

Continuous time-varying kriging for spatial prediction of functional data: An environmental application

Ramón Giraldo^{1,2}, Pedro Delicado¹ and Jorge Mateu^{3,*}

¹Universitat Politècnica de Catalunya, Barcelona, Spain.

²Universidad Nacional de Colombia, Bogotá, Colombia.

³Universitat Jaume I, Castellón, Spain.

Abstract

Spatially correlated functional data is present in a wide range of environmental disciplines and, in this context, efficient prediction of curves is a key issue. We present an approach for spatial prediction based on the functional linear point-wise model adapted to the case of spatially correlated curves. First, a smoothing process is applied to the curves by expanding the curves and the functional parameters in terms of a set of Fourier basis functions. The number of basis functions is chosen by cross-validation. Then, the spatial prediction of a curve is obtained as a point-wise linear combination of the smoothed data. The prediction problem is solved by estimating a linear model of coregionalization to set the spatial dependence among the fitted coefficients. We extend an optimization criterion used in multivariable geostatistics to the functional context. The method is illustrated by smoothing and predicting temperature curves measured at 35 Canadian weather stations.

*Corresponding author. E-mail: mateu@mat.uji.es. Fax:+34-964-728429

Keywords: Basis functions; Coregionalization linear model; Cross-validation; Functional linear point-wise model; Ordinary kriging.

1 Introduction

In many fields of environmental sciences such as agronomy, ecology, meteorology or monitoring of contamination and pollution, the observations consist of samples of random functions. For example in meteorology when curves of climatological variables are obtained in weather stations of a country (Ramsay and Silverman, 2005), or when solar radiation is monitored in both space and time over a region, and smoothing methods are used to fit each time series (Bodas-Salcedo et al., 2003). Since the beginning of nineties, Functional Data Analysis (FDA) (Ramsay and Dalzell, 1991) is used to model this type of data. From the FDA point of view, each curve corresponds to one observation, that is, the basic unit of information is the entire observed function rather than a string of numbers (Ramsay and Silverman, 2001). Functional versions for many branches of statistics have been given. Examples of such methods include exploratory analysis (Ramsay and Silverman, 2005), analysis of variance (Cuevas et al., 2004; Delicado, 2007), regression (Cardot et al., 1999, 2007), non-parametric methods (Ferraty and Vieu, 2006) or multivariate techniques (Ferraty and Vieu, 2003). An overview of statistical methods for analyzing functional data are shown in Ramsay and Silverman (2005) and recent developments in this field are given in special issues of several journals (González-Manteiga and Vieu, 2007; Valderrama, 2007).

The standard statistical techniques for modeling functional data are focused on independent functions. However, in several disciplines of applied sciences there exists an increasing interest for modeling correlated functional data: it is the case when samples of functions are observed over a discrete set of time points (*temporally*

correlated functional data) or when these functions are observed in different sites of a region (*spatially correlated functional data*). In these cases the methodologies above mentioned could not be appropriate as they do not incorporate the dependence among functions into the analysis. For this reason some statistical methods for modeling correlated variables such as time series (Box and Jenkins, 1976) or geostatistical analysis (Cressie, 1993) have been adapted to the functional context.

An example of modeling temporally correlated functional data is shown in Ruiz-Medina et al. (2007). These authors consider an autoregressive Hilbertian model of order one to represent the dynamic of a sequence of functional data. For spatially correlated functional data, Yamanishi and Tanaka (2003) develop a regression model that enables studying the relation among variables over time and space combining both geographically weighted regression (Brunsdon et al., 1998) and functional multiple regression (Ramsay and Silverman, 2005). Baladandayuthapani et al. (2008) show an alternative for analyzing an experimental design with a spatially correlated functional response. They both use a hierarchical model and a Bayesian approach. Contributions of Yamanishi and Tanaka (2003) and Baladandayuthapani et al. (2008) give the possibility of including spatial dependence among curves into the standard functional analysis, such as functional multiple regression and functional analysis of variance.

When the objective is to perform spatial prediction of functional data several approaches based on kriging and cokriging predictors have been considered. Goulard and Voltz (1993) is a pioneer work in this context. They propose three geostatistical approaches to predict curves: a curve kriging approach and two multivariate approaches based on cokriging on either discrete data or on coefficients of the parametric models that have been fitted to the observed curves. Giraldo et al. (2007) give a non-parametric approach to solve the first approach considered by Goulard and Voltz (1993). The predictor in the first proposal of Goulard and Voltz (1993) as well as that considered by Giraldo et al. (2007) has the same form as the classical

ordinary kriging predictor but considering curves instead of one-dimensional data, that is, each curve is weighted by an scalar parameter. In this paper we consider the problem of spatial prediction of functional data by weighting each observed curve by a functional parameter. This approach was mentioned in Goulard and Voltz (1993) but was not developed there. The modeling approach we present is a hybrid between ordinary kriging and the functional linear concurrent (*point-wise*) model such as shown in Ramsay and Silverman (2005). We propose a solution based on basis functions. The curves as well as the functional parameters are expanded in terms of a set of basis functions. Thus, the problem turns into estimating the coefficients of these basis functions for each functional parameter. In order to give a solution, we use a linear model of coregionalization for estimating the covariances among coefficients of each curve. An essential step in our proposal is to choose the number of basis functions. We consider here two alternative criteria based on cross-validation analysis and the minimum sum of squared errors (SSE).

The plan of the paper is as follows. Section 2 presents the data set to be analyzed. Section 3 introduces the predictor and the parameter estimation. Application of the proposed methodology to the data set considered is given in Section 4. The paper ends with a brief discussion and suggestions for further research.

2 Data set: Canadian temperature

Spatial prediction of meteorological data is an important input for many types of models including hydrological or those of regeneration, growth, and mortality of forest ecosystem. In particular, the modeling of spatially correlated temperature data is of interest, among others, for predicting microclimate conditions in mountainous terrain, resource management, calibration of satellite sensors or for studying the “greenhouse effect”. Many methods have been developed and used for doing spatial prediction of temperatures. However, at our best knowledge all of these ignore its functional character. Here we use a well-known meteorological data set

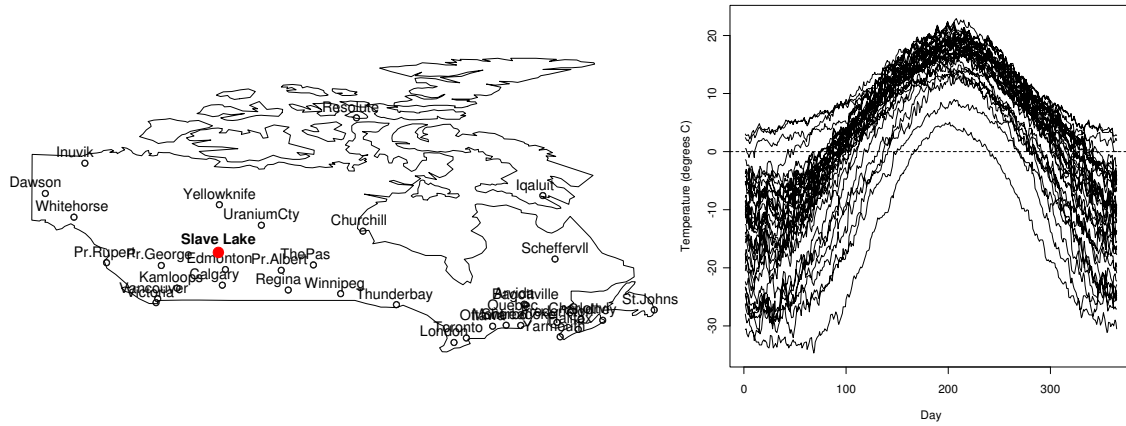


Figure 1: Averages (over 30 years) of daily temperature curves (*right panel*) observed at 35 Canadian weather stations (*left panel*).

in FDA consisting of daily temperature and precipitation measurements recorded at 35 weather stations of Canada (Ramsay and Dalzell, 1991; Ramsay and Silverman, 2005). These authors use Fourier basis functions for constructing curves from discrete data. They apply functional principal components and functional linear models to describe the modes of variability in temperature curves, and for establishing the influence of temperature on precipitation. We specifically use the temperature values of this data set to provide an applied context for our proposal. In particular we analyze information of daily temperature averaged over the years 1960 to 1994 (February 29th combined with February 28th) (Figure 1, right panel). The data for each station were obtained from Ramsay and Silverman's home page (<http://www.functionaldata.org/>). The geographical coordinates of weather stations (Figure 1, left panel) were obtained from the Meteorological Service of Canada (<http://www.climate.weatheroffice.ec.gc.ca/climateData/>).

3 Point-wise kriging for curves

In this section we introduce some notation and assumptions, and present the modeling scheme including the predictor proposed and the minimization criterion considered. We also provide a method for estimating the parameters of the model.

3.1 Notation and assumptions

Let $\{\chi_s(t), t \in T, s \in D \subset \mathbf{R}^d\}$ be a random function defined on some compact set T of \mathbf{R} . Assume we observe a sample of curves $\chi_{s_i}(t)$, for $t \in T$ and $s_i \in D, i = 1, \dots, n$. It is usually assumed that these curves belong to a separable Hilbert space H of square integrable functions defined on T . We assume for each $t \in T$ that we have a second-order stationary and isotropic random process, that is, the mean and variance functions are constant and the covariance depends only on the distance among sampling sites. Formally, we assume that:

- $E(\chi_s(t)) = m(t)$, for all $t \in T, s \in D$.
- $\text{Cov}(\chi_{s_i}(t), \chi_{s_j}(u)) = C(h; t, u)$, $s_i, s_j \in D, t, u \in T, h = \|s_i - s_j\|$. If $t = u$, $\text{Cov}(\chi_{s_i}(t), \chi_{s_j}(t)) = C(h; t)$.
- $\frac{1}{2}\text{V}(\chi_{s_i}(t) - \chi_{s_j}(u)) = \gamma(h; t, u)$, $s_i, s_j \in D, t, u \in T, h = \|s_i - s_j\|$. If $t = u$, $\frac{1}{2}\text{V}(\chi_{s_i}(t) - \chi_{s_j}(t)) = \gamma(h; t)$.

The function $\gamma(h; t)$, as a function of h , is called variogram of $\chi(t)$. We propose to use a family of point-wise linear predictors for $\chi_{s_0}(t)$, $t \in T$, given by

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i(t) \chi_{s_i}(t), \quad \lambda_1(t), \dots, \lambda_n(t) : T \rightarrow \mathbf{R}, \quad (1)$$

that was previously mentioned in Goulard and Voltz (1993) without further development. For each $t \in T$, the predictor (1) has the same expression as an ordinary kriging predictor. In the rest of the paper this predictor is called point-wise linear predictor for functional data (PWKFD). This modeling approach is

coherent with the functional linear concurrent model (FLCM) (Hastie and Tibshirani, 1993; Ramsay and Silverman, 2005) in which the influence of each covariate on the response is *simultaneous* or *point-wise*. FLCM is defined as $Y(t) = \alpha(t) + \beta_1(t)X_1(t) + \cdots + \beta_q(t)X_q(t) + \epsilon(t)$. In this model the response $Y(t)$ and each covariate $X_j(t), j = 1, \dots, q$, are functions of the same argument and $X_j(t)$ only influences $Y(t)$ through its value at time t (Ramsay and Silverman, 2005). Estimation of functional parameters $\alpha(t), \beta_j(t), j = 1, \dots, q$, is carried out by solving $\text{Min}_{\alpha(\cdot), \dots, \beta_q(\cdot)} E \|\hat{\mathbf{Y}}(t) - \mathbf{Y}(t)\|^2$ (Ramsay and Silverman, 2005). In our context the covariates are the observed curves at n sites of a region and the functional response is an unobserved function on an unsampled location. Consequently, our optimization problem is $\text{Min}_{\lambda_1(\cdot), \dots, \lambda_n(\cdot)} E \|\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\|^2$ or equivalently, by using Fubini's Theorem,

$$\text{Min}_{\lambda_1(\cdot), \dots, \lambda_n(\cdot)} \int_T E (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2 dt.$$

If we consider the stationarity assumptions above given, the problem becomes

$$\text{Min}_{\lambda_1(\cdot), \dots, \lambda_n(\cdot)} \int_T V (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) dt. \quad (2)$$

In a classical univariate geostatistical setting we assume that the observations are realizations of a random field $\{Z(s) : s \in D, D \in \mathbf{R}^d\}$. The kriging predictor is defined as $\sum_{i=1}^n \lambda_i Z(s_i)$, and the best linear unbiased predictor (BLUP) is obtained by minimizing $\sigma_{s_0}^2 = V(\hat{Z}(s_0) - Z(s_0))$ subject to $\sum_{i=1}^n \lambda_i = 1$. On the other hand in multivariable geostatistics (Myers, 1982; Ver Hoef and Cressie, 1993; Wackernagel, 1995) the data consist of $\{\mathbf{Z}(s_1), \dots, \mathbf{Z}(s_n)\}$, that is, we have observations of a spatial vector-valued process $\{\mathbf{Z}(s) : s \in D\}$, where $\mathbf{Z}(s) \in \mathbf{R}^m$ and $D \in \mathbf{R}^d$. In this context $V(\hat{\mathbf{Z}}(s_0) - \mathbf{Z}(s_0))$ is a matrix, and the BLUP of m variables on an unsampled location s_0 can be obtained by minimizing $\sigma_{s_0}^2 = \sum_{i=1}^m V(\hat{Z}_i(s_0) - Z_i(s_0))$ subject to constraints that guarantee unbiasedness conditions, that is, minimizing the trace of the mean-squared prediction error matrix subject to some restrictions given by the unbiasedness condition (Myers, 1982). The optimization problem given in (2)

is an extension of the minimization criterion given by Myers (1982) to the functional context, by replacing the summation by an integral and the random vectors $[Z_1(s_0), \dots, Z_m(s_0)]$ and $[\hat{Z}_1(s_0), \dots, \hat{Z}_m(s_0)]$ by the functional variables $\boldsymbol{\chi}(t)$ and $\hat{\boldsymbol{\chi}}(t)$, respectively, with $t \in T$. The predictor (1) is unbiased if $E(\hat{\boldsymbol{\chi}}_{s_0}(t)) = m(t)$, for all $t \in T$, that is, if $\sum_{i=1}^n \lambda_i(t) = 1$ for all $t \in T$. Consequently, in order to find the BLUP, the n functional parameters in the predictor proposed are given by the solution of the following optimization problem

$$\underset{\lambda_1(\cdot), \dots, \lambda_n(\cdot)}{\text{Min}} \int_T V(\hat{\boldsymbol{\chi}}_{s_0}(t) - \boldsymbol{\chi}_{s_0}(t)) dt, \text{ s.t. } \sum_{i=1}^n \lambda_i(t) = 1, \text{ for all } t \in T. \quad (3)$$

3.2 A solution based on basis functions

We assume that each observed function can be expressed in terms of K basis functions by

$$\boldsymbol{\chi}_{s_i}(t) = \sum_{l=1}^K a_{il} B_l(t) = \mathbf{a}_i^T \mathbf{B}(t), \quad i = 1, \dots, n. \quad (4)$$

Taking into account that $\boldsymbol{\chi}_{s_i}(t), i = 1, \dots, n$, are random functions with spatial dependence, we assume that the matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nK} \end{pmatrix} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_K)_{(n \times K)}$$

forms a K multivariable random field with $E(\boldsymbol{\alpha}_i) = \mathbf{v}_i$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1K} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1} & \Sigma_{K2} & \cdots & \Sigma_{KK} \end{pmatrix}_{(K \times n) \times (K \times n)} \quad (5)$$

where $\Sigma_{ij} = C(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)_{(n \times n)}$. The coefficients a_{ij} are assumed a realization of the spatial random field $\boldsymbol{\alpha}_j$, $j = 1, \dots, K$. We propose to use multivariable geostatistics (Wackernagel, 1995) and specifically a linear model of coregionalization (LMC) for estimating the matrix (5). In order to establish the unbiasedness condition and for carrying out the parameter estimation in (1) we further expand each functional parameter $\lambda_i(t)$ by

$$\lambda_i(t) = \sum_{l=1}^K b_{il} B_l(t) = \mathbf{b}_i^T \mathbf{B}(t). \quad (6)$$

Therefore, using (4) and (6) the predictor in equation (1) is given by

$$\begin{aligned} \hat{\chi}_{s_0}(t) &= \sum_{i=1}^n \mathbf{b}_i^T \mathbf{B}(t) \mathbf{a}_i^T \mathbf{B}(t) \\ &= \sum_{i=1}^n \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{a}_i. \end{aligned} \quad (7)$$

Using (6) and by expanding the constant function 1 by means of $\sum_{l=1}^K c_l B_l(t) = \mathbf{c}^T \mathbf{B}(t) = 1$, the unbiasedness constraint can be expressed as

$$\sum_{i=1}^n \mathbf{b}_i^T \mathbf{B}(t) = \mathbf{c}^T \mathbf{B}(t), \quad \forall t, \Leftrightarrow \sum_{i=1}^n \mathbf{b}_i = \mathbf{c},$$

or more specifically by

$$\sum_{i=1}^n b_{i1} = c_1, \dots, \sum_{i=1}^n b_{iK} = c_K. \quad (8)$$

Developing the variance in the objective function (3) we have

$$\begin{aligned}
V(\hat{\boldsymbol{\chi}}_{s_0}(t) - \boldsymbol{\chi}_{s_0}(t)) &= V(\hat{\boldsymbol{\chi}}_{s_0}(t)) + V(\boldsymbol{\chi}_{s_0}(t)) - 2C(\hat{\boldsymbol{\chi}}_{s_0}(t), \boldsymbol{\chi}_{s_0}(t)) \\
&= V\left(\sum_{i=1}^n \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{a}_i\right) + \mathbf{B}^T(t) V(\mathbf{a}_0) \mathbf{B}(t) \\
&\quad - 2 \sum_{i=1}^n \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) C(\mathbf{a}_i, \mathbf{a}_0) \mathbf{B}(t) \\
&= \sum_{i=1}^n \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) V(\mathbf{a}_i) \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{b}_i \\
&\quad + 2 \sum_{i < j} \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) C(\mathbf{a}_i, \mathbf{a}_j) \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{b}_j \\
&\quad + \mathbf{B}^T(t) V(\mathbf{a}_0) \mathbf{B}(t) \\
&\quad - 2 \sum_{i=1}^n \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) C(\mathbf{a}_0, \mathbf{a}_i) \mathbf{B}(t). \tag{9}
\end{aligned}$$

In equation (9), for $i < j$, $i, j = 0, 1, \dots, n$, we have

$$V(\mathbf{a}_i) = \begin{pmatrix} \text{var}(a_{i1}) & \text{cov}(a_{i1}, a_{i2}) & \cdots & \text{cov}(a_{i1}, a_{iK}) \\ \text{cov}(a_{i2}, a_{i1}) & \text{var}(a_{i2}) & \cdots & \text{cov}(a_{i2}, a_{iK}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(a_{iK}, a_{i1}) & \text{cov}(a_{iK}, a_{i2}) & \cdots & \text{var}(a_{iK}) \end{pmatrix}_{(K \times K)}$$

and

$$C(\mathbf{a}_i, \mathbf{a}_j) = \begin{pmatrix} \text{cov}(a_{i1}, a_{j1}) & \text{cov}(a_{i1}, a_{j2}) & \cdots & \text{cov}(a_{i1}, a_{jK}) \\ \text{cov}(a_{i2}, a_{j1}) & \text{cov}(a_{i2}, a_{j2}) & \cdots & \text{cov}(a_{i2}, a_{jK}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(a_{iK}, a_{j1}) & \text{cov}(a_{iK}, a_{j2}) & \cdots & \text{cov}(a_{iK}, a_{jK}) \end{pmatrix}_{(K \times K)}$$

If we define

$$\begin{aligned}
\mathbf{Q}_i &= \int_T (\mathbf{B}(t) \mathbf{B}^T(t) V(\mathbf{a}_i) \mathbf{B}(t) \mathbf{B}^T(t)) dt, \\
\mathbf{Q}_{ij} &= \int_T (\mathbf{B}(t) \mathbf{B}^T(t) C(\mathbf{a}_i, \mathbf{a}_j) \mathbf{B}(t) \mathbf{B}^T(t)) dt, \\
\mathbf{D} &= \int_T \mathbf{B}^T(t) V(\mathbf{a}_0) \mathbf{B}(t) dt,
\end{aligned}$$

and

$$\mathbf{J}_i = \int_T (\mathbf{B}(t)\mathbf{B}^T(t)C(\mathbf{a}_0, \mathbf{a}_i)\mathbf{B}(t)) dt,$$

and by considering K Lagrange multipliers $\mathbf{m}^T = (m_1, \dots, m_K)$, the objective function (3) can be expressed as

$$\min_{\mathbf{b}_1, \dots, \mathbf{b}_n, \mathbf{m}} \sum_{i=1}^n \mathbf{b}_i^T \mathbf{Q}_i \mathbf{b}_i + 2 \sum_{i < j} \mathbf{b}_i^T \mathbf{Q}_{ij} \mathbf{b}_j + \mathbf{D} - 2 \sum_{i=1}^n \mathbf{b}_i^T \mathbf{J}_i + 2\mathbf{m}^T \left(\sum_{i=1}^n \mathbf{b}_i - \mathbf{c} \right). \quad (10)$$

Taking $\boldsymbol{\beta} = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T, \mathbf{m}^T)^T_{(K(n+1) \times 1)}$, the expression (10) is given by

$$\text{Min}_{\boldsymbol{\beta}} \boldsymbol{\beta}^T \mathbf{Q} \boldsymbol{\beta} + \mathbf{D} - 2\boldsymbol{\beta}^T \mathbf{J} \quad (11)$$

where

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_{12} & \cdots & \mathbf{Q}_{1n} & \mathbf{I} \\ \mathbf{Q}_{21} & \mathbf{Q}_2 & \cdots & \mathbf{Q}_{2n} & \mathbf{I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{Q}_{n1} & \mathbf{Q}_{n2} & \cdots & \mathbf{Q}_n & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix}, \mathbf{J} = \begin{pmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \\ \vdots \\ \mathbf{J}_n \\ \mathbf{c} \end{pmatrix}. \quad (12)$$

The identity matrix in (12) is of order K . Minimizing equation (11) with respect to $\boldsymbol{\beta}$ we obtain

$$2\mathbf{Q}\boldsymbol{\beta} - 2\mathbf{J} = \mathbf{0} \Rightarrow \mathbf{Q}\boldsymbol{\beta} = \mathbf{J} \Rightarrow \hat{\boldsymbol{\beta}} = \mathbf{Q}^{-1}\mathbf{J}. \quad (13)$$

In practice, we start estimating both a LMC for the multivariable random field $\mathbf{A} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_K)$ and the matrix in equation (5). Subsequently, we can calculate the matrices \mathbf{Q} and \mathbf{J} in equation (12). Replacing these matrices in equation (13) we can estimate $\mathbf{b}_i, i = 1, \dots, n$ and consequently the functional parameters given in (6). On the other hand, a plug-in estimation of the integrated prediction variance $\sigma_{s_0}^2 = \int_T V(\hat{\boldsymbol{\chi}}_{s_0}(t) - \boldsymbol{\chi}_{s_0}(t)) dt$ is given by

$$\hat{\sigma}_{s_0}^2 = \hat{\boldsymbol{\beta}}^T \mathbf{Q} \hat{\boldsymbol{\beta}} + \mathbf{D} - 2\hat{\boldsymbol{\beta}}^T \mathbf{J}, \quad (14)$$

where the matrix \mathbf{D} is calculated by using $\hat{V}(\mathbf{a}_0)$, which is obtained from the fitted LMC. The integrated prediction variance $\hat{\sigma}_{s_0}^2$ is a measure of the uncertainty in the prediction of a whole curve. Based on the estimated parameters and using equation (9), a point-wise prediction variance function can also be estimated.

3.3 Choosing the number of basis functions

Let us assume that functions $\chi_{s_i}(t)$, $i = 1, \dots, n$, defined on T have been observed at points t_1, \dots, t_M and we want to approximate them through a basis functions. We thus should choose an appropriate order of expansion K , taking into account that on one hand, a large K causes overfitting and on the other hand, if we take K too small we may miss important aspects of the function that we are estimating (Ramsay and Silverman, 2005). Two procedures based on cross-validation analysis are considered in this paper. The first one uses a classical non-parametric cross-validation analysis, and the second one takes into account the cross-validation predictions obtained by PWKFD.

A simple way of establishing an appropriate K is calculating the cross-validation SSE in a classical non-parametric sense. Let $\tilde{\chi}_{s_i}^{(j)}(t_j)$ be the estimated function at t_j by means of equation (4) when the datum $\chi_{s_i}(t_j)$ has been temporarily suppressed from the sample. Then for each K , the cross-validation SSE is calculated by

$$SSE_{NP} = \sum_{i=1}^n SSE_{NP}(i) = \sum_{i=1}^n \sum_{j=1}^M (\tilde{\chi}_{s_i}^{(j)}(t_j) - \chi_{s_i}(t_j))^2. \quad (15)$$

The strategy is to add basis functions until SSE_{NP} decreases substantially. We shall call this method *non-parametric cross-validation*.

In the context of spatially correlated functional data, when we use a basis function to fit a sample of a function $\chi_{s_i}(t)$, the goal is not to predict new values of this particular function, but to predict a whole function $\chi_{s_0}(t)$ at an unvisited site s_0 . Therefore another alternative for choosing the number of basis functions is taking into account the cross-validation prediction errors obtained with a specific predictor

of functional data. We call this procedure *functional cross-validation*. We perform a *leaving-one-out* cross-validation analysis where each data location is removed from the data set, and a smoothed function at this location is predicted using both the remaining smoothed functions and the PWKFD predictor (7). Now for each K we calculate the SSE by

$$SSE_F = \sum_{i=1}^n SSE_F(i) = \sum_{i=1}^n \sum_{j=1}^M (\hat{\chi}_{s_i}^{(i)}(t_j) - \chi_{s_i}(t_j))^2 \quad (16)$$

where $\hat{\chi}_i^{(i)}(t_j)$ is the PWKFD prediction on s_i evaluated at t_j , $j = 1, \dots, M$, by leaving the site s_i temporarily out of the sample. Again, we add basis functions until SSE_F decreases considerably. Once the value of K is selected, we perform spatial prediction by PWKFD at unvisited sites.

4 Application: Spatial prediction of Canadian temperature curves

In this section we illustrate our approach by using the Canadian temperature data set described in Section 2. We initially select an appropriate number of basis functions. In a second stage we perform prediction at an unvisited site using the proposed predictor and describe the results from a practical point of view. Finally, we compare our results with those obtained with the predictor proposed by Giraldo et al. (2007). This comparison is given in terms of SSE_F values resulting from functional cross-validation analysis.

When data are periodic, Fourier basis with an even number of basis functions is the most appropriate choice (Ramsay and Silverman, 2005). A Fourier basis with 65 basis functions is the most frequently used to expand the Canadian temperature data (Ramsay and Silverman, 2005). We use the criteria described in the previous section for exploring the number of functions that could be used for smoothing the observed discrete data set. Although we can expand in terms of a Fourier basis with

an infinite number of sinusoids, we take as limit 365 because this is the number of discrete data for each site in our data set. Frequencies greater than 365 in this case will distort the signal. This is known as the problem of *aliasing* (Lfeachor and Jervis, 1993). Figure 2 shows the relation between K and SSE_{NP} obtained by non-parametric (left panel) and SSE_F obtained by functional (right panel) cross-validation. We can observe that in both cases the SSE values decrease significantly until the number of basis functions is around 50. Then the rate of decreasing is small. For instance, SSE_{NP} decreases 57 % when K is between 5 and 55, whereas this percentage is 69 % for K varying between 5 and 245 (the minimum SSE_{NP} attained). In the same sense SSE_F decreases 5.2 % for K in the interval 5-55, and 6.8 % for K varying between 5 and 365. In summary, SSE values indicate that there is not advantage in using a value of K much larger than 55. As mentioned before, Ramsay and Silverman (2005) use 65 Fourier basis functions for smoothing the data set here considered. Our results also suggest that this number of basis functions could be appropriate. In consequence a pragmatic choice of K is 65. Therefore in the following we assume that the data to be analyzed correspond to the temperatures curves obtained after smoothing each discrete data set by means of a Fourier basis with 65 functions.

Initially PWKFD was used to predict a temperature curve at an unvisited site with coordinates -114.5813 (eastings) and 55.73 (northings). This site corresponds to Slave Lake station (Figure 1). In a first stage of the analysis, and using the library `gstat` of R language (Pebesma, 2004), a LMC was fitted to the multivariable random field $\mathbf{A} = (\alpha_1, \dots, \alpha_{65})$ composed by the coefficients of the Fourier basis used for smoothing each observed sample. We assume stationarity for each process α_j , $j = 1, \dots, 65$. All single (direct) variograms and cross-variograms were modeled as linear combination of nugget and exponential models. Based on the fitted LMC, the matrices \mathbf{Q} and \mathbf{J} given in (12) were estimated and used to solve the system (13) to finally find \mathbf{b}_i and the functional parameters $\lambda_i(t)$, $i = 1, \dots, n$. Figure 3 (left

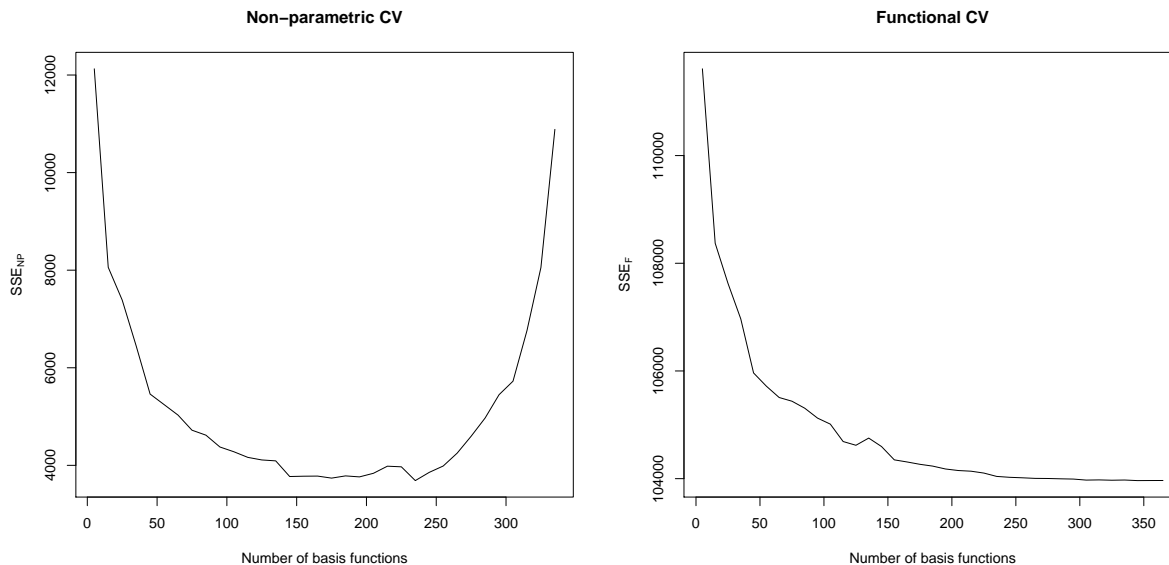


Figure 2: Sum squared errors obtained by non-parametric cross-validation (SSE_{NP} , *left panel*) and functional cross-validation (SSE_F , *right panel*).

panel) shows a plot of the estimated functional parameters. We note that an estimated functional parameter is much greater than the others (functional parameter with values around 0.6). This corresponds to Edmonton, the closest station to Slave Lake (Figure 1). Other stations near to Slave Lake, and therefore with influence on the prediction, are Yellowknife (weights around 0.2, Figure 3), Uranium City and Pr George (values around 0.1, Figure 3). The curves corresponding to the sites furthest from Slave Lake receive almost a weight of zero (Figure 3, left panel). This result is coherent with the kriging philosophy, that is, sites closer to the prediction location have greater influence than others more far apart. The sum of the estimated functional parameters is equal to 1 for all t (Figure 3). With this result we verify graphically that the system (8) guarantees the unbiasedness constraint. A plot of the temperature prediction at Slave Lake is shown also in Figure 3 (right panel). It is remarkable that the predicted curve shows a seasonal behavior similar to the smoothed curves. In addition, predicted values are consistent with real values recorded for this

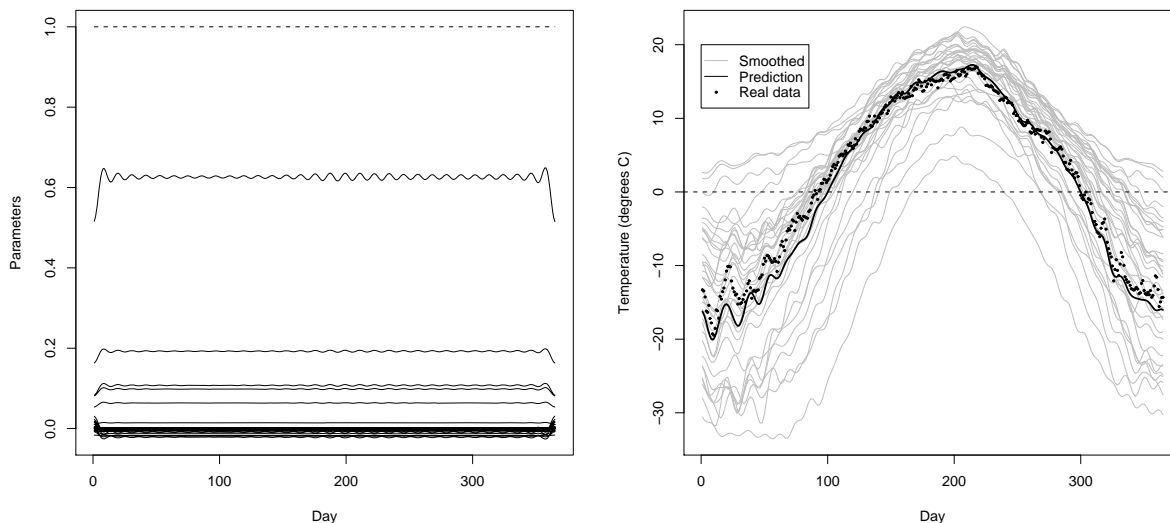


Figure 3: Estimated functional parameters (*left, dark lines*), sum of functional estimated parameters (*left, dotted line*), smoothed temperature curves (*right, clear lines*), temperature prediction function at Slave Lake (*right, dark line*) and real temperature values at Slave Lake (*right, circles*).

weather station (<http://www.climate.weatheroffice.ec.gc.ca/climateData/>).

To verify the goodness-of-fit of the proposed predictor, we use the functional cross-validation results obtained with 65 Fourier basis functions. Each individual smoothed curve $\chi_{s_i}(t)$, $i = 1, \dots, 35$, was temporarily removed, and further predicted from the remaining ones by means of PWKFD. A comparison between predicted and smoothed curves (Figures 3 and 4) shows that the predictions have the same temporal behavior that the smoothed curves. Note also that the latter curves have less variance, in particular in wintertime (where the Canadian weather is more variable, Figure 4). This is not surprising, since on one hand, kriging is itself a smoothing method and on the other hand, the data set includes weather stations with very different temperatures (Figure 1). For instance, Resolute and Iqaluit in the Arctic (Figure 1) as well as Inuvik in the northwest (200 kilometers from Arctic

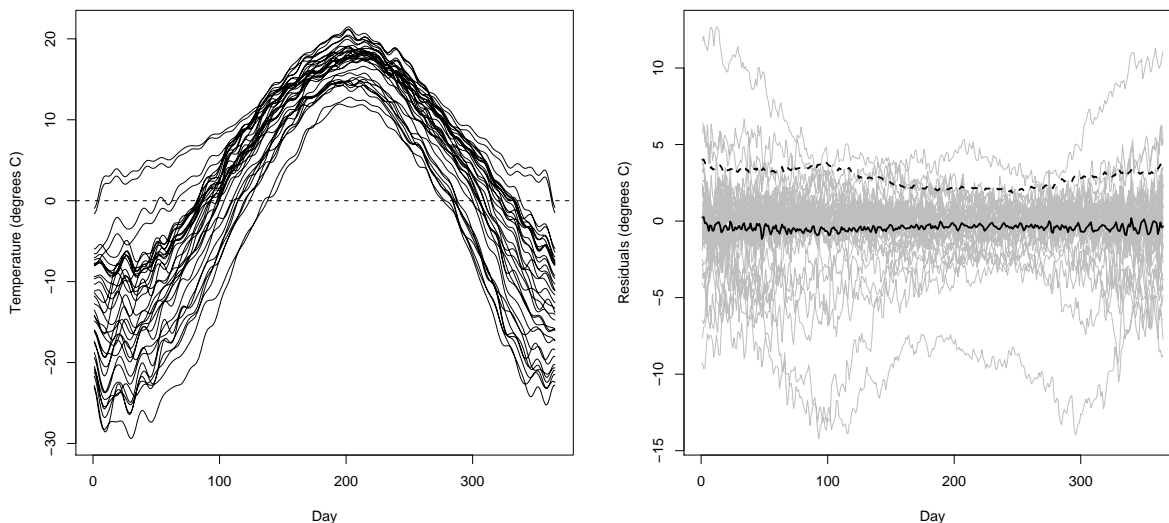


Figure 4: *Right panel:* Point-wise kriging predictions based on cross-validation. *Left panel:* Cross-validation residuals (*clear lines*), residual mean (*dark line*) and residual standard deviation (*dashed line*) for the Canadian temperature data set.

circle) with very cold winters and short summers are significantly different from other stations as Victoria, Vancouver or Prince Rupert in the southwest of the country (Figure 1) which have a temperate climate with mild winters and summers.

Figure 4 (right panel) shows cross-validation residuals. The predictions are plausible in a high proportion of sites (those having residuals around zero). However there are some stations with large positive or negative residual curves. This is due to the fact that the temperature curves at Resolute, Inuvik, Iqaluit, Dawson, Churchill, Prince Rupert and St Johns are not well predicted because they have extreme temperature values, and are spatially very separated from the remaining ones (Figure 1). As a example of this phenomenon, we can compare predictions for Bagottville, Edmonton, Resolute and Prince Rupert stations (Figure 5). We observe a good fit for Bagottville and Edmonton which are close to other weather stations (Figure 1), whereas for Resolute, the farthest station considered in our data set, and

for Prince Rupert with an oceanic climate, the difference between smoothed and predicted curves is greater than 10 degrees Celsius several days or months of the year.

Following with the cross-validation residual analysis, we note in Figure 4 (right panel) that, although there are outliers, the residual mean indicates that the predictions are unbiased (mean around zero). We can also observe that the variation on the prediction is lower in the summer (days 100 and 300) than in the winter (Figure 4, right panel) as a consequence of the reasons above mentioned.

Cross-validation prediction variances were estimated by using equation (14). As in multivariable kriging, this statistic depends only on estimations of simple (direct) and cross-covariances, that is, depends on the distance between the prediction site and the sampling locations, and does not take into account the observed values. The further the prediction site, the greater the prediction variance. This result is clearly highlighted in the map of prediction variances (Figure 6) which shows that weather stations located in the Arctic or in the northwest have greater variances. This result should be interpreted carefully because two close stations could have very different climatic conditions, but this result will not be reflected on the prediction variance. As an example in this sense we can observe that Pr Rupert (Figure 6) has a small prediction variance (compared with Resolute, Inuvik or Iqaulit) because it is close to Victoria, Vancouver or Pr George. However, the temperature curve is not well predicted in this site (Figure 5) due to the fact that Pr George has significantly different temperature values. In any case, the cross-validation results show that the predictions by PWKFD are close to the smoothed curves and therefore this method can be considered a valid technique for performing spatial prediction of functional data.

A particular case of the predictor (1) is obtained by considering $\lambda_i(t) = \lambda_i$ for all $i = 1, \dots, n$, that is, by carrying out spatial prediction of functional data by means of the predictor

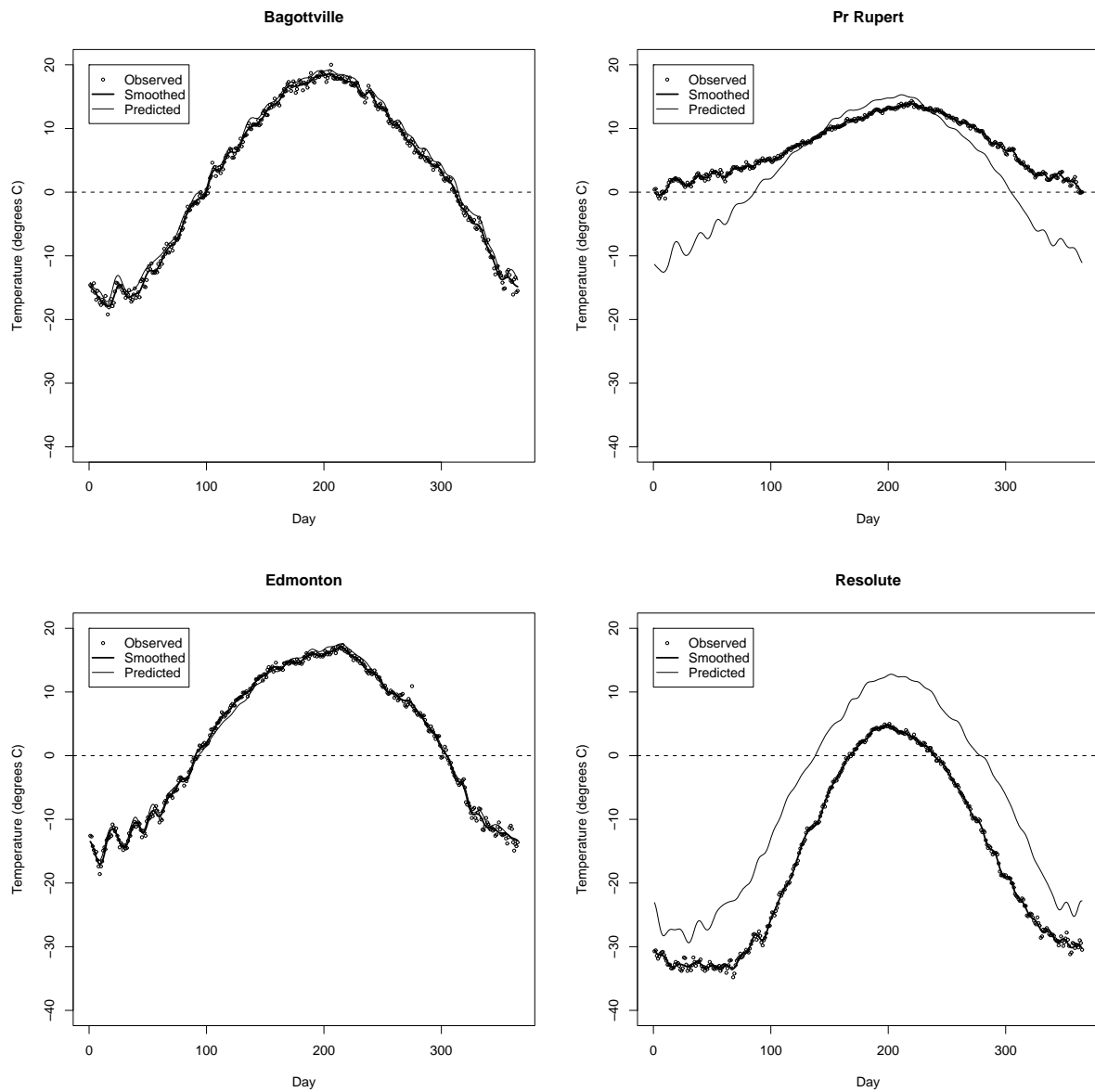


Figure 5: Temperature curves (observed, smoothed and predicted by point-wise kriging) at four Canadian weather stations.

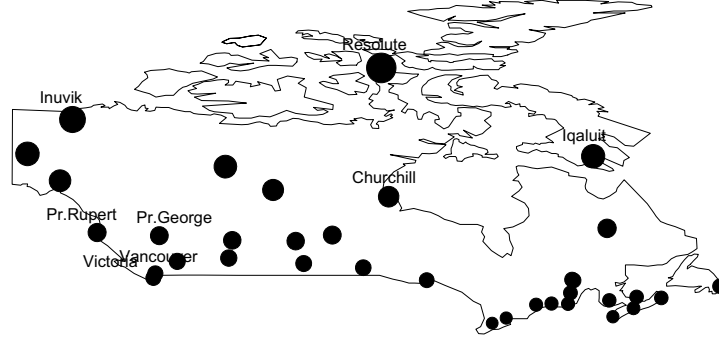


Figure 6: Point-wise kriging prediction variances of cross-validation analysis. Circle size is proportional to the prediction variance.

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i \chi_{s_i}(t), \quad \lambda_1, \dots, \lambda_n \in \mathbf{R}. \quad (17)$$

This predictor was initially considered by Goulard and Voltz (1993) and more recently by Giraldo et al. (2007) which named this method as ordinary kriging for function-valued spatial data (OKFD). We use the Canadian data set analyzed in this section for comparing OKFD and PWKFD, in terms of graphical outputs and SSE_F values. The SSE_F values for OKFD were also calculated with 65 basis functions. Table 1 summarizes the comparative cross-validation SSE_F results.

The predictions at Slave Lake (Figures 3 and 7) as well as the predictions by functional cross-validation (Figures 4 and 7) obtained with the two predictors are graphically similar. However, the sum of SSE_F values (Table 1) indicates that PWKFD has better performance. A detailed analysis of the summary statistics given in Table 1 shows that PWKFD is better than OKFD, in particular when we perform prediction at the farthest stations. We note that there are small differences between the two methods in terms of minimum or median values. This indicates that both methods have a similar performance on well-predicted stations as Edmonton or Bagottville. The differences between these methods are essentially due to their performance on stations as Resolute, Inuvik or Iqaluit. With both methods the

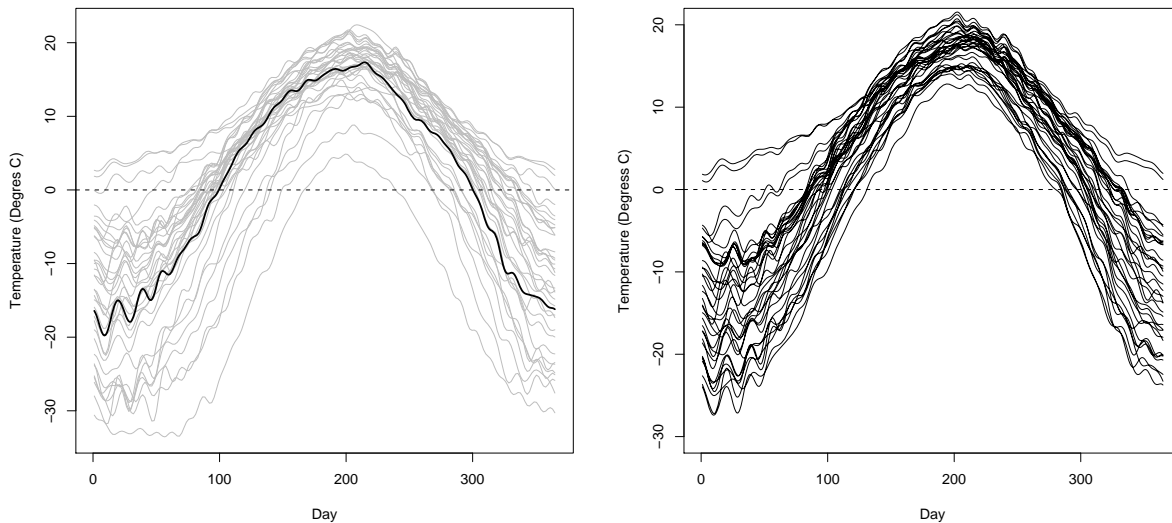


Figure 7: Prediction on Slave Lake (*left panel*) and cross-validation predictions (*right panel*) by means of OKFD.

greatest SSE_F value corresponds to Resolute station. According to the maximum SSE_F values (Table 1) we have much less error on the prediction in this site by using PWKFD. An analogous result is achieved at other further stations as Inuvik or Iqaluit.

5 Conclusions and further research

We have shown a kriging methodology for functional data. We have considered basis expansion as a way to represent the observed functions. A minimization criterion given in multivariable geostatistics has been adapted to the functional context. Our approach was applied to a climatological data set. The cross-validations results show a good performance of the proposed predictor, and indicate from a descriptive point of view that this one can be taken as a valid method for modeling spatially correlated functional data. In addition, the predictor proposed performed better

Table 1: Summary statistics of cross-validation $SSE_F(i), i = 1, \dots, 35$, values. OKFD: Ordinary kriging for function-valued spatial data; PWKFD: Point-wise kriging for functional data.

Statistic	OKFD	PWKFD
Minimum	135.1	154.7
Median	586.6	597.8
Mean	5004.0	3033.0
Maximum	91806.8	32770.0
Standard deviation	15536	6173
Sum	175140	106155

than the based on ordinary kriging for function-valued spatial data.

There is still a long way of research necessary for spatial prediction of functional data. More complex procedures can be considered by replacing functional parameters $(\lambda_i(t), t \in T)$, by double indexed functional parameters $(\lambda_i(s, t), s, t \in T)$ which would be an extension from multivariable geostatistics to functional geostatistics. Some preliminary works for this approach are showed in Giraldo et al. (2008) and Nerini and Monestiez (2008). Models for doing spatial prediction based on information of several functional variables, that is, two or more functional variables observed at each sampling location could also be considered. Further attention should be given to the use of other basis system to get functional data from discrete observations.

Acknowledgements

Research partially supported by the Spanish Ministry of Education and Science and FEDER through grants MTM2007-62923 and MTM2006-09920. The first author thanks the FUNDACION CAROLINA for financial support through a PhD scholarship.

References

- Baladandayuthapani, V., B. Mallick, M. Hong, J. Lupton, N. Turner, and R. Carroll (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics* *64*, 64–73.
- Bodas-Salcedo, A., E. López-Baeza, F. Martínez, J. Mateu, and F. Montes (2003). Spatio-temporal modeling and prediction of solar radiation. *Journal of Geophysical Research* *108*(D24), 1877. doi:10.1029/2003JD003903.
- Box, G. and G. Jenkins (1976). *Time Series Analysis*. New York: Holden Day.
- Brunsdon, C., S. Fotheringham, and M. Charlton (1998). Geographically weighted regression-modelling spatial non-stationarity. *The Statistician* *47*, 431–443.
- Cardot, H., C. Crambes, A. Kneip, and P. Sarda (2007). Smoothing spline estimators in functional linear regression with errors in variables. *Computational Statistics & Data Analysis* *51*(10), 4832–4848.
- Cardot, H., F. Ferraty, and P. Sarda (1999). Functional linear model. *Statistics and Probability Letters* *45*, 11–22.
- Cressie, N. (1993). *Statistic for Spatial Data*. New York: John Wiley & Sons.
- Cuevas, A., M. Febrero, and R. Fraiman (2004). An anova test for functional data. *Computational Statistics & Data Analysis* *47*, 111–122.
- Delicado, P. (2007). Functional k -sample problem when data are density functions. *Computational Statistics* *22*(3), 391–410.
- Ferraty, F. and P. Vieu (2003). Curves discrimination. A non parametric functional approaches. *Computational Statistics & Data Analysis* *44*, 161–173.
- Ferraty, F. and P. Vieu (2006). *Non Parametric Functional Data Analysis. Theory and Practice*. New York: Springer.

- Giraldo, R., P. Delicado, and J. Mateu (2007). Geostatistics for functional data: An ordinary kriging approach. Technical Report <http://hdl.handle.net/2117/1099>, Universitat Politècnica de Catalunya. Submitted to *Environmental and Ecological Statistics*.
- Giraldo, R., P. Delicado, and J. Mateu (2008). Functional kriging: Total model. Long abstract sent to the eighth international geostatistics congress. Santiago, Chile. <http://www.geostats2008.com>.
- González-Manteiga, W. and P. Vieu (2007). Statistics for functional data. *Computational Statistics and Data Analysis* 51, 4788–4792.
- Goulard, M. and M. Voltz (1993). Geostatistical interpolation of curves: A case study in soil science. In A. Soares (Ed.), *Geostatistics Tróia '92*, Volume 2, pp. 805–816. Kluwer Academic Press.
- Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, 55, 755–796.
- Lfeachor, E. and B. Jervis (1993). *Digital Signal Processing, A Practical Approach. 1st Edition*. New York: Addison-Wesley.
- Myers, D. (1982). Matrix formulation of co-kriging. *Mathematical Geology* 14(3), 249–257.
- Nerini, D. and P. Monestiez (2008). A cokriging method for spatial functional data with applications in oceanology. Long summary sent to the first international workshop on functional and operational statistics. Toulouse, France. <http://www.lsp.ups-tlse.fr/staph/IWFOS2008/>.
- Pebesma, E. (2004). Multivariable geostatistics in S. The gstat package. *Computers & Geosciences* 30, 683–691.

- Ramsay, J. and C. Dalzell (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B*, 53, 539–572.
- Ramsay, J. and B. Silverman (2001). Functional data analysis. In *International Encyclopedia of the Social & Behavioral Sciences*, pp. 5822–5828. Elsevier Science.
- Ramsay, J. and B. Silverman (2005). *Functional Data Analysis. Second edition*. New York: Springer.
- Ruiz-Medina, M. D., S. Salmerón, and J. M. Angulo (2007). Kalman filtering from POP-based diagonalization of ARH(1). *Computational Statistics & Data Analysis* 51, 4994–5008.
- Valderrama, M. (2007). An overview to modelling functional data. *Computational Statistics* 22, 331–334.
- Ver Hoef, J. and N. Cressie (1993). Multivariable spatial prediction. *Mathematical Geology* 25(2), 219–240.
- Wackernagel, H. (1995). *Multivariable Geostatistics: An Introduction with Applications*. Berlin: Springer-Verlag.
- Yamanishi, Y. and Y. Tanaka (2003). Geographically weighted functional multiple regression analysis: A numerical investigation. *Journal of Japanese Society of Computational Statistics* 15, 307–317.