012; Jeong and Jun, 2015; and Porcu *et al.*, 2016, are some examples), without having fully resolved a number of relevant issues in the univariate field.

Given the long list of "recent" models cited above, one might think that, as in the spatial-only case, there is a "standard" list of permissible models; however, this is currently far from being the case. The new families of spacetime covariance models are not exempt from problems and, in practice, the selection of one of them to represent the space-time dependencies existing in the phenomenon under study continues to be a challenge. Given the complexity of these "new" models, their visualization is recommended, as well as an analysis of how they evolve when the model parameters change; this is core information in the selection of the appropriate model to capture the spatio-temporal dependencies. However, visualization is not an easy topic and constitutes another of the pending questions regarding spatio-temporal covariogram analysis. By way of example, Huang and Sun (2017) state that "it is challenging to visualize and assess separability and full symmetry from spatio-temporal observations".

In conclusion, spatio-temporal interactions usually take very complicated forms. Although recent years have witnessed the development of a long list of spatio-temporal covariance models overcoming separability, isotropy, stationarity and full symmetry, easily-implementable and interpretable analytical procedures are still needed. The ingredients required for the construction of these covariances are readily available (for instance, it is easy to find a list of Bernstein functions, and even easier to find increasing concave functions), but a new effort must be made to provide researchers with new classes of covariance functions that are valid for a wide range of cases of real phenomena, and that are manageable and easily interpretable. The list of "standard" covariance models (as in the spatialonly case) still needs a lot of work and, consequently, constitutes one of the open questions in this area. As highlighted by Wikle (2017) in his book review of Montero et al. (2015), "to date, there are no general spatio-temporal covariance or semivariogram models that can adequately describe the complex dynamics that are exhibited by the multivariate, often nonlinear, spatio-temporal processes that govern most real-world processes".

Second, when dealing with spatio-temporal data, the solution of STK equations can involve considerable computational burden which might lead to failure in the prediction procedure. More specifically, the calculation of the inverse of covariance matrices becomes a crucial problem. This is known as the "big n problem" (Banerjee *et al.*, 2004), which arises very frequently when

dealing with massive spatio-temporal datasets. Therefore, it is no surprise that recent literature emphasizes the use of approximation methods and new methodologies for dealing with massive spatio-temporal data sets. Geostatistics proposes functional kriging (FK) as a promising alternative, but this is not the only one. However, this issue will be addressed in the next section. Following Wikle (2017), the new observation platforms have brought a huge number of observations with complex error structures (and, often, interest in huge numbers of prediction locations) to bear on the problem. Without substantial modification, classical geostatistical methods are unlikely to be suitable for these problems. Overviews of other recent methods that can deal with these issues can be found in Cressie and Wikle (2011), Hodges (2014), Banerjee *et al.* (2015) and Shaddick and Zidek (2015).

Third, in case studies and practical applications, it has been found that STK prediction exhibits a very rapid reversion to the mean in locations near the border of the spatial region and/or close to the end of the time period considered (and, of course, in out-of-sample prediction). This does not come as a surprise, since kriging is an interpolator procedure; however, it is certainly disappointing and some additional research is called for on this topic.

# 5. PRESENT AND FUTURE: FUNCTIONAL GEOSTATISTICS

Functional geostatistics (FG), and more specifically FK, can be considered as the geostatistical alternative to STK aimed at overcoming the "big n problem" when dealing with large spatial or spatio-temporal datasets. This does not mean that it is necessarily the best alternative; only that is an alternative from the geostatistical perspective.

According to Montero and Fernández-Avilés (2017b), the idea behind the functional alternative is to reduce the spatio-temporal problem to a spatial-only one, so that the observations of the phenomenon for the instants of time considered in the analysis at a particular spatial location are substituted with a smooth curve from which said observations are supposedly drawn. This smooth curve is called a functional datum, or functional observation (Levitin *et al.*, 2007). The spatio-temporal point data problem is thus reduced to a spatial-only geostatistical (or kriging) exercise with functional data instead of traditional point data. Since the spatial kriging equations are well-known, the only problem is the construction of functional data and how to deal with them in statistical terms.

In formal terms (Ferraty and Vieu, 2006), a random variable,  $\mathcal{X}$ , is called a functional variable if it takes values in an infinite-dimensional space (or functional space). An observation of that variable,  $\mathcal{X}$ , is called a functional observation. A functional dataset  $\chi_1, \chi_2, \dots, \chi_N$  is the observations of *N* functional variables  $\chi_1, \chi_2, \dots, \chi_N$  identically distributed as  $\mathcal{X}$ .

The construction of the functional data from the observed point data (the first step in FK) is usually carried out using cubic B-splines (Peng *et al.*, 2006; Franco-Villoria and Ignaccolo, 2015). Another alternative pioneered by Eilers and Marx (1996) is penalized splines (Ranalli *et al.*, 2016; Aguilera-Morillo *et al.*, 2017). This approach is recommended because when including a penalization, the subjectivism induced by the number of knots used and where they are located disappears. In addition, when using penalized splines, the representation of the spline as a mixed model is also recommended (Currie and Durbán, 2002; Eilers *et al.*, 2015). This way, there is no need to select the smoothing parameter a priori, via cross-validation or information criteria, because the above transformation allows the smoothing parameter to be estimated (REML estimation is recommended) together with the other model parameters, which is a great advantage. Other less common choices are LOESS (local regression) smoothers or harmonic functions.

The point kriging predictor is adapted to the functional case in order to make predictions of a functional datum at a non-observed spatial location  $\mathbf{s}_0$ . Accordingly, the FK predictor adopts the expression:  $\hat{\chi}_{\mathbf{s}_0} = \sum_{i=1}^N \lambda_i \chi_{\mathbf{s}_i}$ . The weights  $\lambda_i$  are obtained from the adaptation of the spatial-only kriging equations to the functional case. For example, in the case of second-order or intrinsically stationary stochastic RFs, the functional ordinary kriging (FOK) equations are as follows (Giraldo, 2009):

$$\begin{cases} \sum_{j=1}^{N} \lambda_j \int_T \gamma_{s_i, s_j}(t) dt + \alpha = \int_T \gamma_{s_i, s_0}(t) dt, \quad \forall i = 1, \dots, N \\ \sum_{i=1}^{N} \lambda_i = 1 \end{cases}, \quad (9)$$

with  $\int_{T} \gamma_{\mathbf{s}_{i},\mathbf{s}_{j}}(t) dt$  and  $\int_{T} \gamma_{\mathbf{s}_{i},\mathbf{s}_{0}}(t) dt$  denoting the trace-semivariogram function of the process evaluated in  $\mathbf{h} = \|\mathbf{s}_{i} - \mathbf{s}_{j}\|$  (left-hand side of the first *N*-1 FOK equations) and  $\mathbf{h} = \|\mathbf{s}_{i} - \mathbf{s}_{0}\|$  (right-hand side of those equations), respectively (in the functional case, there is a functional semivariogram; see Giraldo, 2009; Menafoglio *et al.*, 2013), and  $\alpha$  is the Lagrange multiplier.

The prediction trace-variance is given by:

c ...

$$\sigma_{FOK}^{2}\left(\mathbf{s}_{0}\right) = \int_{T} Var\left(\hat{\chi}_{\mathbf{s}_{0}}\left(t\right) - \chi_{\mathbf{s}_{0}}\left(t\right)\right) dt = \sum_{i=1}^{N} \lambda_{i} \int_{T} \gamma_{\mathbf{s}_{i},\mathbf{s}_{0}}\left(t\right) dt - \alpha .$$
(10)

Note that two "complex" new terms appear in the FK equations: the tracesemivariogram and prediction trace-variance. Both are adaptations to the functional context of the minimization criterion (minimizing the trace of the mean-squared prediction variance matrix) given by Myers (1983) in multivariate geostatistics. The adaptation consists in replacing the sum with an integral.

For practical purposes, specific values for  $\int_T \gamma_{s_i,s_j}(t) dt$  and  $\int_T \gamma_{s_i,s_0}(t) dt$  are needed. The steps to obtain them are as follows. First, the usual empirical semivariogram is adapted to the functional case as follows:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{i,j \in N(\mathbf{h})} \int_{T} \left( \chi_{\mathbf{s}_{i}}(t) - \chi_{\mathbf{s}_{j}}(t) \right)^{2} dt \,.$$
(11)

Second, the functions  $\chi_{s_i}(t)$  are expanded in terms of some basis functions,  $B_k(t)$ :

$$\chi_{\mathbf{s}_i}(t) = \sum_{k=1}^{K} c_{ik} B_k(t) = \mathbf{c}'_i \mathbf{B}(t), \quad i = 1, 2, \dots, N \quad ,$$
(12)

where *K* is the dimension of the expansion and  $c_{ik}$  are numerical coefficients. When using cubic splines with a common base of B-splines (as recommended in Giraldo, 2009; Romano *et al.*, 2010) the calculation of  $\int_{T} (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt$  can be simplified as follows:

$$\int_{T} \left( \chi_{\mathbf{s}_{i}}(t) - \chi_{\mathbf{s}_{j}}(t) \right)^{2} dt = \int_{T} \left( \mathbf{c}_{i}' \mathbf{B}(t) - \mathbf{c}_{j}' \mathbf{B}(t) \right)^{2} dt = \int_{T} \left( \left( \mathbf{c}_{i}' - \mathbf{c}_{j}' \right) \mathbf{B}(t) \right)^{2} dt$$
$$= \left( \mathbf{c}_{i}' - \mathbf{c}_{j}' \right)^{\prime} \left( \int_{T} \mathbf{B}(t) \mathbf{B}'(t) dt \right) \left( \mathbf{c}_{i}' - \mathbf{c}_{j}' \right)$$
$$= \left( \mathbf{c}_{i}' - \mathbf{c}_{j}' \right)^{\prime} \mathbf{W} \left( \mathbf{c}_{i}' - \mathbf{c}_{j}' \right),$$
(13)

because, in such a case, the Gramian matrix **W** depends only on the knots and is thus constant for each pair of locations. In practice, **W** is computed as follows:

$$\mathbf{W} = \begin{pmatrix} \sum_{\pi=1}^{T} (b_{1}^{i=\pi})^{2} & \sum_{\pi=1}^{T} b_{1}^{i=\pi} b_{2}^{i=\pi} & \sum_{\pi=1}^{T} b_{1}^{i=\pi} b_{3}^{i=\pi} & \cdots & \sum_{\pi=1}^{T} b_{1}^{i=\pi} b_{K}^{i=\pi} \\ \sum_{\pi=1}^{T} b_{2}^{i=\pi} b_{1}^{i=\pi} & \sum_{\pi=1}^{T} (b_{2}^{i=\pi})^{2} & \sum_{\pi=1}^{T} b_{2}^{i=\pi} b_{3}^{i=\pi} & \cdots & \sum_{T=1}^{T} b_{2}^{i=\pi} b_{K}^{i=\pi} \\ \sum_{\pi=1}^{T} b_{3}^{i=\pi} b_{1}^{i=\pi} & \sum_{\pi=1}^{T} b_{3}^{i=\pi} b_{2}^{i=\pi} & \sum_{\pi=1}^{T} (b_{3}^{i=\pi})^{2} & \cdots & \sum_{\pi=1}^{T} b_{3}^{i=\pi} b_{K}^{i=\pi} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{\pi=1}^{T} b_{K}^{i=\pi} b_{1}^{i=\pi} & \sum_{\pi=1}^{T} b_{K}^{i=\pi} b_{2}^{i=\pi} & \sum_{\pi=1}^{T} b_{K}^{i=\pi} b_{3}^{i=\pi} & \cdots & \sum_{\pi=1}^{T} (b_{K}^{i=\pi})^{2} \end{pmatrix},$$
(14)

where  $b_k^{t=\pi}$  represents the value of the *k*-th basis (*k*=1,2,...*K*) at instant of time  $\pi$ 

 $(\pi=1,2,...T)$ . Third, the vectors of coefficients  $\mathbf{c}'_i$ , i=1,2,...,N, are obtained by ordinary least squares (OLS) or weighted least squares (WLS). Fourth, for each pair of functional data, the corresponding point in the semivariogram cloud is computed. Fifth, for each lag distance, **h**, the semivariance is computed, that is, the semi-average of the corresponding points in the semivariogram cloud (Eq. 13). The set of semivariances obtained is the empirical semivariogram. Finally, after these steps, a theoretical (valid) semivariogram must be fitted to the empirical one to be employed in the FK equations (9).

Having outlined the main features of FK, it should be pointed out that, in spite of the limited use of this technique due to the complexity of the procedure, there is a misuse of FOK, which is applied in situations where stationarity does not apply. The functional universal kriging (FUK) equations have been derived in Caballero-Guardo (2011), but FUK shows the same problems as its point counterpart, UK, which strongly limits its use in practice.

A pending question in FK is the issue that functional prediction reflects in reality the inertia of the RF at the prediction point, but part of the variability is lost. In fact, the extent of such a loss depends on the decisions involved in the smoothing process (this is why the representation in form of a mixed model is strongly recommended). Another pending matter is a more user-friendly explanation of both the trace-semivariogram and the trace-prediction variance. It is often the case that both are computed and used but it is clear that the researcher does not understand what he or she is doing. Finally, the show must go on in the field of FK prediction and an effort should be made to consolidate the interesting avenues of research proposed in Giraldo (2009) and Fernández-Avilés (2010) on the use of functional weights in both FK and functional cokriging equations.

Lastly, as outlined in section 4, FG, and more specifically FK, is not the only alternative to STK to overcome the "big n problem" when dealing with large spatial or spatio-temporal datasets. A review of these alternatives can be seen in chapter 12 of Banerjee et al. (2015), but, as mentioned in the previous section, Cressie and Wikle (2011), Hodges (2014) and Shaddick and Zidek (2015) are also recommended references including overviews of modern methods that deal with this issue (as well as with other relevant issues in the field of spatiotemporal data: random effects parameterizations, dynamic spatiotemporal models and hierarchical Bayesian inference). In broad terms, these alternative approaches can be classified as those that seek approximations to the exact likelihood, and those based on models which can handle fitting with large values of n. The approximate likelihood approaches include (i) spectral methods, lattice and conditional independence methods, Markov field approximations on regular grids (more specifically the integrated nested Laplace approximation), pseudolikelihood approximations, the variational Bayes algorithm and covariance tapering (the underlying idea is the use of compactly supporting covariance

functions; see Gneiting, 2002b; Mateu *et al.*, 2013). The option of devising models for large spatial and spatio-temporal datasets includes low rank models, the essential idea of which is to replace the spatial or spatio-temporal process with a dimension-reducing process; and predictive process models, which is a class of models based on the idea that every spatial or spatio-temporal process induces a predictive process model (in fact induces many of them) which projects realizations of the "parent" process on to a lower-dimensional subspace, thereby reducing the computational burden.

# **BIBLIOGRAPHY REFERENCES**

- AGUILERA-MORILLO, M.C.; DURBÁN, M. and AGUILERA, A.M. (2017). "Prediction of functional data with spatial dependence: a penalized approach". *Stochastic Environmental Research and Risk Assessment*, 31, pp. 7-22.
- ARMSTRONG, M. and WACKERNAGEL, H. (1988). "The influence of the covariance function on the kriged estimator". Sciences de la Terre, Scientifique, Informatique, Géologique, 27(II), pp. 245-62.
- BANERJEE, S.; CARLIN, B.P. and GELFAND, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* Boca Raton, FL.: Chapman & Hall/CRC Press.
- BANERJEE, S.; CARLIN, B.P. and GELFAND, A.E. (2015). *Hierarchical Modeling and Analysis for Spatial Data* (second edition). Boca Raton, FL: Chapman & Hall/CRC Press.
- BOUROTTE, M.; ALLARD D. and PORCU E. (2016). "A flexible class of non-separable cross-covariance functions for multivariate space-time data". *Spatial Statistics*, 18, pp. 125-146.
- CABALLERO GUARDO, W.J. (2011). *Kriging Universal para Datos Funcionales*. Cartagena de Indias: Universidad Nacional de Colombia.
- CHEN, L.; FUENTES, M. and DAVIS, J.M. (2006). "Spatial temporal statistical modelling and prediction of environmental processes". J.S Clark and A.E. Gelfand (Eds.): *Hierarchical Modelling for the Environmental Sciences (*pp. 121–144). Oxford: Oxford University Press.
- CHILÈS, J.P. and DELFINER, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Chichester: John Wiley & Sons.
- CRESSIE, N. (1993). Statistics for Spatial Data. New York: John Wiley & Sons.
- CRESSIE, N. (2015). Statistics for Spatial Data (revised edition). Chichester: John Wiley & Sons.
- CRESSIE, N. and HUANG, C. (1999). "Classes of nonseparable, spatiotemporal stationary covariance functions". *Journal of the American Statistical Association*, 94, pp. 1330-40.
- CRESSIE, N. and WIKLE, C.K. (2011). *Statistics for Spatio-Temporal Data*. Chichester: John Wiley & Sons.
- CURRIE, I.D. and DURBÁN, M. (2002). "Flexible smoothing with P-splines: a unified approach". *Statistical Modelling*, 2, pp. 333-349.

- DE CESARE, L.; MYERS, D. and POSA, D. (1997). "Spatial-temporal modeling of SO2 in Milan district". In E.Y. Baafi and N.A. Schofield (eds): *Geostatistics Wollongong'96*, vol. 2 (pp. 1031-1042). Dordrecht: Kluwer Academic Publishers.
- DE IACO, S.; MYERS, D.E. and POSA, D. (2001). "Space-time analysis using a general product-sum model". *Statistics & Probability Letters*, 52, pp. 21-28.
- DE IACO, S.; MYERS, D.E. and POSA, D. (2002a). "Space-time variograms and a functional form for total air pollution measurements". *Computational Statistics & Data Analysis*, 41, pp. 311-28.
- DE IACO, S.; MYERS, D.E. and POSA, D. (2002b). "Nonseparable space-time covariance models: some parametric families". *Mathematical Geology*, 34, pp. 23-42.
- DIMITRAKOPOULOS, R. and LUO, X. (1994). "Spatiotemporal modeling: covariances and kriging system". In R. Dimitrakopoulos (ed.): *Geostatistics for the Next Century* (pp. 88-93). Dordrecht: Kluwer Academic Publisher.
- EILERS, P.H.C. and MARX, B.D. (1996). "Flexible Smoothing with 6-splines and Penalties". *Statistical Science*, 11(2), pp. 89-121.
- EILERS, P.H.C., MARX, B.D., DURBÁN M. (2015). "Twenty years of P-splines". SORT, 39(2), pp. 149-186.
- FERNANDEZ-AVILÉS, G. (2010). Spatio-temporal modeling of environmental processes derived from the economic activity, Doctoral Disertation. Toledo: University of Castilla-La Mancha.
- FERNANDEZ-AVILÉS, G. and MONTERO, J.M. (2016). "Spatio-temporal modeling of financial maps from a joint multidimensional scaling-geostatistical perspective". *Expert Systems with Applications*, 60(30), pp. 280-293.
- FERNANDEZ-AVILÉS, G.; MONTERO, J.M. and ORLOV, A. (2012). "Spatial Modeling of Stock Market Comovements". *Finance Research Letters*, 9(4), pp. 202-212.
- FERRATY, F. and VIEU, P. (2006). Non parametric functional data analysis. Theory and practice. New York: Springer.
- FRANCO-VILLORIA, M. and IGNACCOLO, R. (2015). Bootstrap based uncertainty bands for prediction in functional kriging. http://arxiv.org/abs/1505.06966
- FUENTES, M. and SMITH, R. (2001). "A new class of nonstationary spatial models". *Research Report.* Statistics Department, North Carolina State University.
- GIRALDO, R. (2009). *Geostatistical Analysis of Functional Data*. Ph.D. Thesis, Barcelona: Universidad Politécnica de Cataluña.
- GNEITING, T. (2002a). "Stationary covariance functions for space-time data". *Journal of the American Statistical Association*, 97,pp. 590-600.
- GNEITING, T. (2002b). "Compactly supported correlation functions". *Journal of Multivariate Analysis*, 83, pp. 493-508.
- GNEITING, T.; GENTON, M.G. and GUTTORP, P. (2007). "Geostatistical space-time models, stationarity, separability and full symmetry". In B. Finkenstadt, L. Held and V. Isham (eds): *Statistical Methods for Spatio-Temporal Systems* (pp. 151-175). Boca Raton, FL: Chapman and Hall/CRC.
- GOOVAERTS, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.
- HALLEY, E. (1686). "An Historical Account of the Trade Winds, and Monsoons, Observable in the Seas between and Near the Tropicks, with an Attempt to Assign the

Phisical Cause of the Said Winds". *Philosophical Transactions* 1686-1692, 16, pp. 153-168.

- HENGL, T.; HEUVELINK, G.B.M. and STEIN, A. (2003). Comparison of kriging with external drift and regression-kriging, Technical note, ITC. Available at https://www.itc.nl/library/ .../hengl\_comparison.pdf
- HITCZENKO, M. and STEIN, M.L. (2012). "Some Theory for Anisotropic Processes on the Sphere". *Statistical Methodology*, 9, pp. 211-227.
- HODGES, J.S. (2014). Richly Parameterized Linear Models, Additive, Time Series, and Spatial Models Using Random Effects, Boca Raton, FL: CRC Press.
- HUANG, H. and SUN, Y. (2017). "Visualization and Assessment of Spatio-temporal Covariance Properties". arXiv preprint arXiv:1705.01789.
- HYVARINEN, A.; KARHUNEN, J. and OJA, E. (2001). *Independent component analysis*. New York: John Wiley & Sons.
- IP, R.H.L. and LI, W.K. (2016). "Matérn cross-covariance functions for bivariate spatiotemporal random fields" Spatial Statistics, 17, pp. 22-37.
- IP, R.H.L. and LI, W.K. (2017). "On some Matérn covariance functions for spatio-temporal random fields". *Statistica Sinica*, 27, pp. 805-822.
- JEONG, J. and JUN, M. (2015). "A Class of Matérn-like Covariance Functions for Smooth Processes on a Sphere" *Spatial Statistics*, 11, pp. 1-18.
- JOHO, M. and RAHBAR, K. (2002). Joint diagonalization of correlation matrices by using Newton methods with application to blind signal separation. Sensor Array and Multichannel Signal Processing Workshop Proceedings. doi:10.1109/sam.2002. 1191070
- LEUANGTHONG, O. and DEUTSCH, C.V. (2003). "Stepwise Conditional Transformation for Simulation of Multiple Variables". *Mathematical Geology*, 35(2), pp. 155-173.
- LEVITIN, D.J., NUZZO; R.L., BRADLEY, W.V. and RAMSAY J.O. (2007). "Introduction to Functional Data Analysis". Canadian Psychology, 48(3), pp. 135-155.
- LINTING, M.; MEULMAN, J.J.; GROENEN, P.J.F. and VAN DER KOOJJ, A.J. (2007). "Nonlinear Principal Components Analysis: Introduction and Application". *Psychological Methods* 12(3), pp. 336-58.
- MA, C. (2002). "Spatio-temporal covariance functions generated by mixtures". *Mathematical Geology*, 34, pp. 965-75.
- MA, C. (2003a). "Families of spatio-temporal stationary covariance models". Journal of Statistical Planning and Inference, 116, pp. 489-501.
- MA, C. (2003b). "Spatio-temporal stationary covariance models". Journal of Multivariate Analysis, 86, pp. 97-107.
- MA, C. (2003c). "Nonstationary covariance functions that model space-time interactions". Statistics and Probability Letters, 61, pp. 411-419.
- MA, C. (2005a). "Linear combinations of space-time covariance functions and variograms". IEEE Transactions on Signal Processing, 53, pp. 857-64.
- MA, C. (2005b). "Semiparametric spatio-temporal covariance models with the ARMA temporal margin". Annals of the Institute of Mathematical Statistics, 57, pp. 221-233.
- MARTÍNEZ-MARTÍNEZ, J.; BENAVENTE, D.; ORDÓÑEZ, S. and GARCÍA-DEL-CURA, M.Á. (2008). "Multivariate statistical techniques for evaluating the effects of brecciated rock fabric on ultrasonic wave propagation". *International Journal of Rock Mechanics* and Mining Sciences, 45(4),pp. 609-620.

MATHERON, G. (1962). Traité de Géostatistique appliquée. vol. I., Paris: Éditions Technip.

- MATHERON, G. (1989). Estimating and choosing. Berlin: Springer Verlag.
- MATEU, J.; FERNÁNDEZ-AVILÉS, G. and MONTERO, J.M. (2013). "On a class of nonstationary, compactly supported spatial covariance functions". Stochastic. Environmental Research and Risk Assessment, 27(2),pp. 297-309.
- MENAFOGLIO, A.; SECHI, P. and DALLA ROSA, M. (2013). "A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space". *Electronic. Journal of Statistics*, 7, pp. 2209-2240.
- MONTERO, J.M.; LARRAZ, B. and PÁEZ, A. (2009). "Estimating Commercial Property Prices: An Application of Cokriging with Housing Prices as Ancillary Information". *Journal of Geographical Systems*, 11, pp. 407-425.
- MONTERO, J.-M. and FERNÁNDEZ-AVILÉS, G. (2017a). "La importancia de los efectos espaciales en la predicción del precio de la vivienda. Una aplicación geoestadística en España". *Papeles de Economía Española*, 152, pp. 104-124.
- MONTERO, J.-M. and FERNÁNDEZ-AVILÉS, G. (2017b)."Functional kriging prediction of atmospheric particulate matter concentrations in Madrid, Spain: Is the new monitoring system masking potential public health problems?". *Journal of the Cleaner Production* (in press).
- MONTERO, J.-M.; FERNÁNDEZ-AVILÉS, G. and MATEU, J. (2015). Spatial and Spatio-Temporal Geostatistical Modeling and Kriging. Chichester: John Wiley & Sons.
- MONTERO, J.M.; MINGUEZ, R. and FERNÁNDEZ-AVILÉS, G. (2017). "Housing price prediction. Parametric vs. non-parametric spatial hedonic models" in *Journal of Geographical Systems* (in press). Available at https://link.springer.com/article/10. 1007/s10109-017-0257-y
- MUELLER, U.A. and FERREIRA, J. (2012). "The U-WEDGE Transformation Method for Multivariate Geostatistical Simulation". *Mathematical Geosciences*, 44(4), pp. 427-448.
- MUSAFER, G.N.; THOMPSON, M.H.; WOLFF, R.C. and KOZAN, E. (2017). "Nonlinear multivariate spatial modeling using NLPCA and pair-copulas". *Geographical Analysis* (in press). doi:10.1111/gean.12126.
- MYERS, D.E. (1983). "Estimation of linear combinations and co-kriging". *Journal of the International Association of Mathematical Geology*, 15, pp. 633-637.
- PENG, R.D.; DOMINICI, F. and LOUIS, T.A. (2006). "Model choice in time series studies of air pollution and mortality". *Journal of the Royal Statistics Society* A, 169, Part 2: pp. 179-203.
- PORCU, E. and MATEU, J. (2007). "Mixture-based modeling for space-time data". *Environmetrics*, 18, pp. 285-302.
- PORCU, E.; BEVILAQCUA, M. and GENTON, M.G. (2016). "Spatio-Temporal Covariance and Cross-Covariance Functions of the Great Circle Distance on a Sphere". *Journal of the American Statistical Association*, 111(514), pp. 888-898.
- PORCU, E.; GREGORI, P. and MATEU, J. (2006). "Nonseparable stationary anisotropic space-time covariance functions". *Stochastic Environmental Research and Risk Assessment*, 21, pp. 113-122.
- PORCU, E.; MATEU, J. and CHRISTAKOS, G. (2009). "Quasi-arithmetic means of covariance functions with potential applications to space-time data". *Journal of Multivariate Analysis*, 100(8), pp. 1830-44.
- RANALLI, M.G.; ROCCO, G.; LASINIO, G.J.; MORONI, B.; CASTELLINI, S.; CROCCHIANTI, S. and CAPPELLETTI, D. (2016). "Functional exploratory data

analysis for high-resolution measurements of urban particulate matter". *Biometrical Journal*, 58(5), pp. 1229-1247.

- RODRÍGUEZ-ITURBE, I. and MEJIA, J.M. (1974). "The design of rainfall networks in time and space". *Water Resources Research*, 10, pp. 713-28.
- ROMANO, E.; BALZANELLA, A. and VERDE, R. (2010). "A new regionalization method for spatially dependent functional data based on local variogram models: an application on environmental data" Atti delle XLV Riunione Scientifica della Societá Italiana di Statistica, Universitá degli Studi di Padova, Padova, 16-18 giugno 2010. CLEUP, ISBN: 978 88 6129 566 7.
- RONDON, O. (2012). "Teaching Aid: Minimum/Maximum Autocorrelation Factors for Joint Simulation of Attributes". *Mathematical Geosciences*, 44(4), pp. 469-504.
- ROUHANI, S. and HALL, T.J. (1989). "Space-time kriging of groundwater data". In M. Armstrong (Ed.): *Geostatistics* (pp. 639-51). Dordrecht: Kluwer Academic Publishers.
- SHABENBERGER, O. and GOTWAY, C.A. (2005). *Statistical Methods for Spatial Data Analysis.* Boca Raton, FL: CRC Press.
- SHADDICK, G. and ZIDEK, J.V. (2015). Spatio-Temporal Methods in Environmental Epidemiology. Boca Raton, FL: CRC Press.
- SOHRABIAN, B.B. and TERCAN, A.E. (2014). "Introducing minimum spatial crosscorrelation kriging as a new estimation method of heavy metal contents in soils". *Geoderma*, 226/227, pp. 317-331.
- STEIN, M.L. (2005). "Space-time covariance functions". *Journal of the American Statistical Association*, 100, pp. 310-21.
- TOBLER, W.R. (1970). "A computer movie simulation of urban growth in the Detroit region". *Economic Geography*, 46(2), pp. 234-240.
- WACKERNAGEL, H. (2003). Multivariate Geostatistics. Berlin: Springer-Verlag.
- WEBSTER, R. and OLIVER, M.A. (2001). *Geostatistics for Environmental Scientists*. Chichester: John Wiley & Sons.
- WIKLE, C.K. (2017). "Book review: José-María Montero, Gema Fernández-Avilés, and Jorge Mateu: Spatial and Spatio-Temporal Geostatistical Modeling and Kriging John Wiley & Sons, 2015" *Journal of Agricultural, Biological and Environmental Statistics*, 22(3), pp. 420-423.
- XIE, T.; MYERS, D.E. and LONG, A.E. (1995). "Fitting matrix-valued variogram models by simultaneous diagonalization, Part II: application". *Mathematical Geology*, 27(7), pp. 877-888.