ENVIRONMETRICS
*Environmetrics* 2010; **21**: 224–239
Published online 3 July 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.1003

# Statistics for spatial functional data: some recent contributions

P. Delicado[1*,†], R. Giraldo[1,2], C. Comas[1] and J. Mateu[3]

[1]*Universitat Politècnica de Catalunya, Barcelona, Spain*
[2]*Universidad Nacional de Colombia, Bogotá, Colombia*
[3]*Universitat Jaume I, Castellón, Spain*

SUMMARY

Functional data analysis (FDA) is a relatively new branch in statistics. Experiments where a complete function is observed for each individual give rise to functional data. In this work we focus on the case of functional data presenting spatial dependence. The three classic types of spatial data structures (geostatistical data, point patterns, and areal data) can be combined with functional data as it is shown in the examples of each situation provided here. We also review some contributions in the literature on spatial functional data.


KEY WORDS: areal data; functional data analysis; geostatistical data; point processes; smoothing

## 1. INTRODUCTION

Observing and saving complete functions as a result of random experiments is nowadays possible by the development of real-time measurement instruments and data storage resources. For instance, continuous-time clinical monitoring is a common practice today. Ramsay and Silverman (2005) express this stating that random functions are in these cases the *statistical atoms*. Functional data analysis (FDA) deals with the statistical description and modeling of samples of random functions. Functional versions for a wide range of statistical tools (ranging from exploratory and descriptive data analysis to linear models and multivariate techniques) have been recently developed. See Ramsay and Silverman (2005) for a general perspective on FDA. Ferraty and Vieu (2006) propose non-parametric metohds for FDA. Special journal issues recently dedicated to this topic (Davidian *et al*., 2004; González-Manteiga and Vieu, 2007; Valderrama, 2007) reveal the interest for this topic in the statistical community.

In this paper we focus on the case of functional data presenting spatial dependence. Given that FDA is a relatively new topic in statistics, the literature on spatial statistics for functional data is not extensive at this moment. Ramsay (2008) includes this topic in the list of the eight problems that represent the most exciting opportunities for research in FDA.

*Correspondence to: P. Delicado, Universitat Politècnica de Catalunya, Barcelona, Spain.
†E-mail: pedro.delicado@upc.edu

*Received 9 March 2009*
*Accepted 6 April 2009*

Following Cressie (1993) for the definition of spatial processes and Ferraty and Vieu (2006) for that of functional random variables, we can define a *spatial functional process* as

$$\left\{\boldsymbol{\chi}_s : s \in D \subseteq \mathbb{R}^d\right\} \tag{1}$$

where $s$ is a generic data location in the $d$-dimensional Euclidean space ($d$ is usually equal to 2), the set $D \subseteq \mathbb{R}^d$ can be fixed or random, and $\boldsymbol{\chi}_s$ are *functional random variables*, defined as random elements taking values in an infinite dimensional space (or *functional space*). Typically $\boldsymbol{\chi}_s$, for every fixed $s$, is a real function from $[a, b] \subseteq \mathbb{R}$ to $\mathbb{R}$.

As it happens in univariate or multivariate spatial data analysis, the nature of the set $D$ allows to classify spatial functional data. *Geostatistical functional data* appear when $D$ is a fixed subset of $\mathbb{R}^d$ with positive volume and $n$ points $s_1, \ldots, s_n$ in $D$ are chosen to observe the random functions $\boldsymbol{\chi}_{s_i}$, $i = 1, \ldots, n$. We say that we have a *functional marked point pattern*, when a complete function is observed at each point generated by a standard point process. *Functional areal data* (or *functional data in lattice*) correspond to the case of $D$ being a fixed and countable set. Usually there is a bijection between $D$ and a partition of a geographical area and, for any $s \in D$, $\boldsymbol{\chi}_s$ is a summary function of an event happened at the part of the area corresponding to $s$ by this bijection.

The rest of the paper is organized as follows. Sections 2, 3, and 4 are devoted to each type of spatial functional data, respectively. To fix ideas, each section starts with a real example of the corresponding spatial functional data. Then we review the existing literature and we also present a summary of the research we are actually running on each kind of data. The paper ends with some concluding remarks.

## 2. GEOSTATISTICAL FUNCTIONAL DATA

Ramsay and Silverman (2005) introduce the Canadian temperature data set as one of their main examples of functional data. For 35 weather stations, the daily temperature was averaged over a period of 30 years. The resulting functions, $\chi_{s_i}$, $i = 1, \ldots, 35$ are shown in the right panel of Figure 1. In Giraldo *et al.* (2009a) the coordinates of these stations were added to the original functional data set. The location of the stations are shown in the left panel of Figure 1.

The goal in this example is the prediction of the daily temperature function $\boldsymbol{\chi}_{s_0}$, the value of the functional random process at $s_0$, where $s_0$ is an unsampled location in Canada. Note that the objective is to predict a complete function $\boldsymbol{\chi}_{s_0} : [0, 365) \to \mathbb{R}$, and not a particular value of a one-dimensional variable. In this sense, functional geostatistics is close to multivariate spatial prediction (Ver Hoef and Cressie, 1993).

We assume that the functional random process (1) is second-order stationary and isotropic, that is, the mean and variance functions are constant and the covariance depends only on the distance between sampling points (however, the methodology could also be developed without assuming these conditions). In addition, consider that for every fixed $t_0 \in [a, b]$, the finite-dimensional section $\boldsymbol{\chi}_s(t_0)$ is a scalar random function defined on some probability space. Formally, we assume that

- $\mathbb{E}(\boldsymbol{\chi}_s(t)) = m(t)$ and $\mathbb{V}(\boldsymbol{\chi}_s(t)) = \sigma^2(t)$ for all $t \in [a, b]$ and all $s \in D$.
- $\mathbb{C}\text{ov}(\boldsymbol{\chi}_s(t), \boldsymbol{\chi}_{s+h}(u)) = C(h; t, u)$, with $h$ the spatial lag vector, and for all $t, u \in [a, b]$. This ensures that the variance of the associated process $C(0; t, u)$ exists and is finite. We use the notation $C(h; t)$ for $C(h; t, t)$.



DOI: 10.1002/env

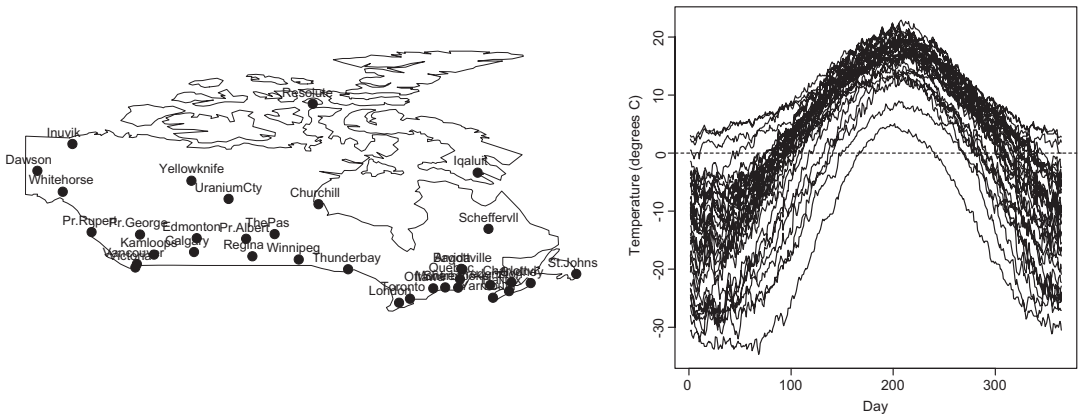


Figure 1. Averages (over 30 years) of daily temperature data (right) observed at 35 Canadian weather stations (left)

These assumptions imply that there exists a mapping $h \mapsto \gamma(h; t, u)$ such that

$$\mathbb{V}\mathrm{ar}(\boldsymbol{\chi}_{s+h}(t) - \boldsymbol{\chi}_s(u)) = \mathbb{E}\big((\boldsymbol{\chi}_{s+h}(t) - \boldsymbol{\chi}_s(u))^2\big) = \gamma(h; t, u) = C(0; t, u) - C(h; t, u)$$

for all $t, u \in [a, b]$. We use the notation $\gamma(h; t)$ for $\gamma(h; t, t)$. This function is called variogram.

The first attempt (to the best of our knowledge) of applying geostatistical interpolation methods to predict functions at unvisited sites was done in the pioneering work of Goulard and Voltz (1993). They consider that curves are only known up to finite set of $M$ values ($\chi_{s_i}(t_j)$, $j = 1, \ldots, M$, $i = 1, \ldots, n$), and a parametric model (assumed to be known) is fitted to them for reconstructing the whole curve. Both, the number of known points for each function and the number of parameters in the parametric model are implicitly assumed to be small. For instance, in the case study analyzed in Goulard and Voltz (1993), each function was measured at $M = 8$ points (a much smaller value than usual values of $M$ in real applications) and the fitted parametric model had four parameters.

Goulard and Voltz (1993) propose three approaches for predicting curves at unvisited sites: two of them are based on a multivariate approach using cokriging, and the other one performs a functional kriging step predicting directly the curves. Their proposals are as follows:

- *Multivariate approach 1: Cokrige first, Fit later Predictor* (CFP). The vector of observed values $(\chi_{s_i}(t_1), \ldots, \chi_{s_i}(t_M))$ is considered as the observation of a $M$-dimensional random variable at site $s_i$. Cokriging is then applied to predict the values of this random vector at the unvisited site $s_0$: $(\hat{\chi}_{s_0}(t_1), \ldots, \hat{\chi}_{s_0}(t_M))$. Then a parametric model $\chi(\cdot; \theta)$, $\theta \in \mathbb{R}^p$ is fitted to the values $(\hat{\chi}_{s_0}(t_1), \ldots, \hat{\chi}_{s_0}(t_M))$ for reconstructing a whole function at $s_0$: $\chi(\cdot; \hat{\theta}_{s_0})$.
- *Multivariate approach 2: Fit first, Cokrige later Predictor* (FCP). First, the parametric model is fitted to the observed curves: $\chi(\cdot; \hat{\theta}_{s_i})$, $i = 1, \ldots, n$. The $p$-dimensional parameter values $\hat{\theta}_{s_1}, \ldots, \hat{\theta}_{s_n}$ are considered as observations of a multivariate random field. Then cokriging is applied to predict the value of the parameter $\theta$ at site $s_0$, say $\hat{\theta}^*_{s_0}$, and $\chi(\cdot; \hat{\theta}^*_{s_0})$ is the resulting predicted function at $s_0$.



DOI: 10.1002/env

- *A Curve Kriging Predictor* (CKP). Goulard and Voltz (1993) define the best linear unbiased predictor for $\boldsymbol{\chi}_{s_0}$ given by

$$\hat{\boldsymbol{\chi}}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \boldsymbol{\chi}_{s_i}(t), \qquad t \in [a, b],\ \lambda_1, \ldots, \lambda_n \in \mathbb{R} \tag{2}$$

where the coefficients $\lambda_i$ are such that $\mathbb{E}(\hat{\boldsymbol{\chi}}_{s_0} - \boldsymbol{\chi}_{s_0}) = 0$ and

$$\mathbb{E}\left[\int_a^b (\hat{\boldsymbol{\chi}}_{s_0}(t) - \boldsymbol{\chi}_{s_0}(t))^2 \,\mathrm{d}t\right] = \int_a^b \mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\chi}}_{s_0}(t) - \boldsymbol{\chi}_{s_0}(t)\right) \,\mathrm{d}t$$

is minimized. Thus the optimization problem to be solved is

$$\min_{\lambda_1,\ldots,\lambda_n} \int_a^b \mathbb{V}\mathrm{ar}(\hat{\boldsymbol{\chi}}_{s_0}(t) - \boldsymbol{\chi}_{s_0}(t)) \,\mathrm{d}t,\ \text{s.t.} \sum_{i=1}^{n} \lambda_i = 1$$

where $\sum_{i=1}^{n} \lambda_i = 1$ is the unbiasedness constraint.
Given that the functions $\boldsymbol{\chi}_{s_i}$ are known only for $M$ values, Goulard and Voltz (1993) propose fitting a parametric model $\chi(\cdot;\theta)$, $\theta \in \mathbb{R}^p$, to these data to obtain $\chi(\cdot;\hat{\theta}_{s_i})$ as an approximation of the whole function $\chi_{s_i}$. Then Equation (2) can be rewritten as

$$\hat{\boldsymbol{\chi}}_{s_0} = \sum_{i=1}^{n} \lambda_i \chi(\cdot;\hat{\theta}_{s_i})$$

and the integrals on $[a, b]$ involved in estimating the coefficients $\lambda_i$ are calculated using $\chi(\cdot;\hat{\theta}_{s_i})$ instead of $\chi_{s_i}$.

Giraldo *et al*. (2008) recover CKP, the third proposal of Goulard and Voltz (1993), overcoming the restrictive assumptions on parametric modeling and small number of observed points per function. Giraldo *et al*. (2008) use the term *ordinary kriging for function-valued data* for this predictor. In particular, they apply a non-parametric fitting pre-processing to the observed functions where the smoothing parameter is chosen by *functional cross-validation*: each functional data location is removed from the data set and a function is predicted at this location using a functional kriging predictor based on the remaining smoothed functions; then the sum of square errors is computed as

$$\mathrm{SSE}_{\mathrm{FCV}} = \sum_{i=1}^{n} \mathrm{SSE}_{\mathrm{FCV}}(i) = \sum_{i=1}^{n} \sum_{j=1}^{M} \left(\hat{\boldsymbol{\chi}}_{s_i}^{(i)}(t_j) - \boldsymbol{\chi}_{s_i}(t_j)\right)^2 \tag{3}$$

where $\hat{\boldsymbol{\chi}}_{s_i}^{(i)}(t_j)$ is the functional kriging prediction on $s_i$ evaluated at $t_j$, $j = 1, \ldots, M$, by leaving the site $s_i$ temporarily out of the sample. The involved smoothing parameters are chosen by minimization of $\mathrm{SSE}_{\mathrm{FCV}}$.

One of the advantages (already noted by Goulard and Voltz, 1993) of the predictor based on Equation (2), is that only the *trace-variogram function* $\gamma(h) = \int_a^b \gamma(h;t,t)\,\mathrm{d}t$, $h \geq 0$, is needed to obtain the optimal coefficients $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$. So the computational cost of the procedure is similar to the case of kriging for one-dimensional data.



DOI: 10.1002/env

In a series of consecutive papers Giraldo *et al.* have proposed other forms of defining functional kriging. A second alternative is to define the predictor of $\boldsymbol{\chi}_{s_0}$ as

$$\hat{\boldsymbol{\chi}}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i(t)\boldsymbol{\chi}_{s_i}(t), \qquad t \in [a,b],\ \lambda_i : [a,b] \mapsto \mathbb{R},\ i = 1,\ldots,n \tag{4}$$

Therefore, in order to find the best linear unbiased predictor, the $n$ functional parameters $\lambda_i(\cdot)$ are given by the solution of the following optimization problem

$$\min_{\lambda_1(\cdot),\ldots,\lambda_n(\cdot)} \int_a^b \mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\chi}}_{s_0}(t) - \boldsymbol{\chi}_{s_0}(t)\right) \mathrm{d}t,\ \ s.t.\ \sum_{i=1}^{n} \lambda_i(t) = 1, \quad \text{for all } t \in [a,b]$$

This version of functional kriging was only mentioned in Goulard and Voltz (1993). Giraldo *et al.* (2009a) developed this approach, called *point-wise functional kriging*. Here the coefficients $\lambda_i(\cdot)$ are functions that have to be determined to have the best linear unbiased predictor.

The optimization problem is solved by fitting non-parametrically the observed functions $\chi_{s_i}(t)$, as well as the parameter functions $\lambda_i(\cdot)$. It is assumed that these functions can be expressed in terms of $K$ basis functions, $B_1(t), \ldots, B_K(t)$, as

$$\boldsymbol{\chi}_{s_i}(t) = \sum_{l=1}^{K} a_{il} B_l(t) = \boldsymbol{a}_i^T \boldsymbol{B}(t), \quad \lambda_i(t) = \sum_{l=1}^{K} b_{il} B_l(t) = \boldsymbol{b}_i^T \boldsymbol{B}(t),\ i = 1,\ldots,n$$

In practice, these expressions are truncated versions of Fourier series (for periodic functions, as it is the case for Canadian temperatures), Karhunen-Loève representations or B-splines expansions. The use of finite expansions allows to reduce the infinite dimensional problem to a multivariate geostatistics problem. Thus this procedure has the same structure as the FPC proposal of Goulard and Voltz (1993).

The choice of $K$, the number of basis functions in the expansions, can be solved by functional cross-validation (as proposed in Giraldo *et al.*, 2008) or using standard methods in non-parametric regression for the choice of the smoothing parameter (see Ramsay and Silverman, 2005 or Wasserman, 2006, for instance). It should be noted that the computational cost of functional cross-validation in point-wise functional kriging is far greater than in ordinary kriging for function-valued data.

A third alternative for functional kriging is allowing the coefficients $\lambda_i$ to be defined in $[a,b] \times [a,b]$. Then, the predictor of $\boldsymbol{\chi}_{s_0}$ is

$$\hat{\boldsymbol{\chi}}_{s_0}(t) = \sum_{i=1}^{n} \int_a^b \lambda_i(t,v)\boldsymbol{\chi}_{s_i}(v)\,\mathrm{d}s, \quad t \in [a,b],\ \lambda_i : [a,b] \times [a,b] \mapsto \mathbb{R},\ i = 1,\ldots,n \tag{5}$$

This modeling approach is coherent with the functional linear model for functional responses (total model) introduced by Ramsay and Silverman (2005). The functional parameter $\lambda_i(t,v)$ in Equation (5) determines the impact of the $i$th observed function at time $t$ on an unobserved function at time $v$. This kriging predictor has been separately proposed by Giraldo *et al.* (2009b) (called as *functional kriging*



DOI: 10.1002/env

(*total model*)) and by Monestiez and Nerini (2008). The functional parameters $\lambda_i(t, v)$ in Equation (5) are estimated taking into account constraints of unbiasedness and minimum prediction variance. Thus the optimization problem to be solved is

$$\underset{\lambda_1(\cdot,\cdot),\ldots,\lambda_n(\cdot,\cdot)}{Min} \int_a^b \mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\chi}}_{s_0}(v) - \boldsymbol{\chi}_{s_0}(v)\right) \mathrm{d}v \; s.t. \; \mathbb{E}(\hat{\boldsymbol{\chi}}_{s_0}(v)) = \mathbb{E}(\boldsymbol{\chi}_{s_0}(v)), \quad \forall v \in [a, b]$$

Observed functions $\boldsymbol{\chi}_{s_i}(t)$ are expanded as before and bivariate functional parameters $\lambda_i(t, v)$ are expressed now as

$$\lambda_i(t, v) = \sum_{j=1}^{K}\sum_{l=1}^{K} c_{jl}^i B_j(t) B_l(v) = \mathbf{B}^T(t)\mathbf{C}_i\mathbf{B}(v)$$

Giraldo *et al*. (2009b) also propose a related functional cokriging technique. They consider the prediction of the random variable $\boldsymbol{\chi}_{s_0}(v)$, that is, the value of the functional spatial process at an unsampled location $s_0$ at a specific time $v \in [a, b]$. Likewise cokriging (Myers, 1982; Bogaert, 1996), these authors propose the *cokriging predictor based on functional data*

$$\hat{\boldsymbol{\chi}}_{s_0}(v) = \sum_{i=1}^{n} \int_a^b \lambda_i^v(t) \boldsymbol{\chi}_{s_i}(t) \, \mathrm{d}t \tag{6}$$

In order to find the best linear unbiased predictor, the $n$ functional parameters in the proposed predictor are given by the solution of the following optimization problem

$$\underset{\lambda_1^v(\cdot),\ldots,\lambda_n^v(\cdot)}{Min} \mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\chi}}_{s_0}(v) - \boldsymbol{\chi}_{s_0}(v)\right), \; s.t. \, \mathbb{E}(\hat{\boldsymbol{\chi}}_{s_0}(v)) = \mathbb{E}(\boldsymbol{\chi}_{s_0}(v))$$

Assume now that we express the observed functions $\chi_{s_i}(t)$ and the functional parameters $\lambda_i^v(t)$ in terms of $K$ basis functions, $B_1(t), \ldots, B_K(t)$. We solve the family of such optimization problems indexed by $v \in [a, b]$, and then we look at the solution $\hat{\lambda}_1^v(t)$ as a function defined for all $(v, t) \in [a, b] \times [a, b]$. Giraldo *et al*. (2009b) prove that this bivariate function coincides with the solution of the functional kriging (total model) optimization problem: $\hat{\lambda}_1^v(t) = \hat{\lambda}_i(t, v)$ for all $(v, t) \in [a, b] \times [a, b], i = 1, \ldots, n$. Therefore the predictor proposed in Equation (5) coincides with the predictor given by the family of Equations (6) indexed by $v \in [a, b]$. The equivalence between both functional kriging predictors is analogous to that of cokriging analysis and multivariable spatial prediction (see Ver Hoef and Cressie, 1993), in the sense that the prediction obtained by cokriging based on functional data at a time is identical to the prediction achieved by functional kriging (total model) at the same time (this is the reason why we only include one of them in the comparative study summarized in Table 1). A distinctive feature between these methodologies is given in terms of their prediction variances. The estimated prediction variance of functional kriging (total model) can be used as a global measure of uncertainty on the prediction of a whole curve, whereas the estimated prediction variance of cokriging based on functional data can be used in a classical sense, that is, we can calculate point-wise confidence intervals for the prediction.

We now illustrate and compare the different reviewed approaches by using the Canadian temperature data set, described at the beginning of this Section. When periodic functional data are smoothed,



DOI: 10.1002/env

Table 1. Summary statistics for the sum of individual square errors obtained by functional cross-validation. The sum row corresponds to $SSE_{FCV}$

| Statistic | Ordinary kriging for function-valued data | Point-wise functional kriging | Functional kriging (Total model) |
|---|---|---|---|
| Minimum | 135.1 | 154.7 | 140.1 |
| Median | 586.6 | 597.8 | 576.6 |
| Mean | 5004.0 | 3033.0 | 3006.0 |
| Maximum | 91806.8 | 32770.0 | 32620.0 |
| St. dev. | 15536.0 | 6173.0 | 6145.0 |
| Sum | 175140.0 | 106155.0 | 105208.0 |

Fourier basis with an even number of basis functions is considered an appropriate choice (Ramsay and Silverman, 2005). A Fourier basis with 65 functions is the most frequently used expansion for the Canadian temperature data (Ramsay and Silverman, 2005) and this is what we use here. The comparison of different prediction methods is done in terms of functional cross-validation. For each method the individual $SSE_{FCV}(i)$ values, defined in Equation (3), are calculated for the 35 weather stations. Table 1 shows the summary statistics of these quantities obtained by the three functional kriging methods: ordinary kriging for function-valued data, point-wise functional kriging, and functional kriging (total model).

Though the differences are small, these summary statistics indicate that functional kriging (total model) has better performance than other predictors. The results shown in Table 1 highlight that including double indexed functional parameters into the analysis implies better predictions. The temperature curve in Resolute (the station in the Arctic) is not well predicted in any case. The worst prediction for this station is obtained by ordinary kriging. However, in other stations this method produces very similar predictions to those obtained by other predictors. In fact, the best prediction is also obtained using ordinary kriging. In summary we can conclude that, for this data set, functional kriging (total model) is the best option for carrying out spatial prediction of functional data. The differences among functional kriging (total model) and other methods are very small in a high proportion of sites. Consequently ordinary kriging and point-wise kriging are also good alternatives for performing spatial prediction of functional data, taking into account that these methods are easier than functional kriging (total model) from practical and computational points of view.

Let us now go over other contributions dealing with geostatistical functional data. As we have previously mentioned, Monestiez and Nerini (2008) also introduce the functional kriging given by Equation (5) simultaneously to Giraldo *et al*. (2009b). Monestiez and Nerini (2008) use the term *cokriging for spatial functional data* (see also Nerini *et al*., 2008). They also use truncated expansions in a functional basis, followed by cokriging over the basis coefficients. The main methodological difference between Giraldo *et al*. (2009b) and Monestiez and Nerini (2008) is that the latter work is based on orthonormal basis functions, whereas orthogonality is not a required condition in the former. Monestiez and Nerini (2008) illustrate their proposal analyzing temperature versus depth profiles in the Antarctic Ocean. An elephant seal is used as sampler. It is followed during a cruise of 4.5 months. Each dive of the elephant seal provides a temperature curve, located at precise coordinates.

Spatio-temporal modeling (Christakos, 2000) is an alternative way to interpolate general functional data. It consists of modeling covariance functions for data evolving in space and time showing light on the dependence between the spatial and temporal components. It plays with the important assumptions



DOI: 10.1002/env

of stationarity or isotropy (Gneiting, 2002; Porcu *et al*., 2007). In a general context, the role of time can be played by the elements $t \in [a, b]$, the interval where the functional data are defined. When using spatio-temporal modeling to predict the functional process $\boldsymbol{\chi}_s$ at an unsampled site $s_0$, the values $\boldsymbol{\chi}_{s_0}(t_i)$ ($t_i$ forming a fine grid in $[a, b]$) must be predicted. The computational cost is high if the grid contains a large number of points. An advantage of spatio-temporal modeling is that it is useful even when different functional data are observed in different subintervals of $[a, b]$.

The question of which approach, geostatistics for funcional data or space-time geostatistics, should be used to analyze a particular data set is an important one with no closed answer. When using geostatistics for functional data we are assuming that the observed data for a fixed site $s$, follows the model $\boldsymbol{\chi}_s(t_j) = \mu_s(t_j) + \varepsilon_s(t_j)$, $j = 1, \ldots, M$, where $\varepsilon_s(t_j)$ are independent zero mean residuals and the main structure of $\boldsymbol{\chi}_s(\cdot)$, as a function of $t \in [a, b]$, is in the mean function $\mu_s(\cdot)$. When using space-time geostatistics, it is assumed implicitly that $\mu_s(\cdot)$ summarizes only a little part of the structure of $\boldsymbol{\chi}_s(\cdot)$, and that the residuals $\varepsilon_s(t_j)$ have a rich covariance structure. We face again the known problem in geostatistics: one person's deterministic mean structure may be another person's correlated error structure (Cressie, 1993, p. 114). In practice we recommend using space-time geostatistics when the number $M$ of observed values of functional data $\boldsymbol{\chi}_s(\cdot)$ is small or when the interest is predicting a specific value $\boldsymbol{\chi}_{s_0}(t_0)$ for an unvisited site $s_0$ and/or unvisited time $t_0$. In other cases (many observations per functional data and interest in prediction of a whole function at an unvisited site), geostatistics for functional data should be the default approach.

An example of using spatio-temporal modeling for interpolating spatial functional data is the work of Bel *et al*. (2008). They perform functional regression on paleo-ecological data. An index of genetic diversity (a real value) is the dependent variable and temperature and precipitation curves (functional data, time going from $-15.000$ to 0 years) are the predictors. The data are georeferenced, thus the spatial dependence is considered when fitting the model. Moreover, the spatial locations of the observed predictors do not match with those of the observed responses. To make them match, curves of temperature and precipitation are firstly estimated on sites where the genetic measures are collected. The interpolation is done by a spatio-temporal kriging, assuming that the covariance function is exponential and separable.

Three more works where spatial functional processes are considered are the following. Dabo-Niang and Yao (2007) propose non-parametric kernel regression with scalar response $Y_s$ and functional predictors, as observations of a continuous spatial process $\boldsymbol{\chi}_s$. The objective is to non-parametrically estimate $E(Y_s|\boldsymbol{\chi}_s)$ taking into account the spatial dependence. Basse *et al*. (2008) propose kernel density estimates for spatial functional random variables, with respect to a dominant measure. Yamanishi and Tanaka (2003) develop a regression model where both response and predictors are functional data, and the relation among variables may change over the space. The model combines geographically weighted regression (Brunsdon *et al*., 1998) and functional multiple regression (Ramsay and Silverman, 2005). Observe that none of these papers consider kriging or spatial interpolation for functional data.

There are some contributions to spatial FDA from a Bayesian perspective. Baladandayuthapani *et al*. (2008) show an alternative for analyzing an experimental design with a spatially correlated functional response. They use Bayesian hierarchical models allowing to include spatial dependence among curves into standard FDA techniques, such as functional multiple regression and functional analysis of variance. Similar possibilities are offered by the Bayesian non-parametric models for functional data proposed by Rodríguez *et al*. (2008, 2009) and Petrone *et al*. (2008). They propose hierarchical models that are extensions of the Dirichlet process mixture of Gaussians. Rodríguez *et al*. (2008) includes a spatial functional data example: a sample of functions relating temperature with depth is collected at 87 locations in the North Atlantic ocean. Nevertheless the model they fit to these data is not explicitly spatial.



DOI: 10.1002/env

## 3. POINT PROCESSES WITH ASSOCIATED FUNCTIONAL DATA

Although, a partial study of a functional marked point pattern can be performed by a point-wise analysis of such functions (applying the mark or pair correlation functions), the use of such partial information can be extremely inefficient for a space-time process involving, for instance, a large period of time. This problem, together with the impossibility to consider partial information of certain curves (for instance spectral curves or curves representing probability density functions) suggests the definition of new second-order characteristics to analyze spatial patterns containing functional data.

Note that although functions associated to each point position depend on time, the point pattern under analysis does not evolve through time. As such, the point pattern is not a space-time point pattern, but a point pattern with associated functions which evolves dynamically. This clearly limits the type of statistical methods to analyze such point patterns. In fact, time is not necessarily a dependent variable for the associated functions, since it would be also possible to have a point pattern involving density functions, spectral curves, or functions representing the soil vertical concentration of a given element, all being examples of point patterns with no temporal component.

Here, we illustrate and analyze a functional marked point pattern based on a data set extensively studied in point process theory, namely *the Spanish town data*. Glass and Tobler (1971) introduced such data to study the distribution of cities on a plain. These data has been frequently used as an example of an inhibitory point process (see, for instance, Ripley, 1977, 1981, or Stoyan *et al*., 1995; this data set is also included in the R library `spatial`). It contains the centers of 69 towns in a 40 miles homogeneous square located in Cuenca (south-east of Madrid). We have been able to identify the names of 67 out of the 69 towns by visual inspection of a 2007 map of Cuenca. Only 56 of them correspond to actual town councils, and for them it was possible to find statistical information on the web page of the Spanish Institute of Statistics (www.ine.es). We have focused on the demographic evolution from 1900 to 2007. Complete information was available for 53 towns (the other three towns started to be independent town councils after 1900). Figure 2 shows the spatial position of towns along with the relative population (fixing 100 in 1900 and taking logs) for the 53 towns over time.

The objectives of studying point processes with a functional mark are essentially the same as in other marked point processes. The most important question is that of knowing if there is spatial dependence in the functional marks. Note that the point pattern is a standard one and only the type of mark is different.

Comas *et al*. (2008) define a counterpart version of the mark correlation function assuming functional data. Let $h(\cdot, \cdot)$ be a test functional involving two functions (for instance, $h(f, g)$ could be a similarity measure between functions $f$ and $g$), and let $\lambda^{(2)}(r)$, $r \in \mathbb{R}^+$, be the usual second-order product density for the stationary and isotropic point process $\Psi$, Let $\lambda_f^{(2)}(r)$ be the counterpart version of this density for a functional marked point process, that is, $\lambda_f^{(2)}(r)$ is the density associated to the second-order functional factorial moment measure (when this measure is absolutely continuous with respect to the Lebesgue measure)

$$\alpha_f^{(2)}(A_1 \times A_2) = E\left[\sum_{s_1, s_2 \in \Psi}^{\neq} h(\boldsymbol{\chi}_{s_1}, \boldsymbol{\chi}_{s_2}) I_{A_1 \times A_2}(s_1, s_2)\right]$$



DOI: 10.1002/env

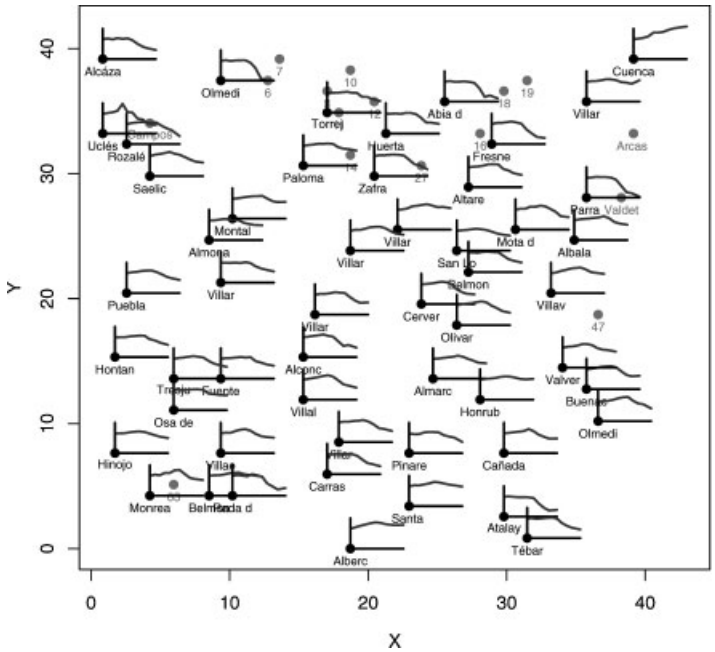


Figure 2. Spanish town data. Demographic evolution (in logs) from 1900 to 2007. There is no available information for 16 towns. This figure is available in color online at www.interscience.wiley.com/journal/env

where $A_1, A_2 \subseteq \mathbb{R}^2$. The *functional mark-correlation function* (Comas *et al*., 2008) is defined as

$$g_f(r) = \frac{\lambda_f^{(2)}(r)}{\lambda^{(2)}(r) E[h(\boldsymbol{\chi}_{s_1}, \boldsymbol{\chi}_{s_2})]}$$

where $r = \|s_1 - s_2\|$. Comas *et al*. (2008) propose to estimate $g_f(r)$ in the observation window $W$ by

$$\hat{g}_f(r) = \frac{1}{2\pi r \hat{\lambda}_p^2 |W|} \sum_{s_1, s_2 \in \psi}^{\neq} \frac{h(\boldsymbol{\chi}_{s_1}, \boldsymbol{\chi}_{s_2}) K(\|s_1 - s_2\| - r)}{\hat{E}[h(\boldsymbol{\chi}_{s_1}, \boldsymbol{\chi}_{s_2})] e(s_1, \|s_1 - s_2\|)}$$

where $\psi$ is the observed point pattern, $\hat{\lambda}_p$ is an estimator of the point intensity, $K(\cdot)$ is a kernel function that is non-negative and symmetric with respect to the origin, and $e(\cdot)$ is a factor to correct for edge-effects. In our example we have used the test function

$$h(\boldsymbol{\chi}_{s_1}, \boldsymbol{\chi}_{s_2}) = \int_a^b (\chi_{s_1}(t) - \bar{\chi}(t))(\chi_{s_2}(t) - \bar{\chi}(t))\, \mathrm{d}t$$

where $\bar{\chi}(t)$ is the average function over the observed functions. This is a kind of correlation between functions. Therefore, if $\hat{g}_f(r) > 1$, then pairs of functions at distances $r$ are more similar than the average,



DOI: 10.1002/env

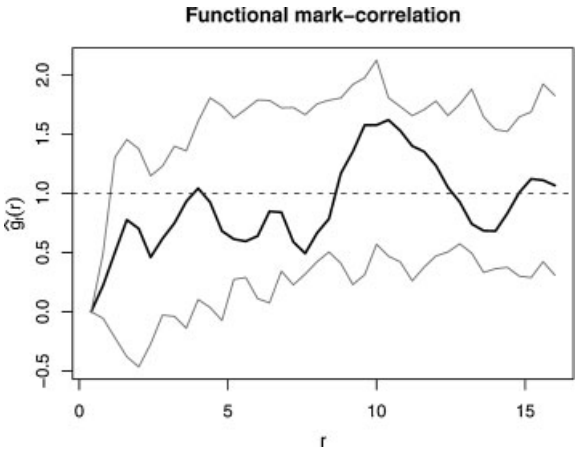


Figure 3. Estimation of the functional mark-correlation function $g_f(r)$ for the demographic evolution in 53 Spanish towns. Red lines are point-wise 95% confidence bands for the null hypothesis of no spatial dependence between marks, based on 1000 permutations. This figure is available in color online at www.interscience.wiley.com/journal/env

suggesting positive correlation, whilst $\hat{g}_f(r) < 1$ corresponds to the opposite case, and $\hat{g}_f(r) = 1$ implies spatial independence between functions.

Figure 3 shows the resulting functional mark correlation function for the Spanish town functional marked point pattern (see Figure 2). This function suggests that there is no spatial dependence between marks, so we can conclude that the way the population size has evolved from 1900 to now in different towns is spatially independent. It must be noted that the sample size is small ($n = 53$ in this case), and thus a larger sample size could help to discover potential dependencies between functional marks.

We finish this section talking about two more works involving spatial point patterns and functional data. Mateu *et al*. (2007) consider the problem of detecting features via local indicators of spatial association (LISA) functions. They estimate a LISA function at each event of a point pattern. Thus they have a functional marked point process, where the marks are the LISA functions. Distances from the observed LISA functions to their expected values (if the underlying process is Poisson) are calculated. A mixture of two distributions is fitted to the set of observed distances. Both mixture components are identified as feature and noise, respectively, and a classification is considered.

Illian *et al*. (2005) use functional principal component analysis for analyzing spatial point processes. They consider spatial point patterns of ecological plant communities in which a very large number of points exist for many different plant species. L-functions (or the pair-correlation functions) for each species are computed. Functional principal component analysis of these functions is used to analyze and cluster the plant species by their spatial behavior. Note that functional marked point processes are not used here.

## 4. FUNCTIONAL AREAL DATA

We finally present an example of functional areal data consisting in population pyramids for 38 neighborhoods in Barcelona (Spain). The 38 areas constitute an official division of the municipality of Barcelona (they are called *statistical zones*). The information about population in these areas for



DOI: 10.1002/env

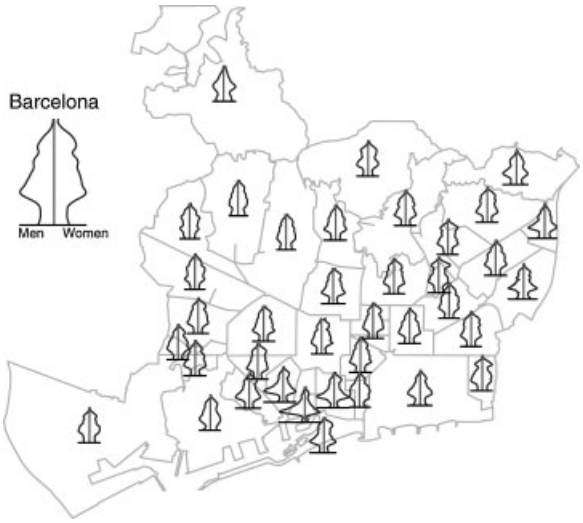


Figure 4. Smoothed population pyramids for 38 statistical zones of Barcelona. This figure is available in color online at www.interscience.wiley.com/journal/env

2005 (first of January) has been obtained from the Department of Statistics at Barcelona Municipal Council (www.bcn.cat/estadistica/angles/index.htm). The total population in the areas goes from 1243 to 101 100, with quartiles 16 450, 36 810, and 64 180. Data on population, classified by sex and age (in intervals of 1 year), has been used to build population pyramids for each statistical zone. Spline smoothing has been used to obtain the smoothed population pyramids shown in Figure 4.

The aims for this kind of spatial FDA are similar to those concerning univariate or multivariate areal data: detection of the spatial dependence (via spatial autocorrelation testing), identification of spatial clusters, and to modeling the spatial dependence (via spatial regression models, for instance) are probably the main ones. Thus it is natural to adapt descriptive and inferential techniques used for univariate and multivariate areal data to the case of having a function at each area. For instance, Delicado and Broner (2008) propose a local spatial autocorrelation test based on dissimilarities between areas. The dissimilarity is computed from the information available at each area. Thus it can be applied to any kind of observable characteristic, provided that a dissimilarity (or distance) can be defined between any pair of observations. Delicado and Broner (2008) design an algorithm to identify spatial clusters based on these local tests. The graphical representation of the clustering is a distance-based version of LISA maps (Anselin, 1995). The proposed algorithm works in five steps (see Delicado and Broner, 2008 for more details):

Distance-based LISA maps algorithm

Step 1: Detect global outliers.
Step 2: Using the distance-based local spatial autocorrelation test, mark areas significantly similar to their neighbors. Detect also the areas that are significantly different from their neighbors (spatial outliers).
Step 3: Mark the non-marked areas that are similar to a neighbor marked area.



DOI: 10.1002/env

Step 4: Identify spatial clusters by applying any standard clustering algorithm to the areas marked at Steps 2 and 3.
Step 5: Draw a map using different colors to identify global outliers, spatial outliers, different clusters in the set of marked areas, and the rest of non-marked areas.

When working with population pyramids, that are particular cases of density functions, an appropriate distance between them is the symmetric version of the Kullback-Leibler divergence

$$d_{KL}(f_i, f_j) = \int_a^b \log\left(\frac{f_i(x)}{f_j(x)}\right) f_i(x)\,\mathrm{d}x + \int_a^b \log\left(\frac{f_j(x)}{f_i(x)}\right) f_j(x)\,\mathrm{d}x$$

Figure 5 shows the resulting distance-based LISA map. There are three areas that have been detected as global outliers (colored in red). Among them, one having the lowest extension is a neighborhood receiving many immigrant in the last years (mainly young men), and it is reflected directly in its population pyramid. The reasons why the other two are global outliers are not so clear. The areas in white do not belong to any spatial cluster. In this data set there are no spatial outliers. There are three spatial clusters, colored in blue, light blue, and magenta, respectively. The main characteristics of these clusters are summarized by their average population pyramid, as shown in Figure 6. Blue and light blue clusters are very similar (and they are also similar to the average of the whole city), differing only in the amount of old woman. Magenta cluster corresponds to a younger population where probably young immigrant males have been recently incorporated.

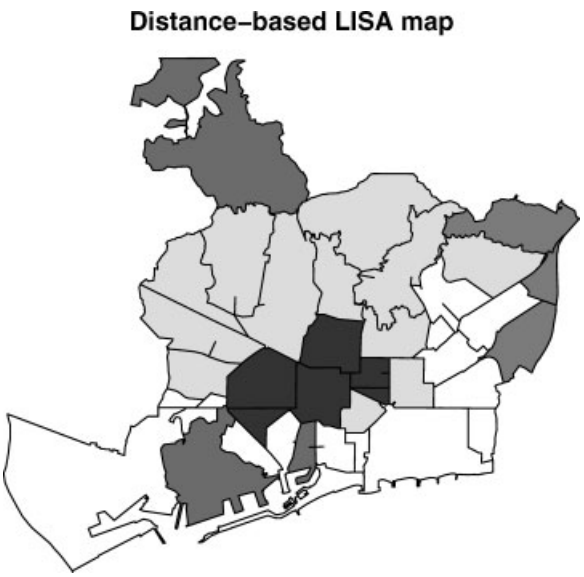


Figure 5. Distance-based LISA map for population pyramids in Barcelona. This figure is available in color online at www.interscience.wiley.com/journal/env



DOI: 10.1002/env

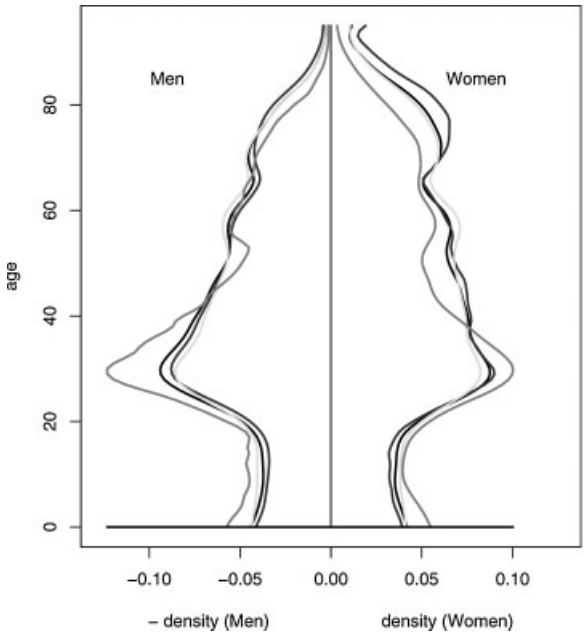


Figure 6. Average population pyramids for the three clusters (colored in blue, light blue, and magenta) detected by the distance-based LISA map algorithm. The black pyramid corresponds to the whole city. This figure is available in color online at www.interscience.wiley.com/journal/env

## 5. CONCLUDING REMARKS

Functional data with spatial dependence is a new topic that offers the possibility of combining knowledge from spatial statistics and FDA. We believe that this combination has a very promising future in both applied and theoretical sides of statistics. In fact, it is already a fertile field of research. Multivariate spatial statistical tools can be generalized to be valid for functional data. Specially well suited are those based on distances between observed features. Some contributions on geostatistics, point processes and areal data with functional observations have been presented, as well as examples of each type of functional data with spatial dependence. Geostatistics for functional data is at this moment the most developed topic.

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministerio de Educación y Ciencia and FEDER grants MTM2006-09920 and MTM2007-62923. R. Giraldo thanks FUNDACIÓN CAROLINA for financial support through a PhD scholarship. C. Comas was supported during 2008 by a Beatriu de Pinós contract from the Comissionat per a la Universitat i Recerca del Dep. d' Innovació, Universitats i Empresa de la Generalitat de Catalunya.
The authors are grateful to the Associate Editor and the referees whose comments and suggestions have clearly improved an earlier version of the manuscript.



DOI: 10.1002/env

## REFERENCES


Anselin L. 1995. Local indicators of spatial association-lisa. *Geographical Analysis* **2**: 93–115.

Baladandayuthapani V, Mallick B, Hong M, Lupton J, Turner N, Caroll R. 2008. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinoginesis. *Biometrics* **64**: 64–73.

Basse M, Diop A, Dabo-Niang S. 2008. Mean squares properties of a class of kernel density estimates for spatial functional random variables. *Technical Report*, Université Gaston Berger de Saint-Louis, Senegal.

Bel L, Bar-Hen A, Cheddadi R, Petit R. 2008. Spatio-temporal functional regression on paleo-ecological data. arXiv:0807.2588v1. Submitted.

Bogaert P. 1996. Comparison of kriging techniques in a space-time context. *Mathematical Geology* **28**(1): 73–86.

Brunsdon C, Fotheringham S, Charlton M. 1998. Geographically weighted regression-modelling spatial non-stationarity. *The Statistician* **47**: 431–443.

Christakos G. 2000. *Modern Spatiotemporal Geostatistics*. Oxford University Press: New York.

Comas C, Delicado P, Mateu J. 2008. Analysing spatial point patterns with associated functional data. In *Statistics for Spatio-temporal Modelling. Proceedings of the 4th International Workshop on Spatio-temporal Modelling (METMA-4)*, 157–163.

Cressie N. 1993. *Statistics for Spatial Data*. John Wiley & Sons: New York.

Dabo-Niang S, Yao AF. 2007. Kernel regression estimation for continuous spatial processes. *Mathematical Methods for Statistics* **16**: 298–317.

Davidian M, Lin X, Wang J. 2004. Introduction: emerging issues in longitudinal and functional data analysis. *Statistica Sinica* **14**: 613–614.

Delicado P, Broner S. 2008. Distance-based LISA maps for multivariate lattice data. *Technical Report* http://hdl.handle.net/2117/2421, Universitat Politècnica de Catalunya. Preprint.

Ferraty F, Vieu P. 2006. *Non Parametric Functional Data Analysis. Theory and Practice*. Springer: New York.

Giraldo R, Delicado P, Mateu J. 2008. Ordinary kriging for function-valued spatial data. *Technical Report* http://hdl.handle.net/2117/1099, Universitat Politècnica de Catalunya. Submitted.

Giraldo R, Delicado P, Mateu J. 2009a. Continuous time-varying kriging for spatial prediction of functional data: an environmental application. *Journal of Agricultural, Biological, and Environmental Statistics (JABES)*. Accepted for publication.

Giraldo R, Delicado P, Mateu J. 2009b. From multivariable to functional geostatistics: spatial prediction of temperature curves. Submitted.

Glass L, Tobler WR. 1971. Uniform distribution of objects in a homogeneous field: cities on a plain. *Nature* **233**: 67–68.

Gneiting T. 2002. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* **97**: 590–600.

González-Manteiga W, Vieu P. 2007. Editorial: statistics for functional data. *Computational Statistics and Data Analysis* **51**: 4788–4792.

Goulard M, Voltz M. 1993. Geostatistical interpolation of curves: a case study in soil science. In *Geostatistics Tróia'92* (Vol. 2), Soares A (ed.). Kluwer Academic Press: Boston; 805–816.

Illian J, Benson E, Crowford J, Staines H. 2005. Principal component analysis for spatial point processes—assessing the appropriateness of the approach in an ecological context. In *Case Studies in Spatial Point Process Modeling*, Baddeley A, Gregori P, Mateu J, Stoica R, Stoyan D (eds). Number 185 in Lecture Notes in Statistics, Springer-Verlag.

Mateu J, Lorenzo G, Porcu E. 2007. Detecting features in spatial point processes with clutter via local indicators of spatial association. *Journal of Computational and Graphical Statistics* **16**: 968–990.

Monestiez P, Nerini D. 2008. A cokriging method for spatial functional data with applications in oceanology. In *Functional and Operatorial Statistics. Proceedings of the First International Workshop on Functional and Operational Statistics*, Dabo-Niang S, Ferraty F (eds). Springer, Toulouse, France, June 2008, 237–242.

Myers D. 1982. Matrix formulation of Co-kriging. *Mathematical Geology* **14**(3): 249–257.

Nerini D, Monestiez P, Manté C. 2009. Cokriging for spatial functional data. *Journal of Multivariate Analysis*. DOI:10.1016/j.jmva.2009.03.005

Petrone S, Guindani M, Gelfand A. 2008. Hybrid dirichlet mixture models for functional data. Unpublished.

Porcu E, Mateu J, Bevilacqua M. 2007. Covariance functions which are stationary or nonstationary in space and stationary in time. *Statistica Neerlandica* **61**: 58-–382.

Ramsay J. 2008. Fda problems that i like to talk about. Personal communication.

Ramsay J, Silverman BW. 2005. *Functional Data Analysis* (2nd edn). Springer: New York.

Ripley BD. 1977. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B* **39**: 172–212.

Ripley BD. 1981. *Spatial Statistics*. Wiley: New York.

Rodríguez A, Dunson DB, Gelfand A. 2008. The nested dirichlet process. *Journal of the American Statistical Association* **103**: 1131–1144. With discussion.

Rodríguez A, Dunson DB, Gelfand A. 2009. Nonparametric functional data analysis through Bayesian density estimation. *Biometrika* **96**(1): 149–162.

Stoyan D, Kendall W, Mecke J. 1995. *Stochastic Geometry and its Applications*. John Wiley & Sons: New York.

Valderrama MJ. 2007. Editorial: an overview to modelling functional data. *Computational Statistics* **22**: 331–334. Special issue on Modelling Functional Data in Practice.
Ver Hoef J, Cressie N. 1993. Multivariable spatial prediction. *Mathematical Geology* **25**(2): 219–240.
Wasserman L. 2006. *All of Nonparametric Statistics*. Springer: New York.
Yamanishi Y, Tanaka Y. 2003. Geographically weighted functional multiple regression analysis: a numerical investigation. *Journal of Japanese Society of Computational Statistics* **15**: 307–317.



DOI: 10.1002/env