UNIVERSIDAD CARLOS III DE MADRID working papers

UC3M Working Papers
Statistics and Econometrics
15-12
ISSN 2387-0303
June 2015

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-48

PENALIZED FUNCTIONAL SPATIAL REGRESSION

M. Carmen Aguilera-Morillo[a], María Durbán[a] and Ana M. Aguilera[b]

Abstract

This paper is focus on spatial functional variables whose observations are realizations of a spatio-temporal functional process. In this context, a new smoothing method for functional data presenting spatial dependence is proposed. This approach is based on a P-spline estimation of a functional spatial regression model. As alternative to other geostatistical smoothing methods (kriging and kernel smoothing, among others), the proposed P-spline approach can be used to estimate the functional form of a set of sample paths observed only at a finite set of time points, and also to predict the corresponding functional variable at a new location within the plane of study. In order to test the good performance of the proposed method, two simulation studies and an application with real data have been developed, and the results have been compared with functional kriging.

Keywords: Functional data, functional spatial regression, P-splines.

[a] Department of Statistics, Universidad Carlos III de Madrid.
[b] Department of Statistics and O. R., Universidad de Granada.

Acknowledgements: financial support from the project P11-FQM-8068 from Consejería de Innovación, Ciencia y Empresa. Junta de Andalucía, Spain and the projects MTM2013-47929-P and MTM 2011-28285-C02-C2 from Secretaría de Estado Investigación, Desarrollo e Innovación, Ministerio de Economía y Competitividad, Spain.

Penalized functional spatial regression

M. Carmen Aguilera-Morillo, María Durbán and Ana M. Aguilera

June 18, 2015

Abstract

This paper is focus on spatial functional variables whose observations are realizations of a spatio-temporal functional process. In this context, a new smoothing method for functional data presenting spatial dependence is proposed. This approach is based on a P-spline estimation of a functional spatial regression model. As alternative to other geostatistical smoothing methods (kriging and kernel smoothing, among others), the proposed P-spline approach can be used to estimate the functional form of a set of sample paths observed only at a finite set of time points, and also to predict the corresponding functional variable at a new location within the plane of study. In order to test the good performance of the proposed method, two simulation studies and an application with real data will be developed and the results will be compared with functional kriging.

Keywords: Functional data, functional spatial regression, P-splines.

1 Introduction

This work is focused on the estimation of functional data with spatial dependence. This problem has been approached by different authors in the context of geostatistical techniques. The first notions about this topic can be found in Goulard and Voltz (1993), where multivariate approaches were used to predict curves at unvisited spatial sites. A more recent collection of geostatistical tools for spatial functional data can be seen in Giraldo (2009); Delicado et al. (2009). In general, the most used technique to predict functional data with spatial dependence is functional kriging. In Giraldo et al. (2010) a continuous time-varying kriging was proposed and applied to environmental data. A formal version of ordinary kriging for functional data (OKFD) was developed by Giraldo et al. (2011), and implemented in the R package *geofd*. Recently, a universal kriging predictor for functional data with spatial dependence belonging to a Hilbert space was proposed in Menafoglio et al. (2013).

In the context of spatial data an alternative to geostatistical techniques are the spatial regression models. A popular approach consists of using

Penalized-splines (Eilers and Marx, 1996). They are based on the use of a rich basis for regression and a penalty (based on differences of adjacent coefficients) to control the smoothness of the fit. This methodology has been successfully applied to both, functional and spatial data in different contexts. In functional data analysis P-splines were used for smoothing the sample curves (Aguilera and Aguilera-Morillo, 2013a) and estimating different FDA models as PCA (Aguilera and Aguilera-Morillo, 2013b) or functional regression (Marx and Eilers, 1999; AguileraMorillo et al., 2013, among others). In the case of spatial data, Lee and Durban (2009) and Ugarte et al. (2009) used P-splines for smoothing spatially correlated count data, Lee and Durban (2011) extended their use to the case of spatio-temporal data, and more recently, Sangalli et al. (2013) proposed a spatial regression model for data distributed over irregularly shaped spatial domains.

Our aim is to use spatial smoothing regression techniques within a functional data approach to provide a new method to estimate (or predict at unvisited sites) functional data with spatial dependence. From the formal definition of spatio-temporal functional data, which is given in Section 2, a spatial regression model is extended to the functional context in Section 3. The idea is to consider the functional regression model for functional response and scalar covariates (Faraway, 1997; Ramsay and Silverman, 2005; Chiou et al., 2004) by using the spatial information as regressors. So, a mixture of functional regression model for functional response and spatial regression will yield the proposed functional spatial regression model. In practice, functional data are usually observed with some error or noise. To overcome this problem, Ramsay and Silverman (2005) considered a penalized version of functional regression for functional response by introducing a continuous penalty (based on the second order squared derivatives of the parameter functions) in the least squares fitting, and Reiss et al. (2010) used a penalized generalized least squares criterion based on basis representation. In this paper, we will adapt the idea developed in Eilers et al. (2006), and combine the two-dimensional penalty used for spatial regression with the one proposed in Ramsay and Silverman (2005) to obtain a three dimensional P-spline penalty. Hereinafter, this method will be called penalized functional spatial regression model (PFSRM).

Finally, the proposed PFSRM is compared with OKFD in two simulation studies in Sect. 5. An application to the Canadian maritime weather data is presented in Section 6. Canadian maritime weather is a well known example of functional data, which in most cases have been consider as a set of independent curves related to daily temperature and precipitation at 35 different locations in Canada averaged over 1960 to 1994 (Ramsay and Silverman, 2005). But this is a clear example of functional data presenting spatial dependence and in this sense was studied in Delicado et al. (2009); Menafoglio et al. (2013). The conclusions about these studies in Section 7 close the paper.

2 Basis expansion of spatio-temporal functional data

In this work, we are focus on spatial functional variables whose observations are realizations of a spatio-temporal stochastic process (Delicado et al., 2009) given by

$$\left\{X\left(s,t\right):s\in S\subseteq\mathbb{R}^{2},t\in T\subseteq\mathbb{R}\right\},\tag{1}$$

where $s=(u,v)$ is a generic data location in the spatial domain $S=U\times V$, U,V and T are real intervals, and for each fixed spatio-temporal position (s,t), $X\left(s,t\right)$ is a real random variable defined on a probabilistic space $(\Omega,\mathcal{A},P)$.

Let us suppose that the realizations of this spatio-temporal process are square integrable functions on the spatio-temporal domain $U\times V\times T$, so that each sample function $x(s,t)$ belongs to the Hilbert space $L^{2}\left(U\times V\times T\right)$ defined by

$$L^{2}\left(U\times V\times T\right)=\left\{f:U\times V\times T\longrightarrow\mathbb{R}:\int_{U}\int_{V}\int_{T}f^{2}\left(u,v,t\right)du\ dv\ dt<\infty\right\},$$

with the usual scalar product given by

$$\langle f,g\rangle=\int_{U}\int_{V}\int_{T}f\left(u,v,t\right)g\left(u,v,t\right)du\ dv\ dt,\ \forall f,g\in L^{2}\left(U\times V\times T\right).$$

In general, the spatio-temporal process given by (1) can be seen as a functional random variable on the space $L^{2}\left(U\times V\times T\right)$

$$\begin{array}{rccl} X: & \Omega & \rightarrow & L^{2}\left(S\times T\right) \\ & \omega & \rightarrow & X_{w}: \end{array}\qquad\begin{array}{ccc} S\times T & \rightarrow & \mathbb{R} \\ (s,t) & \rightarrow & X_{w}\left(s,t\right). \end{array}$$

Then, for each fixed location $s\in S$, the realizations of the this spatio-temporal process $x(s,.)$ are square integrable functions on the temporal domain T. That is, each sample curve $x(s,.)$ belongs to the Hilbert space $L^{2}\left(T\right)$ defined by

$$L^{2}\left(T\right)=\left\{f:T\longrightarrow\mathbb{R}:\int_{T}f^{2}\left(t\right)dt<\infty\right\},$$

with the usual scalar product given by

$$\langle f,g\rangle=\int_{T}f\left(t\right)g\left(t\right)dt,\ \forall f,g\in L^{2}\left(T\right).$$

On the other hand, for each fixed time point $t\in T$, the realizations of this spatio-temporal process $x(.,t)$ are square integrable functions on the spatial domain $S=U\times V$. That is, each sample surface $x(.,t)$ belongs to the Hilbert space $L^{2}\left(U\times V\right)$ defined by

$$L^{2}\left(U\times V\right)=\left\{f:U\times V\longrightarrow\mathbb{R}:\int_{U}\int_{V}f^{2}\left(u,v\right)dudv<\infty\right\},$$

with the usual scalar product given by

$$\langle f, g \rangle = \int_U \int_V f(u,v)\, g(u,v)\, du\ dv,\ \forall f, g \in L^2(U \times V).$$

In practice, spatio-temporal data are observed at a finite set of spatial locations $\{s_i = (u_i, v_i) : i = 1, \ldots, n\}$ and a finite set of time points $\{t_j : j = 1, \ldots, m\}$ that could be fixed or random. Then, the sample information is given by a matrix $Y = (y_{ij})_{n \times m}$ with y_{ij} being the observed value of the spatio-temporal functional variable at location s_i and time point t_j. Because of this, the first step in FDA is to reconstruct the true functional form of data from discrete spatio-temporal observations.

In this section, we extend the usual basis expansion approach for representing curves to the case of spatio-temporal functions that depend on three continuous arguments.

Let us consider three univariate basis $\left\{\phi_h^T(t) : t \in T; h = 1, \ldots, p\right\}$, $\left\{\phi_k^U(u) : u \in U; k = 1, \ldots, q\right\}$ and $\left\{\phi_l^V(v) : v \in V; l = 1, \ldots, r\right\}$. Then, we assume that the realizations of the spatio-temporal functional variable X belong to the $qrp-$dimensional tensor function space generated by the basis

$$\left\{\phi_k^U(u)\phi_l^V(v)\phi_h^T(t) : k = 1, \ldots, q; l = 1, \ldots, r; h = 1, \ldots, p\right\}.$$

That is,

$$x(s,t) = \sum_{k=1}^{q}\sum_{l=1}^{r}\sum_{h=1}^{p} a_{klh}\phi_k^U(u)\phi_l^V(v)\phi_h^T(t). \tag{2}$$

This means that for all spatial locations, the associated sample curves belong to the finite-dimension space generated by the basis $\left\{\phi_h^T : h = 1, \ldots, p\right\}$, so that they admit the basis expansion

$$x(s,.) = \sum_{h=1}^{p} a_h(s)\phi_h^T,$$

where the basis coefficients are realizations of a multivariate spatial process given by

$$a_h(s) = \sum_{k=1}^{q}\sum_{l=1}^{r} a_{klh}\phi_k^U(u)\phi_l^V(v).$$

For each time point, the associated sample surfaces belong to the tensor function space generated by the basis $\left\{\phi_k^U\phi_l^V : k = 1, \ldots, q; l = 1, \ldots, r\right\}$ so that can be expressed as

$$x(.,t) = \sum_{k=1}^{q}\sum_{l=1}^{r} a_{kl}(t)\phi_k^U\phi_l^V$$

where the basis coefficients are realizations of a multivariate stochastic process given by

$$a_{kl}(t) = \sum_{h=1}^{p} a_{klh}\phi_h^T(t).$$

Then, the matrix $X = (x_{ij})_{n\times m}$ whose entries are the values of the spatio-temporal functional variable X at the sampling points is given by

$$x_{ij} = x(s_i, t_j) = \sum_{k=1}^{q}\sum_{l=1}^{r}\sum_{h=1}^{p} a_{klh}\phi_k^U(u_i)\phi_l^V(v_i)\phi_h^T(t_j),$$

that can be equivalently written in matrix form as

$$X = \left(\Phi^U \odot \Phi^V\right) A\Phi^{T\prime}, \tag{3}$$

where $\Phi^U = (\Phi^U_{ik})_{n\times q}$ with $\Phi^U_{ik} = \phi^U_k(u_i)$, $\Phi^V = (\Phi^V_{il})_{n\times r}$ with $\Phi^V_{il} = \phi^V_l(v_i)$, $\Phi^T = (\Phi^T_{jh})_{m\times p}$ with $\Phi^T_{jh} = \phi^T_h(t_j)$, $A = \left(a_{(kl)h}\right)_{qr\times p}$ is the matrix comprising the basis coefficients and $\odot$ denotes the row-wise Khatri-Rao product Rao and Rao (1998) so that $\Phi^U \odot \Phi^V = \left(\left(\Phi^U \odot \Phi^V\right)_{i(kl)}\right)_{n\times qr}$ with entries $\left(\Phi^U \odot \Phi^V\right)_{i(kl)} = \phi^U_k(u_i)\phi^V_l(v_i)$.

Taking into account the following property (Harville, 1997)

$$vec(ABC) = (C' \otimes A)vec(B),$$

with $\otimes$ denoting the Kronecker product, the matrix expression (3) can be vectorized as

$$vec(X) = \left(\Phi^T \otimes \left(\Phi^U \odot \Phi^V\right)\right) vec(A).$$

Once the basis coefficients in A are estimated from the discrete observations y_{ij}, the spatio-temporal functional variable can be estimated at unobserved locations and times (s_0, t_0) by using the model

$$\hat{x}(s_0, t_0) = \sum_{k=1}^{q}\sum_{l=1}^{r}\sum_{h=1}^{p} \hat{a}_{klh}\phi_k^U(u_0)\phi_l^V(v_0)\phi_h^T(t_0).$$

This way, we can obtain the complete curve of temporal evolution of the variable for not sampled geographical locations, and the complete surface of spatial evolution of the variable for any time point in the temporal domain.

3 Penalized functional spatial regression model

Let us suppose that we have a sample of non-independent functions (spatial dependence) $\{y_i(t) : t \in T, i = 1, \ldots, n\}$ given by

$$y_i(t) = x(s_i, t) + \epsilon_i(t),$$

which have been observed with error at a finite set of time points $\{t_j : j = 1, \ldots, m\}$ for each geographical location s_i. That is, the sample observations are given by

$$y_{ij} = y_i(t_j) = x(s_i, t_j) + \epsilon_i(t_j), \quad i = 1, \ldots, n; j = 1, \ldots, m.$$

In this work we propose to estimate the basis coefficients in Equation (2) by introducing the spatial variability through the following functional spatial regression model:

$$y(t) = Z\alpha(t) + \epsilon(t),\ \forall t \in T, \tag{4}$$

where $y(t) = (y_1(t), \ldots, y_n(t))'$ is the vector of response functions, $Z = (z_{ik})_{n\times qr} = \Phi^U \odot \Phi^V$ is the two dimensional B-spline basis for the geographical position, $\alpha(t) = (\alpha_1(t), \ldots, \alpha_{qr}(t))'$ is the vector of parameter functions to be estimated and $\epsilon(t) = (\epsilon_1(t), \ldots, \epsilon_n(t))'$ the vector of error terms.

Let us assume a basis representation for the functional response $y(t) = C\theta^T(t)$, with $C = (c_{ij})_{n\times s}$ being the matrix of basis coefficients and $\theta^T(t) = (\theta_1^T(t), \ldots, \theta_s^T(t))'$ the vector of basis functions, and a basis representation for the the functional parameters $\alpha(t) = A\phi^T(t)$, where $A = (a_{(kl)h})_{qr\times p}$ is the matrix of basis coefficients and $\phi(t)^T = (\phi_1^T(t), \ldots, \phi_p^T(t))'$ is the vector of basis functions. Then, the model given in Equation (4) can be rewritten as follows

$$C\theta^T(t) = ZA\phi^T(t) + \epsilon(t),\ \forall t \in T.$$

In order to estimate this model in an accurate way, a roughness penalty is introduced in the least squares fitting criterion, so that

$$\begin{aligned} PSSE(y,\alpha) &= \int (C\theta^T(t) - ZA\phi^T(t))'(C\theta^T(t) - ZA\phi^T(t))dt + \\ &+ PEN_d^T + PEN_d^{U,V}, \end{aligned} \tag{5}$$

where PEN_d^T denotes the d-order penalty for the time and $PEN_d^{U,V}$ is the d-order penalty for the space. Both penalties can be expressed in terms of d-order difference operators Δ_d (Eilers et al., 2006), so that

$$\begin{aligned} PEN_d^T &= vec(A)'[\lambda_1(\Delta_d^{T'}\Delta_d^T \otimes I_q \otimes I_r)]vec(A) \\ PEN_d^{U,V} &= vec(A)'[\lambda_2(I_p \otimes \Delta_d^{U'}\Delta_d^U \otimes I_r) + \lambda_3(I_p \otimes I_q \otimes \Delta_d^{V'}\Delta_d^V)]vec(A). \end{aligned}$$

In this context, $\Delta_d^T, \Delta_d^U, \Delta_d^V$ are matrices of d-order differences, λ_1, λ_2, and λ_3 are the smoothing parameters and the operator $vec(A)$ creates a column vector from any matrix A by stacking the column vectors of A.

Interchanging the integration and summation operations implied by the matrix products, and taking into account the following properties (Harville, 1997)

$$\begin{aligned} vec(A)'(D \otimes B)vec(C) &= trace(A'BCD') \\ trace(A'AB) &= trace(ABA'), \end{aligned} \tag{6}$$

the equation (5) can be re-written as

$$\begin{aligned} PSSE(y,\alpha) &= trace(C'C\Psi^{\theta^T\theta^T}) + trace(Z'ZA\Psi^{\phi^T\phi^T}A') \\ &- 2trace(A\Psi^{\theta^T\phi^T}C'Z) + \lambda_1 trace(A(\Delta_d^{T'}\Delta_d^T)A') \\ &+ \lambda_2 trace(A'(\Delta_d^{U'}\Delta_d^U \otimes I_q)A) + \lambda_3 trace(A'(I_q \otimes \Delta_d^{V'}\Delta_d^V)A), \end{aligned} \tag{7}$$

with $\Psi^{\theta^T\theta^T} = \int \theta^T\theta^T$, $\Psi^{\phi^T\phi^T} = \int \phi^T\phi^T$, and $\Psi^{\theta^T\phi^T} = \int \theta^T\phi^T$ being the inner product matrixes according to the different basis functions. Next step is to compute the derivatives of Equation (7) with respect to A. By considering some matrix algebra properties Harville (1997) (see Appendix for more details), we can see that A satisfies the matrix system of linear equations given by

$$Z'ZA\Psi^{\phi^T\phi^T} + \lambda_1 A(\Delta_d^{T'}\Delta_d^T) + \lambda_2(\Delta_d^{U'}\Delta_d^U \otimes I_r)A + \lambda_3(I_q \otimes \Delta_d^{V'}\Delta_d^V)A = Z'C\Psi'^{\theta^T\phi^T}. \tag{8}$$

In order to get the solution to A, the Kronecker product is used to express Equation (8) in conventional matrix algebra (see Appendix for more details), so that Equation (8) can be re-written as follows

$$\left[\Psi^{\phi^T\phi^T} \otimes (Z'Z) + PEN_d^{T,U,V}\right] vec(A) = vec(Z'C\Psi'^{\theta^T\phi^{T'}}),$$

where $PEN_d^{T,U,V}$ is a P-spline penalty developed by Eilers et al. (2006), which is given by

$$\begin{aligned} PEN_d^{T,U,V} &= \lambda_1\left(\triangle_d^{T'}\triangle_d^T \otimes I_q \otimes I_r\right) + \lambda_2\left(I_p \otimes \triangle_d^{U'}\triangle_d^U \otimes I_r\right) + \\ &+ \lambda_3\left(I_p \otimes I_q \otimes \triangle_d^{V'}\triangle_d^V\right). \end{aligned}$$

Finally, A is given by

$$vec(A) = \left[\Psi^{\phi^T\phi^T} \otimes (Z'Z) + PEN_d^{T,U,V}\right]^{-1} vec(Z'C\Psi'^{\theta^T\phi^T}).$$

4 Smoothing parameters selection

The three smoothing parameters involved in this problem $(\lambda_1, \lambda_2, \lambda_3)$ are simultaneously selected by minimizing the following generalized cross validation error

$$GCVE(\lambda_1, \lambda_2, \lambda_3) = \frac{\sum_{i=1}^n SSE_i}{(n - trace(H))^2},$$

where

$$SSE_i = \sum_{j=1}^m (y(s_i, t_j) - \hat{y}(s_i, t_j))^2$$

and

$$H = (\Phi^T \otimes Z)\left[\Psi^{\phi^T\phi^T} \otimes (Z'Z) + PEN_d^{T,U,V}\right]^{-1}(\Phi^{T'} \otimes Z'),$$

with $PEN_d^{T,U,V}$ being the three-dimensional P-spline penalty of order d described above.

5 Simulation studies

In order to check the forecasting performance of the proposed penalized functional spatial regression model (PFSRM) two different simulation studies have been considered. The first one considers non equally spaced spatial locations and random errors added only at the time dimension. The second one was simulated into a grid of equally spaced spatial sites and random errors were added at the two dimensions (space and time). In addition, the predictions are compared with the ones given by ordinary kriging for functional data (OKFDA).

For each of the proposed methods, a leave one out cross validation procedure is proposed to predict each curve at each spatial site, so that the integrated squared error with respect to the original data can be computed as

$$ISE_i \int_T (x(s_i,t) - \hat{y}^{(-i)}(s_i,t))^2 dt,\ i = 1,\ldots,n,$$

with $\hat{y}^{(-i)}(s_i,t)$ being the predicted curve at location s_i when the observation $y(s_i,t)$ is not in the sample, and n the number of spatial sites. Let us observe that in the applications with real data, the ISEs must be computed with respect to the basis expansion of $y(s_i,t)$ estimated from the observed data.

5.1 Simulation study I

This simulation study was first considered in Giraldo et al. (2012). In our case, 225 sites were fixed in a grid according to the coordinates $u = v = (-20,-16,-15,-10,-8,\ -5,-1,1,2,6,10,12,15,16,20)$, on which a set of spatially correlated functional data were simulated at 365 equally spaced time points according to the model

$$Y(s_i,t) = \sum_{k=1}^{15} a_k(s_i)\phi_k(t) + \epsilon_i(t),\ i:1,\ldots,225,$$

where $\phi(t) = (\phi_1(t),\ldots,\phi_{15}(t))$ is a cubic B-spline basis, and each coefficient a_k is a realization of a Gaussian random field whose covariance structure is defined according to the exponential model $C(h) = 2exp(\frac{-h}{8})$, where $h = \|s_i - s_j\|, (i,j = 1,\ldots,225)$ is the Euclidean distance between two sites s_i and s_j. Finally, $\epsilon_i(t)$ is a random error for each t, with $t = 1,\ldots,365$, simulated according a distribution N(0,0.09). The simulated data sets, with and without noise, can be seen in Figure 2. The spatial sites are shown in Figure 1.

The first step was to approximate the sample paths by using regression splines in terms of a basis defined on 40 equally spaced knots. Due to a P-spline penalty is considered for fitting the model, the selection of the number of basis knots in this part of the problem is not too relevant, only a large number of equally spaced knots is needed to get a good fit (Ruppert, 2002).

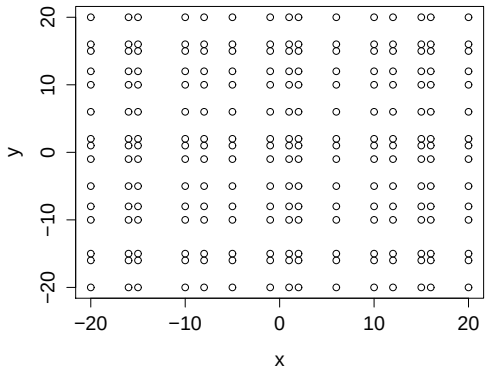


Figure 1: Simulation I: Spatial sites.

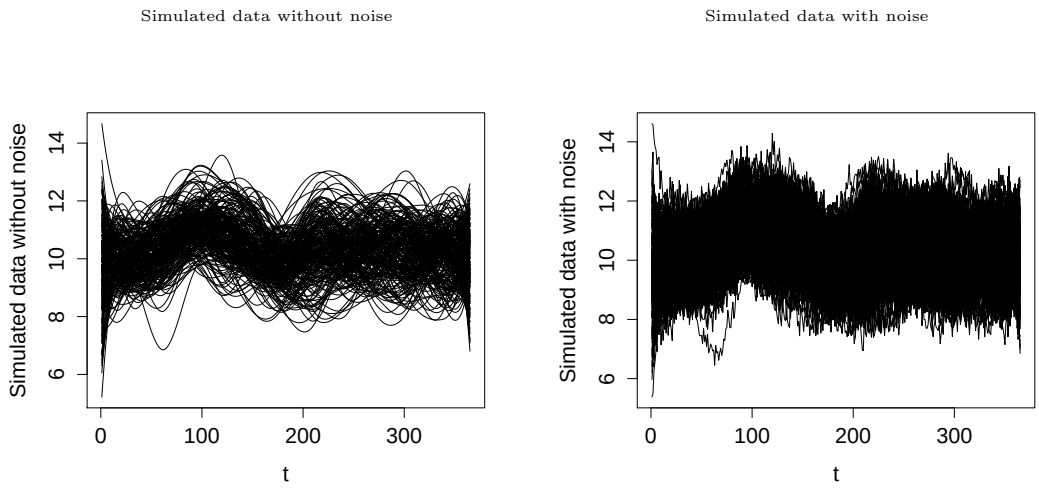


Figure 2: Simulation I: Simulated data without noise (left) and simulated data by adding a random error (right).

	Mean	s.d.	Median	Min	Max	Sum
OKFD	80.09	46.39	68.02	15.33	282.80	18019.13
PFSRM	73.08	42.34	65.08	6.67	268.50	16443.95

Table 1: Simulation I: Summary of ISE's from the cross validation predictions.

In order to check the good performance of the proposed methods, a leave one out cross validation procedure was carried out to obtain the predicted curve at each unvisited spatial site. These predictions are shown in Figure 3. In general, both methods provide good predictions of the true curves, with PFSRM providing the most accurate ones. For two particular cases, the simulated data (with and without noise, grey and black, respectively) together with the predicted curves by OKFD (blue) and PFSRM (red) can be seen in Figure 4. The proposed method PFSRM provides the best results since OKFD loses the trend of the true curves and achieves to non-smooth estimations. On the other hand, the mean functions of the predictions are quite similar in both cases (see Figure 5). However, taking into account the distribution of the ISEs, the differences between the two methods are evident. In Table 1 the main statistics revel that PFSRM provides the lowest ISEs. In Figure 6 the box plots related to the ISEs from OKFD and PFSRM are shown. It is clear that the proposed method presents less variability in its estimations, coming down the median with respect to the OKFD.

5.2 Simulation study II

Now we consider a grid of 225 equally spaced sites which are shown in Figure 8 (top), and a set of 100 equally spaced times at the interval $[0, 1]$. The idea is to simulate a set of spatially correlated functional data according to the model

$$Y(s,t) = [af_1(s,t) + bf_2(s,t) - 0.5]\sin(c\pi * t - 0.2) + \epsilon_1(t) + \epsilon_2(s),$$

where

$$f_1(s,t) = e^{\left(\frac{-(u-0.2)^2}{5} - \frac{(v-0.5)^2}{3} - \frac{(t-0.5)^2}{4} - 1\right)},$$

$$f_2(s,t) = e^{\left(-\frac{(u-0.3)^2}{4} - \frac{(v-0.7)^2}{2} - \frac{(t-0.4)^2}{6}\right)},$$

$a = 1$, $b = 0.9$, $c = 1.2$, $\epsilon_1(t)$ are independent random error at each t $(t = 1, \ldots, 100)$ (for each spatial location $s_i = (u_i, v_i)$, $i = 1, \ldots 225$) simulated according to the distribution $N(0, \sigma_{\epsilon_1})$, and $\epsilon_2(s)$ is for each fixed s, $t = 1, \ldots, 100$ a Gaussian random field for the given covariance parameters $(\sigma^2_{\epsilon_2}, \phi = 0.1)$, so that the covariance function can be written as $C(h) = \sigma^2_{\epsilon_2} \times \rho(h)$, being $\rho(h)$ a positive definite correlation function given by $\rho(h) = exp(-(h/\phi)^2)$ and h the distance between two spatial locations.

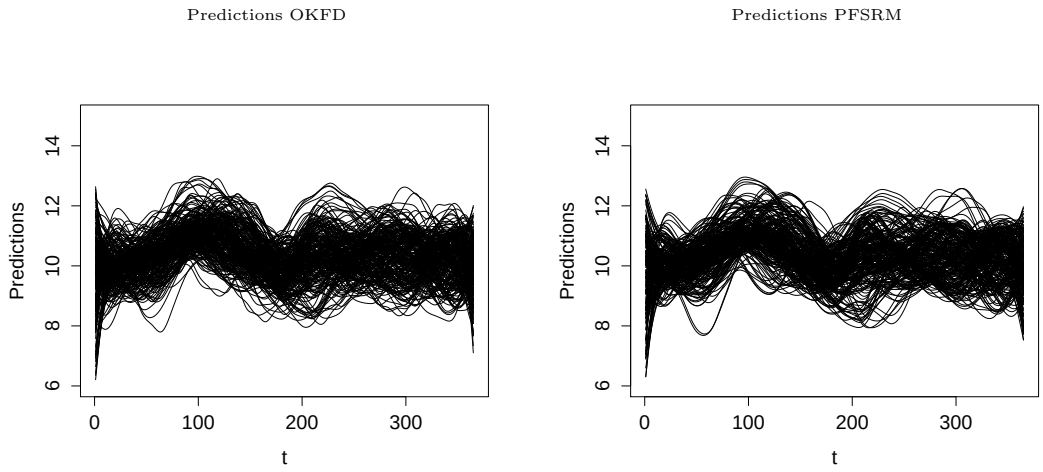


Figure 3: Simulation I: Predicted curves by using OKFD and PFSRM.

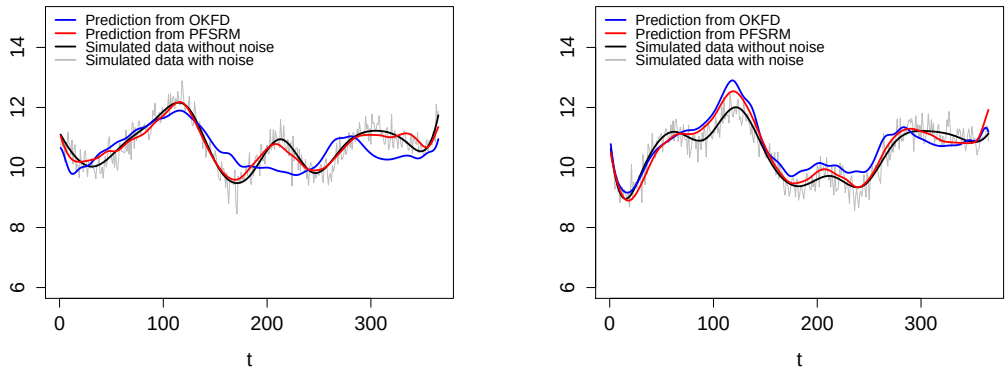


Figure 4: Simulation I: Predictions of two sample curves.

For $\sigma^2_{\epsilon_1} = 0.01$ and $\sigma^2_{\epsilon_2} = 0.05$ the simulated sample paths can be seen in Figure 8. The spatial locations are displayed in Figure 7.

The initial approximation of the sample curves was carried out by using regression splines in terms of a basis defined on 20 equally spaced knots.

In this study the differences between the two methods are more evident than in the first simulation study. According to the predictions obtained with each method (see Figure 9), in both cases the noise was not totally avoided. But it is clear that OKFD loses the trend of the original data, providing in some cases, predictions far from the true ones. On the other hand, PFSRM achieves predictions with similar slope to the original data. In Figure 10 two of the predicted curves for each method are shown. In both cases the predictions from OKFD are noisy and far from the true ones. The mean

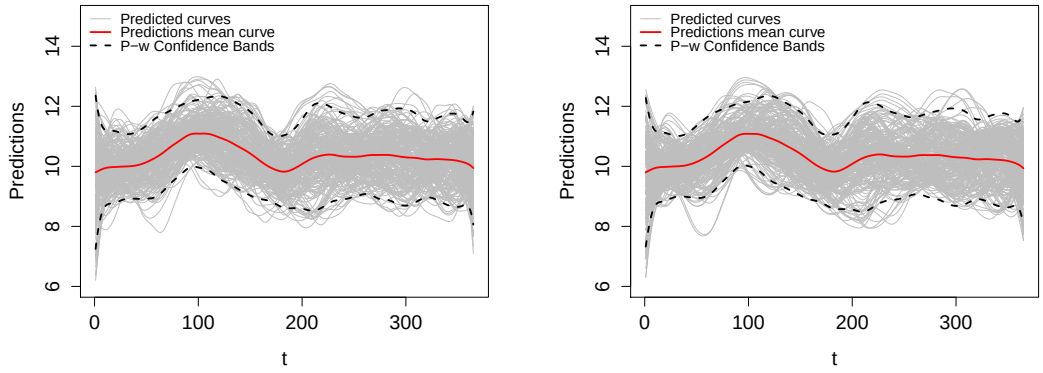


Figure 5: Simulation I: Mean function of the predictions and point wise confidence bands according to the mean $\pm$ 2s.d. by OKFD (left) and PFSRM (right).

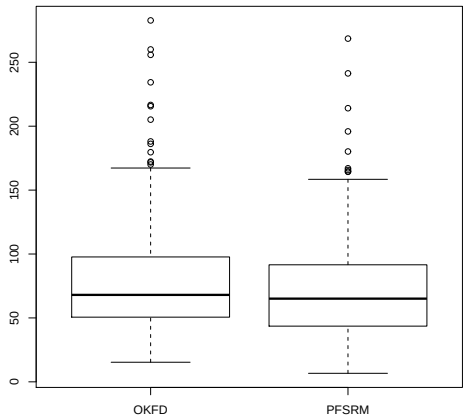


Figure 6: Simulation I: Box plot related to the ISE of the predictions.

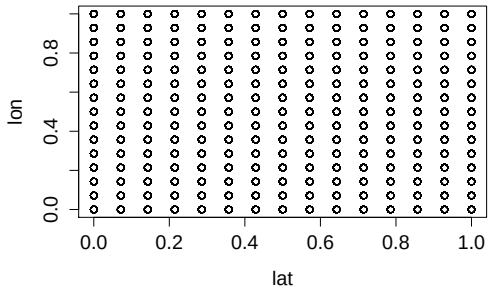


Figure 7: Simulation II: Spatial sites.

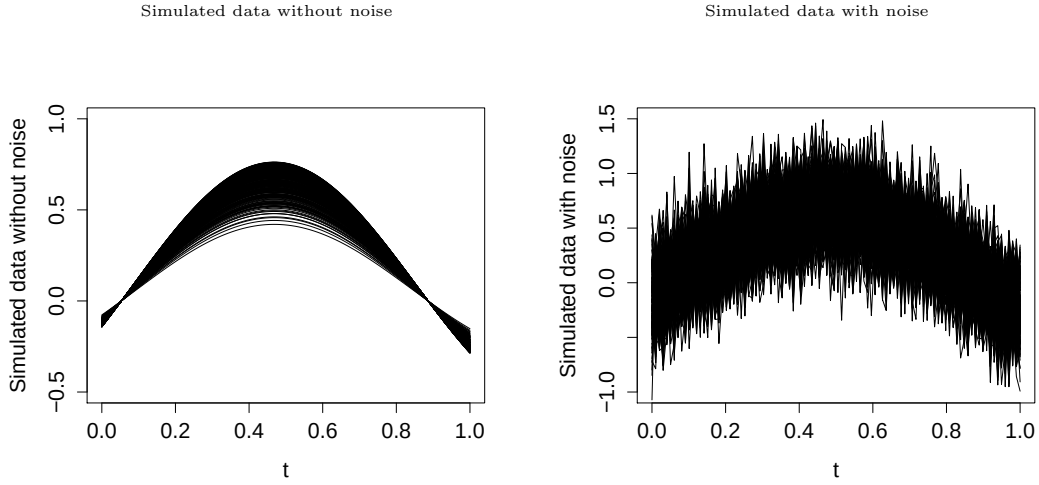


Figure 8: Simulation II: Simulated data without noise (left) and simulated data with noise (right).

functions for the predicted curves, displayed in Figures 11 and 12 (left), do not show clearly these differences. But the summary of statistics in Table 2 and the box plots of the ISEs in Figure 12 highlight the great differences between them. OKFD and PFSRM have similar median, but its variability is completely different. Therefore, we can conclude that PFSRM provides the best results under both scenarios.

6 Application to Canadian maritime weather data

In this study we use averages (over 30 years) of daily temperature curves observed at 35 Canadian Maritime weather stations. This is a clear example

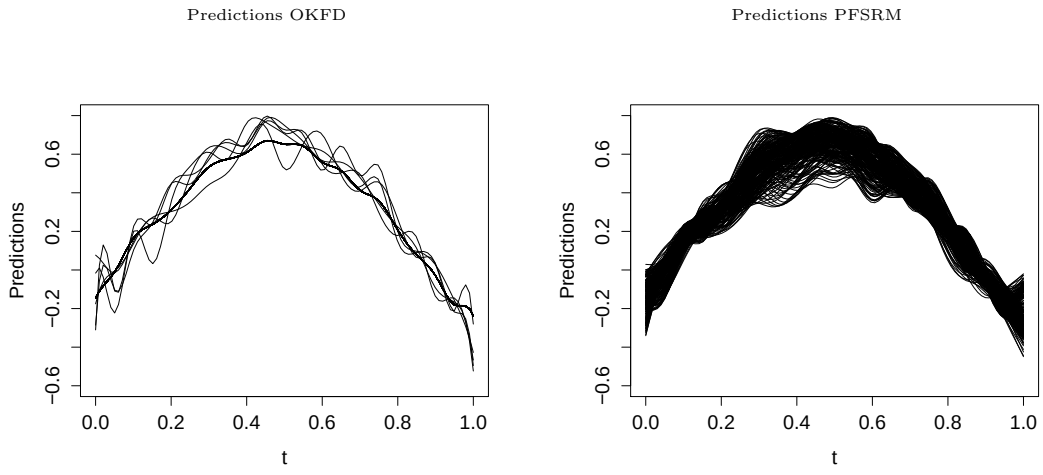


Figure 9: Simulation II: Predicted curves by using OKFD and PFSRM.

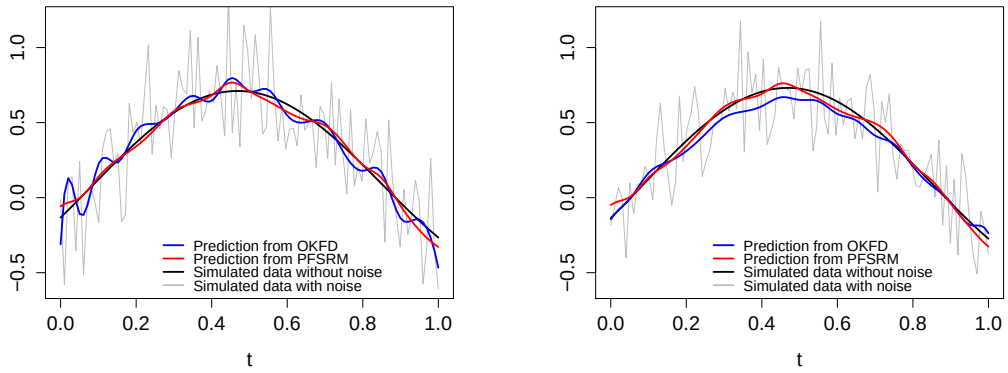


Figure 10: Simulation II: Predictions of two sample curves.

	Mean	s.d.	Median	Min	Max	Sum
OKFD	0.2997	0.3305	0.2017	0.0410	2.3950	67.4437
PFSRM	0.2399	0.0924	0.2248	0.0754	0.5834	53.9865

Table 2: Simulation II: Summary of ISE's from the cross validation predictions.

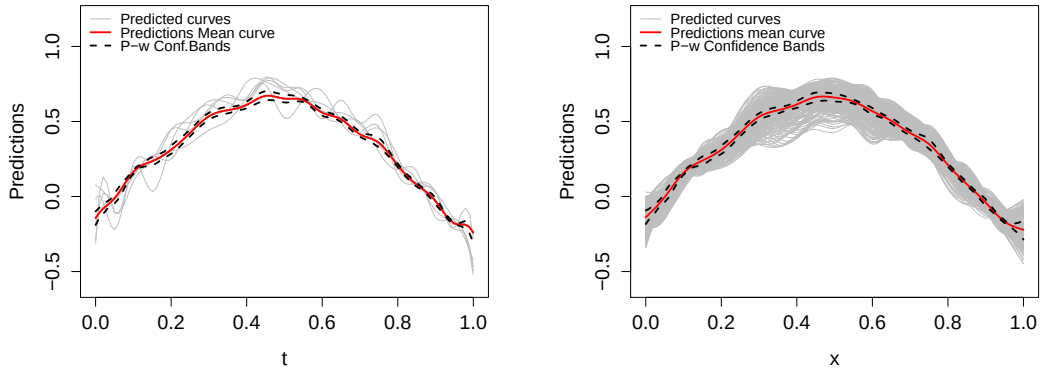


Figure 11: Simulation II: Mean function of the predictions and point wise confidence bands according to the mean ± 2s.d. by OKFD (left) and PFSRM (right).

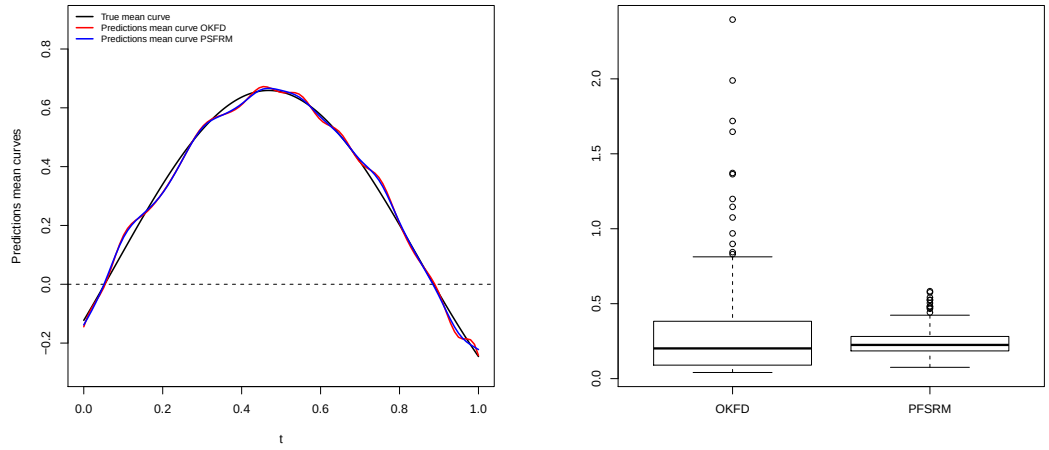


Figure 12: Simulation II: Mean curves of the predictions joint to the true mean function of the simulated smoothed data (left) and box plot related to the ISE of the predictions (right).

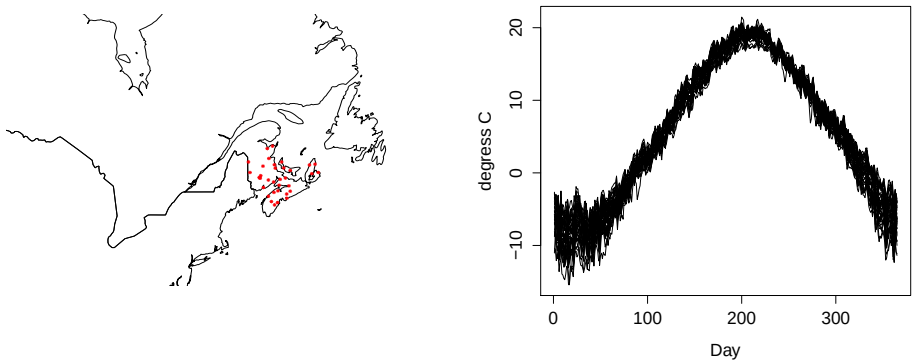
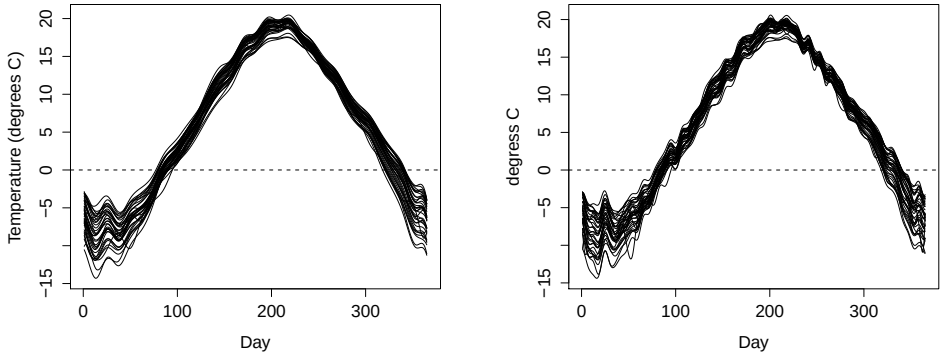


Figure 13: Application: Averages (over 30 years) of daily temperature curves observed at 35 Canadian Maritime weather stations.

of functional data presenting spatial dependence, since curves located at closer geographical locations will be similar to other there are further apart. The raw data set together with the map with the geographical locations are shown in Figure 13.

The first step for both methods is to consider the basis representation of the raw sample paths in terms of cubic B-spline basis functions. In order to get more general conclusions, different number of basis functions have been considered for the initial basis representation of the sample paths, exactly 33 (Case 1) and 65 (Case 2). The regression splines fitted in the two cases are displayed in Figure 14.

In order to get the predicted curve on each geographical site a leave-one-out cross validation procedure was carried out. The predicted curves obtained by OKFD and PFSRM next to their mean curve and point wise confidence bands (according to the mean $\pm$ two times the standard deviation) can be seen in Figure 15. In both cases (Case 1 and Case 2), the spatial basis is made by considering 6 knots for each marginal basis. It can be seen that when the dimension of the basis for fitting the regression splines increases (Case 2), the predictions provided by OKFD are more noisy than in Case 1. By contrast, PFSRM provides similar predictions independently of the number of basis functions used to fit the initial regression splines. In that sense, PFSRM is more robust than OKFD with respect to the dimension of the initial B-spline expansions in the time domain. This is an important advantage of our method, since the selection of the number of initial bases is not as relevant as in functional kriging. For two Canadian Maritime provinces, the predicted temperature curves by OKFD (blue) and PFSRM (red) are plotted together with the observed temperature curves in Figure 16. Independently of the basis dimension, PFSRM provides smoother and more accurate predicted curves than OKFD and also maintains the trend of the raw data.



Case 1 Case 2

Figure 14: Application: Regression splines fitted from the temperature raw data by using 33 and 65 cubic B-spline basis functions (Case 1 and 2, respectively).

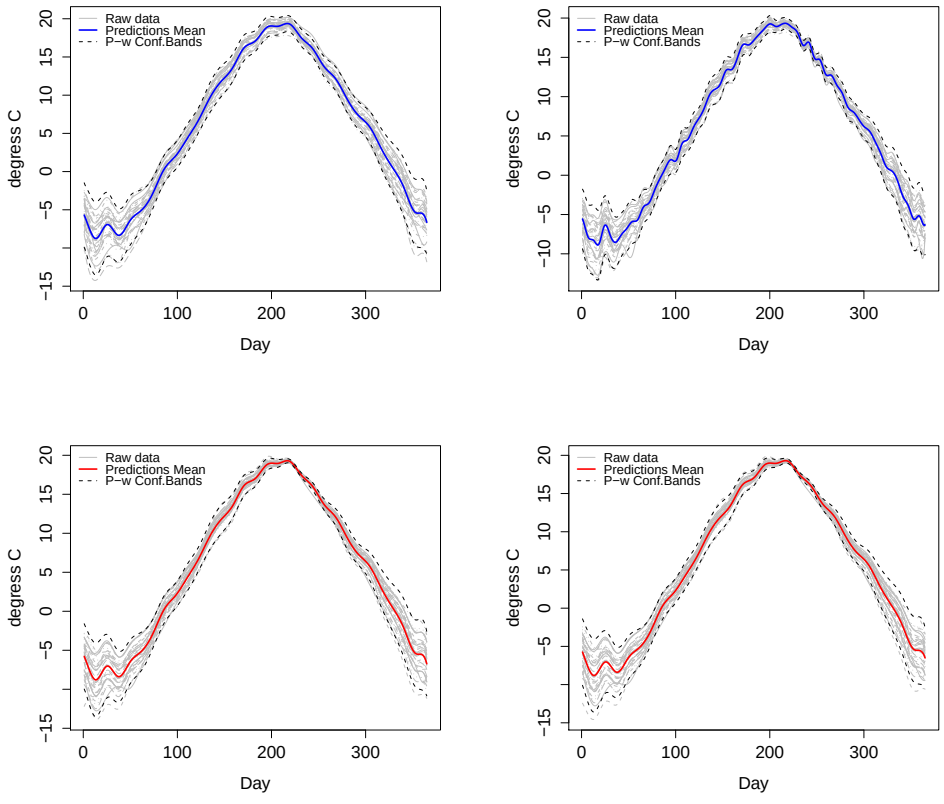
	Case 1		Case 2	
	OKFD	PFSRM	OKFD	PFSRM
Median	320.1	244.5	253.4	240.1
Mean	391.2	307.8	299.5	309.2
s.d.	308.4	181.2	178.4	186.8

Table 3: Application: The median, the mean and the standard deviation of the SSEs (with respect to the observed data) obtained in the cross validation for Cases 1 and 2.

In order to compare the prediction ability of the two methods, the box plots related to the SSE's (with respect to the observed data) obtained by cross validation for Cases 1 and 2 can be seen in Figure 17. Also, the mean, the standard deviation and the median of the 35 SSE's are summarized in Table 3. Again, independently of the dimension of the basis used in the initial regression splines, the lowest values of the Median of the SSE's are always obtained by PFSRM.

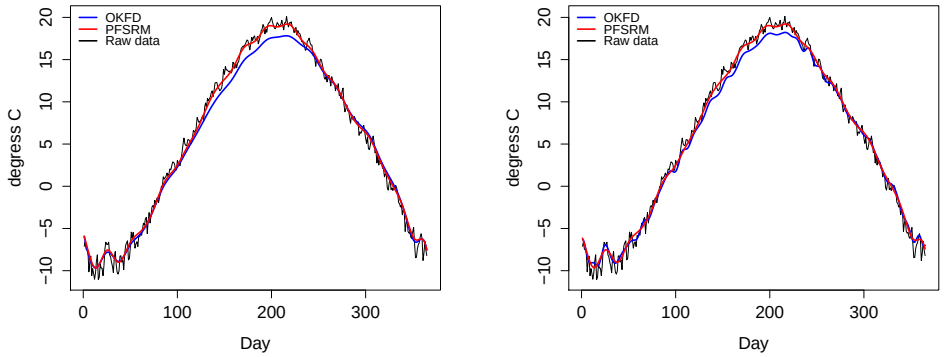
7 Conclusions

The aim of this paper is to provide a new tool to predict spatially dependent functional data as alternative to the geostatistical techniques, such as functional kriging. From a formal definition of spatio-temporal functional data, which was presented in Section 2, a penalized estimation of a functional spatial regression model has been proposed in this paper, by introducing a three-dimensional P-spline penalty at the least squares fitting criterion (Sec-



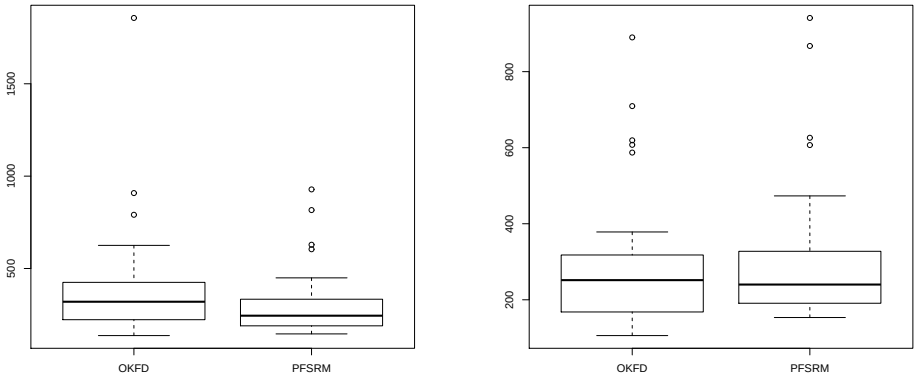
Case 1 Case 2

Figure 15: Application: Predicted curves (grey) by OKFD (at the top) and PFSRM (at the bottom) from the regression splines of the temperature raw data (using 33 (Case 1) and 65 (Case 2) cubic B-spline basis functions) join to its mean curve (blue and red line) and the point wise confidence bands according to the mean ± two times the standard deviation (black and dashed line).



Case 1 Case 2

Figure 16: Application: The predicted curve by OKFD (blue) and PFSRM (red) from the regression splines of the temperature raw data (using 33 (Case 1) and 65 (Case 2) cubic B-spline basis functions) and the observed temperature curve (black) in two of the 35 Canadian Maritimes provinces.



Case 1 Case 2

Figure 17: Application: Box plots related to the SSEs (with respect to the raw data) obtained in the cross validation for Cases 1 and 2.

tion 3).

In order to compare the proposed method with functional kriging on different scenarios, two different simulation schemes have been carried out. The first considers non equally spaced spatial locations and random errors added only at the time dimension. The second one was simulated into a grid of equally spaced spatial sites and random errors were added at all dimensions (space and time). Also, an application to real data has been presented taking into account two cases: Case 1 and Case 2, where 33 and 65 basis functions were respectively used to fit the initial regression splines. From the two simulation studies, it is clear that PFSRM provides the most accurate predictions with less variability and coming down the ISEs median with respect to OKFD. The second simulation study highlights the differences according to the variability in their predictions, so that OKFD loses the trend of the original data by providing predictions that, in some cases, are far from the true. The application revels that PFSRM is more robust that OKFD, in the sense that the last one is quite sensitive to the number of the basis function to be used in the initial fit of the functional data (a lower number of basis functions leads to worse predictions with higher ISEs).

Summarizing, we can be concluded that the proposed PFSRM is a robust and a computationally efficient alternative to the existing geostatistical techniques in order to predict functional data with spatial dependence.

Acknowledgements

This research has been funded by project P11-FQM-8068 from Consejería de Innovación, Ciencia y Empresa. Junta de Andalucía, Spain and the projects MTM2013-47929-P and MTM 2011-28285-C02-C2 from Secretaría de Estado Investigación, Desarrollo e Innovación, Ministerio de Economía y Competitividad, Spain.

References

Aguilera, A. M. and Aguilera-Morillo, M. C. (2013a). Comparative study of different b-spline approaches for functional data. *Mathematical and Computer Modelling*, 58:1568–1579.

Aguilera, A. M. and Aguilera-Morillo, M. C. (2013b). Penalized pca approaches for b-spline expansions of smooth functional data. *Applied Mathematics and Computation*, 219:7805–7819.

AguileraMorillo, M. C., Aguilera, A. M., Escabias, M., and Valderrama, M. J. (2013). Penalized spline approaches for functional logit regression. *Test*, 22(2):251–277.

Chiou, J. M., Müller, H. G., and Wang, J. L. (2004). Functional response models. *Statistica Sinica*, 14:659–677.

Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2009). Statistics for spatial functional data: some recent contributions. *Environmetrics*, 21:224–239.

Eilers, P. and Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121.

Eilers, P. H. C., Currie, I., and Durban, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, 50:61–76.

Faraway, J. J. (1997). Regression Analysis for a Functional Response. *Technometrics*, 39(3):254–261.

Giraldo, R. (2009). *Geostatistical Analysis of Functional Fata, PhD thesis*. Universitat Politècnica de Catalunya.

Giraldo, R., Delicado, P., and Mateu, J. (2010). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(1):66–82.

Giraldo, R., Delicado, P., and Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, 18(3):411–426.

Giraldo, R., Mateu, J., and Delicado, P. (2012). geofd: An R Package for Function-Valued Geostatistical Prediction. *Revista Colombiana de Estadstica*, 35(3):385–407.

Goulard, M. and Voltz, M. (1993). Geostatistical interpolation of curves: A case study in soil science. In *Geostatistics Tria 92*, volume 5 of *Quantitative Geology and Geostatistics*, pages 805–816. Springer Netherlands.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer-Verlag New York, Inc.

Lee, D. and Durban, M. (2009). Smooth-car mixed models for spatial count data. *Computational Statistics and Data Analysis*, 53:2968–2977.

Lee, D. and Durban, M. (2011). Pspline anova type interaction models for spatio temporal smoothing. *Statistical Modelling*, 11:49–69.

Marx, B. D. and Eilers, P. H. C. (1999). Generalized linear regression on sampled signals and curves. a p-spline approach. *Technometrics*, 41(1):1–13.

Menafoglio, A., Secchi, P., and Dalla Rosa, M. (2013). A universal kriging predictor for spatially dependent functional data of a hilbert space. *Electronic Journal of Statistics*, 7:2209–2240.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis (Second Edition)*. Springer-Verlag.

Rao, C. and Rao, M. (1998). *Matrix Algebra and Its Applications to Statistics and Econometrics*. World Scientific Publishing Co. Pte. Ltd.

Reiss, P. T., Huang, L., and Mennes, M. (2010). Fast Function-on-Scalar Regression with Penalized Basis Expansions. *The International Journal of Biostatistics*, 6(1):1–28.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11:735–757.

Sangalli, L., Ramsay, J., and Ramsay, T. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society Ser. B, Statistical Methodology*, 75(4):1–23.

Ugarte, M. D., Goicoa, T., Militino, A. F., and Durban, M. (2009). Spline smoothing in small area trend estimation and smoothing. *Computational Statistics and Data Analysis*, 53:3616–3629.

Appendix

Derivatives of Equation (7) with respect to A

By considering the following properties (Harville, 1997)

$$\frac{\partial trace(XAX')}{\partial X} = X(A+A') \quad (9)$$

$$\frac{\partial trace(X'AX)}{\partial X} = (A+A')X \quad (10)$$

$$\frac{\partial trace(XA)}{\partial X} = A' \quad (11)$$

we have that

$$\begin{aligned}
\frac{\partial trace(C'C\Psi^{\theta^T\theta^T})}{\partial A} &= 0 \\
\frac{\partial trace(Z'ZA\Psi^{\phi^T\phi^T}A')}{\partial A} &\overset{(9)}{=} Z'ZA(\Psi^{\phi^T\phi^T}+\Psi'^{\phi^T\phi^T}) \\
&\overset{(symmetry)}{=} 2Z'ZA\Psi^{\phi^T\phi^T} \\
\frac{\partial -2trace(A\Psi^{\theta^T\phi^T}C'Z)}{\partial A} &\overset{(11)}{=} -2Z'C\Psi'^{\theta^T\phi^T} \\
\frac{\partial \lambda_1 trace(A(\Delta_d^{T'}\Delta_d^T)A')}{\partial A} &\overset{(9)}{=} 2\lambda_1 A(\Delta_d^{T'}\Delta_d^T) \\
\frac{\partial \lambda_2 trace(A'(\Delta_d^{U'}\Delta_d^U\otimes I_r)A)}{\partial A} &\overset{(10)}{=} 2\lambda_2(\Delta_d^{U'}\Delta_d^U\otimes I_r)A \\
\frac{\partial \lambda_3 trace(A'(I_q\otimes\Delta_d^{V'}\Delta_d^V)A)}{\partial A} &\overset{(10)}{=} 2\lambda_3(I_q\otimes\Delta_d^{V'}\Delta_d^V)A.
\end{aligned}$$

Then, A satisfies the matrix system of linear equations given by

$$Z'ZA\Psi^{\phi^T\phi^T}+\lambda_1 A(\Delta_d^{T'}\Delta_d^T)+\lambda_2(\Delta_d^{U'}\Delta_d^U\otimes I_r)A+\lambda_3(I_q\otimes\Delta_d^{V'}\Delta_d^V)A = Z'C\Psi'^{\theta^T\phi^T}.$$

Using the Kronecker product to express Equation (8) in conventional matrix algebra

$$\begin{aligned}
vec(Z'ZA\Psi^{\phi^T\phi^T}) &\overset{(6)}{=} (\Psi^{\phi^T\phi^T}\otimes(Z'Z))vec(A) \\
vec(\lambda_1 A(\Delta_d^{T'}\Delta_d^T)) &= vec(\lambda_1(I_q\otimes I_r)A(\Delta_d^{T'}\Delta_d^T)) \\
&\overset{(6)}{=} \lambda_1(\Delta_d^{T'}\Delta_d^T\otimes I_q\otimes I_r)vec(A) \\
vec(\lambda_2(\Delta_d^{U'}\Delta_d^U\otimes I_r)A) &= vec(\lambda_2(\Delta_d^{U'}\Delta_d^U\otimes I_r)AI_p) \\
&\overset{(6)}{=} \lambda_2(I_p\otimes\Delta_d^{U'}\Delta_d^U\otimes I_r)vec(A) \\
vec(\lambda_3(I_q\otimes\Delta_d^{V'}\Delta_d^V)A) &= vec(\lambda_3(I_q\otimes\Delta_d^{V'}\Delta_d^V)AI_p) \\
&\overset{(6)}{=} \lambda_3(I_p\otimes I_q\otimes\Delta_d^{V'}\Delta_d^V)vec(A).
\end{aligned}$$

Then, Equation (8) can be re-written as follows

$$\left[\Psi^{\phi^T\phi^T} \otimes (Z'Z) + PEN_d^{T,U,V}\right] vec(A) = vec(Z'C\Psi'^{\theta^T\phi^T}).$$