Environ Ecol Stat (2011) 18:411–426
DOI 10.1007/s10651-010-0143-y

Ordinary kriging for function-valued spatial data

R. Giraldo · P. Delicado · J. Mateu

Received: 22 June 2007 / Revised: 10 March 2010 / Published online: 19 May 2010

Abstract In various scientific fields properties are represented by functions varying over space. In this paper, we present a methodology to make spatial predictions at non-data locations when the data values are functions. In particular, we propose both an estimator of the spatial correlation and a functional kriging predictor. We adapt an optimization criterion used in multivariable spatial prediction in order to estimate the kriging parameters. The curves are pre-processed by a non-parametric fitting, where the smoothing parameters are chosen by cross-validation. The approach is illustrated by analyzing real data based on soil penetration resistances.

Keywords Cross-validation · Functional data · Non-parametric curve fitting · Ordinary kriging · Soil penetration resistance · Trace-variogram

1 Introduction

The number of problems and the range of disciplines where the data are functions is recently increasing. This data may be generated by a large number of measurements (over time, for instance), or by automatic recordings of a quantity of interest. Since beginning of the nineties, Functional data analysis (FDA) has been used in order to model this kind of data. Since the pioneer work by Deville (1974), and more recently

R. Giraldo · P. Delicado
Universitat Politècnica de Catalunya, Barcelona, Spain

R. Giraldo
Universidad Nacional de Colombia, Bogotá, Colombia

J. Mateu (✉)
Department of Mathematics, University Jaume I, Campus Riu Sec, 12071 Castellón, Spain
e-mail: mateu@mat.uji.es

with the work by Ramsay and Silverman (2005), the statistical community has shown an increasing interest in developing models for functional data. Functional versions for a wide range of statistical tools have been given. Examples of such methods include exploratory and descriptive data analysis (Ramsay and Silverman 2005), linear models (Cardot et al. 1999; Ramsay and Silverman 2005), non-parametric methods (Ferraty and Vieu 2006) or multivariate techniques (Goulard and Voltz 1993; Voltz and Goulard 1994; Silverman 1995; Ferraty and Vieu 2003).

In various disciplines, it is common that data have both spatial and functional components (Goulard and Voltz 1993). In Agronomy, for instance, prior to planting, penetration resistance measurements are made at the nodes of a grid overlain on the study area (Chan et al. 2006). In this case, and though penetration resistance is only measured at some depths, it is possible to consider it as a functional variable after a smoothing or interpolation process have been applied. Other examples are given when daily cycles of oxygen are measured in different points of a study zone (Mancera and Vidal 1994) or when temperature or precipitation records are made at various weather stations of a country (Ramsay and Silverman 2005).

In the same way that some statistical methods have been generalized to be also useful within the FDA context, geostatistical methods can be adapted to this type of problem and model data with both spatial and functional components. This modeling approach can certainly be useful to predict functions based on observed spatially referenced curves. The pioneering work of Goulard and Voltz (1993) was the first attempt (at our knowledge) to apply geostatistical interpolation methods to predict functions at non-data sites. They proposed three methods: two of them were based on a multivariate approach using cokriging, and the other one used a functional kriging step predicting directly the curves. Goulard and Voltz (1993) assumed that the functions were only known at a finite set of points, and a parametric model was fitted to them for reconstructing the whole curve. In this context the parametric model was assumed to be known and both, the number of known points for each function and the number of parameters in the parametric model, were assumed to be small.

In this paper, we extend the contributions of Goulard and Voltz (1993) overcoming the restrictive assumptions on parametric modeling and small number of observed points per function. In particular, we propose to apply a non-parametric fitting preprocess to the observed functions (in this paper we use B-spline smoothing, but other approaches are possible) where the smoothing parameter is chosen by what we call *functional cross-validation*. We believe that our approach is in complete agreement with present trends in FDA, and in particular, with non-parametric functional estimation methodology. Our proposal for using kriging to predict functions formally coincides with the functional kriging introduced in Goulard and Voltz (1993), but our non-parametric method differs significantly (for instance, data representation) and we deal with the additional problem of choosing the smoothing parameters (the keystone of non-parametric methods). Our predictor is based on the basic tenet of functional data analysis, that is, functions are single entities rather than a sequence of individual observations (Ramsay and Silverman 2005).

The paper is organized as follows. In Sect. 2, we introduce functional notation and some known results, and we summarize the proposals of Goulard and Voltz (1993). In particular, the predictor and the optimization criterion are stated here. In Sect. 3

we focus on measuring the spatial correlation when data are functions. The potential limitations of Goulard and Voltz's proposals are addressed in Sect. 4, and our non-parametric approach is proposed. An application of the proposed methodology to an agronomical data set is considered in Sect. 5. Conclusions and discussion on further topics of research are given in Sect. 6. A final Appendix shows the proofs of technical results.

2 Ordinary kriging based on curves

Ferraty and Vieu (2006) define a *functional random variable* as a random variable taking values in a space of functions. In this paper we are dealing with functional random processes (i.e. functional random fields or functional random functions), $\{\boldsymbol{\chi}(s), \quad s \in D \subseteq \mathbb{R}^d\}$, whose realizations (or values) are functions defined on $T = [a, b] \subseteq \mathbb{R}$ and assumed to belong to

$$L_2(T) = \left\{ f : T \to \mathbb{R}, \text{ such that } \int_T f(t)^2 dt < \infty \right\}.$$

Note that $L_2(T)$ is a Hilbert space with the inner product $\langle f, g \rangle = \int_T f(t)g(t)dt$.

Additionally, for every fixed $t_0 \in T$, $\{\chi_{t_0}(s), \quad s \in D \subseteq \mathbb{R}^d\}$ is a scalar-valued random process (i.e random field or random function) defined on $\mathbb{R}^d$ with values in $\mathbb{R}$. Note that $\chi_t(s)$ is a possible scalar value of $\boldsymbol{\chi}(s)$ evaluated at a particular location s and temporal instant t.

We assume that $\{\boldsymbol{\chi}(s), \quad s \in D \subseteq \mathbb{R}^d\}$ is weakly stationary in the following sense:

- For every fixed $t_0 \in T$, $\mathbb{E}\chi_{t_0}(s)$ does not depend on $s \in D$, and the function $t_0 \mapsto \mathbb{E}\chi_{t_0}(s)$ is measurable on T.
- For every fixed $t_0, t_0' \in T$, and $s, s' \in D$

  $$\mathbb{V}\text{ar}(\chi_{t_0'}(s') - \chi_{t_0}(s)) = \mathbb{E}\left(\left(\chi_{t_0'}(s') - \chi_{t_0}(s)\right)^2\right)$$

  exists and does depend on s and s' only through their difference $h = s' - s$. Then the function

  $$\gamma_{t_0', t_0}(h) = \frac{1}{2}\mathbb{V}\text{ar}\left(\chi_{t_0'}(s+h) - \chi_{t_0}(s)\right)$$

  is called the variogram and we use the notation $\gamma_{t_0}(h)$ for $\gamma_{t_0, t_0}(h)$.
- For every fixed $t_0, t_0' \in T$, and $s, s' \in D$, $\mathbb{C}\text{ov}\left(\chi_{t_0'}(s'), \chi_{t_0}(s),\right)$ depends on s and s' only through their difference $h = s' - s$. Then we can define the function $C_{t_0', t_0}(h) := \mathbb{C}\text{ov}\left(\chi_{t_0'}(s+h), \chi_{t_0}(s),\right)$. We shall assume that $C_{\cdot,\cdot}(h) : T \times T \to \mathbb{R}$ is square integrable. This ensures that the variance of the associated process $C_{\cdot,\cdot}(0)$ exists and is finite.

By construction, the covariance function defined under this setting is positive definite (Berg and Forst 1975) and the corresponding variogram is conditionally negative definite. Throughout this paper we use standard covariance and variogram models with the possibility of geometric anisotropies. However, we assume in this paper that the covariance functions and variograms are isotropic. Finally, observe that $Z(s) := \int_T \chi_t(s)\mathrm{d}t$ defines a scalar-valued random field.

In practice, each particular function coming from a realization of $\boldsymbol{\chi}(s)$ is only known or observed at a finite set of time points, and thus data of the form $\chi_t(s)$ with $t \in T$ is usually represented as a finite set of pairs (t_j, y_{ij}), $t_j \in T$, $j = 1, \ldots, M$ and $y_{ij} = \chi_{t_j}(s_i)$ (if there is no observational noise) or $y_{ij} = \chi_{t_j}(s_i) + \varepsilon_j$ (if noise is present), ε_j having zero mean. The set of points $\{t_j\}_{j=1}^{M} \subset T$ can be considered the same for all the functions in a functional data set, and usually they form a fine evenly spaced grid in T. Nowadays, in many problems based on real data the number M of observed values for each function is usually in the order of several hundreds or thousands (see, as an example, Friman et al. (2004) analyzing fMRI data, or Vandenberghe et al. (2005) monitoring water quality in a river). In such cases interpolation methods (if there is no observation noise) or non-parametric smoothing methods (in the opposite case) are commonly used to smooth out the data. Note that kriging equations can be slightly modified to obtain a smoothing predictor (Cressie 1990; Wahba 1990) and thus a close connection arises between the kriging variance and the smoothing error obtained using splines.

Note that, actually, the (complete) functions are not observed, instead only points on the graphs of the functions are observed. However, when the number of data points in a function is large (as it is the case in this paper), for simplicity we talk about "observed or measured functions".

Our goal is the prediction of $\chi_t(s_0)$ at a non-data location s_0 and for a set of temporal instants t. Note that in our approach we want to predict a function, and not just a scalar. In this sense our goal is close to multivariable spatial prediction (Ver Hoef and Cressie 1993). An even more general framework can be found in Tolosana-Delgado (2005), who apply geostatistics to compositional data.

Goulard and Voltz (1993) also consider the same problem of geostatistical interpolation of curves. They consider that curves are only known by a finite set of their points: $\chi_{t_j}(s_i)$, $j = 1, \ldots, M, i = 1, \ldots, n$. In their case study, each function was measured at $M = 8$ points (a much smaller value than usual values of M in many applications). Goulard and Voltz (1993) present three approaches for predicting curves at non-data locations:

- *Multivariate approach 1: Cokrige first, Fit later (CFP).* The vector of observed values $(\chi_{t_1}(s_i), \ldots, \chi_{t_M}(s_i))$ is considered as the observation of a M-dimensional random variable at site s_i. Cokriging is used to predict the values of this random vector at the non-data location s_0: $\left(\hat{\chi}_{t_1}(s_0), \ldots, \hat{\chi}_{t_M}(s_0)\right)$. The predicted values are then interpolated to generate the predicted function.
- *Multivariate approach 2: Fit first, Cokrige later (FCP).* First, a parametric model is fitted to the observed values: $\chi_t(s_i; \hat{\theta}_{s_i})$, $i = 1, \ldots, n$. The p-dimensional parameter values $\hat{\theta}_{s_1}, \ldots, \hat{\theta}_{s_n}$ are considered as observations of a multivariate random

variable. Then cokriging is used to predict the value of the parameter θ at site s_0, say $\hat{\theta}^*_{s_0}$, and $\chi_t\left(s_0; \hat{\theta}^*_{s_0}\right)$ is the resulting predicted value at s_0.

- *A curve kriging approach (CKP).* Goulard and Voltz (1993) define the best linear unbiased predictor (BLUP) for $\boldsymbol{\chi}(s_0)$ given by

$$\hat{\boldsymbol{\chi}}(s_0) = \sum_{i=1}^{n} \lambda_i \boldsymbol{\chi}(s_i), \quad \lambda_1, \ldots, \lambda_n \in \mathbb{R}, \tag{1}$$

where the coefficients λ_i are such that

$$\mathbb{E}\left(\hat{\boldsymbol{\chi}}(s_0) - \boldsymbol{\chi}(s_0)\right) = 0 \text{ and } \mathbb{E}\left[\int_T \left(\hat{\chi}_t(s_0) - \chi_t(s_0)\right)^2 dt\right] \text{ is minimized.}$$

Given that the values of the functions of the random process $\boldsymbol{\chi}(s_i)$ are known only at M points in T, Goulard and Voltz (1993) fitted a parametric model $\chi_t(\cdot;\theta), \theta \in \mathbb{R}^p$, to these data to obtain $\chi_t(\cdot;\hat{\theta}_{s_i})$ as an approximation of $\chi_t(s_i)$. Equation (1) can be then rewritten as

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

and the integrals on T involved in estimating the coefficients λ_i are calculated using the parametric fitted model instead of $\boldsymbol{\chi}(s_i)$ (see Sect. 3).

We should note that the spline is a special case of a Radial Basis function (RBF) interpolator, and in turn, using a RBF is equivalent to kriging (sometimes called *dual kriging*). Then, the smoothing spline is a special case of cokriging.

In Sect. 4 we discuss the three methods of Goulard and Voltz (1993) when M is large, or when a parametric fitting is not available. We present a non-parametric alternative to curve kriging prediction (CKP). We use the family of linear predictors for $\boldsymbol{\chi}(s_0)$ as in Eq. (1). Note that it has the same form as the classical ordinary kriging predictor, but using curves instead of scalars. The predicted curve is a linear combination of data curves. The kriging coefficients or weights λ_i in Eq. (1) reflect the influence of the data curves. Curves at those locations closer to the prediction point will naturally have greater influence than others more far apart.

Using the same expression (1) as Goulard and Voltz (1993) for the kriging predictor of $\boldsymbol{\chi}(s_0)$, the concept of best linear unbiased predictor can be approached in a different way. In multivariable geostatistics (Myers 1982; Ver Hoef and Cressie 1993; Wackernagel 1995, 1998), the BLUP of p variables on a non-data location s_0 is obtained by minimizing $\sigma^2_{s_0} = \sum_{j=1}^{p} \mathbb{V}\text{ar}\left(\hat{Z}_j(s_0) - Z_j(s_0)\right)$, that is, minimizing the trace of the mean-squared prediction error matrix (Myers 1982). We thus adopt here an extension of the minimization criterion given by Myers (1982) to the functional context, by replacing the summation by an integral. Consequently, in order to find

the BLUP, the n parameters λ_i in the kriging predictor (1) of $\boldsymbol{\chi}(s_0)$ are given by the solution of the following optimization problem

$$\min_{\lambda_1,\ldots,\lambda_n} \int_T \mathbb{V}\mathrm{ar}\left(\hat{\chi}_t(s_0) - \chi_t(s_0)\right) dt, \ \text{s.t.} \sum_{i=1}^{n} \lambda_i = 1, \tag{2}$$

where $\sum_{i=1}^{n} \lambda_i = 1$ is the unbiasedness constraint. Observe that unbiasedness and Fubini's Theorem imply that

$$\begin{aligned}\int_T \mathbb{V}\mathrm{ar}\left(\hat{\chi}_t(s_0) - \chi_t(s_0)\right) dt &= \int_T \mathbb{E}\left[\left(\hat{\chi}_t(s_0) - \chi_t(s_0)\right)^2\right] dt \\ &= \mathbb{E}\left[\int_T \left(\hat{\chi}_t(s_0) - \chi_t(s_0)\right)^2 dt\right].\end{aligned}$$

Therefore the objective function in (2) coincides with that proposed by Goulard and Voltz (1993) in the definition of CKP.

Solving the optimization problem in (2), and taking into account that $\gamma_t(h) = C_t(0) - C_t(h)$, as previously stated, the optimal weights λ_i can be found as the solution of the linear system

$$\begin{aligned}&\begin{pmatrix} \int_T \gamma_t(\|s_1 - s_1\|)dt & \cdots & \int_T \gamma_t(\|s_1 - s_n\|)dt & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \int_T \gamma_t(\|s_n - s_1\|)dt & \cdots & \int_T \gamma_t(\|s_n - s_n\|)dt & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} \\ &= \begin{pmatrix} \int_T \gamma_t(\|s_0 - s_1\|)dt \\ \vdots \\ \int_T \gamma_t(\|s_0 - s_n\|)dt \\ 1 \end{pmatrix}.\end{aligned} \tag{3}$$

Note that the objective function in Eq. (2) involves variances of differences between the predictor $\hat{\boldsymbol{\chi}}(s_0)$ and the target $\boldsymbol{\chi}(s_0)$ evaluated always at the same value $t \in T$.

We name the function $\gamma(h) = \int_T \gamma_t(h)dt$ *trace-variogram*. Details on its estimation can be found in Sect. 3. Simple algebra derivation shows that the prediction trace-variance of the functional ordinary kriging based on the trace-variogram is given by

$$\begin{aligned}\sigma^2_{s_0} &= \int_T \mathbb{V}\mathrm{ar}\left(\hat{\chi}_t(s_0) - \chi_t(s_0)\right) dt = \sum_{i=1}^{n} \lambda_i \int_T \gamma_t(\|s_i - s_0\|)dt - \mu \\ &= \sum_{i=1}^{n} \lambda_i \gamma(\|s_i - s_0\|) - \mu.\end{aligned} \tag{4}$$

The parameter defined in Eq. (4) should be considered as a global uncertainty measure, in the sense that it is an integrated version of the classical point-wise prediction variance of ordinary kriging. Under a specified trace-variogram model, we can use estimations of this parameter to identify those zones presenting greater uncertainty on the predictions.

3 Estimating the trace-variogram

To solve the system in expression (3), an estimator of the trace-variogram is needed. As we are assuming that the random process is weakly stationary, and thus $\mathbb{V}\text{ar}(\chi_t(s_i) - \chi_t(s_j)) = \mathbb{E}\left[\left(\chi_t(s_i) - \chi_t(s_j)\right)^2\right] = \gamma_t(h)$, using Fubini's theorem, we have that

$$\gamma(h) = \frac{1}{2}\mathbb{E}\left[\int_T (\chi_t(s_i) - \chi_t(s_j))^2 dt\right], \quad \text{for } s_i, s_j \in D \text{ with } h = \|s_i - s_j\|.$$

The following is a modification of the classical moment estimator

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|}\sum_{i,j\in N(h)}\int_T (\chi_t(s_i) - \chi_t(s_j))^2 dt, \tag{5}$$

where $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\}$, and $|N(h)|$ is the number of distinct elements in $N(h)$. For irregularly spaced data there are generally not enough observations separated by exactly h. Then $N(h)$ is modified to $\{(s_i, s_j) : \|s_i - s_j\| \in (h-\varepsilon, h+\varepsilon)\}$, with $\varepsilon > 0$ being a small value.

Once we have estimated the trace-variogram for a sequence of K values h_k, we fit a parametric model $\gamma_\alpha(h)$ (any of the classical and widely used models such as spherical, Gaussian, exponential or Matérn could well be used) to the points $\left(h_k, \hat{\gamma}(h_k)\right)$, $k = 1, \ldots, K$ by ordinary least squares (OLS) or weighted least squares (WLS) (see, for instance, Cressie 1993). The fitted variogram model is used to obtain the kriging coefficients and to compute the trace variance using Eq. (4)

4 A non-parametric approach

In this section we consider the case where each data curve is known at a large number of points but fitting a function to each is not satisfactory. In this case the second method of Goulard and Voltz (1993), CFP, is computationally expensive. When the data function can be fitted with a parametric model, the alternative FCP method of Goulard and Voltz (1993) is feasible, because in this case the cokriging step involves p-dimensional vectors, where p is the number of parameters in the parametric model. However, fitting a function to a data curve is not always satisfactory, and an alternative non-parametric fitting could do the job.

So we limit ourselves to using the curve kriging predictor (CKP) proposed by Goulard and Voltz (1993), introduced in Sect. 2. In particular, we need to compute

integrals of the form $\int_T (\chi_t(s_i) - \chi_t(s_j))^2 dt$. If a parametric model is available, then Goulard and Voltz (1993) replace this integral by $\int_T (\chi_t(s_i; \hat{\theta}_{s_i}) - \chi_t(s_j; \hat{\theta}_{s_j}))^2 dt$. In this paper we replace the fitted parametric function $\chi_t(\cdot; \hat{\theta}_\cdot)$ by its non-parametric counterpart. We use cubic B-splines for smoothing but other nonparametric methods might be used instead. Good references for these methods can be found in Green and Silverman (1994), Ramsay and Silverman (2005) or Wasserman (2006).

Given a set of L interior points of $T = [a, b]$, say $a < \tau_1 < \cdots < \tau_L < b$ (we also define $\tau_0 = a$ and $\tau_{L+1} = b$), a cubic spline S is a function defined on T such that S is a cubic polynomial in $[\tau_{l-1}, \tau_l]$, $l = 1, \ldots, L+1$ and S has a continuous second derivative in T (in particular, S, S' and S'' are continuous at all τ_l). The points $\tau_l, l = 0, \ldots, L+1$, are called *knots*. It can be shown that the set of cubic splines with knots $\tau_l, l = 0, \ldots, L+1$, is a vector space with dimension $L+4$. Cubic splines commonly used to approximate unknown functions. In fact any set of points $(\tau_l, f_l), l = 0, \ldots, L+1$, can be interpolated using a cubic spline.

A useful system of basis functions for the set of cubic splines with knots $\tau_l, l = 0, \ldots, L+1$ is the set of *cubic B-splines* $B_k(t)$, $k = 1, \ldots, L+4$. The cubic B-splines are nonzero in no more than 4 inter-knots intervals, see Ramsay and Silverman (2005, Sect. 3.5) for details on the shape of the cubic B-splines and other properties. Any cubic spline S with knots $\tau_l, l = 0, \ldots, L+1$ can be written in the form

$$S(t) = \sum_{k=1}^{L+4} c_k B_k(t) = \mathbf{c}^T \mathbf{B}(t),$$

where $\mathbf{c}$ is the vector of coefficients c_k and $\mathbf{B}(t)$ is a $(L+4)$-dimensional function with components $B_k(t)$.

In the most general case, assume that a function f defined on a temporal domain T has been observed at points $t_1, \ldots, t_M \in T$, possibly with errors: the values $y_j = f(t_j) + \varepsilon_j$, where ε_j are independent random variables with zero mean. The coefficients in the cubic spline are obtained as the solution of

$$\min_{\mathbf{c} \in \mathbf{R}^{L+4}} \sum_{j=1}^{M} (y_j - S(t_j))^2 + \eta \int_T (S''(t))^2 dt. \tag{6}$$

The parameter η is a *smoothing parameter* that controls the trade-off between the fit to the observed data and the smoothness of the approximating cubic spline. If η goes to infinity, the spline solution of (6) would approach the least square regression line. η is also known as *roughness penalty parameter*. In our approach there are additional parameters that affect the smoothing properties of the cubic spline approximation: the number and location of the interior knots. In the present work we are always using evenly spaced interior knots in T. Therefore we work with two smoothing parameters: η and L. The degree of smoothing is an increasing function of η and a decreasing function of L.

Note that each of the estimators of the function f depends on one or more smoothing parameters. Indeed, the smoothing parameter choice is the most difficult step, and

cross-validation (CV) is one possible way to choose the smoothing parameter. In the case of cubic splines with L interior knots, CV is defined as follows. For $j = 1, \ldots, M$, let $S^{(j)}_{L,\eta}$ be the solution of the problem (6) when the observation (t_j, y_j) is temporarily suppressed, and the parameter values L and η are used to fit the rest of the data. We define

$$CV(L, \eta) = \sum_{j=1}^{M} \left(y_j - S^{(j)}_{L,\eta}(t) \right)^2 .$$

Then the smoothing parameters (L, η) are chosen to minimize $CV(L, \eta)$.

Coming back to the context of spatially correlated functional data, the aim when fitting a cubic spline to an observed function $\boldsymbol{\chi}(s_i)$ is not to predict new values of this particular function, but to predict a whole function $\boldsymbol{\chi}(s_0)$ at a non-data location s_0. Therefore we propose a different way of choosing the smoothing parameters in this context, and we call it *functional cross-validation (FCV)*. This leave-one-out cross-validation method minimizes the function

$$FCV(L, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{M} \left(\chi_{t_j}(s_i) - \tilde{\chi}_{t_j}(s_i)^{(i)} \right)^2 \tag{7}$$

for $L \in [L_{\min}, L_{\max}]$ and $\eta \in [\eta_{\min}, \eta_{\max}]$, and where $\tilde{\chi}_{t_j}(s_i)^{(i)}$ is the prediction on s_i evaluated at t_j for $j = 1, \ldots, M$, by leaving the site s_i temporarily out of the sample. In particular, the minimization procedure is as follows:

1. For $i = 1, \ldots, n$, repeat:
   (a) Temporarily suppress the data at site s_i.
   (b) For $i' \neq i$, fit a cubic spline to $\boldsymbol{\chi}(s'_i)$ using Eq. (6) and smoothing parameters (L, η). Let $\tilde{\boldsymbol{\chi}}(s'_i)$ be the smoothed function.
   (c) Use Eq. (5) to estimate the empirical trace-variogram from the data set $\tilde{\boldsymbol{\chi}}(s'_i)$, $i' = 1, \ldots, n$, $i' \neq i$, and then fit a model for the trace-variogram, as described in Sect. 3.
   (d) Solve the system (3) with the trace-variogram estimated in the previous step and the data set $\tilde{\boldsymbol{\chi}}(s'_i)$, $i' = 1, \ldots, n$, $i' \neq i$, to predict the function at site $s_0 = s_i$. Let $\tilde{\boldsymbol{\chi}}(s_i)^{(i)}$ be the resulting function.
   (e) Compute a measure of distance between $\boldsymbol{\chi}(s_i)$ and $\tilde{\boldsymbol{\chi}}(s_i)^{(i)}$ at the values $t_1, \ldots, t_M$: $\mathrm{SSE}(i) = \sum_{j=1}^{M} \left(\chi_{t_j}(s_i) - \tilde{\chi}_{t_j}(s_i)^{(i)} \right)^2$.
2. Define $\mathrm{FCV}(L, \eta) = \sum_{i=1}^{n} \mathrm{SSE}(i)$.

Then the optimal values (L^*, η^*) are used with the cubic B-splines. Finally fit a model for the trace-variogram and use it to predict functions at non-data locations using the kriging equations in (3).

Some comments are in order. When t_j, $j = 1, \ldots, M$, are equally spaced the quantity $\mathrm{SSE}(i)$ is (up to a multiplicative constant) an approximation to the integral $\int_T (\chi_t(s_i) - \tilde{\chi}_t(s_i)^{(i)})^2 dt$.

The estimation of the empirical trace-variogram using Eq. (5) involves the computation of integrals that, in the case of fitting cubic splines with a common basis of B-splines, can be simplified to give

$$\int_T \left(\tilde{\chi}_t(s_i) - \tilde{\chi}_t(s_j)\right)^2 dt = \int_T \left(\mathbf{c_i}^T\mathbf{B}(t) - \mathbf{c_j}^T\mathbf{B}(t)\right)^2 dt = \int_T \left((\mathbf{c_i} - \mathbf{c_j})^T\mathbf{B}(t)\right)^2 dt$$
$$= (\mathbf{c_i} - \mathbf{c_j})^T \left(\int_T \mathbf{B}(t)\mathbf{B}^T(t)dt\right)(\mathbf{c_i} - \mathbf{c_j})^T$$
$$= (\mathbf{c_i} - \mathbf{c_j})^T\mathbf{W}(\mathbf{c_i} - \mathbf{c_j})^T.$$

The matrix $\mathbf{W}$ depends only on the knots, so it is common for all the sites s_i. Similar reasoning allows to write the penalty term in Eq. (6) as $\mathbf{c_i D c_i}^T$, where the term (l, k) in matrix $\mathbf{D}$ is $\int_T B_l''(t)B_k''(t)dt$, only depending on the knots.

To minimize computational costs we use the same value of L and η for all data locations.

5 Data analysis: penetration resistance curves

Before sowing seeds in a field it is common to measure the soil penetration resistance (Chan et al. 2006). Figure 1 shows 32 data locations (in UTM coordinates) in an experimental plot at the National University of Colombia, together with penetration resistance profiles.

At each data location, penetration resistance (MPa) measurements were made at 334 depths between 0 and 45 cm. The objective is to predict penetration resistance curves at non-data locations. Cokriging (the CFP approach of Goulard and Voltz 1993) or space-time kriging are alternatives to functional kriging. The latter procedure is preferred as estimating and fitting a Linear Coregionalization Model (Wackernagel 1995, 1998) in the two former cases is highly computational time demanding. In addition, functional kriging is easier to use.

The data set consisted of 32 penetration curves with values known at a finite number of points in the range of depths (see Fig. 2, left panel). This plot suggests that there is a certain degree of observational noise, and thus it is not easy to find a parametric model for these curves. So we smooth the observed functions using B-splines basis functions, as proposed in Sect. 4.

Note that penetration values must be non-negative, but the BLUP in (1) does not guarantee that the predicted curve values are always non-negative, even if the observed curves are (this is because some λ_i in the solution to the system of Eq. (3) could be negative). In this case, some kind of (possibly linear) transformation should be used. However, in our case all the predicted curves were non-negative, and therefore we worked with the original penetration values.

The functional cross-validation process described in Sect. 4 was used to find the optimal smoothing parameters: L, the number of interior knots, and η, the roughness penalty parameter. Based on a preliminary exploration of the data set, we set

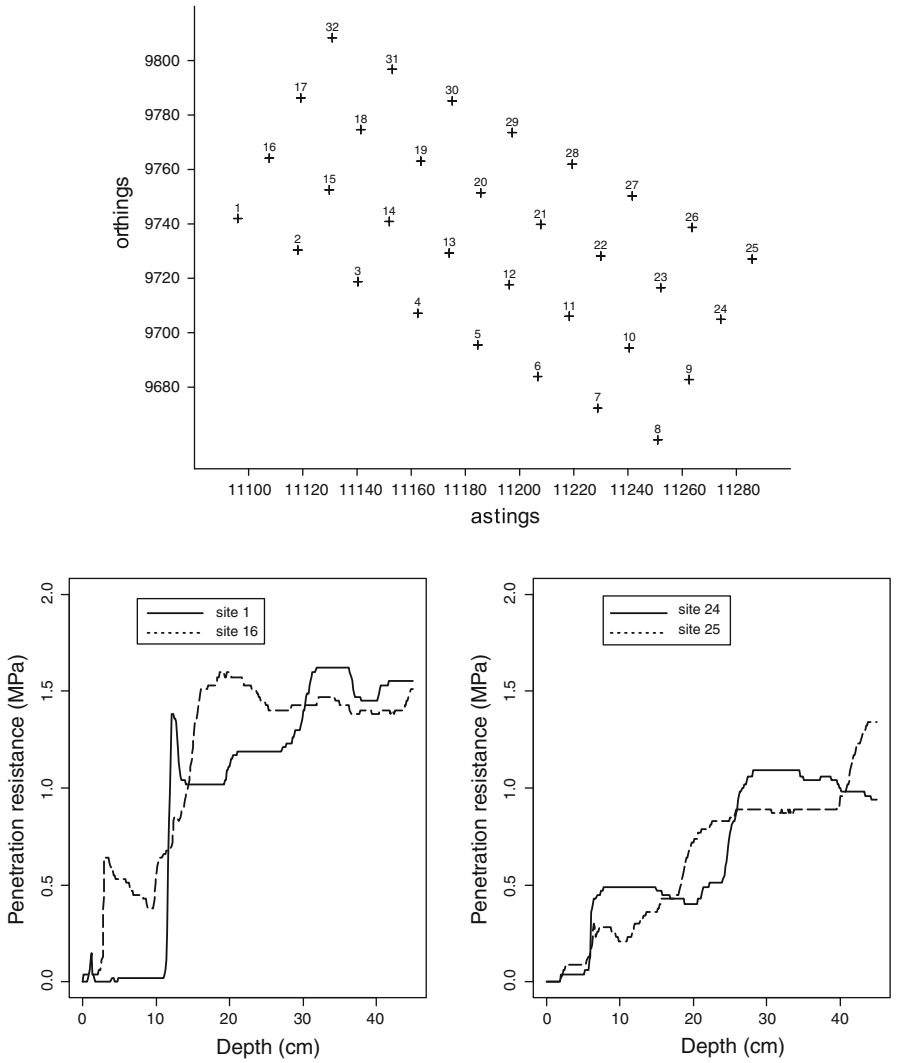

Fig. 1 Data locations (the coordinate system uses UTM) and some observed penetration resistance curves. Data are measured at the Marengo Experimental Station (National University of Colombia) during 2004

$\{6, \ldots, 11\}$ and $\{0, 1, 10, 10^2, 10^3\}$ the range of possible values for L and η, respectively. Figure 3 shows the contour plot of the function $FCV(L, \eta)$, using a logarithmic scale for η. The values $L^* = 10$ and $\eta^* = 0$ minimized the function $FCV(\cdot, \cdot)$. The set of smoothed functions is shown in Fig. 2, right panel.

A spherical model was fitted to the empirical trace-variogram and this was used in step 1(c) of the FCV algorithm. Weighted least squares was used for the fitting, where the squared differences were weighted by the number of points in each lag. The shape of the spherical model was akin to the shape of the empirical trace-variogram, it has been widely used in the literature, and was chosen in terms of its flexibility, simplicity,

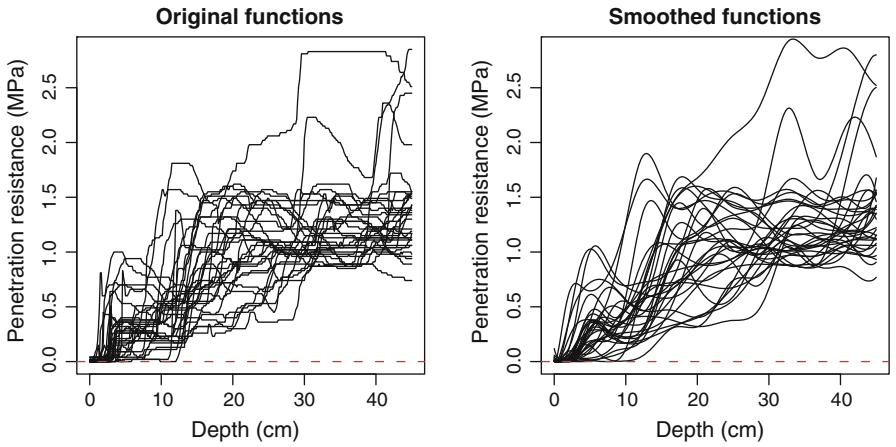

Fig. 2 Set of 32 penetration resistance curves (*left*) and smoothed functions (*right*)

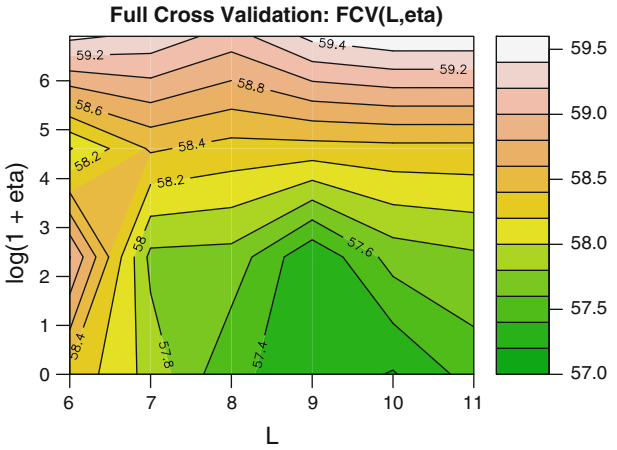

Fig. 3 Contour plot for the function $FCV(L, \eta)$, using a logarithmic scale for η (eta in the graphic)

interpretability of its parameters, and, in particular, it provided the least SSE (sum of squares of errors), and the largest R-square compared to other alternative models, such as exponential or Gaussian. The spherical model has three free parameters: nugget, partial sill (σ^2) and range (ϕ). The fitted trace-variogram, using $L^* = 10$ and $\eta^* = 0$, was $\hat{\gamma}(h) = 1.01 + 2.25(1.5h/39.10 - 0.5(h/39.10)^3)$ for $0 < h \leq 39.10$, and 3.26 for $h > 39.10$.

To illustrate the method, kriging was used to predict the curve at 11179 (easting), 9750 (northing) using UTM coordinates (see Fig. 1). The kriging coefficients λ_i were obtained by solving the system of Eq. (3). Figure 4 (left panel) shows the weights λ_i as a function of the Euclidean distance between data locations s_i and the non-data location s_0. It is clearly seen that the largest weights λ_i correspond to the four sites surrounding s_0. The predicted curve (Fig. 4, thick line in the right panel) indicates that

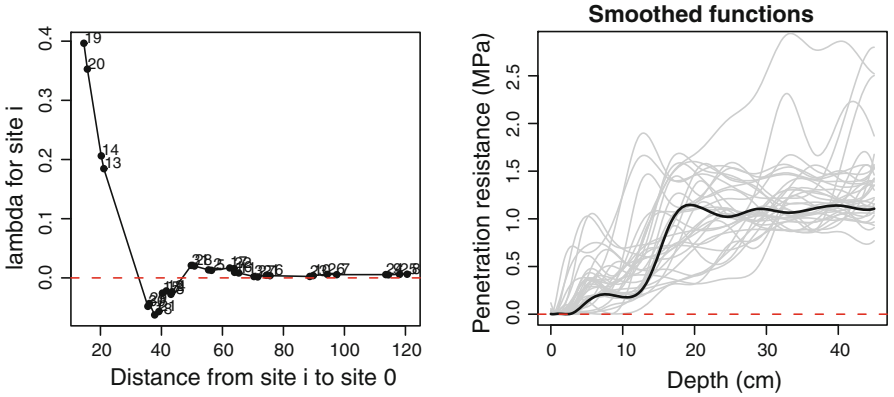

Fig. 4 *Left panel:* Kriging weights λ_i as a function of the Euclidean distance between sites s_i and s_0. *Right panel:* Kriging prediction at a non-data location (*black thick line*) over the set of smoothed curves (*grey lines*)

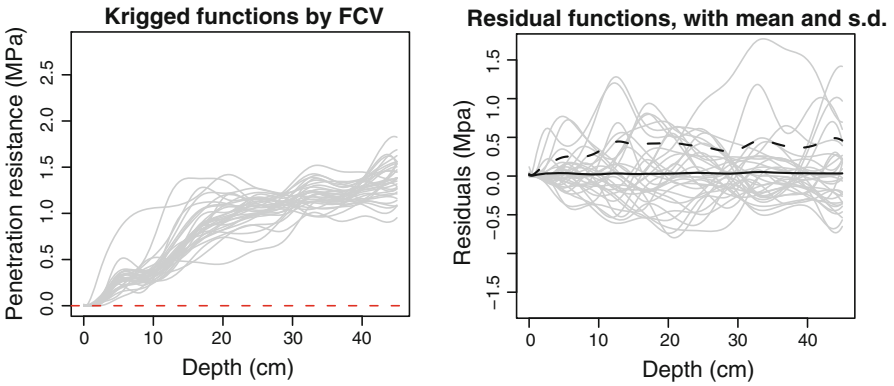

Fig. 5 *Left panel:* Kriging predictions for the 32 original sites based on functional cross-validation. *Right panel:* Functional cross-validation residuals (*grey lines*), residual mean (*black thick line*) and residual standard deviation (*dashed line*)

in this location there is a good soil compaction level, because the predicted penetration resistance is less than 2 MPa, which is considered the critical limit for root growth (Chan et al. 2006).

The functional cross-validation method (FCV) described in Sect. 4, and used to choose the smoothing parameters, is also useful to compare observed and predicted curves, as it defines a measure of distance between these two curves. Indeed, the $SSE(i)$ can be considered an approximation to the mean integrated squared error (MISE) (Myers 1991). Note that when using cross-validation the idea is that the predicted values should be "close" to the observed values (using the leave-one-out technique). However there is not a single best way to quantify "closeness". In the case of numerical valued random functions, there are at least six different statistics (Myers 1991). In this paper we chose FCV.

Figure 5 shows a graphical comparison between observed and predicted curves (using FCV). The predicted curves are smoother and present less variability than the original ones. This was not surprising since kriging is itself a smoothing method (thus the variance decreases), and also because there is a significant high variability amongst penetration resistance values for some particular depth levels (see right panel of Fig. 4). In our case, though noting that several senses of smoothness could be considered, the fitted functions satisfy both of the following conditions: (a) a particular predicted function is smoother than the corresponding observed function, and (b) the variability among the set of predicted functions is less than the variability between the observed functions. A detailed analysis of functional cross-validation residuals indicates that the residual mean is very close to zero (Fig. 5) and the uncertainty on predictions is approximately constant for depths greater than 15 cm. (see residual standard deviation in Fig. 5).

6 Conclusions and discussion

We have enhanced the curve kriging predictor (CKP) proposed by Goulard and Voltz (1993) by introducing a non-parametric smoothing step. In particular, we have introduced functional cross-validation to automatically choose the smoothing parameters. More complex procedures than CKP can be considered by replacing the scalar coefficients λ_i, $i = 1, \ldots, n$ in Eq. (1) by functional coefficients $\lambda_i(t), t \in T$ (Goulard and Voltz 1993 mentioned this possibility but they did not develop it), or even by double indexed functional coefficients $\lambda_i(s, t), s, t \in T$, and using integrals over T as a way to extend the definition of linear combinations. These extensions are parallel to regression models with functional responses (see, Ramsay and Silverman 2005, Chapters 14, 16), and could be considered as extensions of the cokriging predictor (Ver Hoef and Cressie 1993) to the functional context.

In this paper we have used the usual moment variogram estimator together with weighted least squares for fitting the variogram model. Alternative methods of estimating the empirical trace-variogram, for instance, by using robust estimators (Cressie 1993) or kernel-based estimation methods (Yu et al. 2007) could be considered.

Note that soils are structured in layers, and so it is not likely that penetration resistance is a continuous function of soil depth. This can be seen in some of the curves in Figs. 1 and 2 where a number of discontinuities can be observed at some depths. It would seem that wavelets instead of B-splines might better capture the behavior of the curves. As considered in the paper, a basis function system is a set of known functions B_K that are mathematically independent of each other having the property that we can approximate any function arbitrarily well by a linear combination of a sufficiently large number K of these functions (Ramsay and Silverman 2005). Here we perform smoothing using B-splines but any other basis functions (Fourier, wavelets) or non-parametric regression methods could also be used. In particular, the wavelet transform replaces the Fourier transform sinusoidal waves by a family generated by translations and dilations of a window called a wavelet. In contrast to Fourier series, wavelets expansions cope well with discontinuities or rapid changes in behavior. Wavelets are families of orthonormal basis functions that can be used to represent other functions parsimoniously. Giraldo (2009) considers the use of wavelets and par-

ticularly the Haar wavelet transform for analyzing the penetration resistance data set. In particular, a Haar wavelet 3-level transform to the square root of penetration values is fitted. Giraldo (2009) compares the prediction on an unvisited site, the cross-validation predictions, and the cross-validation residuals obtained with both the Haar wavelet transform and B-splines and concluded that both smoothing methods show similar performances, although the greater smoothness degree provided by B-splines reflected in slightly better predictions.

In the spirit of the open source philosophy, the authors make available their source code through the web site http://www.docentes.unal.edu.co/rgiraldoh/, where a suite of *R* functions to perform spatial prediction for functional data can be found. Our code makes use of the *R* libraries *fda* and *geoR* to develop particular functions such as

```
okfd=function(coord, data, argvals=seq(0,1,len=dim(data)[1]),
argnames=c("argument","sites","values"),new.coord=NULL,nbasis=
max(50,dim(data)[1]),lambda =0,max.dist.variogram=NULL,nugget.fix=NULL)
```

or

```
okfd.CV=function(coord, data, argvals=seq(0,1,len=dim(data)[1]),
argnames=c("argument","sites","values"),array.nbasis=max(50, dim
(data)[1]), array.lambda = 0, max.dist.variogram=NULL,nugget. fix=NULL)
```

to perform spatial prediction and cross-validation, respectively.

Acknowledgments Research partially supported by the Spanish Ministry of Education and Science, and FEDER through grants MTM2007-62923 and MTM2009-13985-C02-01. We also would like to acknowledge COLCIENCIAS for supporting the project "*Manejo sostenible del suelo en sistemas de cultivos transitorios bajo la concepcion de agricultura de precision*", led by the research group "Desarrollo Sostenible y Gestion Ambiental" from National University of Colombia, Bogota. We thank the Editor and two anonymous referees whose comments and suggestions have helped to improve earlier versions of the manuscript. Our colleagues Porcu and Gregori are acknowledged with thanks.

References

Berg C, Forst G (1975) Potential theory on locally compact abelian groups. Ergebnisse der Mathematik und ihrer Grenzgebiete. Band 87. Springer, Berlin

Cardot H, Ferraty F, Sarda P (1999) Functional linear model. Stat Probab Lett 45:11–22

Chan K, Oates A, Swan A, Hayes R, Dear B, Peoples M (2006) Agronomic consequences of tractor wheel compaction on a clay soil. Soil Tillage Res 89:13–21

Cressie NAC (1990) Reply to G. Wahba's Letter to the Editor. The American Statistician 44:256–258

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

Deville J (1974) Méthodes statistiques et numeriques de l' analyse harmonique. Ann. Insee 15:3–104

Ferraty F, Vieu P (2003) Curves discrimination. A non parametric functional approaches. Comput Stat Data Anal 44:161–173

Ferraty F, Vieu P (2006) Non parametric functional data analysis. Theory and practice. Springer, New York

Friman O, Borga M, Lundberg P, Knutsson H (2004) Detection and detrending in fMRI data analysis. Neuroimage 22:645–655

Giraldo R (2009) Geostatistical Analysis of Functional Data. PhD Thesis, Universitat Politecnica de Barcelona, Spain

Goulard M, Voltz M (1993) Geostatistical interpolation of curves: a case study in soil science. In: Soares A (ed) Geostatistics Tróia '92. Kluwer, Doedretcht, pp 805–816

Green PJ, Silverman BW (1994) Nonparametric regression and generalized linear models: a roughness penalty approach. Chapman and Hall, London

Mancera J, Vidal L (1994) Florecimiento de microalgas relacionado con mortandad masiva de peces en el complejo lagunar Ciénaga Grande de Santa Marta, Caribe colombiano. Anales del Instituto de Investigaciones Marinas 23:103–117

Myers D (1982) Matrix formulation of co-kriging. Mathematical Geology 14:249–257
Myers D (1991) Interpolation and estimation with spatially located data. Chemometr Intel Lab Syst 11:209–228
Ramsay J, Silverman B (2005) Functional data analysis. 2. Springer, New York
Silverman B (1995) Incorporating parametric effects into functional principal components. J R Stat Soc Ser B 57:673–689
Tolosana-Delgado R (2005) Geostatistics for constrained variables. Positive data, compositions and probabilities. Applications to environmental hazard monitoring. PhD Thesis, Universitat de Girona, Spain
Vandenberghe V, Goethals PLM, van Griensven A, Meirlaen J, De Pauw N, vanrolleghem P, Bauwens W (2005) Application of automated measurement stations for continuous water quality monitoring of the dender river in Flanders, Belgium. Environ Monit Assess 108:85–98
Ver Hoef J, Cressie NAC (1993) Multivariable spatial prediction. Math Geol 25:219–240
Voltz M, Goulard M (1994) Spatial interpolation of soil moisture retention curves. Geoderma 62:109–123
Wackernagel H (1995) Multivariable geostatistics: an introduction with applications. Springer, Berlin
Wackernagel H (1998) Principal components analysis for autocorrelated data: a geostatistical perspective. Technical Report 22/98/G, Centre de Geostatistique-Ecole des Mines de Paris
Wahba G (1990) Letter to the Editor. Am Stat 44:255–256
Wasserman L (2006) All of nonparametric statistics. Springer, Berlin
Yu K, Mateu J, Porcu E (2007) A kernel-based method for nonparametric estimations of variograms. Stat Neerlandica 61:173–197

Author Biographies

R. Giraldo graduated in Statistics from the National University of Colombia in 1993. In 2003, he received the MSc degree in Statistics at the same University. He then received the PhD degree in Statistics and Operation Research from the Technical University of Catalonia, Spain in 2009. Since 1998, he has been an Associate Professor at the National University of Colombia. His research activities are focused on environmetrics, geostatistics, and functional data analysis.

P. Delicado graduated in Mathematics from the University Complutense of Madrid, Spain, in 1990. He joined the Department of Statistics at University Carlos III of Madrid as a PhD candidate. He received the PhD degree from Carlos III of Madrid in 1995. From 1995 to 2009, he was an Assistant Professor in the Department of Economics at the University Pompeu Fabra, Barcelona. Since 2000, he has been an Associate Professor in University Politécnica de Catalunya, Barcelona. His research activities are focused on nonparametric statistics, multivariate data analysis, functional data analysis and spatial statistics.

J. Mateu graduated in Mathematical Sciences at University of Valencia (Spain) and fulfilled the PhD at both University of Valencia and University of Lancaster (UK) in 1998. He joined the Department of Mathematics at University Jaume I of Castellon as a lecturer in 1998 and then as an associate professor in 2000. He is actually full professor of Statistics at the Department of Mathematics of University Jaume I of Castellon. His main fields of interest are stochastic processes in wide sense with a particular focus on spatial and space-time geostatistical and point processes.