This is a postprint version of the following published document:

Aguilera-Morillo, M.C., Durbán, M. & Aguilera, A.M.
Prediction of functional data with spatial dependence:
a penalized approach. (2017). *Stochastic
Environmental Research and Risk Assessment*, 31, 7–
22

DOI: <https://doi.org/10.1007/s00477-016-1216-8>

© 2016, Springer-Verlag Berlin Heidelberg

Prediction of functional data with spatial dependence: a penalized approach

M. Carmen Aguilera-Morillo · María Durbán · Ana M. Aguilera.

Received: date / Accepted: date

Abstract The scope of this paper are spatial functional variables whose observations are realizations of a spatio-temporal functional process. In this context, as alternative to other geostatistical techniques (kriging, kernel smoothing, among others), a new method to approximate the functional form of a set of spatially correlated sample paths (observed only at a finite set of time points) and also to predict the corresponding curve at new spatial locations is proposed. In order to test the good performance of the proposed method, two simulation studies and an application with real data have been carried out. Finally, the results were compared with functional kriging.

Keywords Spatio-temporal functional data · Spatial correlation · P-spline penalty

1 Introduction

Functional Data Analysis (FDA) is currently a very active statistical research topic both from 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

---

M. Carmen Aguilera-Morillo
University Carlos III de Madrid
Tel.: +916249179
Fax: +916249177
E-mail: maguiler@est-econ.uc3m.es

María Durbán
University Carlos III de Madrid

Ana M. Aguilera
University of Granada.

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

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 [29] 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. And at the same time new books appear on these FDA topics [13, 19, 20, 36, 41].

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 [29] 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 Figure 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 these type can be seen in [10], [40] and [22].

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 [18], 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 [14, 6]. In general, the most used technique to predict functional data with spatial dependence is functional kriging. In [15] 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 [16], and implemented in the R package *geofd* [17]. Recently, different versions of universal kriging predictor for functional data with spatial dependence were proposed in [28, 4]. Kriging with external drift has also been extended for introducing exogenous variables with spatially correlated functional data [21].

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 [8]. 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 [1] and estimating different FDA models as PCA [2] or functional regression [26, 3], among others. In the case of spatial data, [24] and [38] used P-splines for smoothing spatially correlated count data, [25] extended their use to the case of spatio-temporal data, and more recently, [35] 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 [12]. On the other hand, functional approaches based on autoregressive Hilbertian processes were considered in [33, 32].

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 [39]. 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 [7, 23].

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 unvisited sites. From the formal definition of spatio-temporal functional data, which is given in Section 2, a penalized functional regression model is extended for predicting spatially correlated functional data in Section 2.1. The idea is to consider the functional regression model for functional response and scalar covariates [11, 29, 5] 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, [29] 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 [31] used a penalized generalized least squares criterion based on a basis representation. In this paper, we will adapt the idea developed in [9], and combine the two-dimensional penalty used for spline spatial regression with the one proposed in [29] 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 Section 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 to 1994 [29]. But this is a clear example of functional data presenting spatial dependence and in this sense was studied in [6,28]. The conclusions about these studies close the paper in Section 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 spatio-temporal functional variable (stochastic process)

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

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 $(\Omega, \mathcal{A}, P)$.

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 $y_j = y_i(t_j) \quad i = 1, \dots, n; j = 1, \dots, m$.

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

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

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,

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

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

$$X = (\Phi^U \odot \Phi^V) A \Phi^{T'}, \quad (3)$$

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 $(\Phi^U \odot \Phi^V)_{i(kl)} = \phi_k^U(u_i) \phi_l^V(v_i)$ [30].

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 not sampled 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 Equation (2) by introducing the spatial variability through the following functional spatial regression model:

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

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 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 Equation (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

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

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 [9], so that

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

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

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

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

3 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

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

where

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

and

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

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 oversmooth the spatial component of the model and AIC performed as well as GCVE.

4 Simulation studies

In order to test the good 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 on a grid and random errors added only at the time dimension. The second one was simulated by considering non-regular spatial locations and the random errors were added at the two dimensions (space and time) throw a spatio-temporal Gaussian process. In addition, the results are compared with an other powerful predictor, the ordinary kriging for functional data (OKFD) developed by [16].

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

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

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, and n the number of spatial locations. 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.

4.1 Simulation study I

This simulation study was first considered in [17]. 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

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

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

Fig. 2 Simulation I: Spatial locations.

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

defined according to the exponential model $C(h) = 2\exp(-\frac{h}{8})$, where $h = \|s_i - s_j\|$, $(i, j = 1, \dots, 225)$ is the Euclidean distance between two sites s_i and s_j . Finally, $\epsilon_i(t)$ is a random error for each t , with $t = 1, \dots, 365$, simulated according a distribution $N(0, 0.09)$. The spatial locations are shown in Figure 2. The simulated data sets, with and without noise, can be seen in Figure 3.

The first step was to approximate the sample curves by using regression splines. In order to check the relation between the forecasting performance and the dimension of the initial approximation of the sample paths, regression splines in terms of a cubic B-spline basis defined on 12 and 40 equally spaced knots have been considered. 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 Figure 4. The regression splines of all curves can be seen in Figure 5. In that case, regression splines on 15 basis functions get a perfect approximation to the true data (without noise). On the other hand,

Fig. 4 Simulation I: A sample paths without noise (dashed line) together with the noisy sample path (grey line) and its basis representation (solid line) on a Cubic B-spline basis defined on 12 (left) and 40 (right) equally spaced knots.

Fig. 5 Simulation I: Basis representation of the sample curves by using regression splines in terms of a basis defined on 12 (left) and 40 (right) equally spaced knots.

by considering 43 basis functions the regression splines are obviously noisier, but not too different than the true data. So, in this study the dimension of the basis does not seem to be an important problem.

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 (due to a P-spline penalty is considered for fitting the model, to consider 1 basis knot at each 5 observation knots is enough [34]). Regard to the penalty, a 2-order penalty has been considered.

Fig. 6 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 grid, in the middle). The initial sample curves were approximated by using regression splines on 15 B-spline basis functions.

		Mean	s.d.	Median	Min	Max	Sum
15 Basis functions	OKFD	80.62	47.13	68.60	16.87	278.70	18139.18
	PFSRM	72.94	42.82	64.02	3.132	257.80	16411.87
43 Basis functions	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 by OKFD and PFSRM by considering 15 and 43 basis functions at the initial regression splines fitting.

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 non-sampled spatial location. Also, we are interest on testing if the performance varies depending on the spatial location. So, in Figure 6 the predicted curves at two outlying spatial locations (A and C) and a more central spatial location (B) have been displayed (locations A, B and C are highlighted in Figure 2). In this sense, we can see that even in the more favorable case for OKFD (working on the initial regression splines closer to the true data without noise, that is using 15 basis functions), PFSRM provides the best predicted curves closer to the shape of the true data, independently of the predicted spatial location. It 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.

On the other hand, in Figure 7 the 225 predicted curves on the regression splines fitted with 43 basis are displayed joint to the mean predictions curves (red lines) and the point wise confidence bands according to the mean $\pm$ two times the standard deviation. The predictions are similar in both methods. However, taking into account the multiple box plot related to the distribution of the ISEs (Figure 8) 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 ISEs, independently of the number of basis used at the initial fitting of the regression splines.

Fig. 7 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). The initial sample curves were approximated by using regression splines on 43 B-spline basis functions.

Fig. 8 Simulation I: Box plot related to the ISE of the predictions by OKFD and PFSRM, considering 15 and 43 basis functions at the initial regression splines fitting.

4.2 Simulation study II

Let us consider a set of 80 non regular spatial locations, which are displayed in Figure 9, 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

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

Fig. 9 Simulation II: Spatial locations.

where

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

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

with $s = (u, v)$ denoting the pair of coordinates of the spatial locations, $a \sim \text{Uniform}[0.5, 2]$, $b \sim \text{Uniform}[0.5, 1]$, $c \sim \text{Uniform}[1.5, 2]$ 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 [37], with parameters ($\rho_0 = 0.1, \sigma^2 = 0.25, \zeta = 0.9, \rho_1 = 0.1, \gamma = 2, \alpha = \pi/4, \mu_x = 0.2, \mu_y = -0.2, \tau^2 = 0.01$). The simulated sample paths with and without error can be seen in Figure 10.

The first step was to carry out the basis representation of the sample curves by using regression splines in terms of a cubic B-spline basis defined on 10 and 20 equally spaced knots. In Figure 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 Figure 12. Obviously, a higher dimension for the basis provides noisier sample curves. It could be a disadvantage for OKFD. Because of this, 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.

In this study the differences between the two methods are also proved. 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 Figure 13.

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

Fig. 11 Simulation II: A sample paths without noise (dashed line) together with the noisy sample path (grey line) and its basis representation (solid line) on a Cubic B-spline basis defined on 10 (left) and 20 (right) equally spaced knots.

		Mean	s.d.	Median	Min	Max	Sum
13 Basis functions	OKFD	3.3000	2.6879	2.4510	0.2672	12.3500	263.9705
	PFSRM	3.0690	2.8618	2.0950	0.1355	12.0000	245.5324
23 Basis functions	OKFD	4.0250	3.1965	3.0360	0.4638	17.5700	321.9834
	PFSRM	3.3870	3.0223	2.5190	0.1391	12.3000	270.9940

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 fitting.

Fig. 12 Simulation II: Basis representation of the sample curves by using regression splines in terms of a basis defined on 10 (left) and 20 (right) equally spaced knots.

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 Figure 14 we can check that even in the more favorable case for OKFD (working on the initial regression splines closer to the true data without noise, that is using 13 basis functions) PFSRM provides the most accurate predictions. In order to test if the performance of the predictions vary depending on the spatial location, two outlying spatial locations (A and C) and a central spatial location (B) have been considered. Locations A, B and C are highlighted in Figure 9. Accordingly, non-relevant results have been observed.

On the other hand, if 23 basis functions are considered for fitting the initial regression splines, PFSRM achieves the smoothest predicted curves (see Figure

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). Locations A, B and C are highlighted in Figure 9. The initial sample curves were approximated by using regression splines on 13 B-spline basis functions.

Fig. 15 Simulation II: 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). The initial sample curves were approximated by using regression splines on 23 B-spline basis functions.

15). In the same figure, the mean curve of the predictions (red line) join to the point wise confidence bands (dashed line) according to the mean $\pm$ two times the standard deviation are displayed, being OKFD which provides the noisier mean curve and the predictions with more variability.

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, mean 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 bases dimension.

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

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 presenting spatial dependence, since curves located at closer geographical locations will be similar to other there are further apart (see 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 Figure 16.

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 17. 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 18. Independently of the basis

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 $\pm$ two times the standard deviation (black and dashed line).

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 the observed data) obtained by cross validation for Cases 1 and 2 can be seen in Figure 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 Figure 20 the residual curves have been displayed, highlighting the residuals related 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.

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 Maritimes provinces.

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.

	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.

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.

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 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 (Section 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 random errors added only at the time dimension. The second one was simulated by considering non-regular spatial locations, and the random errors were added at all dimensions (space and time) by means of a spatio-temporal Gaussian process. Furthermore, an application to 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 its 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). Regarding 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 non-sampled spatial 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 more favorable case for OKFD (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, mean 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 so sensitive 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), reducing the mean and the median of the ISEs with respect to OKFD, independently of the number of basis used at the initial fitting of the regression splines.

The application also proves the robustness of PFSRM versus OKFD, so that a lower number of basis functions leads to higher ISEs from OKFD, mean while the ISEs from PFSRM are more balanced in both cases.

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

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, MTM 2011-28285-C02-C2 and MTM 2014-52184-P from Secretaría de Estado Investigación, Desarrollo e Innovación, Ministerio de Economía y Competitividad, Spain. Also, we want to thanks Giraldo et al. by providing the R code related to Functional Kriging.

References

1. Aguilera, A. M. and Aguilera-Morillo, M. C., Comparative study of different B-spline approaches for functional data, *Mathematical and Computer Modelling*, 58, 1568-1579 (2013)
2. Aguilera, A. M. and Aguilera-Morillo, M. C., Penalized PCA approaches for B-spline expansions of smooth functional data, *Applied Mathematics and Computation*, 219, 7805-7819 (2013)
3. AguileraMorillo, M. C., Aguilera, A. M., Escabias, M. and Valderrama, M. J., Penalized spline approaches for functional logit regression, *Test*, 22, 251-277 (2013)
4. Caballero, W., Giraldo, R. and Mateu, J. A universal kriging approach for spatial functional data, *Stochastic Environmental Research and Risk Assessment*, 27, 15531563 (2013)
5. Chiou, J. M., Müller, H. G. and Wang, J. L., Functional response models, *Statistica Sinica*, 14, 659-677 (2004)

6. Delicado, P., Giraldo, R., Comas, C. and Mateu, J., Statistics for spatial functional data: some recent contributions, *Environmetrics*, 21, 224-239 (2009)
7. Dubrule, O., Comparing Kriging and Splines, *Computer and Geosciences*, 10 (2-3), 327-338 (1984)
8. Eilers, P. H. C. and Marx, B., Flexible smoothing with B-splines and penalties, *Statistical Science*, 11, 89-121 (1996)
9. Eilers, P. H. C., Currie, I. and Durban, M., Fast and compact smoothing on large multidimensional grids, *Computational Statistics and Data Analysis*, 50, 61-76 (2006)
10. Escabias, M., Aguilera, A.M. and Valderrama, M.J. Modeling Environmental Data by Functional Principal Component Logistic Regression, *Environmetrics*, 16, 95107 (2005)
11. Faraway, J. J., Regression Analysis for a Functional Response, *Technometrics*, 39, 254-261 (1997)
12. Fernández-Pascual, R.M., Espejo, R. and Ruiz-Medina, M.D. Moment and Bayesian Wavelet Regression from Spatially Correlated Functional Data, *Stochastic Environmental Research and Risk Assessment*, in press (2015)
13. Ferraty, F. and Vieu, P. Nonparametric functional data analysis. Springer-Verlag (2006)
14. Giraldo, R., Geostatistical Analysis of Functional Data, PhD thesis. Universitat Politècnica de Catalunya, Catalunya (2010)
15. Giraldo, R., Delicado, P. and Mateu, J., Continuous time-varying kriging for spatial prediction of functional data: An environmental application, *Journal of Agricultural, Biological and Environmental Statistics*, 15, 66-82 (2010)
16. Giraldo, R., Delicado, P. and Mateu, J., Ordinary kriging for function-valued spatial data, *Environmental and Ecological Statistics*, 18, 411-426 (2011)
17. Giraldo, R., Mateu, J. and Delicado, P., geofd: An R Package for Function-Valued Geostatistical Prediction, *Revista Colombiana de Estadística*, 35, 385-407 (2012)
18. Goulard, M. and Voltz, M., Geostatistical Interpolation of Curves: A Case Study in Soil Science, 805-816. Springer, Netherlands (1993)
19. Horvath, L. and Kokoszka, P. Inference for functional data with applications. Springer-Verlag (2012)
20. Hsing, T. and Eubank, R., Theoretical foundations of functional data analysis with an introduction to linear operators. Wiley, 2015
21. Ignaccolo, R., Mateu, J. and Giraldo, R., Kriging with External Drift for Functional Data for Air Quality Monitoring, 28, 1171-1186 (2014)
22. Kaufman, C.G. and Sain, S.R., Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions, *Bayesian Analysis*, 5 (1), 123150 (2010)
23. Laslett, G.M., Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications, *Journal of the American Statistical Association*, 89 (426), 391-400 (1994)
24. Lee, D.J. and Durban, M., Smooth-Car mixed models for spatial count data, *Computational Statistics and Data Analysis*, 53, 2968-2977 (2009)
25. Lee, D.J. and Durban, M., Pspline ANOVA type interaction models for spatio temporal smoothing, *Statistical Modelling*, 11, 49-69 (2011)
26. Marx, B. D. and Eilers, P. H. C., Generalized linear regression on sampled signals and curves. A P-spline approach, *Technometrics*, 41, 1-13 (1999)
27. Harville, D. A., Matrix Algebra From a Statistician's Perspective. Springer-Verlag New York, Inc. (1997)
28. Menafoglio, A., Secchi, P. and Dalla Rosa, M., A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space, *Electronic Journal of Statistics*, 7, 2209-2240 (2013)
29. Ramsay, J. O. and Silverman, B. W., Functional data analysis (First Edition). Springer-Verlag (1997)
30. Rao, C.R. and Rao, M.B., Matrix Algebra and Its Applications to Statistics and Econometrics. World Scientific Publishing Co. Pte. Ltd. (1998)
31. Reiss, P. T., Huang, L. and Mennes, M., Fast Function-on-Scalar Regression with Penalized Basis Expansions, *The International Journal of Biostatistics*, 6, 1-28 (2010)
32. Ruiz-Medina, M.D., Espejo, R.M., M.D. Ugarte and Militino, A.F. Functional Time Series Analysis of SpatioTemporal Epidemiological Data, *Stochastic Environmental Research and Risk Assessment*, 28, 943954 (2014)

33. Ruiz-Medina, M.D. and Espejo, R.M. Spatial autoregressive functional plug-in prediction of ocean surface temperature, *Stochastic Environmental Research and Risk Assessment*, 26, 335344 (2012)
34. Ruppert, D., Selecting the number of knots for penalized splines, *Journal of Computational and Graphical Statistics*, 11, 735-757 (2002)
35. Sangalli, L. M., Ramsay, J. O. and Ramsay, T. O., Spatial spline regression models, *Journal of the Royal Statistical Society Ser. B, Statistical Methodology*, 75, 1-23 (2013)
36. Shi, J.Q. and Choi, T. Gaussian process regression analysis for functional data. CRC press, Chapman and Hall (2011)
37. Sigrist, F., Kuensch, H. R. and Stahel, W. A., spate: An R Package for Spatio-Temporal Modeling with a Stochastic Advection-Diffusion Process, *Journal of Statistical Software*, 63, 1-23 (2015)
38. Ugarte, M. D., Goicoa, T., Militino, A. F. and Durban, M., Spline Smoothing in small area trend estimation and smoothing, *Computational Statistics and Data Analysis*, 53, 3616-3629 (2009)
39. Yakowitz, S.J. and Szidarovsky, F., A Comparison of Kriging with Non-parametric Regression Methods, *Journal of Multivariate Analysis*, 16, 21-53 (1985).
40. Zhang, J-T. and Chen, J., Statistical Inference for Functional Data, *The Annals of Statistics*, 35 (3), 1052-1079 (2007)
41. Zhang, J-T., Analysis of variance for functional data. CRC press, Chapman and Hall (2014)

A Appendix

Taking into account the following properties [27]

$$\begin{aligned} \text{vec}(A)'(D \otimes B)\text{vec}(C) &= \text{trace}(A'BCD') \\ \text{trace}(A'AB) &= \text{trace}(ABA'), \end{aligned} \quad (1)$$

the equation (5) can be rewritten as

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

with $\Psi = \int \phi^T \phi^T$ being the inner product matrix between the basis functions. Next step is to compute the derivatives of Equation (2) with respect to A . By considering the following properties [27],

$$\frac{\partial \text{trace}(XAX')}{\partial X} = X(A + A') \quad (3)$$

$$\frac{\partial \text{trace}(X'AX)}{\partial X} = (A + A')X \quad (4)$$

$$\frac{\partial \text{trace}(XA)}{\partial X} = A' \quad (5)$$

we have that

$$\begin{aligned}
\frac{\partial \text{trace}(C' C \Psi)}{\partial A} &= 0 \\
\frac{\partial \text{trace}(Z' Z A \Psi A')}{\partial A} &\stackrel{(3)}{=} Z' Z A (\Psi + \Psi') \\
&\stackrel{(\text{symmetry})}{=} 2 Z' Z A \Psi \\
\frac{\partial -2 \text{trace}(A \Psi C' Z)}{\partial A} &\stackrel{(5)}{=} -2 Z' C \Psi' \\
\frac{\partial \lambda_1 \text{trace}(A' (\Delta_d^{U'} \Delta_d^U \otimes I_q) A)}{\partial A} &\stackrel{(4)}{=} 2 \lambda_1 (\Delta_d^{U'} \Delta_d^U \otimes I_q) A \\
\frac{\partial \lambda_2 \text{trace}(A' (I_p \otimes \Delta_d^{V'} \Delta_d^V) A)}{\partial A} &\stackrel{(4)}{=} 2 \lambda_2 (I_p \otimes \Delta_d^{V'} \Delta_d^V) A \\
\frac{\partial \lambda_3 \text{trace}(A (\Delta_d^{T'} \Delta_d^T) A')}{\partial A} &\stackrel{(3)}{=} 2 \lambda_3 A (\Delta_d^{T'} \Delta_d^T).
\end{aligned}$$

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

$$Z' Z A \Psi + \lambda_1 (\Delta_d^{U'} \Delta_d^U \otimes I_q) A + \lambda_2 (I_p \otimes \Delta_d^{V'} \Delta_d^V) A + \lambda_3 A (\Delta_d^{T'} \Delta_d^T) = Z' C \Psi'. \quad (6)$$

In order to get the solution to A , the Kronecker product is used to express Equation (6) in conventional matrix algebra

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

Then, Equation (6) can be re-written as follows

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

where $\text{PEN}_d^{U,V,T}$ is a P-spline penalty developed by [9], which is given by

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

Finally, A is given by

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