Continuous Time-Varying Kriging for Spatial Prediction of Functional Data: An Environmental Application

Author(s): R. Giraldo, P. Delicado and J. Mateu

Source: *Journal of Agricultural, Biological, and Environmental Statistics*, March 2010, Vol. 15, No. 1 (March 2010), pp. 66-82

Published by: Springer

Stable URL: https://www.jstor.org/stable/20778451

Continuous Time-Varying Kriging for Spatial Prediction of Functional Data: An Environmental Application

R. GIRALDO, P. DELICADO, and J. MATEU

Spatially correlated functional data are 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 pointwise 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 basis functions. The number of basis functions is chosen by cross-validation. Then, the spatial prediction of a curve is obtained as a pointwise 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.

Key Words: Basis functions; Coregionalization linear model; Cross-validation; Functional linear pointwise 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 early 1990s, functional data analysis (FDA) (Ramsay and Dalzell 1991) has been used to model this type of data. From the FDA point of view, each curve corresponds to one observation, that is, the basic

R. Giraldo is Associate Professor, Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia. P. Delicado is Associate Professor of Statistics, Department of Statistics and Operation Research, Universitat Politècnica de Catalunya, E-08034 Barcelona, Spain. J. Mateu is Professor of Statistics, Department of Mathematics, Universitat Jaume I, E-12071 Castellón, Spain (E-mail: E-mail: *mateu@mat.uji.es*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 15, Number 1, Pages 66–82
DOI: 10.1007/s13253-009-0012-z

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, Febrero, and Fraiman 2004; Delicado 2007), regression (Cardot, Ferraty, and Sarda 1999; Cardot et al. 2007), nonparametric methods (Ferraty and Vieu 2006), or multivariate techniques (Ferraty and Vieu 2003). An overview of statistical methods for analyzing functional data is shown in the work of Ramsay and Silverman (2005), and recent developments in this field are given in special issues of several journals (González-Manteiga and View 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 in modeling correlated functional data: this 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 above-mentioned methodologies may not be appropriate as they do not incorporate 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 the work of Ruiz-Medina, Salmerón, and Angulo (2007). These authors considered an autoregressive Hilbertian model of order 1 to represent the dynamic of a sequence of functional data. For spatially correlated functional data, Yamanishi and Tanaka (2003) developed a regression model that enables the relation among variables over time and space to be studied, combining both geographically weighted regression (Brunsdon, Fotheringham, and Charlton 1998) and functional multiple regression (Ramsay and Silverman 2005). Baladandayuthapani et al. (2008) showed an alternative for analyzing an experimental design with a spatially correlated functional response. They used both a hierarchical model and a Bayesian approach. Contributions of Yamanishi and Tanaka (2003) and Baladandayuthapani et al. (2008) made it possible to include 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) were pioneer workers in this context. They proposed three geostatistical approaches to predict curves: a curve kriging approach and two multivariate approaches based on cokriging on either discrete data or coefficients of the parametric models that have been fitted to the observed curves. Giraldo, Delicado, and Mateu (2007) gave a nonparametric approach to solving 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, Delicado, and Mateu (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 a scalar parameter. In this article we consider the problem of spatial prediction of functional data by weighting each observed curve by a functional parameter. This approach was mentioned in the work of 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 (*pointwise*) model such as shown by Ramsay and Silverman (2005). We propose a solution based on basis functions. Both the curves and the functional parameters are expanded in terms of a set of basis functions. Thus, the problem becomes one of estimating the coefficients of these basis functions for each functional parameter. To provide a solution, we use a linear model of coregionalization for estimating the covariances among the coefficients of each curve. An essential step in our proposal is the choice of the number of basis functions. Here we consider here a criterion based on cross-validation analysis.

We illustrate our proposal with a real meteorological example. A well-known dataset analyzed through FDA techniques consists of daily temperatures recorded at 35 Canadian weather stations spread over all the Canadian territory. Ramsay and Silverman (2005) used this dataset to provide examples of modeling in functional data. However, the very large spatial area covered by this dataset makes it difficult to apply statistical spatial tools based on stationarity and isotropy. For instance, distances (in the order of thousands of kilometers) in latitude imply bigger weather differences than similar distances in longitude. An exception to this is when sites are near the coast, due to the substantial difference between coastal/island and continental climates. Taking into account that we assume stationarity and isotropy in the model we introduce here, we analyze a homogeneous small area in Canada (the Maritime Provinces), thereby assuming the above-mentioned effects to be negligible. For this region we define a dataset similar in structure to that used by Ramsay and Silverman (2005).

The plan of the article is as follows. Section 2 presents the dataset to be analyzed. Section 3 introduces the predictor and the parameter estimation. Section 4 gives an application of the proposed methodology to the considered dataset. The article ends with a brief discussion and suggestions for further research.

2. DATASET: CANADA'S MARITIME PROVINCES TEMPERATURES

Spatial prediction of meteorological data is an important input for many types of models, including hydrological models or those for regeneration, growth, and mortality of a forest ecosystem. In particular, the modeling of spatially correlated temperature data is of interest, among other examples, 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 carrying out spatial prediction of temperatures. However, to the best of our knowledge all these methods ignore its functional character.

The Maritime Provinces cover a region of Canada consisting of three provinces: Nova Scotia (NS), New Brunswick (NB), and Prince Edward Island (PEI). They occupy just over 1% of Canada's land surface. Here we provide an applied context for our proposal by using a dataset consisting of daily mean temperature measurements recorded at 35 weather

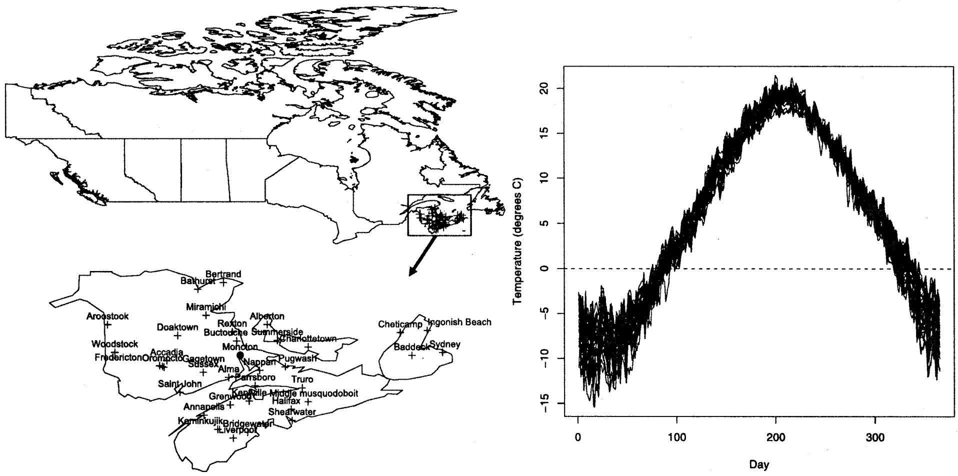

Figure 1. Averages (over the years 1960 to 1994) of daily mean temperature curves (*right panel*) observed at 35 Canadian weather stations (*left panel*). The black dot corresponds to Moncton station, used in Section 4 for temperature curve prediction.

stations located in these provinces (Figure 1, left panel). Given their locations on Canada's Atlantic coast, the Maritime Provinces have a cool, temperate climate: cold continental air masses from the northwest alternate with warmer, humid maritime air from the southwest (Stanley 2002). Their temperature is characterized by cool summers and mild winters, with a much smaller annual temperature range than that recorded in other Canadian regions. In particular, we analyze information of daily mean temperatures averaged over the years 1960 to 1994 (February 29th combined with February 28th) (Figure 1, right panel). The mean of the averaged daily temperatures in the considered dataset varies from −9.4°C in wintertime to 19.3°C in summer. The data for each station were obtained from the Meteorological Service of Canada (*http://www.climate.weatheroffice.ec.gc.ca/climateData/*). The geographical coordinates in decimal degrees of the weather stations (Figure 1, left panel) were obtained from the database of geographic coordinate information (*http://www.tageo.com*).

This dataset has the same structure as the well-known Canadian weather dataset introduced by Ramsay and Silverman (2005), but it covers a much smaller and more homogeneous area. Thus the effects of longitude, latitude, and coastal and inland climates are negligible, and the conditions of stationarity and isotropy are therefore plausible. In accordance with Ramsay and Silverman (2005), we use Fourier basis functions for constructing curves from the observed dataset.

3. POINTWISE 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 that we observe a sample of curves $\chi_{s_i}(t)$, for $t \in T$ and $s_i \in D, i = 1, \ldots, n$. It is usually assumed that these curves belong to a separable Hilbert space $\mathbf{H}$ 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 $t \in T, s \in D$.
- $\operatorname{cov}(\chi_{s_i}(t), \chi_{s_j}(u)) = C(h; t, u)$, $s_i, s_j \in D, t, u \in T$, $h = \|s_i - s_j\|$, the Euclidean distance. If $t = u$, $\operatorname{cov}(\chi_{s_i}(t), \chi_{s_j}(t)) = C(h; t)$.
- $\frac{1}{2}V(\chi_{s_i}(t) - \chi_{s_j}(u)) = \gamma(h; t, u)$, $s_i, s_j \in D, t, u \in T$, $h = \|s_i - s_j\|$. If $t = u$, $\frac{1}{2}V(\chi_{s_i}(t) - \chi_{s_j}(t)) = \gamma(h; t)$.

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

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i(t)\chi_{s_i}(t), \qquad \lambda_1(t), \ldots, \lambda_n(t) T :\rightarrow \mathbf{R}, \tag{3.1}$$

which was mentioned by Goulard and Voltz (1993) without further development. For each $t \in T$, the predictor (3.1) has the same expression as an ordinary kriging predictor. This predictor is here called the pointwise linear predictor for functional data (PWKFD). This modeling approach is consistent 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 *pointwise*. 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 $X_j(t), j = 1, \ldots, q$, 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 $\alpha(t), \beta_j(t), j = 1, \ldots, q$, is carried out by solving $\min_{\alpha(\cdot),\beta_1(\cdot),\ldots,\beta_q(\cdot)} E\|\hat{\mathbf{Y}}(t) - \mathbf{Y}(t)\|^2$ (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, the objective function is

$$E\left\|\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\right\|^2 = \int_T E\left(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\right)^2 dt,$$

by Fubini's theorem. The predictor (3.1) is unbiased if $E(\hat{\chi}_{s_0}(t)) = m(t)$, for all $t \in T$, that is, if $\sum_{i=1}^{n} \lambda_i(t) = 1$ for all $t \in T$. In this case $E(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2 = V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))$. Therefore, to find the best linear unbiased predictor (BLUP), the n functional parameters in the proposed predictor are given by the solution of the following optimization problem:

$$\min_{\lambda_1(\cdot),\ldots,\lambda_n(\cdot)} \int_T V\left(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\right) dt, \quad \text{s.t.} \sum_{i=1}^{n} \lambda_i(t) = 1 \quad \text{for all } t \in T. \tag{3.2}$$

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 BLUP is obtained by minimizing $\sigma_{s_0}^2 = V(\hat{Z}(s_i) - Z(s_i))$ 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 $\{\mathbf{Z}(s_1), \ldots, \mathbf{Z}(s_n)\}$, 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 the 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, that is, by 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 (3.2) 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 $[Z_1(s_0), \ldots, Z_m(s_0)]$ and $[\hat{Z}_1(s_0), \ldots, \hat{Z}_m(s_0)]$ by the functional variables $\chi(t)$ and $\hat{\chi}(t)$, respectively, with $t \in T$.

3.2. A SOLUTION BASED ON BASIS FUNCTIONS

We assume that each observed function can be expressed in terms of K basis functions, $B_1(t), \ldots, B_K(t)$, by

$$\chi_{s_i}(t) = \sum_{l=1}^{K} a_{il} B_l(t) = \mathbf{a}_i^T \mathbf{B}(t), \qquad i = 1, \ldots, n. \tag{3.3}$$

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

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

forms a K multivariable random field with $E(\boldsymbol{\alpha}_i) = \boldsymbol{v}_{i\,(n\times 1)}$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1K} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1} & \Sigma_{K2} & \cdots & \Sigma_{KK} \end{pmatrix}_{(K\times n)\times(K\times n)}, \tag{3.4}$$

where $\Sigma_{ij} = C(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)_{(n\times n)}$. The coefficients a_{ij} are assumed to be a realization of the spatial random field $\boldsymbol{\alpha}_j$, $j = 1, \ldots, K$. We propose the use of multivariable geostatistics (Wackernagel 1995), and specifically a linear model of coregionalization (LMC) for estimating the matrix (3.4). To establish the unbiasedness condition, and for carrying out the parameter estimation in (3.1), we further expand each functional parameter $\lambda_i(t)$ as

$$\lambda_i(t) = \sum_{l=1}^{K} b_{il} B_l(t) = \mathbf{b}_i^T \mathbf{B}(t). \tag{3.5}$$

Therefore, by using (3.3) and (3.5), the predictor in (3.1) is given by

$$\begin{aligned}\hat{\chi}_{s_0}(t) &= \sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{B}(t) \mathbf{a}_i^T \mathbf{B}(t) \\ &= \sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{a}_i. \end{aligned} \tag{3.6}$$

Using (3.5) 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

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

or more specifically by

$$\sum_{i=1}^{n} b_{i1} = c_1, \qquad \ldots, \qquad \sum_{i=1}^{n} b_{iK} = c_K. \tag{3.7}$$

Developing the variance in the objective function (3.2), we have

$$\begin{aligned} V\big(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)\big) &= V\big(\hat{\chi}_{s_0}(t)\big) + V\big(\chi_{s_0}(t)\big) - 2C\big(\hat{\chi}_{s_0}(t), \chi_{s_0}(t)\big) \\ &= V\left(\sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{a}_i\right) + \mathbf{B}^T(t) V(\mathbf{a}_0) \mathbf{B}(t) \\ &\quad - 2 \sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) C(\mathbf{a}_i, \mathbf{a}_0) \mathbf{B}(t) \\ &= \sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) V(\mathbf{a}_i) \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{b}_i \\ &\quad + 2 \sum_{i<j} \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) C(\mathbf{a}_i, \mathbf{a}_j) \mathbf{B}(t) \mathbf{B}^T(t) \mathbf{b}_j \\ &\quad + \mathbf{B}^T(t) V(\mathbf{a}_0) \mathbf{B}(t) - 2 \sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{B}(t) \mathbf{B}^T(t) C(\mathbf{a}_0, \mathbf{a}_i) \mathbf{B}(t). \end{aligned} \tag{3.8}$$

In (3.8), for $i < j$, $i, j = 0, 1, \ldots, n$, we have

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

and

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

If we define

$$\mathbf{Q}_i = \int_T \left(\mathbf{B}(t)\mathbf{B}^T(t)V(\mathbf{a}_i)\mathbf{B}(t)\mathbf{B}^T(t)\right)dt,$$

$$\mathbf{Q}_{ij} = \int_T \left(\mathbf{B}(t)\mathbf{B}^T(t)C(\mathbf{a}_i,\mathbf{a}_j)\mathbf{B}(t)\mathbf{B}^T(t)\right)dt,$$

$$\mathbf{D} = \int_T \mathbf{B}^T(t)V(\mathbf{a}_0)\mathbf{B}(t)\,dt,$$

and

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

and by considering K Lagrange multipliers $\mathbf{m}^T = (m_1,\ldots,m_K)$, the optimization problem (3.2) can be expressed as

$$\min_{\mathbf{b}_1,\ldots,\mathbf{b}_n,\mathbf{m}} \sum_{i=1}^{n} \mathbf{b}_i^T\mathbf{Q}_i\mathbf{b}_i + 2\sum_{i<j}\mathbf{b}_i^T\mathbf{Q}_{ij}\mathbf{b}_j + \mathbf{D} - 2\sum_{i=1}^{n}\mathbf{b}_i^T\mathbf{J}_i + 2\mathbf{m}^T\left(\sum_{i=1}^{n}\mathbf{b}_i - \mathbf{c}\right). \tag{3.9}$$

Taking $\boldsymbol{\beta} = (\mathbf{b}_1^T,\ldots,\mathbf{b}_n^T,\mathbf{m}^T)^T_{(K(n+1)\times 1)}$, the expression (3.9) is given by

$$\min_{\boldsymbol{\beta}} \boldsymbol{\beta}^T\mathbf{Q}\boldsymbol{\beta} + \mathbf{D} - 2\boldsymbol{\beta}^T\mathbf{J}, \tag{3.10}$$

where

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_{12} & \cdots & \mathbf{Q}_{1n} & \mathbf{I} \\ \mathbf{Q}_{21} & \mathbf{Q}_2 & \cdots & \mathbf{Q}_{2n} & \mathbf{I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{Q}_{n1} & \mathbf{Q}_{n2} & \cdots & \mathbf{Q}_n & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix}, \qquad \mathbf{J} = \begin{pmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \\ \vdots \\ \mathbf{J}_n \\ \mathbf{c} \end{pmatrix}. \tag{3.11}$$

The identity matrix $\mathbf{I}$ in (3.11) is of order K. By minimizing (3.10) with respect to $\boldsymbol{\beta}$ we obtain

$$2\mathbf{Q}\boldsymbol{\beta} - 2\mathbf{J} = \mathbf{0} \quad\Rightarrow\quad \mathbf{Q}\boldsymbol{\beta} = \mathbf{J} \quad\Rightarrow\quad \hat{\boldsymbol{\beta}} = \mathbf{Q}^{-1}\mathbf{J}. \tag{3.12}$$

In practice, we start estimating both an LMC for the multivariable random field $\mathbf{A} = (\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_K)$ and the matrix in (3.4). Subsequently, we can calculate the matrices $\mathbf{Q}$ and $\mathbf{J}$ in (3.11). By replacing these matrices in (3.12), we can estimate $\mathbf{b}_i$, $i = 1,\ldots,n$, and consequently the functional parameters given in (3.5). On the other hand, a plug-in estimation of the integrated prediction variance $\sigma^2_{s_0} = \int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))\,dt$ is given by

$$\hat{\sigma}^2_{s_0} = \hat{\boldsymbol{\beta}}^T\mathbf{Q}\hat{\boldsymbol{\beta}} + \mathbf{D} - 2\hat{\boldsymbol{\beta}}^T\mathbf{J}, \tag{3.13}$$

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 $\hat{\sigma}^2_{s_0}$ is a measure of the uncertainty in the prediction of a whole curve. On the basis of the estimated parameters and by using (3.8), a pointwise prediction variance function can also be estimated.

3.3. CHOOSING THE NUMBER OF BASIS FUNCTIONS

Let us assume that functions $\chi_{s_i}(t)$, $i = 1, \ldots, n$, defined on T have been observed at points $t_1, \ldots, t_M$ and we wish to approximate them through a basis function. We should thus 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). We consider a nonparametric cross-validation analysis to choose the number K of basis functions.

A simple way of establishing an appropriate K is by calculating the cross-validation SSE in a classical nonparametric sense. Let $\tilde{\chi}_{s_i}^{(j)}(t_j)$ be the estimated function at t_j by means of (3.3) 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

$$NPCV(K) = \sum_{i=1}^{n}\sum_{j=1}^{M}\left(\tilde{\chi}_{s_i}^{(j)}(t_j) - \chi_{s_i}(t_j)\right)^2. \tag{3.14}$$

The strategy is to minimize $NPCV(K)$ in K. We shall call this method *nonparametric cross-validation.*

3.4. EVALUATION CRITERION OF A KRIGING PREDICTOR

In any prediction problem, the ideal model evaluation consists of splitting the whole dataset into two parts: a training (estimation) sample for model fitting and a test (validation) sample for model evaluation. This approach, however, is not efficient unless the sample size is large. The idea behind cross-validation proposals is to recycle data by switching the roles of training and test samples.

In the context of spatially correlated functional data, where the goal is to predict a whole function $\chi_{s_0}(t)$ at an unvisited site s_0, *leave-one-out* cross-validation works as follows: each data location is removed from the dataset and a smoothed function is predicted at this location using a functional kriging predictor (PWKFD (3.6) in our case) based on the remaining smoothed functions. We calculate the SSE by

$$SSE_{FCV} = \sum_{i=1}^{n} SSE_{FCV}(i) = \sum_{i=1}^{n}\sum_{j=1}^{M}\left(\hat{\chi}_{s_i}^{(i)}(t_j) - \chi_{s_i}(t_j)\right)^2, \tag{3.15}$$

where $\hat{\chi}_i^{(i)}(t_j)$ is the PWKFD prediction on s_i evaluated at t_j, $j = 1, \ldots, M$, by leaving the site s_i temporarily out of the sample. We call this procedure *functional cross-validation.*

4. APPLICATION: SPATIAL PREDICTION OF TEMPERATURE CURVES

In this section we illustrate our approach by using the temperature dataset 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

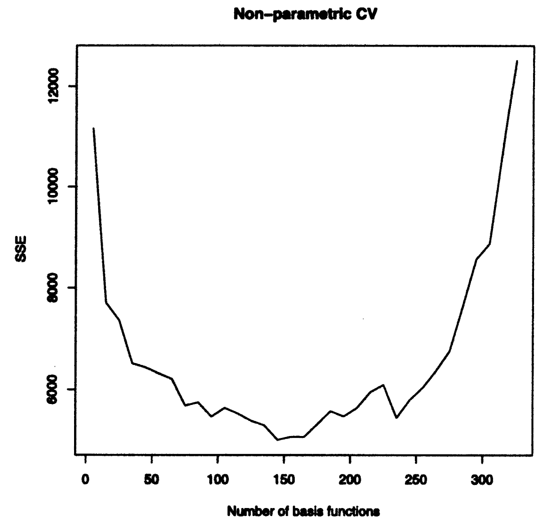

Figure 2. Sum of squared errors as a function of K obtained by nonparametric cross-validation (function $NPCV(K)$).

the predictor proposed by Giraldo, Delicado, and Mateu (2007). This comparison is given in terms of SSE_{FCV} 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 used to expand the Canadian temperature data (Ramsay and Silverman 2005). We use the criterion described in Section 3.3 for selecting the number of functions to be used for smoothing the observed discrete dataset. Although we can expand in terms of a Fourier basis with an infinite number of sinusoids, we take 365 as the limit because this is the number of discrete data for each site in our dataset. 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 $NPCV(K)$ obtained by nonparametric cross-validation. We can observe that the SSE values decrease significantly until the number of basis functions is around 60. Then the rate of decrease is small. In particular, $NPCV(K)$ decreases by 44% when K is between 5 and 65, whereas this percentage is 55% when K varies between 5 and 145 (where the minimum of $NPCV(K)$ is attained). Thus SSE values indicate that there is no advantage in using a value of K much larger than 65. As mentioned before, Ramsay and Silverman (2005) also used 65 Fourier basis functions for smoothing the Canadian temperature dataset. Our results suggest that this number of basis functions is also appropriate for our dataset. In consequence, a pragmatic choice for K is 65. Therefore, in the following we assume that the data to be analyzed correspond to the temperature curves obtained after smoothing each discrete dataset by means of a Fourier basis with 65 functions.

PWKFD is used to predict a temperature curve at an unvisited site with coordinates −64.69 (easting) and 45.10 (northing). This site corresponds to the Moncton station (Figure 1). Moncton is situated in southeastern New Brunswick, and at the geographic center of the Maritime Provinces (Figure 1). The climate of this station is more continental than

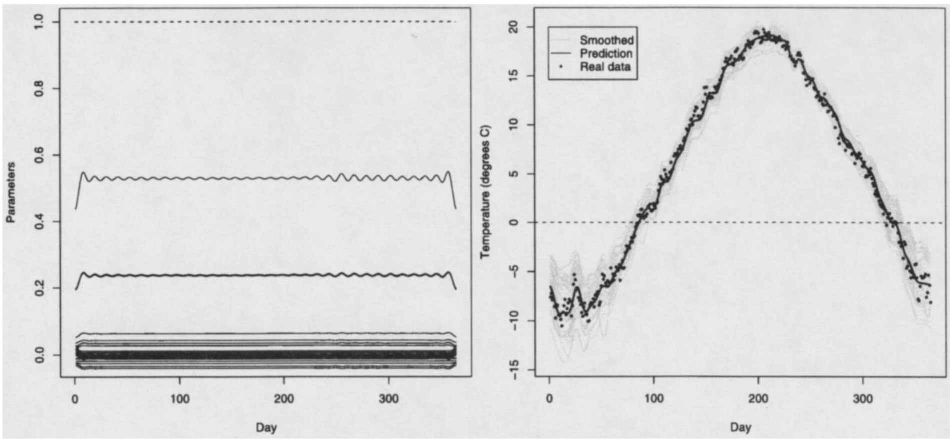

Figure 3. Estimated functional parameters (*left, dark lines*), sum of functional estimated parameters (*left, dotted line*), smoothed temperature curves (*right, light lines*), temperature prediction function at Moncton (*right, dark line*), and real temperature values at Moncton (*right, circles*).

maritime during the summer and winter seasons, whereas maritime influences tend to temper the transitional seasons of spring and autumn. It should be noted that the latitude does not create large climatic variations in this region (Stanley 2002) and consequently it plays no role in the modeling.

In the first stage of the analysis, and using the library `gstat` of R language (Pebesma 2004), a LMC was fitted to the multivariable random field $\mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{65})$ consisting of the coefficients of the Fourier basis used for smoothing each observed sample. We assume stationarity for each process $\boldsymbol{\alpha}_j$, $j = 1, \ldots, 65$. All single (direct) variograms and cross-variograms are modeled as a linear combination of nugget and exponential models. Based on the fitted LMC, the matrices $\mathbf{Q}$ and $\mathbf{J}$ given in (3.11) are estimated and used to solve the system (3.12) finally to find $\mathbf{b}_i$ and the functional parameters $\lambda_i(t)$, $i = 1, \ldots, n$. Figure 3 (left 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 0.5). This corresponds to Bouctouche (NB), the closest station to Moncton (Figure 1). Other stations near Moncton, and therefore with influence on the prediction, are Nappan (NS), and Alma (NB) with weights around 0.2 in Figure 3. The curves corresponding to the sites furthest from Moncton receive a weight of almost zero (Figure 3, left panel). This result is consistent with the kriging philosophy, that is, sites closer to the prediction location have greater influence than others farther 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 (3.7) guarantees the unbiasedness constraint. A plot of the temperature prediction at Moncton is also shown in Figure 3 (right panel). The predicted curve is obtained essentially as a weighted sum of smoothed temperature curves corresponding to Bouctouche, Nappan, and Alma stations. The predicted curve shows a seasonal behavior similar to the smoothed curves. In addition, predicted values are consistent with real values recorded for this weather station (Figure 3). Prediction errors (difference between observed and predicted values in Figure 3) vary from $-1.98°C$ to $2.28°C$. The mean of the errors

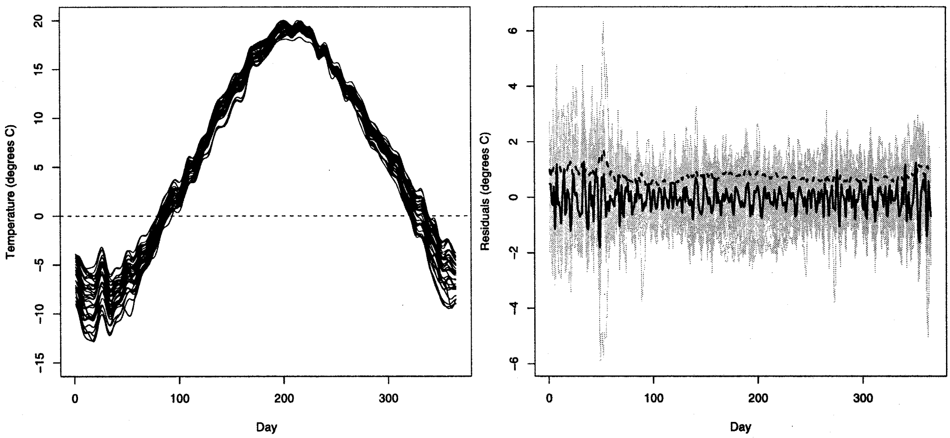

Figure 4. *Left panel*: Pointwise kriging predictions based on cross-validation. *Right panel*: Cross-validation residuals (*light lines*), residual mean (*dark line*), and residual standard deviation (*dashed line*) for the Canadian Maritime Provinces temperature dataset.

is −0.0053°C. These results provide evidence that the prediction using PWKFD is quite close to the observed data. Temperature values for the Moncton station are recorded from *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 $\chi_{s_i}(t), i = 1, \ldots, 35$, is 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 as 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, because kriging is itself a smoothing method.

Figure 4 (right panel) shows cross-validation residuals. The predictions are plausible in a high proportion of sites (those having residuals around zero). There are few stations with large positive or negative residuals. These are particularly obtained in wintertime (residuals higher than 4°C and lower than −4°C at the beginning and at the end of the year). This is due to the fact that the temperature curves show more variability in this season. The greatest residuals are obtained in Bertrand (NB) and Bathurst (NB) (Figure 5). The prediction at Bertrand is most influenced by the curve at Bathurst, and vice versa because the distance between these stations is smaller than the distance to the remaining ones (Figure 1). The temperature curves at Bertrand and Bathurst have a very similar behavior throughout the time considered (Figure 5). However, for some days (19, 49 to 57, 353, 354, and 356) the difference between the temperature values is greater than 4°C. For this reason we obtain high residuals at these stations.

Continuing with the cross-validation residual analysis, in Figure 4 (right panel) we note 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

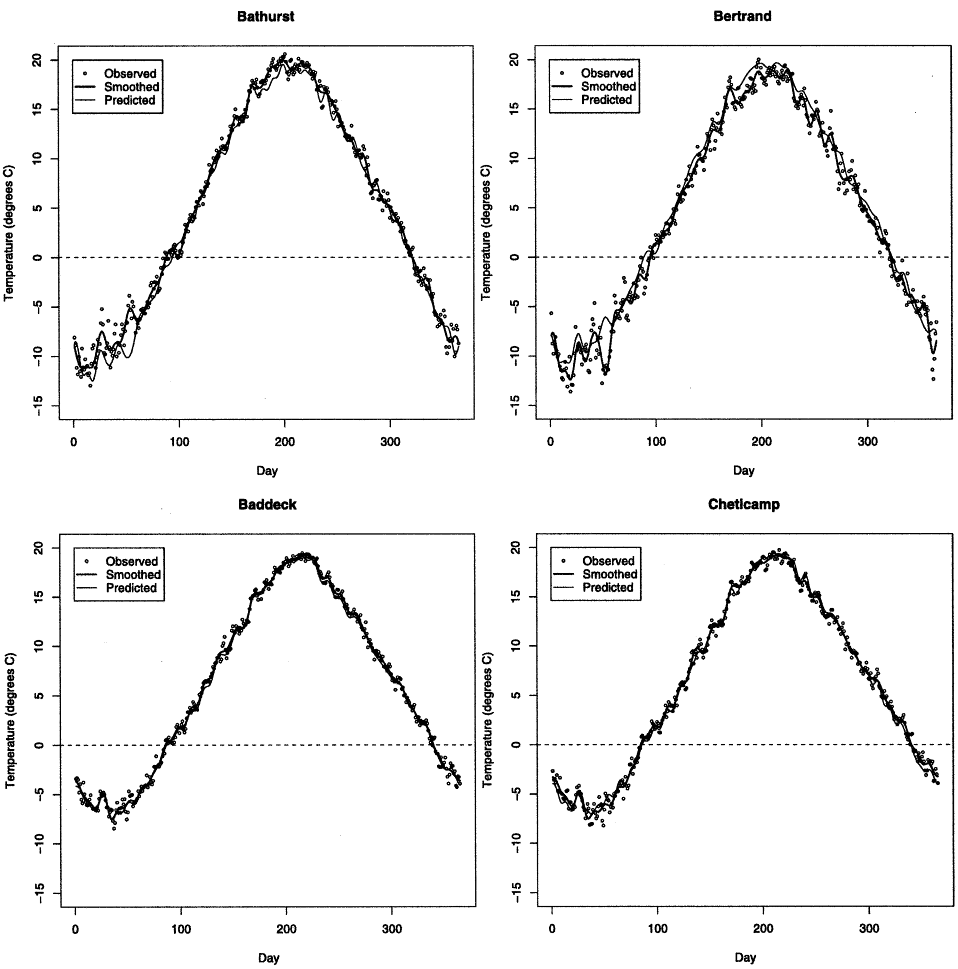

Figure 5. Temperature curves (observed, smoothed, and predicted by pointwise kriging) at four weather stations in Canada's Maritime Provinces.

in summer (days 100 and 300) than in winter (Figure 4, right panel) as a consequence of the reasons given previously.

Cross-validation prediction variances are estimated by using (3.13). As in multivariable kriging, this statistic depends only on estimations of simple (direct) and cross-covariances, that is, it depends on the distance between the prediction site and the sampling locations, and does not take into account the observed values, that is, the uncertainty on the prediction is directly related to the sampling configuration. 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 the most widely separated weather stations in our dataset, such as those located in Nova Scotia's north coast (Sydney, Baddeck, Cheticamp, and Ingonish beach), Aroostook (NB), and Bertrand (NB), have greater variances. This result should be interpreted carefully because at a weather station with a high prediction variance we may find low prediction errors. This is the case of Baddeck (NS) and Cheticamp (NS). There

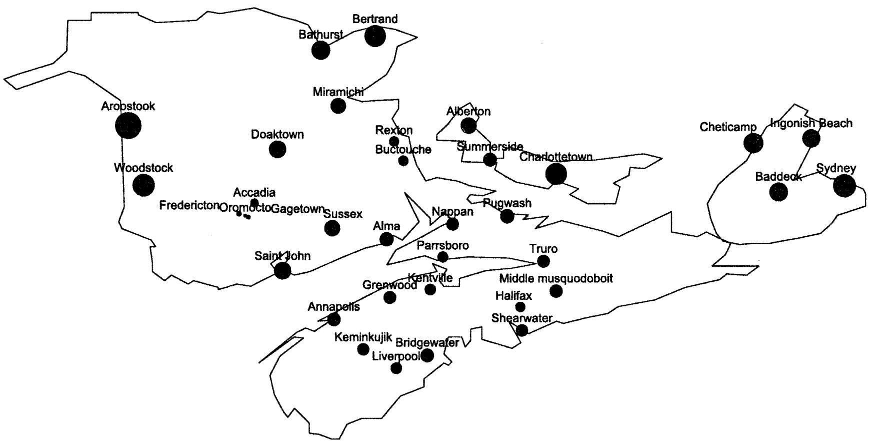

Figure 6. Pointwise kriging prediction variances of cross-validation analysis. Circle diameter is proportional to the prediction variance.

we have a relatively high prediction variance (Figure 6) with a good prediction (Figure 5). In general, 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 (3.1) is obtained by considering $\lambda_i(t) = \lambda_i$ for all $i = 1, \ldots, n$, that is, by carrying out spatial prediction of functional data by means of the predictor

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \chi_{s_i}(t), \qquad \lambda_1, \ldots, \lambda_n \in \mathbf{R}. \tag{4.1}$$

This predictor was initially considered by Goulard and Voltz (1993) and more recently by Giraldo, Delicado, and Mateu (2007), who named this method as ordinary kriging for function-valued spatial data (OKFD). We use the Maritime Provinces dataset analyzed in this section for comparing OKFD and PWKFD, in terms of graphical outputs and SSE_{FCV} values. The SSE_{FCV} values for OKFD are also calculated with 65 basis functions. Table 1 summarizes the comparative cross-validation SSE_{FCV} results.

Table 1. Summary statistics of cross-validation $SSE_{FCV}(i), i = 1, \ldots, 35$, values. OKFD: ordinary kriging for function-valued spatial data; PWKFD: pointwise kriging for functional data.

Statistic	OKFD	PWKFD
Minimum	103.7	104.4
Median	253.4	252.7
Mean	299.5	299.2
Maximum	890.8	902.1
Standard deviation	178.4	175.4
Sum	10,483.9	10,471.3

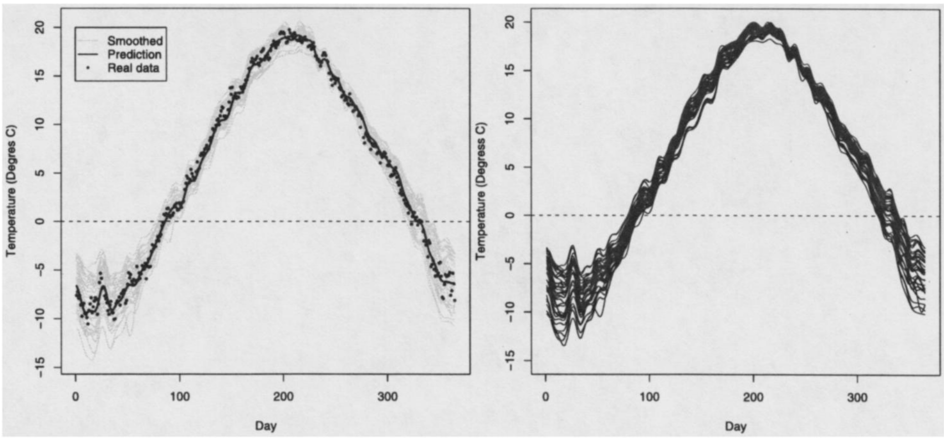

Figure 7. Temperature prediction at Moncton (*left panel*) and cross-validation predictions (*right panel*) by means of OKFD.

The predictions at Moncton (Figures 3 and 7 (left panel)) as well as the predictions by functional cross-validation (Figures 4 and 7 (right panel)) obtained with the two predictors are graphically similar. The summary statistics given in Table 1 show that there are small differences between the two methods. At Baddeck (NB) and Bertrand (NB) the best and the worst predictions, respectively, are obtained. The minimum and maximum SSE_{FCV} values (Table 1) indicate that in both cases OKFD has better behavior on the prediction at these stations. However, the sum and the standard deviation of SSE_{FCV} values (Table 1) show that the PWKFD method has a better performance in general. A detailed analysis of SSE_{FCV} values obtained for each station indicates that the differences between the two methods are due specially to the predictions at the farthest stations. With OKFD we obtain better predictions at Aroostook (NB), Shearwater (NS), Liverpool (NS), Woodstook (NB), and Bertrand (NB), whereas with PWKFD we get better prediction at Sydney (NS), Bathurst (NB), Saint John (NB), Bridgewater (NS), and Keminkujik (NS). At the remaining stations there are very small differences in predictions.

5. CONCLUSIONS AND FURTHER RESEARCH

We show a kriging methodology for functional data. We consider basis expansion as a way to represent the observed functions. A minimization criterion given in multivariable geostatistics is adapted to the functional context. Our approach is applied to a climatological dataset. The cross-validation results show a good performance of the proposed predictor, and indicate from a descriptive point of view that it can be adopted as a valid method for modeling spatially correlated functional data. In addition, the predictor proposed presents better performance than that based on ordinary kriging for function-valued spatial data.

There remains a great deal of research still required for spatial prediction of functional data. More complex procedures can be considered by replacing functional parameters $(\lambda_i(t), t \in T)$ with 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 work for this approach was given by Giraldo, Delicado, and Mateu (2008) and Nerini and Monestiez (2008). Models for carrying out 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 systems for obtaining functional data from discrete observations.

ACKNOWLEDGEMENTS

Research supported in part 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 Ph.D. scholarship.

[Received July 2008. Revised November 2008. Published Online January 2010.]

REFERENCES

Baladandayuthapani, V., Mallick, B., Hong, M., Lupton, J., Turner, N., and Caroll, R. (2008), "Bayesian Hierarchical Spatially Correlated Functional Data Analysis With Application to Colon Carcinoginesis," *Biometrics*, 64, 64–73.

Bodas-Salcedo, A., López-Baeza, E., Martínez, F., Mateu, J., and Montes, F. (2003), "Spatio-Temporal Modeling and Prediction of Solar Radiation," *Journal of Geophysical Research*, 108, 8777.

Box, G., and Jenkins, G. (1976), *Time Series Analysis*, New York: Holden Day.

Brunsdon, C., Fotheringham, S., and Charlton, M. (1998), "Geographically Weighted Regression-Modelling Spatial Non-Stationarity," *The Statistician*, 47, 431–443.

Cardot, H., Crambes, C., Kneip, A., and Sarda, P. (2007), "Smoothing Spline Estimators in Functional Linear Regression With Errors in Variables," *Computational Statistics & Data Analysis*, 51, 4832–4848.

Cardot, H., Ferraty, F., and Sarda, P. (1999), "Functional Linear Model," *Statistics and Probability Letters*, 45, 11–22.

Cressie, N. (1993), *Statistic for Spatial Data*, New York: Wiley.

Cuevas, A., Febrero, M., and Fraiman, R. (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, 391–410.

Ferraty, F., and Vieu, P. (2003), "Curves Discrimination. A Non Parametric Functional Approach," *Computational Statistics & Data Analysis*, 44, 161–173.

——— (2006), *Non Parametric Functional Data Analysis. Theory and Practice*, New York: Springer.

Giraldo, R., Delicado, P., and Mateu, J. (2007), "Geostatistics for Functional Data: An Ordinary Kriging Approach," technical report, Universitat Politècnica de Catalunya, available at *http://hdl.handle.net/2117/1099*.

——— (2008), "Functional Kriging: Total Model," long abstract sent to The Eighth International Geostatistics Congress, Santiago, Chile.

González-Manteiga, W., and View, P. (2007), "Statistics for Functional Data," *Computational Statistics & Data Analysis*, 51, 4788–4792.

Goulard, M., and Voltz, M. (1993), "Geostatistical Interpolation of Curves: A Case Study in Soil Science," in *Geostatistics Tróia'92*, Vol. 2, ed. A. Soares, Dordrecht: Kluwer Academic, pp. 805–816.

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society, Ser. B*, 55, 755–796.

Lfeachor, E., and Jervis, B. (1993), *Digital Signal Processing, a Practical Approach* (1st ed.), New York: Addison-Wesley.

Myers, D. (1982), "Matrix Formulation of Co-Kriging," *Mathematical Geology*, 14, 249–257.

Nerini, D., and Monestiez, P. (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.

Pebesma, E. (2004), "Multivariable Geostatistics in S. The gstat Package," *Computers & Geosciences*, 30, 683–691.

Ramsay, J., and Dalzell, C. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society, Ser. B*, 53, 539–572.

Ramsay, J., and Silverman, B. (2001), "Functional Data Analysis," in *International Encyclopedia of the Social & Behavioral Sciences*, Oxford: Elsevier Science, pp. 5822–5828.

——— (2005), *Functional Data Analysis* (2nd ed.), New York: Springer.

Ruiz-Medina, M. D., Salmerón, S., and Angulo, J. M. (2007), "Kalman Filtering From POP-Based Diagonalization of ARH(1)," *Computational Statistics & Data Analysis*, 51, 4994–5008.

Stanley, D. (2002), *Canada's Maritime Provinces* (1st ed.), Marybirnong: Lonely Planet Publications.

Valderrama, M. (2007), "An Overview to Modelling Functional Data," *Computational Statistics*, 22, 331–334.

Ver Hoef, J., and Cressie, N. (1993), "Multivariable Spatial Prediction," *Mathematical Geology*, 25, 219–240.

Wackernagel, H. (1995), *Multivariable Geostatistics: An Introduction With Applications*, Berlin: Springer-Verlag.

Yamanishi, Y., and Tanaka, Y. (2003), "Geographically Weighted Functional Multiple Regression Analysis: A Numerical Investigation," *Journal of Japanese Society of Computational Statistics*, 15, 307–317.