

Prediction of functional data with spatial dependence: a penalized approach

M. Carmen Aguilera-Morillo¹ · María Durbán¹ · Ana M. Aguilera²

Published online: 5 February 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract This paper is focus on spatial functional variables whose observations are a set of spatially correlated sample curves obtained as realizations of a spatio-temporal stochastic process. In this context, as alternative to other geostatistical techniques (kriging, kernel smoothing, among others), a new method to predict the curves of temporal evolution of the process at unsampled locations and also the surfaces of geographical evolution of the variable at unobserved time points is proposed. In order to test the good performance of the proposed method, two simulation studies and an application with real climatological data have been carried out. Finally, the results were compared with ordinary functional kriging.

Keywords Spatial functional data · Spatial correlation · P-spline penalty · Functional regression

1 Introduction

Functional data analysis (FDA) is currently a very active statistical research topic from both the theoretical and applied viewpoint. The functional data are set of functions obtained as independent realizations of a functional random variable that takes values in a functional space defined on a continuous domain. In most cases functional data observations are curves that correspond to the evolution of a scalar variable over time but also surfaces representing the evolution of a scalar variable on space can be seen as

functional data. The name FDA is due to the pioneer book by Ramsay and Silverman (1997) that contains an excellent collection of the main FDA methodologies and interesting motivating examples. From that moment FDA had a brilliant development with more than 8000 references in academic google during the twenty first century. Although the key tool in FDA is still principal component analysis, other techniques of multivariate analysis like canonical correlation, discriminant and cluster analysis were also investigated. Recently, research on FDA is oriented to regression models, non-parametric estimation, robust estimation, Bayesian estimation, inference and so on. We can say that almost any statistical method is being extended for analyzing functional data. At the same time new books appear on these FDA topics (Ferraty and Vieu 2006; Horvath and Kokoszka 2012; Hsing and Eubank 2015; Shi and Choi 2011; Zhang 2013).

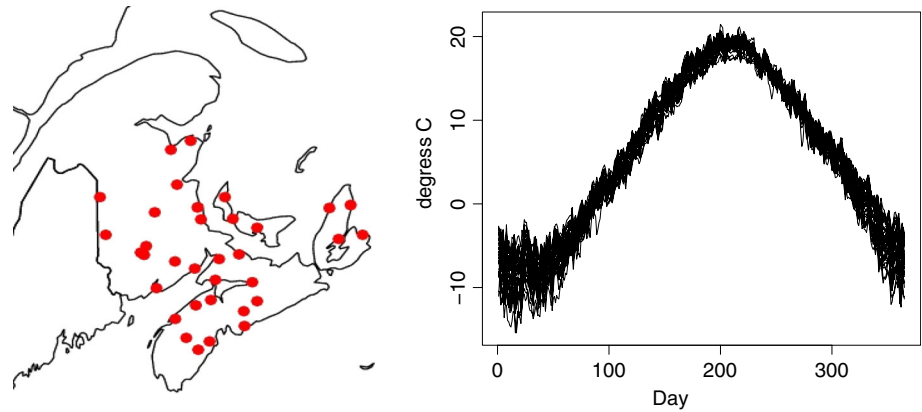
This work is focused on the analysis of univariate functional data with spatial dependence. The sample information is given by a set of curves associated to different geographical locations on a spatial domain. Let us consider as an illustration the Canadian temperature data set introduced by Ramsay and Silverman (1997) that has been analyzed in a lot of papers by using different FDA methodologies. These data are the daily temperature (averaged over 30 years) for 35 weather stations in Canada so that we have a set of 35 spatially correlated curves that represent the temporal evolution of temperatures in different geographical sites (spatio-temporal functional variable). The raw data set together with the map with the geographical locations are shown in Fig. 1. Spatially correlated functional data are very common in environmental applications and are analyzed in many cases by using FDA approaches that does not take into account the spatial dependence structure. Some interesting applications of

✉ M. Carmen Aguilera-Morillo
maguiler@est-econ.uc3m.es

¹ University Carlos III de Madrid, Leganés, Madrid, Spain

² University of Granada, Granada, Spain

Fig. 1 Averages (over 30 years) of daily temperature curves observed at 35 Canadian Maritime weather stations



these type can be seen in Escabias et al. (2005), Zhang and Chen (2007) and Kaufman and Sain (2010).

The aim of this work is to develop a method for modeling the mean of a spatio-temporal functional variable from its discrete observations at a finite set of time points and locations in the temporal and spatial domains, respectively. The estimated model will provide the prediction of the curves of temporal evolution of the variable at unobserved locations and also the prediction of the surface of geographical evolution of the variable at unobserved time points.

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 unsampled spatial sites. A more recent collection of geostatistical tools for spatial functional data can be seen in Giraldo (2010) and 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* (Giraldo et al. 2012). Recently, different versions of universal kriging predictor for functional data with spatial dependence were proposed in Menafoglio (2013) and Caballero et al. (2013). Kriging with external drift has also been extended for introducing exogenous variables with spatially correlated functional data (Ignaccolo et al. 2014).

In the context of spatial data an alternative to geostatistical techniques are the nonparametric 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

FDA 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. A wavelet regression approach for estimating the field of ocean temperature at different depths is introduced in Fernandez-Pascual (2015). On the other hand, functional approaches based on autoregressive Hilbertian processes were considered in Ruiz-Medina et al. (2012, 2014).

Univariate kriging and spline smoothers were compared in several papers without reaching an unanimous conclusion. The major objection to kriging is the assumption of stationarity that could not be right for some types of spatial structure. From simulations where spline regression predicts better than kriging when the data contains trends of various types, some papers conclude that non-parametric regression is more robust than kriging because takes into account spatial structure that geostatistics does not (Yakowitz and Szidarovsky 1985). Other papers conclude that kriging never performs worse than splines and has the potential to outpredict splines when data are not sampled on a grid (Dubrule 1984; Laslett 1994).

Our aim is to use spatial smoothing regression techniques within a functional regression approach to provide a new method to predict functional data with spatial dependence at unsampled locations. From the formal definition of spatio-temporal functional data, which is given in Sect. 2, a penalized functional regression model is extended for predicting spatially correlated functional data in Sect. 2.1. The idea is to consider the functional regression model for functional response and scalar covariates

(Faraway 1997; Ramsay and Silverman 1997; Chiou et al. 2004) by using the spatial information as regressors. So, a mixture of functional regression model for functional response and penalized spline 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 (1997) 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 (2010) used a penalized generalized least squares criterion based on a basis representation. In this paper, we will adapt the idea developed in Eilers et al. (2006), and combine the two-dimensional penalty used for spline spatial regression with the one proposed in Ramsay and Silverman (1997) to obtain a three dimensional P-spline penalty. Hereinafter, this method will be called penalized functional spatial regression model (PFSRM).

Finally, the prediction accuracy of PFSRM is compared with OKFD in two simulation studies in Sect. 4. An application to the Canadian Maritime weather data is presented in Sect. 5. As we have said before, 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 1960–1994 (Ramsay and Silverman 1997). But this is a clear example of functional data presenting spatial dependence and in this sense was studied in Delicado et al. (2009) and Menafoglio (2013). The conclusions about these studies close the paper in Sect. 6.

2 Theoretic framework

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

$$y_i(t) = x(s_i, t) + \epsilon_i(t), \quad i = 1, \dots, n,$$

where $\epsilon_i(t)$ are zero mean random errors and $x(s_i, t)$ are observations of a spatial functional variable (stochastic process)

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

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) , $X(s, t)$ is a real random variable defined on a probabilistic space (Ω, \mathcal{A}, P) .

In addition, these sample curves have been observed with error at a finite set of time points $\{t_j: j = 1, \dots, m\}$ for each geographical location $s_i = (u_i, v_i)$, so that, the sample data y_{ij} are given by $y_j = y_i(t_j), i = 1, \dots, n; j = 1, \dots, m$.

Let us also assume that the realizations of this functional variable 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 $L^2(U \times V \times T)$ with the usual scalar product given by

$$\langle f, g \rangle = \int_U \int_V \int_T f(u, v, t)g(u, v, t)dudvdt, \\ \forall f, g \in L^2(U \times V \times T).$$

In order to reconstruct the true functional form of the data from discrete spatio-temporal observations, we extend the usual basis expansion approach for representing curves in FDA to the case of spatio-temporal functions that depend on three continuous arguments.

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

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

That is,

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

Then, the matrix $X = (x_{ij})_{n \times m}$ whose entries are the values of the spatio-temporal functional variable at the sampling points given by $x_{ij} = x(s_i, t_j)$ can be written in matrix form as

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

where $\Phi^U = (\Phi_{ik}^U)_{n \times p}$ with $\Phi_{ik}^U = \phi_k^U(u_i)$, $\Phi^V = (\Phi_{il}^V)_{n \times q}$ with $\Phi_{il}^V = \phi_l^V(v_i)$, $\Phi^T = (\Phi_{jh}^T)_{m \times r}$ with $\Phi_{jh}^T = \phi_h^T(t_j)$, $A = (a_{(kl)h})_{pq \times r}$ is the matrix comprising the basis coefficients and \odot denotes the row-wise Khatri–Rao product so that $\Phi^U \odot \Phi^V = ((\Phi^U \odot \Phi^V)_{i(kl)})_{n \times pq}$ with entries $(\Phi^U \odot \Phi^V)_{i(kl)} = \phi_k^U(u_i)\phi_l^V(v_i)$ (Rao and Rao 1998).

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 replacing in model (2). This way, we can obtain the complete curve of temporal evolution of the variable for unsampled geographical locations, and the complete surface of spatial evolution of the variable for any

time point in the temporal domain.

2.1 Penalized functional spatial regression model

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

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

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

Let us consider the basis representation for the functional response $y(t) = C\phi^T(t)$ and for the functional parameters $\alpha(t) = A\phi^T(t)$, with $C = (c_{ih})_{n \times r}$ and $A = (a_{(kl)h})_{pq \times r}$ being the corresponding matrices of basis coefficients and $\phi^T(t) = (\phi_1^T(t), \dots, \phi_r^T(t))'$ being the vector of basis functions. Then, the model given in Eq. (4) can be rewritten as follows

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

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

$$PSSE(y, \alpha) = \int (C\phi^T(t) - ZA\phi^T(t))' (C\phi^T(t) - ZA\phi^T(t)) dt + \text{vec}(A)' [PEN_d^{U,V,T}] \text{vec}(A), \tag{5}$$

where the operator $\text{vec}(A)$ creates a column vector from any matrix A by stacking the column vectors of A, and $PEN_d^{U,V,T}$ denotes the d-order P-spline penalty for the space and time. This penalty can be expressed in terms of d-order difference operators Δ_d (Eilers et al. 2006), so that

$$PEN_d^{U,V,T} = \lambda_1 (\Delta_d^U \Delta_d^U \otimes I_q \otimes I_r) + \lambda_2 (I_p \otimes \Delta_d^V \Delta_d^V \otimes I_r) + \lambda_3 (I_p \otimes I_q \otimes \Delta_d^T \Delta_d^T).$$

In this context, $\Delta_d^U, \Delta_d^V, \Delta_d^T$ are matrices of d-order differences, λ_1, λ_2 , and λ_3 are the smoothing parameters.

Interchanging the integration and summation operations implied by the matrix products, and computing the derivatives with respect to A in the resulting equation (see Appendix for further details), finally A is given by

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

where $\Psi = \int \phi^T \phi^T$ is the inner product matrix between the basis functions.

3 Selection of parameters

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

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

where

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

and

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

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

Minimization of the GCVE can become computationally demanding in this case, since we need to search for three smoothing parameters. In order to speed up the computational burden, instead of using an optimization routine, we selected a 3d-array and performed a grid search. We also checked the performance of other criteria such as AIC and BIC, and we found that, in this case BIC tended to over-smooth the spatial component of the model and AIC performed as well as GCVE.

On the other hand, the dimension of the basis in the three spatio-temporal directions must also be selected. Taking into account that the degree of smoothing is controlled by the smoothing parameter, the number and location of knots is not crucial for fitting a P-spline. Generally, the knots of a P-spline are equally spaced and the number of knots must be sufficiently large to fit the data and not so large that computation time is unnecessarily high. Two algorithms for automatic selection of the number of knots by using generalized cross validation were considered in Ruppert (2002). In general, authors select the dimensions of the basis on the rule considered by Ruppert (2002), which proposes to use one definition knot by each five observation knots, approximately.

4 Simulation studies

In order to test the good performance of the proposed PFSRM, two different simulation studies have been developed. The first one considers non equally spaced

spatial locations on a grid and independent random errors. The second one was simulated by considering non-regular spatial locations and the random errors were added at the two dimensions (space and time) through a spatio-temporal Gaussian process. In addition, the results are compared with a powerful geostatistical predictor, the OKFD developed by Giraldo et al. (2011). Let us observe that with both methods, PFSRM and OKFD, the first step is to approximate the true functional form of the sample curves in terms of basis functions. In this paper we will use regression splines in terms of cubic B-splines basis functions.

For each method, a leave-one-out cross validation procedure is considered to predict each curve at each spatial location. The integrated squared error of prediction, with respect to the original data, can be computed as

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

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.

4.1 Simulation study I

This simulation study was first considered in Giraldo et al. (2012). In our case, 225 spatial locations 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

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

where $\phi(t) = (\phi_1(t), \dots, \phi_{15}(t))$ 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 $h = \|s_i - s_j\|$, ($i, j = 1, \dots, 225$) is the Euclidean distance

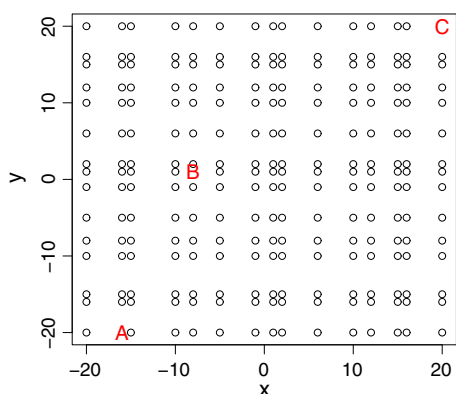


Fig. 2 Simulation I: spatial locations

between two sites s_i and s_j . Finally, $\epsilon(s_i, t)$ are independent random errors for each t , with $t = 1, \dots, 365$, simulated according to a distribution $N(0, 0.09)$. The spatial locations are shown in Fig. 2. The simulated data sets, with and without noise, can be seen in Fig. 3.

The first step for applying both methods, PFSRM and OKFD, was to approximate the sample curves by using regression splines in terms of a cubic B-spline basis of dimension 15 and considering equally spaced knots. As an example, a sample path without noise (dashed line) together with the noisy sample path (grey line) and its basis representation (solid black line) are displayed in Fig. 4. In this case, regression splines on 15 basis functions get a perfect approximation to the true data (without noise).

In the model fitting (second step), the two-dimensional basis for the space was achieved by considering 6 basis knots for each marginal cubic B-spline basis. Regard to the penalty, a 2-order penalty has been considered. For OKFD, the variograms were selected as a linear combination of nugget and exponential models.

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 unsampled spatial location. From the multiple box plot related to the distribution of the ISE's (Fig. 5) and the statistics summary given in Table 1, it can be concluded that PFSRM achieves the lowest values for the mean, the standard deviation and the median of the prediction errors. On the other hand, in Fig. 6 the 225 predicted curves are displayed joint to the mean of the prediction curves (red lines) and the point wise confidence bands according to the mean ± 2 times the standard deviation.

In order to test if the performance varies depending on the spatial location, the predicted curves at two outlying spatial locations (A and C) and a more central spatial location (B) have been displayed in Fig. 7 (locations A–C are highlighted in Fig. 2). We can see that PFSRM provides the best predicted curves closer to the shape of the true data, independently of the predicted spatial location. It is interesting to highlight that the worst prediction from OKFD was in the most distant location C. So, the performance of OKFD is not equal across the grid areas.

Finally, in Fig. 8 the residual curves provided by the two compared methods have been displayed. It can be seen that the mean residual curve is zero in both cases, and in some spatial locations OKFD achieves larger residuals than our method.

4.2 Simulation study II

Let us now consider a set of 80 non regular spatial locations, which are displayed in Fig. 9, and a set of 100

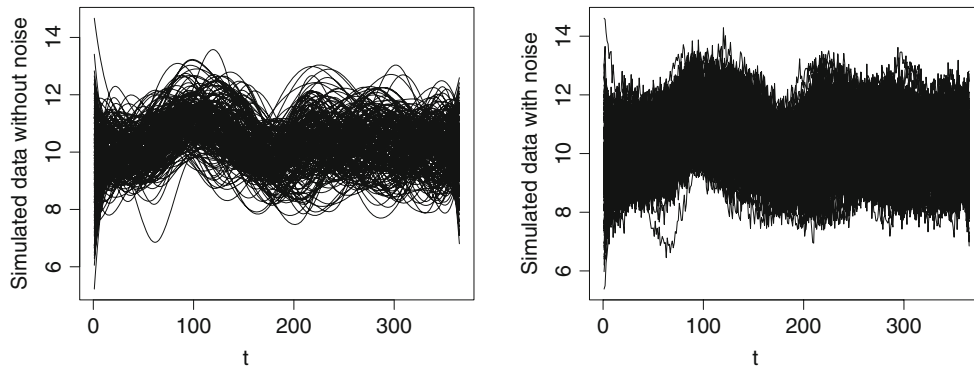


Fig. 3 Simulation I: simulated data without noise (left) and with noise (right)

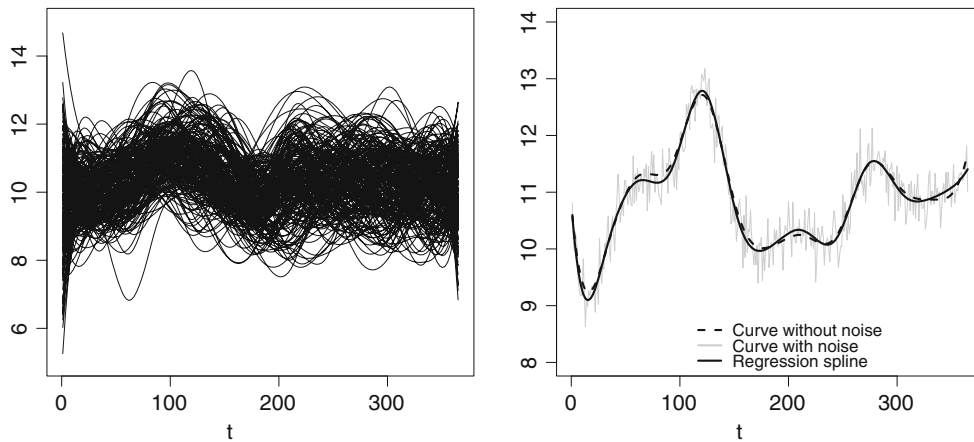


Fig. 4 Simulation I: basis representation of the sample curves by using regression splines in terms of a basis of cubic B-splines of dimension 15 (left). A sample path without noise (dashed line) together with the noisy sample path (grey line) and its basis representation (solid line) (right)

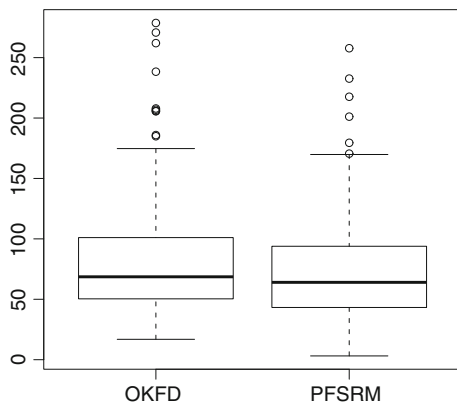


Fig. 5 Simulation I: box plot related to the ISE's of the predictions provided by OKFD and PFSRM

equally spaced times at the interval $[0, 1]$. The idea is to simulate a set of spatially correlated functional data according to the model

$$Y(s_i, t) = [a_i f_1(s_i, t) + b_i f_2(s_i, t) - 0.5] \sin(c_i \pi * t - 0.2) + \epsilon(s_i, t), \quad i = 1, \dots, 80,$$

Table 1 Simulation I: summary of ISE's from the cross-validation predictions provided by OKFD and PFSRM

	Mean	s.d.	Median	Min.	Max.	Sum
OKFD	80.62	47.13	68.60	16.87	278.70	18,139.18
PFSRM	72.94	42.82	64.02	3.132	257.80	16,411.87

where

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

$$f_2(s, t) = e^{\left(-\frac{(u-0.3)^2}{4} - \frac{(v-0.7)^2}{2} - \frac{(t-0.4)^2}{6} - 1\right)},$$

with $s = (u, v)$ denoting the pair of coordinates of the spatial locations, a_i, b_i , and c_i randomly simulated from $a \sim \text{Uniform}[0.5, 2]$, $b \sim \text{Uniform}[0.5, 1]$, and $c \sim \text{Uniform}[1.5, 2]$, and $\epsilon(s, t)$ being the error term corresponding to an observation of a spatio-temporal Gaussian process defined through a stochastic partial differential equation (Sigrist et al. 2015), with parameters $(\rho_0 = 0.1,$

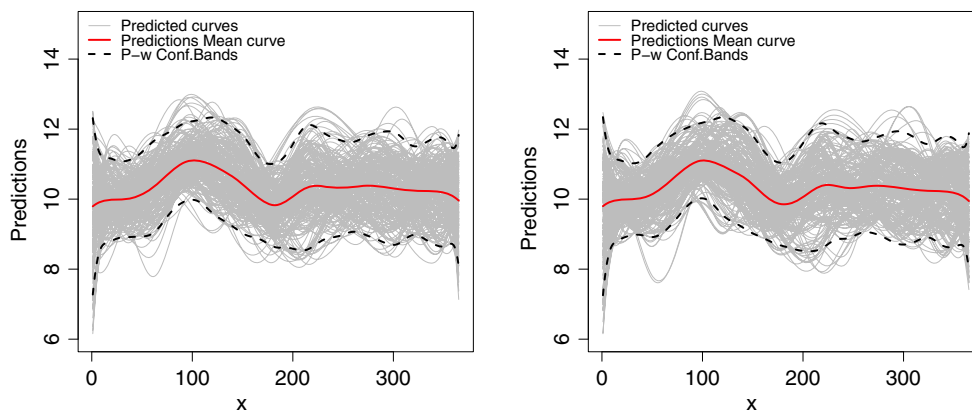


Fig. 6 Simulation I: mean curve of the predictions and point wise confidence bands according to the mean ± 2 times the standard deviation of the predicted curves by OKFD (left) and PFSRM (right)

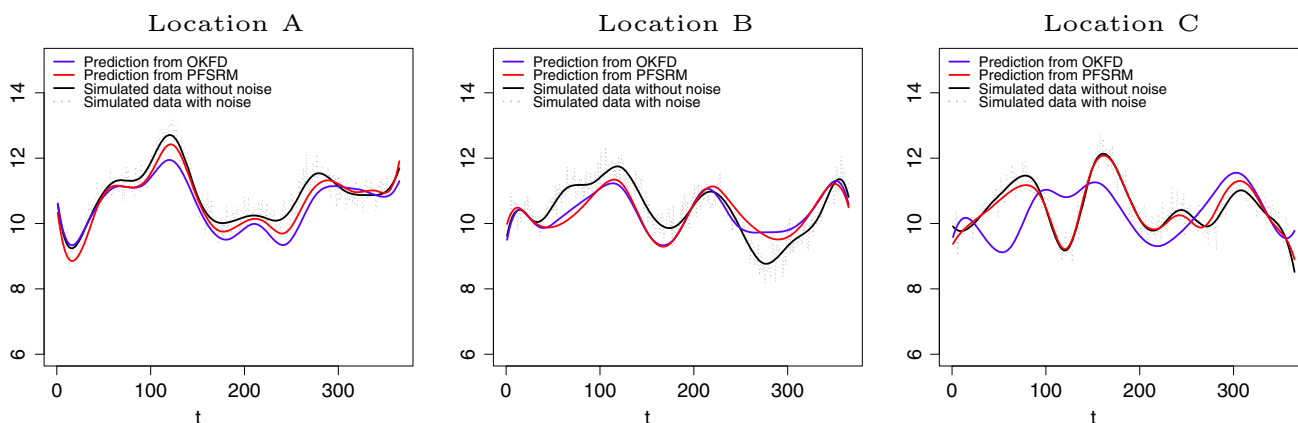


Fig. 7 Simulation I: predicted curves by OKFD in two outlying spatial locations (locations A and C, left and right, respectively) and a more central location B in the middle of the grid

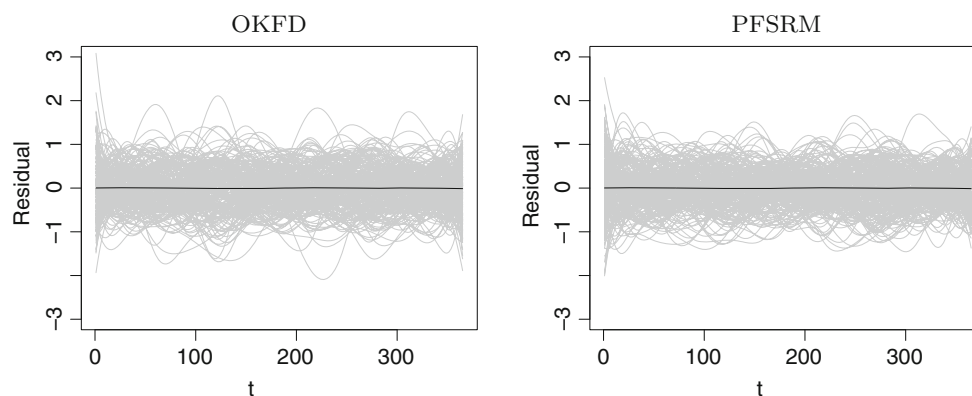


Fig. 8 Simulation I: residuals (grey lines) and the mean curve of the residuals (black line) from OKFD and PFSRM

$\sigma^2 = 0.25, \zeta = 0.9, \rho_1 = 0.1, \gamma = 2, \alpha = \pi/4, \mu_x = 0.2, \mu_y = -0.2, \tau^2 = 0.01$). The simulated sample paths with and without error can be seen in Fig. 10.

The first step was to approximate the sample curves by using basis representations with B-splines. In order to check the relation between the forecasting performance and

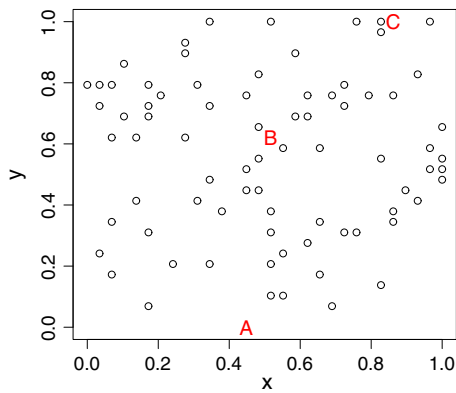


Fig. 9 Simulation II: spatial locations

the dimension of the initial approximation of the sample paths, regression splines in terms of a cubic B-spline basis of dimension 13 and 23 with equally spaced knots have been considered. In Fig. 11 a sample path without noise (dashed line) together with the noisy sample path (grey

line) and its basis representation (solid black line) are shown. The regression splines of all curves can be seen in Fig. 12. Obviously, a higher dimension for the basis provides noisier sample curves and far away from the original ones. The two scenarios (13 and 23 basis functions) are considered and compared. In the model fitting (second step), the two-dimensional basis for the space was achieved by considering 15 basis knots for each marginal basis. As in simulation I, a 2-order P-spline penalty has been considered. For OKFD, the variograms were selected as a linear combination of nugget and exponential models.

In this study the differences in the accuracy of the predictions provided by the two methods are also shown. According to the statistics summary provided in Table 2, the proposed method reduces the value of the mean and the median of the ISE's independently of the number of basis functions used at the first step (regression splines fitting). This fact is also supported by the box plots given in Fig. 13.

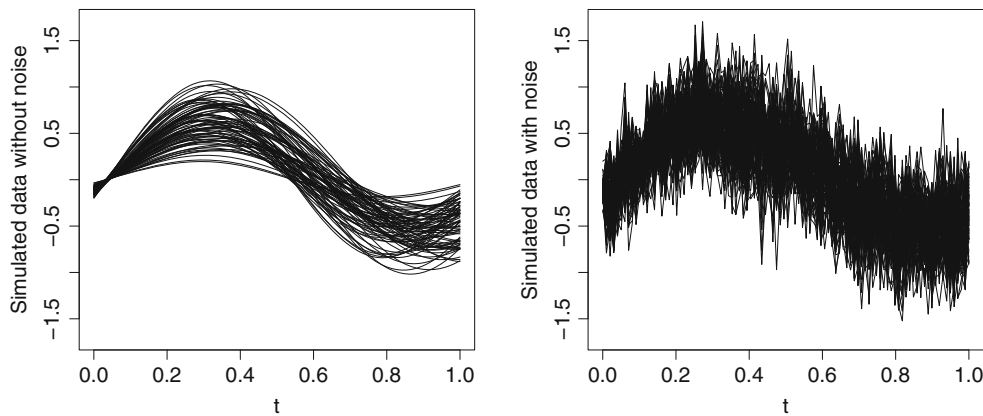


Fig. 10 Simulation II: simulated data without noise (left) and with noise (right)

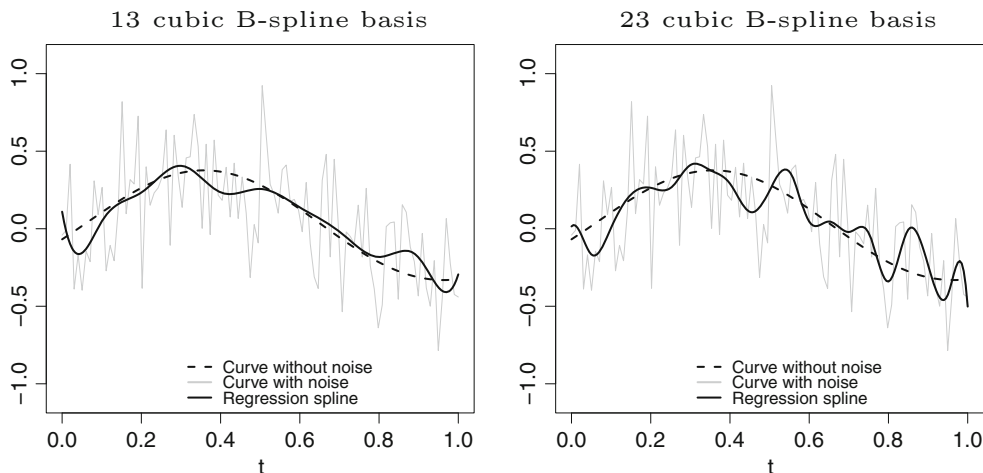


Fig. 11 Simulation II: a sample path without noise (dashed line) together with the noisy sample path (grey line) and its basis representation (solid line) on a cubic B-spline basis of dimension 13 (left) and 23 (right)

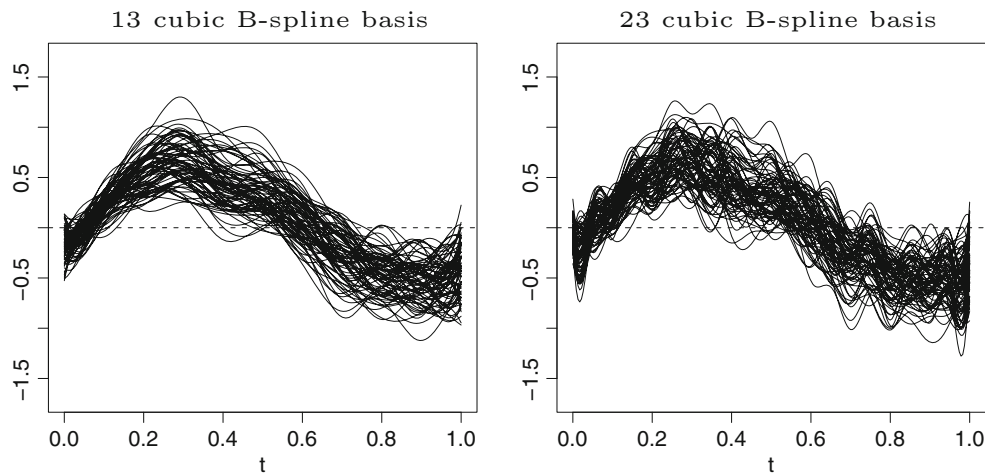


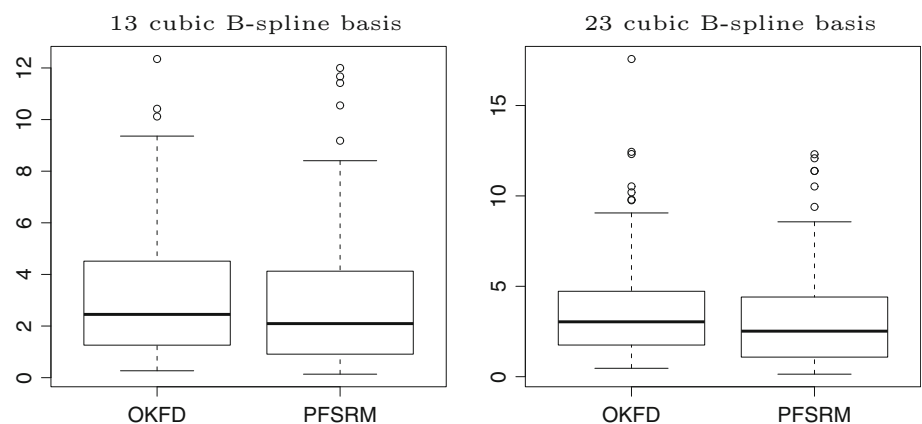
Fig. 12 Simulation II: basis representation of the sample curves by using regression splines in terms of a cubic B-spline basis of dimension 13 (left) and 23 (right)

Table 2 Simulation II: summary of ISE’s from the cross validation predictions by considering 13 and 23 basis functions at the initial regression splines (R-splines) fitting

Initial fitting	Methods	Mean	s.d.	Median	Min.	Max.	Sum
R-splines	OKFD	3.3000	2.6879	2.4510	0.2672	12.3500	263.9705
(13 Basis)	PFSRM	3.0690	2.8618	2.0950	0.1355	12.0000	245.5324
R-splines	OKFD	4.0250	3.1965	3.0360	0.4638	17.5700	321.9834
(23 Basis)	PFSRM	3.3870	3.0223	2.5190	0.1391	12.3000	270.9940
P-splines	OKFD	3.3050	3.0119	2.5440	0.1609	14.3300	264.4087
(23 Basis)	PFSRM	3.0860	2.8937	2.1980	0.1881	12.7900	246.8615

Last couple of rows were obtained by considering P-splines (on 23 basis) at the initial basis representation of the sample curves

Fig. 13 Simulation II: box plot related to the ISE of the predictions by OKFD and PFSRM, considering 13 and 23 basis functions at the initial regression splines fitting



With respect to the predicted curves obtained by each method, in Fig. 14 we can check that even in the most favorable case for OKFD of 13 basis functions, PFSRM provides the most accurate predictions. In order to test if the performance of the predictions varies depending on the spatial location, two outlying spatial locations (A and C) and a central spatial location (B) have been considered.

Locations A–C are highlighted in Fig. 9. Accordingly, PFSRM achieves the predicted curves closest to the true data.

On the other hand, in the Fig. 15 the mean curve of the predictions (red line) join to the point wise confidence bands (dashed line) according to the mean ± 2 times the standard deviation are displayed, with OKFD being the

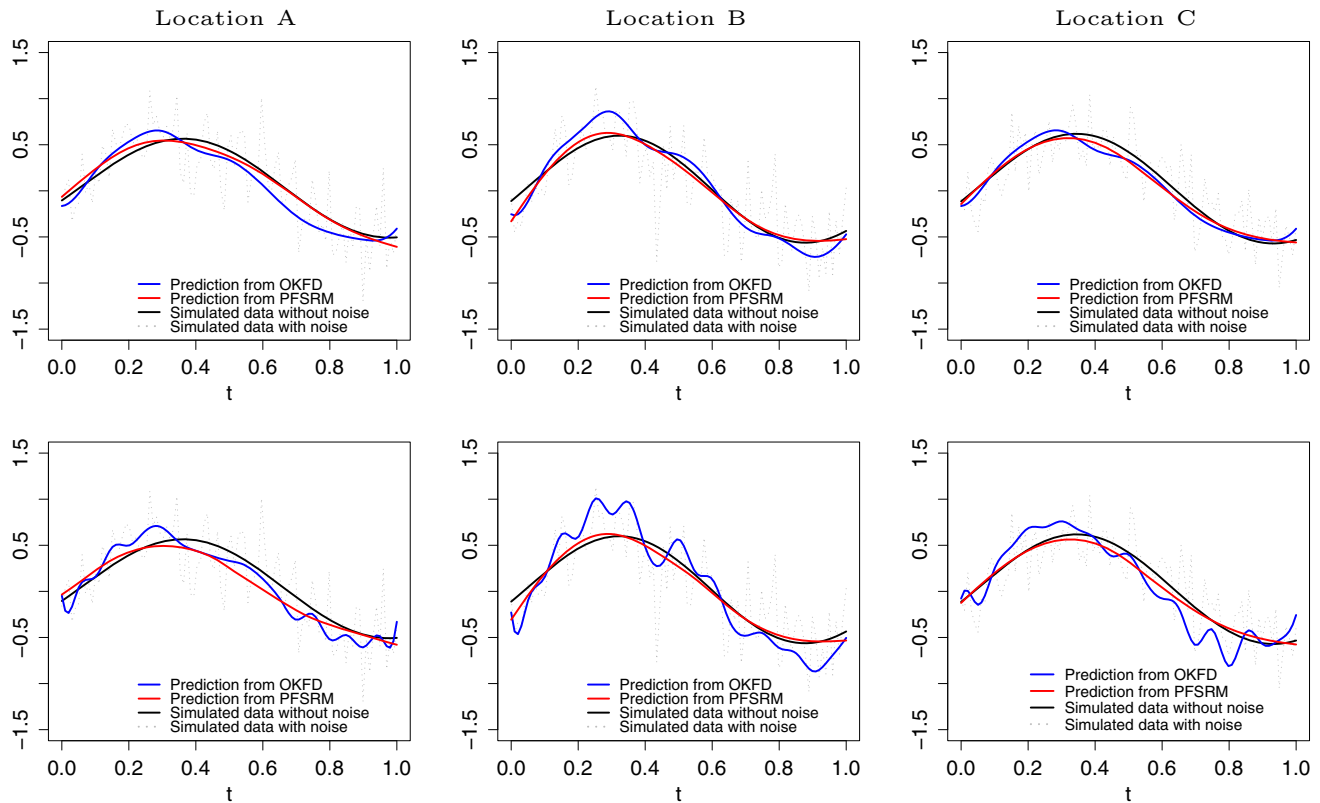


Fig. 14 Simulation II: predicted curves by OKFD and PFSRM in two outlying spatial locations (*locations A and C, left and right, respectively*) and a centered spatial location *B* (in the *middle*).

method that provides the noisier mean curve and the predictions with more variability. Let us also observe that PFSRM achieves the smoothest predicted curves while OKFD provides predicted curves further away from the real ones. Taking into account that OKFD seems to be very sensitive to the dimension of the basis, other possibility could be pre-smoothing the data by using P-splines. The measures of accuracy of the predictions provided by the two methods figure in Table 2. In this case the prediction errors decrease for both methods, but the ones given by PFSRM continue being lower.

Therefore, we can say that independently of the basis dimension used to approximate the sample curves at the beginning, OKFD has not been able to predict curves with the same shape than the true data while PFSRM gets more accurate predictions in both scenarios. Then, PFSRM is more robust than OKFD in the sense that it is not so sensitive to the selection of the dimension of the basis.

5 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 of functional data

Locations A–C are highlighted in Fig. 9. The initial sample curves were approximated by using regression splines on 13 (*first row*) and 23 (*second row*) B-spline basis functions

presenting spatial dependence, since curves located at closer geographical locations will be similar to other there are further apart (see Fig. 1).

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 Fig. 16. As in both simulation studies, a 2-order P-spline penalty was considered. Furthermore, the variograms were selected as a linear combination of nugget and exponential models.

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 ± 2 times the standard deviation) can be seen in Fig. 17. In both cases (Cases 1 and 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 noisier than in Case 1. By contrast, PFSRM provides similar predictions independently of the number of basis functions used to fit

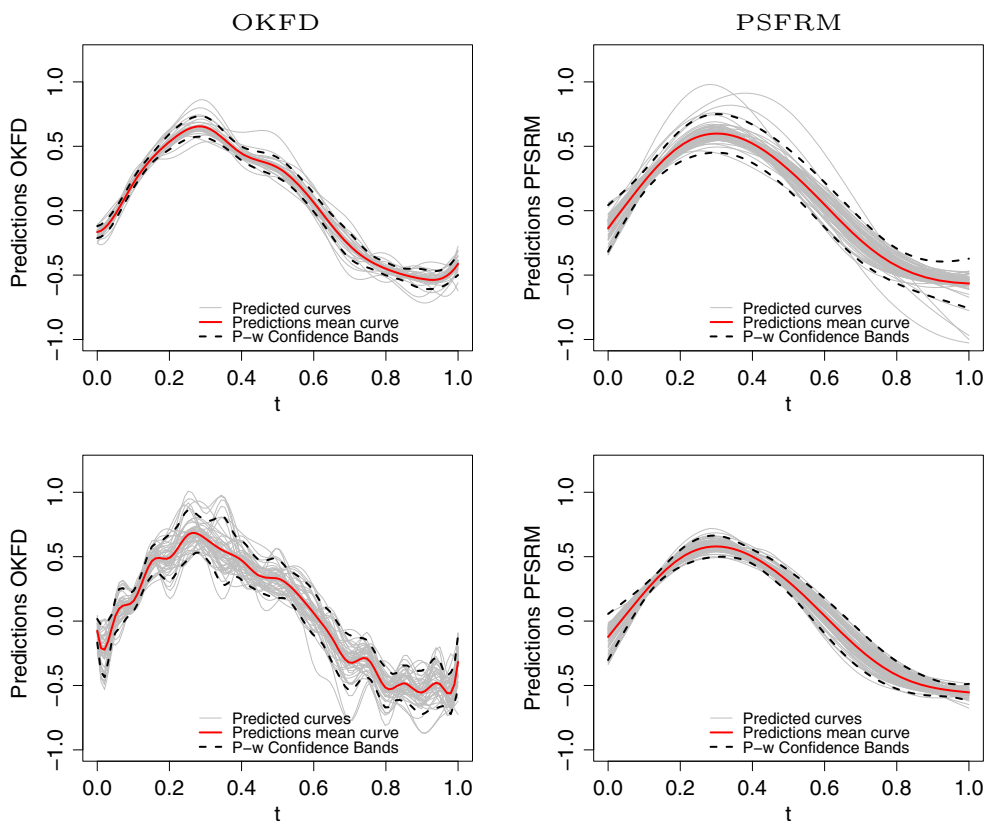


Fig. 15 Simulation II: mean curve of the predictions and point wise confidence bands according to the mean curve ± 2 times the standard deviation of the predicted curves by OKFD and PFSRM. The initial

sample curves were approximated by using regression splines on 13 (first row) and 23 (second row) B-spline basis functions

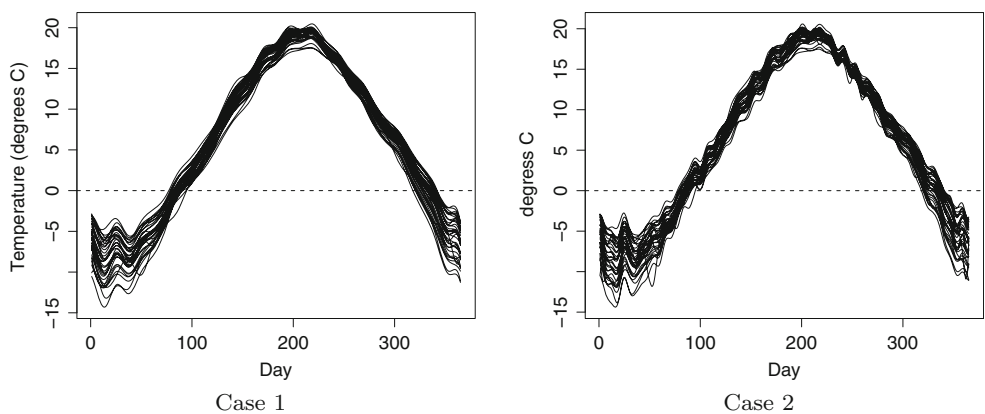


Fig. 16 Application: regression splines fitted from the temperature raw data by using 33 and 65 cubic B-spline basis functions (Cases 1 and 2, respectively)

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 basis functions 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 Fig. 18. Independently of the basis dimension, PFSRM provides smoother and more accurate predicted curves than OKFD and also maintains the trend of the raw data.

In order to compare the prediction ability of the two methods, the box plots related to the SSE's (with respect to

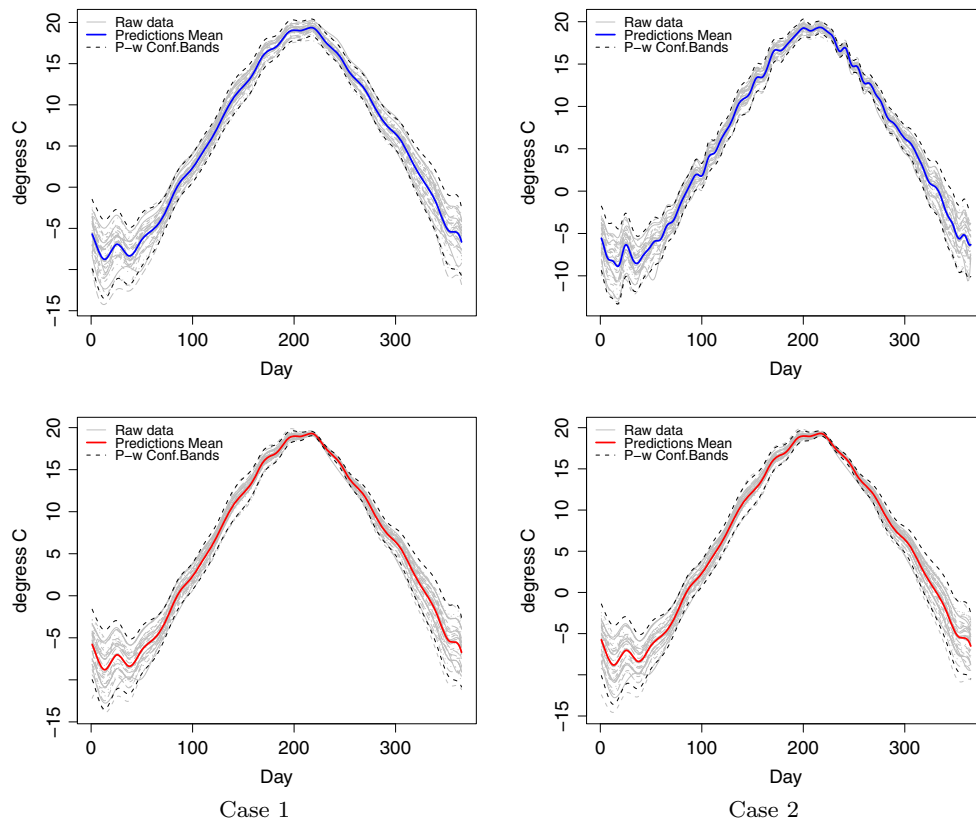


Fig. 17 Application: predicted curves (grey) by OKFD (at the top) and PFSRM (at the bottom) from the regression splines of the temperature raw data [using 33 (Case 1) and 65 (Case 2) cubic B-

spline basis functions] join to its mean curve (blue and red line) and the point wise confidence bands according to the mean ± 2 times the standard deviation (black and dashed line)

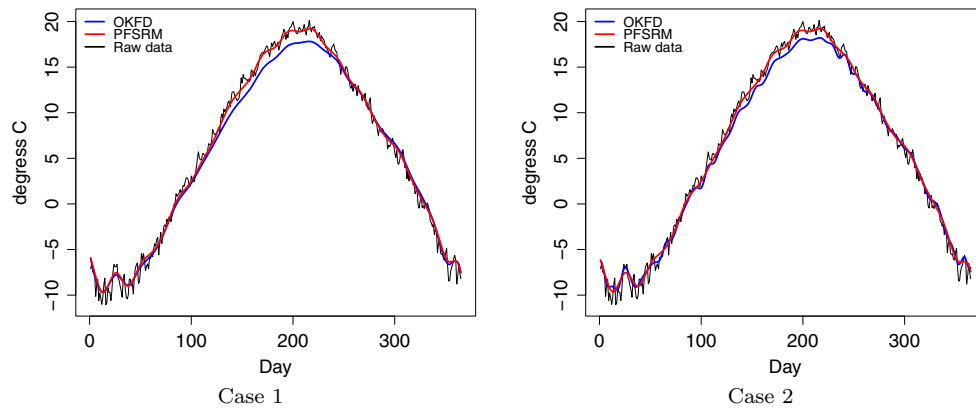


Fig. 18 Application: the predicted curve by OKFD (blue) and PFSRM (red) from the regression splines of the temperature raw data [using 33 (Case 1) and 65 (Case 2) cubic B-spline basis functions]

and the observed temperature curve (black) in two of the 35 Canadian Maritime provinces

the observed data) obtained by cross validation for Cases 1 and 2 can be seen in Fig. 19. 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.

Finally, as goodness-of-fit measure, in Fig. 20 the residual curves have been displayed, highlighting the residuals related

Fig. 19 Application: *box plots* related to the SSEs (with respect to the raw data) obtained in the cross validation for *Cases 1 and 2*

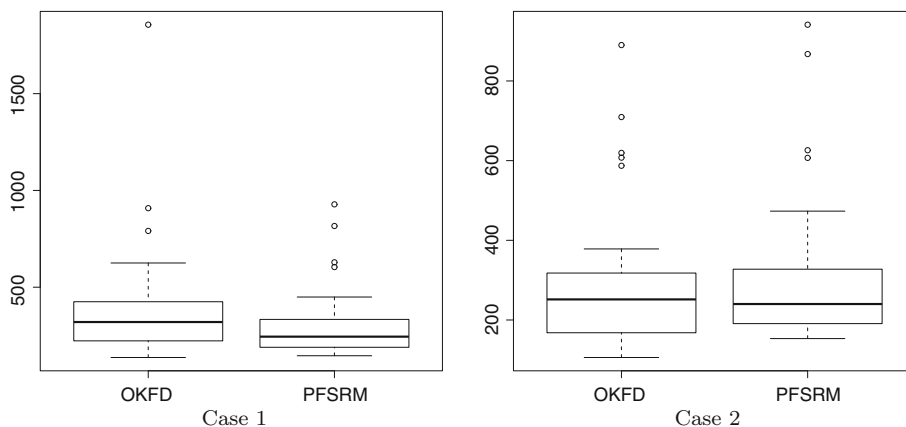


Table 3 Application: the median, the mean and the standard deviation of the SSEs (with respect to the observed data) obtained in the cross validation for *Cases 1 and 2*

	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

to three spatial locations (color and dashed lines). From this figures we can conclude that both methods provide a considerable number of residuals close to zero, but in some locations OKFD provides larger residuals than PFSRM.

6 Conclusions

The aim of this paper is to provide a new tool to predict spatially dependent functional data as alternative to other geostatistical prediction techniques, such as functional kriging. From a formal definition of spatial functional data,

which was presented in Sect. 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 (Sect. 2.1).

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 on a grid and independent random errors. The second one was simulated by considering non-regular spatial locations and random errors simulated from a spatio-temporal Gaussian process. Furthermore, an application to climatological real data has been presented.

In both simulation studies and the application to real data, the first step was to approximate the sample curves from their discrete observations by using regression splines on cubic B-spline basis with different dimensions just for a comparison purpose (the number basis functions at each study was proposed attending to the corresponding data structure). Regard to the P-spline penalty used in the model fitting, in all cases a 2-order penalty was considered.

A leave-one-out cross-validation procedure was carried out to obtain the predicted curve at each unsampled spatial

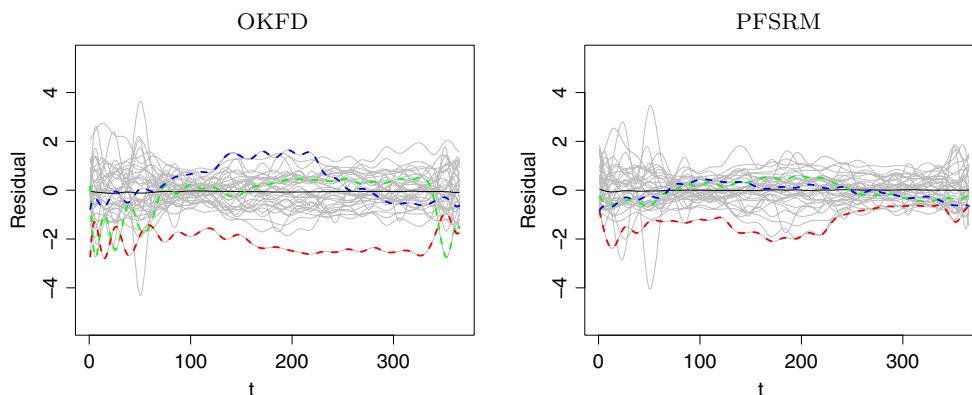


Fig. 20 Application: residuals (*grey lines*) and the mean curve of the residuals (*black line*) from OKFD and PFSRM, considering 33 basis functions at the initial regression splines fitting. In *dashed and color lines* two residuals have been highlighted

location. Also, in order to test if the forecasting performance varies depending on the spatial location, the predicted curves in three outlying spatial locations were analyzed in both simulation studies.

The first simulation study reveals that even in the most favorable (where the initial regression splines are closer to the true data without noise), PFSRM provides the most accurate predicted curves very close to the shape of the true data, independently of the spatial location. By contrast, the forecasting performance of OKFD is not equal across the grid areas.

The second simulation study highlights that independently of the basis dimension used to approximate the sample curves at the beginning, OKFD has not been able to predict curves with the same shape than the true data while PFSRM gets the most accurate predictions in the two considered scenarios. In addition, we can say that PFSRM is more robust than OKFD in the sense that it is not as sensitive as OKFD to the selection of the basis dimension.

In general, from the two simulation studies, it is clear that PFSRM provides the most accurate predicted curves (even in outlying spatial locations) and reduces the mean and the median of the ISE’s with respect to OKFD, independently of the number of basis functions used at the initial smoothing with regression splines.

In addition, one of the advantages of our method is that it can be used when the spatial locations where measurements are taken change from one time point to another, and also the model can easily cope with missing space or time points.

Summarizing, it can be concluded that in order to predict functional data with spatial dependence, the proposed PFSRM is a more accurate and computationally efficient alternative to existing geostatistical predictors as ordinary functional kriging.

Acknowledgments 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, MTM 2011-28285-C02-C2 and MTM 2014-52184-P from Secretaría de Estado Investigación, Desarrollo e Innovación, Ministerio de Economía y Competitividad, Spain. we want to thanks Giraldo et al. by providing the R code related to Functional Kriging. Finally, we also thank the referees for the valuable comments on our manuscript. These comments helped to improve the organization and the understanding of our paper.

Appendix

Taking into account the following properties (Harville 1997)

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

the Eq. (5) can be rewritten as

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

with $\Psi = \int \phi^T \phi^T$ being the inner product matrix between the basis functions. Next step is to compute the derivatives of Eq. (3) with respect to A . By considering the following properties (Harville 1997),

$$\frac{\partial \text{trace}(XAX')}{\partial X} = X(A + A'), \tag{8}$$

$$\frac{\partial \text{trace}(X'AX)}{\partial X} = (A + A')X, \tag{9}$$

$$\frac{\partial \text{trace}(XA)}{\partial X} = A', \tag{10}$$

we have that

$$\frac{\partial \text{trace}(C' C \Psi)}{\partial A} = 0,$$

$$\begin{aligned} \frac{\partial \text{trace}(Z' Z A \Psi A')}{\partial A} &\stackrel{(3)}{=} Z' Z A (\Psi + \Psi') \\ &\stackrel{(symmetry)}{=} 2Z' Z A \Psi, \end{aligned}$$

$$\frac{\partial -2\text{trace}(A \Psi C' Z)}{\partial A} \stackrel{(5)}{=} -2Z' C \Psi',$$

$$\frac{\partial \lambda_1 \text{trace}(A' (\Delta_d^U \Delta_d^U \otimes I_q) A)}{\partial A} \stackrel{(4)}{=} 2\lambda_1 (\Delta_d^U \Delta_d^U \otimes I_q) A,$$

$$\frac{\partial \lambda_2 \text{trace}(A' (I_p \otimes \Delta_d^V \Delta_d^V) A)}{\partial A} \stackrel{(4)}{=} 2\lambda_2 (I_p \otimes \Delta_d^V \Delta_d^V) A,$$

$$\frac{\partial \lambda_3 \text{trace}(A (\Delta_d^T \Delta_d^T) A')}{\partial A} \stackrel{(3)}{=} 2\lambda_3 A (\Delta_d^T \Delta_d^T).$$

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

$$\begin{aligned} Z' Z A \Psi + \lambda_1 (\Delta_d^U \Delta_d^U \otimes I_q) A + \lambda_2 (I_p \otimes \Delta_d^V \Delta_d^V) A \\ + \lambda_3 A (\Delta_d^T \Delta_d^T) = Z' C \Psi'. \end{aligned} \tag{11}$$

In order to get the solution to A , the Kronecker product is used to express Eq. (9) in conventional matrix algebra

$$\begin{aligned} \text{vec}(Z'ZA\Psi) &\stackrel{(1)}{=} (\Psi \otimes (Z'Z))\text{vec}(A), \\ \text{vec}\left(\lambda_1\left(\Delta_d^{U'}\Delta_d^U \otimes I_q\right)A\right) &= \text{vec}\left(\lambda_1\left(\Delta_d^{U'}\Delta_d^U \otimes I_q\right)AI_r\right) \\ &\stackrel{(1)}{=} \lambda_1\left(\Delta_d^{U'}\Delta_d^U \otimes I_q \otimes I_r\right)\text{vec}(A), \\ \text{vec}\left(\lambda_2\left(I_p \otimes \Delta_d^{V'}\Delta_d^V\right)A\right) &= \text{vec}\left(\lambda_2\left(I_p \otimes \Delta_d^{V'}\Delta_d^V\right)AI_r\right) \\ &\stackrel{(1)}{=} \lambda_2\left(I_p \otimes \Delta_d^{V'}\Delta_d^V \otimes I_r\right)\text{vec}(A), \\ \text{vec}\left(\lambda_3A\left(\Delta_d^{T'}\Delta_d^T\right)\right) &= \text{vec}\left(\lambda_3\left(I_p \otimes I_q\right)A\left(\Delta_d^{T'}\Delta_d^T\right)\right) \\ &\stackrel{(1)}{=} \lambda_3\left(I_p \otimes I_q \otimes \Delta_d^{T'}\Delta_d^T\right)\text{vec}(A). \end{aligned}$$

Then, Eq. (9) can be re-written as follows

$$[\Psi \otimes (Z'Z) + PEN_d^{U,V,T}]\text{vec}(A) = \text{vec}(Z'CP'),$$

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

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

Finally, A is given by

$$\text{vec}(A) = [\Psi \otimes (Z'Z) + PEN_d^{U,V,T}]^{-1}\text{vec}(Z'CP').$$

References

Aguilera AM, Aguilera-Morillo MC (2013) Comparative study of different B-spline approaches for functional data. *Math Comput Model* 58:1568–1579

Aguilera AM, Aguilera-Morillo MC (2013) Penalized PCA approaches for B-spline expansions of smooth functional data. *Appl Math Comput* 219:7805–7819

Aguilera-Morillo MC, Aguilera AM, Escabias M, Valderrama MJ (2013) Penalized spline approaches for functional logit regression. *Test* 22:251–277

Caballero W, Giraldo R, Mateu J (2013) A universal kriging approach for spatial functional data. *Stoch Environ Res Risk Assess* 27:1553–1563

Chiou JM, Müller HG, Wang JL (2004) Functional response models. *Stat Sin* 14:659–677

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

Dubrulle O (1984) Comparing kriging and splines. *Comput Geosci* 10(2–3):327–338

Eilers PHC, Marx B (1996) Flexible smoothing with B-splines and penalties. *Stat Sci* 11:89–121

Eilers PHC, Currie I, Durban M (2006) Fast and compact smoothing on large multidimensional grids. *Comput Stat Data Anal* 50:61–76

Escabias M, Aguilera AM, Valderrama MJ (2005) Modeling environmental data by functional principal component logistic regression. *Environmetrics* 16:95–107

Faraway JJ (1997) Regression analysis for a functional response. *Technometrics* 39:254–261

Fernandez-Pascual RM, Espejo R, Ruiz-Medina MD (2015) Moment and Bayesian wavelet regression from spatially correlated functional data. *Stoch Environ Res Risk Assess*. doi:10.1007/s00477-015-1130-5

Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer, New York

Giraldo R (2010) Geostatistical analysis of functional data. PhD Thesis, Universitat Politècnica de Catalunya, Catalunya

Giraldo R, Delicado P, Mateu J (2010) Continuous time-varying kriging for spatial prediction of functional data: an environmental application. *J Agric Biol Environ Stat* 15:66–82

Giraldo R, Delicado P, Mateu J (2011) Ordinary kriging for function-valued spatial data. *Environ Ecol Stat* 18:411–426

Giraldo R, Mateu J, Delicado P (2012) geofd: an R package for function-valued geostatistical prediction. *Rev Colomb Estad* 35:385–407

Gouillard M, Voltz M (1993) Geostatistical interpolation of curves: a case study in soil science. Springer, Dordrecht, pp 805–816

Harville DA (1997) Matrix algebra from a statistician’s perspective. Springer, New York

Horvath L, Kokoszka P (2012) Inference for functional data with applications. Springer, New York

Hsing T, Eubank R (2015) Theoretical foundations of functional data analysis with an introduction to linear operators. Wiley, Chichester

Ignaccolo R, Mateu J, Giraldo R (2014) Kriging with external drift for functional data for air quality monitoring. *Stoch Environ Res Risk Assess* 28:1171–1186

Kaufman CG, Sain SR (2010) Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Anal* 5(1):123–150

Laslett GM (1994) Kriging and splines: an empirical comparison of their predictive performance in some applications. *J Am Stat Assoc* 89(426):391–400

Lee DJ, Durban M (2009) Smooth-car mixed models for spatial count data. *Comput Stat Data Anal* 53:2968–2977

Lee DJ, Durban M (2011) Pspline ANOVA type interaction models for spatio-temporal smoothing. *Stat Model* 11:49–69

Marx BD, Eilers PHC (1999) Generalized linear regression on sampled signals and curves. A P-spline approach. *Technometrics* 41:1–13

Menafoglio A, Secchi P, Dalla Rosa M (2013) A universal kriging predictor for spatially dependent functional data of a Hilbert Space. *Electron J Stat* 7:2209–2240

Ramsay JO, Silverman BW (1997) Functional data analysis, 1st edn. Springer, New York

Rao CR, Rao MB (1998) Matrix algebra and its applications to statistics and econometrics. World Scientific Publishing Co., Pte. Ltd., Singapore

Reiss PT, Huang L, Mennes M (2010) Fast function-on-scalar regression with penalized basis expansions. *Int J Biostat* 6:1–28

Ruiz-Medina MD, Espejo RM (2012) Spatial autoregressive functional plug-in prediction of ocean surface temperature. *Stoch Environ Res Risk Assess* 26:335–344

Ruiz-Medina MD, Espejo RM, Ugarte MD, Militino AF (2014) Functional time series analysis of spatiotemporal epidemiological data. *Stoch Environ Res Risk Assess* 28:943–954

Ruppert D (2002) Selecting the number of knots for penalized splines. *J Comput Graph Stat* 11:735–757

Sangalli LM, Ramsay JO, Ramsay TO (2013) Spatial spline regression models. *J R Stat Soc B* 75:1–23

Shi JQ, Choi T (2011) Gaussian process regression analysis for functional data. CRC Press, Chapman and Hall, Boca Raton

Sigrist F, Kuensch HR, Stahel WA (2015) spat: an R package for spatio-temporal modeling with a stochastic advection–diffusion process. *J Stat Softw* 63:1–23

- Ugarte MD, Goicoa T, Militino AF, Durban M (2009) Spline smoothing in small area trend estimation and smoothing. *Comput Stat Data Anal* 53:3616–3629
- Yakowitz SJ, Szidarovsky F (1985) A comparison of kriging with non-parametric regression methods. *J Multivar Anal* 16:21–53
- Zhang J-T (2013) *Analysis of variance for functional data*. CRC Press, Chapman and Hall
- Zhang J-T, Chen J (2007) Statistical inference for functional data. *Ann Stat* 35(3):1052–1079